INAUGURAL-DISSERTATION

zur Erlangung der Doktorwürde der

NATURWISSENSCHAFTLICH-MATHEMATISCHEN
GESAMTFAKULTÄT

der

RUPRECHT-KARLS-UNIVERSITÄT
HEIDELBERG

vorgelegt von
Diplom-Mathematiker Diplom-Physiker

Matthias Sebastian Maier

aus Neunkirchen (Saar)

Tag der mündlichen Prüfung: _____

Duality-based adaptivity of model and discretization in multiscale finite-element methods

Gutachter: **Prof. Dr. Dr. h. c. Rolf Rannacher**

Abstract This thesis develops strategies for a posteriori error control of discretization and model errors, as well as adaptation strategies, in the context of multiscale finite-element methods. This is done within the general methodology of the Dual Weighted Residual Method (DWR).

In particular, a reformulation of the Heterogeneous Multiscale Method (HMM) as an abstract model-adaptation framework is introduced that explicitly decouples discretization and model parameters. Based on the framework a sampling-adaptation strategy is proposed that allows for simultaneous control of discretization and model errors with the help of classical refinement strategies for mesh and sampling regions. Further, a model-adaptation approach is derived that interprets model adaptivity as a minimization problem of a local model-error indicator. This allows for the formulation of an efficient post-processing strategy that lifts the requirement of strict a priori knowledge about applicability and quality of effective models.

The proposed framework is tested on an elliptic model problem with heterogeneous coefficients, as well as on an advection-diffusion problem with dominant microscopic transport.

Zusammenfassung Ziel der Arbeit ist die Entwicklung von Strategien zur a posteriori Fehlerkontrolle sowohl des Diskretisierungs-, als auch des Modellfehlers im Rahmen von Finite-Elemente-Mehrskalenmethoden. Zur Anwendung kommt hierbei die „Dual Weighted Residual"-Methode.

Zu diesem Zweck wird zunächst eine Reformulierung der bekannten heterogenen Mehrskalenmethode („Heterogeneous Multiscale Method") als generalisierte Methode eingeführt, in der sowohl Diskretisierungs- und Modellparameter, als auch die konkrete Wahl des effektiven Modells, entkoppelt sind. Basierend auf der Reformulierung wird eine „Sampling"-Adaptionsstrategie eingeführt, die die gleichzeitige Kontrolle von Diskretisierungs- und Modellfehler durch lokale Gitter- und Samplingbereichsverfeinerung ermöglicht. Des Weiteren erlaubt es die Reformulierung, Modelladaptation als ein Minimierungsproblem eines lokalen Fehlerindikators aufzufassen. Die Strategie ein Minimierungsproblem zu lösen hat insbesondere den Vorteil, dass keine strikte a priori Aussage über die Anwendbarkeit und Qualität effektiver Modelle vorliegen muss. Die entwickelten Modelladaptionsstrategien werden sowohl an einem elliptischen Modellproblem mit heterogenen Koeffizienten, als auch an einem Advektions-Diffusions-Problem mit dominantem, mikroskopischem Transport numerisch untersucht.

Contents

1 Introduction

A large class of modeling problems in Physics and Engineering is of *multiscale character*, meaning, that relevant physical processes act on highly different length and time scales. A prominent example in Fluid Mechanics is flow in porous media where the macroscopic flow behavior is highly influenced by different permeabilities varying rapidly on a small length scale.

A direct numerical treatment of problems that exhibit multiscale phenomena usually makes a full resolution of all relevant scales necessary. This implies high computational costs, rendering such a simulation computationally infeasible in practice. One way to avoid the need for a full resolution is the use of *multiscale methods*. These are methods where, generally speaking, an *effective model* is solved on a coarse scale with upscaled, effective parameters that are determined with the help of localized (possibly coupled) sampling problems on a fine scale. The usage of such multiscale schemes can be traced back as early as the 1970s, where upscaling principles for effective parameters in the context of elasticity problems were formulated[56,57]. Multiscale methods have become increasingly popular since then in the engineering[49] and mathematical[45] communities.

Different approaches for modeling multiscale phenomena in the context of finite-element methods exist and have lead to a number of methods introduced over the last years. Most of them either rely on the existence of a periodic or stochastic substructure or on the scale-dependent splitting of variational solution- and test spaces. Most notable are the *Variational Multiscale Method* (VMM) developed by Hughes et al.[63] and Brezzi[29], the *Mixed Multiscale Methods* by Arbogast and Boyd[11] or Chen and Hou[32], the *Two-Scale* or *Generalized Finite-Element Method* by Matache and Schwab[75,76], or the *Multiscale Finite-Element Method* (MsFEM) introduced by Hou and Wu[59], Efendiev et al.[47,48], and variants of these approaches. A mathematically rigorous formulation in the context of finite-element theory and homogenization theory was given by E and Engquist[42,43,45] with the description of the so-called *Heterogeneous Multiscale Method* (HMM). In its original setting the HMM relies on a local periodic substructure of the coefficient matrix and can be viewed as a direct discretization of the underlying homogenization process (see Babuška[13] and Ohlberger[88]).

The use of multiscale methods comes at a significant price with respect to sources of error: Among the usual *discretization error* due to a numerical approximation of the partial differential equation (PDE), multiscale methods typically exhibit additional discretization errors in (localized) fine-scale problems, as well as an inherent *model error* resulting from a *modeling assumption* for *scale separation*. Consequently, when applying a multiscale scheme, there is not only the problem of choice for a number of discretization parameters (such as mesh sizes on macro- and microscale) and possible choices of upscaling principles or microscale models, but also the inherent problem that if only one of the many parameters involved is insufficiently chosen, usually a bad approximation is observed that does not obey any nice convergence behavior known from "smooth" problems.

This consideration makes the idea of *a posteriori error estimation*, where a *quantitative estimate* for the different sources of error is computed by means of a post-processing approach (within an adaptation cycle), highly attractive and a promising field of research.

How to deal with discretization errors is well understood: For the HMM a number of theoretical a priori and a posteriori results are available: An *a priori* error estimation dealing with the discretization errors was first presented by E and Engquist[42,43], E et al.[45], and later, with improved results, by Abdulle[2] and Abdulle and Vilmart[5] (for a nonlinear case) – without an estimation of the underlying modeling error. *A posteriori* error estimation for the discretization errors was later presented by Ohlberger[88], Henning and Ohlberger[53], Henning et al.[55], and Abdulle[3]; *goal-oriented* error estimation results (for discretization errors on macro- and microscale) were formulated by Abdulle and Nonnenmacher[4]. Corresponding a priori and a posteriori results for the VMM ansatz were derived by Larson and Målqvist[67–69].

The novelty in a posteriori error estimation with respect to multiscale methods lies in the possibility for *model adaptivity*. First results for *estimating and controlling the model error* in the context of multiscale schemes were given by Oden and Vemaganti[84–87,91] and Braack and Ern[26].

Goal-oriented adaptivity for multiscale methods A versatile method for a posteriori error control is the *Dual Weighted Residual* (DWR) method developed by Becker and Rannacher[19–21]. It constructs estimates of local error contributions in terms of a target functional, the so called *quantity of interest*, with the help of *dual problems*. While such duality arguments and error identities are also one of the main tools in classical (residual based) a posteriori error estimation for multiscale methods[4,88] the principal idea of the DWR method lies in the

fact that it is possible to numerically approximate the dual solution directly with comparable accuracy than the primal problem resulting in *dual weighted*, and hence improved, local error indicators.

In contrast to PDEs without multiscale behavior the following main problems have to be addressed[72]:

- The dual problem exhibits the same multiscale features as the primal problem and is therefore of the same computational complexity. A reduction of the dual problem that still preserves all multiscale features is necessary. Depending on the information that shall be extracted from the dual problem, it can either be approximated by a coarse, or homogenized dual problem (for discretization errors) or with the help of local reconstructions (to assess the local quality of the current multiscale model for the primal problem).

- Discretization errors on macro- and microscale can be dealt with by classical mesh refinement. However, a good estimate to balance the amount of mesh refinement between scales and, in particular, in comparison to the model adaptation is important.

- Local model adaptation to minimize a high local model error is a hard problem. Locally switching to a better problem is an obvious strategy (with, for example, a full resolution as the "best" available model). But whether such an approach leads to a better model is hard to predict a priori, especially information about quantitative behavior of the better model is unavailable. Furthermore, better models might not always be computationally feasible.

Objective of the thesis The main target of this thesis is the application of the DWR method in the context of multiscale schemes, an investigation about the principal challenges this involves, and the development of *model-adaptation strategies*. Therefore, this thesis introduces a reformulation of the HMM as an abstract model-adaptation framework that explicitly decouples all sources of error (Chapter 3 and 4). Based on this framework, different approaches for model-adaptation strategies are developed and discussed. In particular, a *sampling-adaptation strategy* is proposed that employs model adaptivity by means of locally refining a sampling mesh $\mathbb{T}_\delta(\Omega)$ that holds cell-wise constant, *effective* parameters (Chapter 5). A balancing strategy is introduced for the sampling-adaptation process that simultaneously controls macro- and microscale discretization in addition to the sampling process. Further, a novel model-adaptation approach is derived that interprets model adaptivity as an *optimization problem* (Chapter 6).

This allows for the formulation of an efficient post-processing strategy that can be interpreted as a multiscale approach in its own right (see Section 6.3) and lifts the requirement of strict a priori knowledge about applicability and quality of effective models.

Layout of the thesis In detail, the thesis is organized as follows: *Chapter 2* gives an overview over different multiscale-modeling approaches. Based on the variational, elliptic model problem to find $u^\varepsilon \in H_0^1(\Omega)$ s. t.

$$\int_\Omega A^\varepsilon(x)\nabla u^\varepsilon(x)\cdot\nabla\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x \quad \forall\varphi\in H_0^1(\Omega) \tag{1.1}$$

on the Sobolev space $H_0^1(\Omega)$, the HMM and VMM approaches, as well as its most important variants and related methods, are introduced. Here, $A^\varepsilon(x)\in L^\infty(\Omega)^{d\times d}$ is sought to contain *multiscale features*, i. e. it rapidly oscillates on a small length scale. Further, a computational homogenization approach based on physical upscaling principles is exemplarily examined. The chapter concludes with a discussion about upscaling strategies based on simple averaging principles.

In *Chapter 3* an abstract model-adaptation framework based on the HMM scheme is derived that explicitly decouples discretization and modeling parameters. This is done by defining the concept of an *effective model*

$$\bar{A}^\delta \;:\; \mathbb{T}_\delta(\Omega)\to\mathbb{R}^{d\times d} \tag{1.2}$$

on a so-called *sampling mesh* $\mathbb{T}_\delta(\Omega)$ and by further introducing an *effective equation* to find $u^\delta\in H_0^1(\Omega)$ s. t.

$$\int_\Omega \bar{A}^\delta(x)\nabla u^\delta(x)\cdot\nabla\varphi(x)\,\mathrm{d}x = \int_\Omega f(x)\varphi(x)\,\mathrm{d}x \quad \forall\varphi\in H_0^1(\Omega). \tag{1.3}$$

A priori error estimates for two different finescale reconstruction approaches (simple upscaling by averaging and an HMM ansatz) are discussed.

The corresponding a posteriori error analysis in the DWR framework is given in *Chapter 4*. A key point is the efficient numerical approximation of the *dual equation*

$$\int_\Omega A^\varepsilon(x)\nabla\varphi(x)\cdot\nabla z^\varepsilon(x)\,\mathrm{d}x = \langle j\,,\varphi\rangle \quad \forall\varphi\in H_0^1(\Omega) \tag{1.4}$$

where the functional j denotes a *quantity of interest*. Efficiently approximating the dual solution can be done by different localization strategies. As a novel

localization technique the approximation of the dual problem by *globally* using the same effective model $\bar{A}^\delta$ and *locally reconstructing* finescale features of the full solution z^ε with the help of local cell problems is proposed (Section 4.3). The dual equation leads to an *error identity* for the finite-element approximation U,

$$\begin{aligned}\langle j, u^\varepsilon\rangle - \langle j, U\rangle &= \underbrace{\big(f, z^\delta\big) - \big(A^{\delta,h}\nabla U, \nabla z^\delta\big)}_{=:\theta^H} \\ &+ \underbrace{\big(A^{\delta,h}\nabla U, \nabla z^\delta\big) - \big(A^\delta \nabla U, \nabla z^\delta\big)}_{=:\theta^h} + \underbrace{\big(A^\delta \nabla u^\delta, \nabla z^\varepsilon\big) - \big(A^\varepsilon \nabla u^\delta, \nabla z^\varepsilon\big)}_{=:\theta^\delta}, \quad (1.5)\end{aligned}$$

which is used to derive *decoupled local indicators* for the macroscale discretization error (θ^H), the microscale discretization error (θ^h) and the model error (θ^δ). This split allows to balance error contributions coming from discretization and model.

Based on the model framework and the a posteriori error analysis derived for it, different model-adaptation strategies are introduced and discussed in *Chapter 5*: Within the given framework, two fundamentally different approaches for model adaptivity are possible. The first possibility is to adjust the sampling mesh $\mathbb{T}_\delta(\Omega)$ and the location of associated sampling regions while keeping the same reconstruction process for all sampling regions. The second one consists of "improving" the effective model used for the reconstruction (by various means)[26,84–87,91]. This is typically done while maintaining the same fixed sampling discretization. In particular, *model switching* (i. e. the gradual, local improvement of $\bar{A}^\delta$ from a fixed set of a priori chosen models) is discussed as an example for the second approach. Further, a *sampling-mesh adaptation strategy* is introduced, where the mesh $\mathbb{T}_\delta(\Omega)$ is locally refined in order to improve the effective model.

As a novel approach *Chapter 6* presents a *model-optimization framework* that interprets model adaptivity as an optimization problem to minimize the model error θ^δ in the above error identity:

$$\operatorname*{arg\,inf}_{\bar{A}^\delta \in \mathscr{A}^\delta} \sum_{K\in\mathbb{T}_\delta(\Omega)} \big|\big((\bar{A}^\delta - A^\varepsilon)\nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\big)_K\big|^2 + \alpha_K \big\|\bar{A}^\delta_K - \bar{A}^{\delta,0}_K\big\|^2_{\mathbb{R}^{d\times d}}. \quad (1.6)$$

The optimization problem allows to derive an efficient post-processing strategy for model adaptation. Most importantly, it lifts the requirement of strict a priori knowledge about applicability and quality of effective models.

Numerical results will be given throughout Chapters 3 to 6 to validate the theoretical findings. Especially, the model adaptation strategies derived in Chapters 5 and 6 are tested on prototypical numerical examples. A conclusion and outlook is given in Chapter 7.

2 A survey of multiscale methods

A natural approach to tackle problems with *multiscale character*—beside the obvious method of running an *expensive* computation with full resolution—is to identify and exploit *scale separating* properties. This may happen on a purely mathematical level starting from an already derived microscale model that captures all relevant scales. Consider a sequence of abstract problems:

$$\text{Find } u^\varepsilon \in \mathcal{V}^\varepsilon \text{ s.t. } \mathcal{L}^\varepsilon u^\varepsilon = f^\varepsilon, \tag{2.1}$$

with a sequence of operators $\mathcal{L}^\varepsilon : \mathcal{V}^\varepsilon \to \mathcal{V}^{\varepsilon *}$ and data $f^\varepsilon \in \mathcal{V}^{\varepsilon *}$ associated with a scaling parameter $\varepsilon > 0$. Here, $\mathcal{V}^\varepsilon$ denotes a separable Hilbert space and $\mathcal{V}^{\varepsilon *}$ its dual space. Different methods emerge by either exploiting structure in the (variational) ansatz spaces, or by exploiting structure in the operator.

One possible strategy is to construct sequences of subsequently refined ansatz spaces,

$$\mathcal{V}^\varepsilon = \mathcal{V}^0 \oplus \mathcal{V}^1 \oplus \mathcal{V}^2 \oplus \dots, \tag{2.2}$$

with the idea to *localize* $\mathcal{V}^i$ and only solve for a subspace $\mathcal{V}^0 \oplus \mathcal{V}^1 \oplus \dots \oplus \mathcal{V}^n$ resulting in an efficient multiscale approximation of the variational ansatz space. Examples for approaches of this nature are the *Variational Multiscale Method* introduced by Hughes et al.[63], the *Mixed Multiscale Methods* by Arbogast and Boyd[11] or Chen and Hou[32], the *Two-Scale* or *Generalized Finite-Element Method* by Matache and Schwab[75,76] or the *Multiscale Finite-Element Method* introduced by Hou and Wu[59] and Efendiev et al.[48], and variants thereof.

A different approach comes from a formalization of structure in the operator $\mathcal{L}^\varepsilon$ of the partial differential equation. With the help of *formal* expansions

$$\mathcal{L}^\varepsilon = \mathcal{L}^0 + \varepsilon \mathcal{L}^1 + \varepsilon^2 \mathcal{L}^2 + \dots \tag{2.3}$$

matched with an expansion of the solution

$$u^\varepsilon = u^0 + \varepsilon u^1(x, \tfrac{x}{\varepsilon}) + \varepsilon^2 u^2(x, \tfrac{x}{\varepsilon}) + \dots \tag{2.4}$$

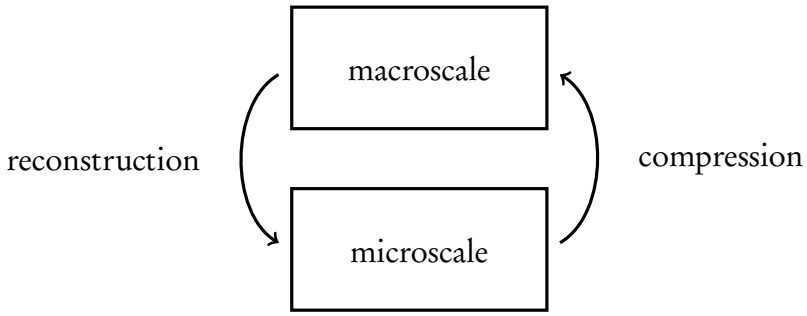


Figure 2.1. The general multiscale modeling approach: macroscale and microscale are coupled by an averaging prescription (*compression*) and in opposite direction a *reconstruction* procedure to recover fine-scale behavior.

it is possible to derive effective equations for the constituents $u^0, u^1, \ldots, u^n$. The idea is that a cut-off of the formal expansion (2.4) is already a good approximation of u^ε (in some notion yet to be defined):

$$u^\varepsilon \approx u^0 + \varepsilon u^1(x, \tfrac{x}{\varepsilon}) + \ldots + \varepsilon^n u^n(x, \tfrac{x}{\varepsilon}). \tag{2.5}$$

The procedure of expanding functions and equations in powers of a scaling parameter ε and match the result in powers of ε is known as asymptotic analysis[22]. A mathematically precise justification is the subject of mathematical *homogenization theory*[7,22,34]. Effective models derived in the context of homogenization theories allow for a reformulation as a numerical multiscale scheme[2,88] such as the *Heterogeneous Multiscale Method* by E and Engquist[42,43].

Mathematical approaches for deriving multiscale descriptions (by starting from an already derived microscale model) are fundamentally different from physical approaches. Those usually begin at the modeling phase and employ upscaling and conservation principles to couple a priori independent (and possibly different) models for micro- and macroscale[49].

The general physical multiscale modeling approach can be summarized as follows: A macroscale and microscale model are coupled by an averaging prescription, so-called *compression*, that describes how to average quantities from the microscale to be used on the macroscale. Further, a *reconstruction* procedure recovers fine-scale behavior from a given macroscale quantity, see Figure 2.1. The result is a *computational homogenization* technique that couples individual, closed microscale problems on so-called *representative volume elements* via boundary conditions and averaging principles to a macroscale. We refer to a discussion by Zohdi and Wriggers[101] for an overview. An example for pure upscaling principles (without a reconstruction) are approaches using simple averaging of

effective parameters under certain microscale distributions, such as the geometric mean value[77] in context of *log-normally* distributed permeabilities[100]. Examples for combined reconstruction and compression principles are the *Method of Virtual Power*[50] which results in the *Hill-Mandel-Homogeneity condition*[57] (see Section 2.3.1).

It is remarkable that given the very different nature and reasoning in model derivation the resulting multiscale methods are usually closely related. Therefore, the aim of this chapter is to give an introduction to the approaches mentioned above and outline their equivalence.

The underlying reasoning is exemplified with typical, established multiscale schemes; in particular, the Variational Multiscale Method for a multiscale scheme derived from ansatz-space splitting, the Heterogeneous Multiscale Method for a multiscale scheme derived by exploiting structure in the operator, and a (first order) computational homogenization scheme as an example of how to apply physical conservation principles for multiscale coupling.

For simplicity, the discussion in this chapter will be based on the following abstract model problem:

Definition 2.1 (Abstract model problem)**.** Let $\mathcal{V}$ be a separable Hilbert space and

$$a : \mathcal{V} \times \mathcal{V} \to \mathbb{R} \tag{2.6}$$

a $\mathcal{V}$-elliptic, bilinear form. For a given linear and continuous functional $f \in \mathcal{V}^*$ find $u \in \mathcal{V}$ such that

$$a(u, \varphi) = \langle f, \varphi \rangle \quad \forall \varphi \in \mathcal{V}, \tag{2.7}$$

where $\langle ., . \rangle$ denotes the duality pairing.

Often, the abstract model problem will be reduced to a *Laplace-like* problem with *heterogeneous* coefficients:

Definition 2.2 (Model problem)**.** Let Ω be a bounded domain and let $A^\varepsilon \in L^\infty(\Omega)^{d\times d}$ with $A^\varepsilon = (A^\varepsilon_{ij})_{ij=1}^d$,

$$A^\varepsilon_{ij} = A^\varepsilon_{ji} \quad \forall i, j = 1, \ldots, d, \tag{2.8}$$

$$\exists \alpha, \beta \in \mathbb{R}^+ \text{ s. t. } \alpha |\xi|^2 \leq \sum_{ij} A^\varepsilon_{ij} \xi_i \xi_j \leq \beta |\xi|^2 \quad \forall \xi \in \mathbb{R}^d \text{ a. e. on } \Omega, \tag{2.9}$$

and define the bilinear form $a(\varphi,\psi)=\left(A^\varepsilon\nabla\varphi,\nabla\psi\right)_{L^2(\Omega)^d}$ for $\varphi,\psi\in H_0^1(\Omega)$. The model problem now reads

$$\left(A^\varepsilon\nabla u^\varepsilon,\nabla\varphi\right)_{L^2(\Omega)^d}=\langle f,\varphi\rangle\quad\forall\varphi\in H_0^1(\Omega). \tag{2.10}$$

Existence and uniqueness of u and u^ε in (2.7) and (2.10) follow by classical *elliptic solution theory*.

Here, the space $L^\infty(\Omega)$ denotes the space of Lebesgue-Borel-measurable functions bounded in the essential supremum norm and $H^1(\Omega)$ is the usual Sobolev space of functions $\varphi\in L^2(\Omega)$ (Lebesgue-Borel-measurable and square integrable) with generalized derivatives ∇u of class $L^2(\Omega)^d$. The Hilbert space $H_0^1(\Omega)$ is the subspace of $H^1(\Omega)$ of functions fulfilling homogeneous Dirichlet boundary conditions. $(.,.)$ is the usual scalar product on $L^2(\Omega)$.

Further, $\varepsilon>0$ is a scaling parameter for fine-scale oscillations. When a certain fine-scale behavior is needed, the precise definition of the functional dependency $\varepsilon\mapsto A^\varepsilon$ will be given. For now, ε shall be interpreted as a superscript indicating an abstract notion of fine-scale oscillations.

Remark 2.3. While (2.7) and (2.10) might look similar, they exhibit—from a multiscale perspective—quite different behavior: The essential difference with respect to multiscale schemes is the fact that (2.7) allows for *nonlocal* coupling effects and responses between u and φ, whereas (2.10) is localized in this regard.

Remark 2.4. A generalization to an advection-diffusion problem in context of the Heterogeneous Multiscale Method is discussed in Section 2.2.5. The application of a Computational Homogenization scheme on an elasticity problem can be found in Section 2.3.1.

2.1 Multiscale methods based on ansatz-space splitting

Multiscale methods based on the splitting of ansatz spaces arose in the context of (modern) variational formulation of partial differential equations and its numerical treatment with the finite-element method. At heart, it is based on the fundamental observation that *subgrid phenomena*, i. e. effects not captured by a given resolution of a numerical discretization, can have two distinct sources[29]: They may either result from a lack of discretization, more precisely, the lack of necessary stability properties of the discretization. Or, they may come from some

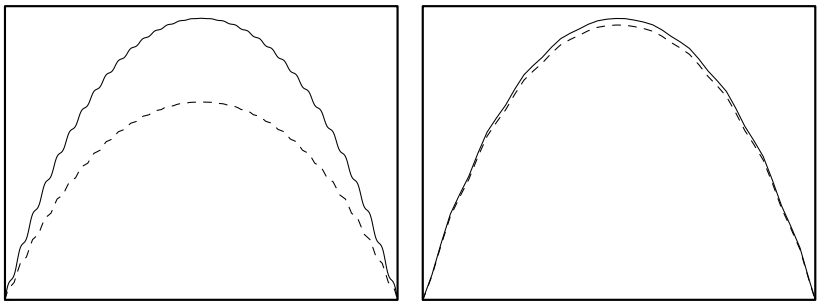


Figure 2.2. Schematic example of a function u with multiscale character: The Ritz projection u^H (dashed line) onto a finite-element space V^H differs greatly in terms of macroscale behavior in case of a coefficient with multiscale behavior (a) compared to a smooth coefficient with mild oscillations (b).

physical (subgrid) phenomena that are present in the model but not captured by the given macroscale resolution.

The stability aspect of numerical discretizations is a well understood field of research, so there is hope that techniques developed for the former are also applicable to the latter. In spirit of this reasoning, Hughes et al.[60,62,63] and Brezzi[29] developed the *Variational Multiscale Method* (VMM) that shall be examined in detail in the following section.

2.1.1 Variational Multiscale Method (VMM)

Assume that the abstract model problem defined in (2.7), find $u \in \mathcal{V}$ such that

$$a(u,\varphi) = \langle f,\varphi\rangle \quad \forall \varphi \in \mathcal{V}, \tag{2.11}$$

is of multiscale character, i. e. the variational formulation (2.11) yields the correct result u for the full space V, but any Ritz projection u^H onto a (classical) finite-element subspace V^H leads to a bad approximation (see Figure 2.2). Now, we assume that $\mathcal{V}$ can be decomposed into a direct sum

$$\mathcal{V} = V^H \oplus \mathcal{V}^f \tag{2.12}$$

consisting of a finite-dimensional finite-element space V^H and a *fine-scale space* $\mathcal{V}^f$. Various different splittings are possible. For the moment we just assume that the splitting is direct in the sense of vector spaces so that subsequent equations are well defined. Additional orthogonality relations are discussed in Section 2.1.4.

Following the notation and general considerations introduced by Brezzi[29], Equation (2.11) can be rewritten as: Find $u^H \in V^H$, $u^f \in \mathcal{V}^f$ such that

$$a(u^H, \varphi^H) + a(u^f, \varphi^H) = \langle f, \varphi^H \rangle \quad \forall \varphi^H \in V^H, \tag{2.13a}$$
$$a(u^f, \varphi^f) + a(u^H, \varphi^f) = \langle f, \varphi^f \rangle \quad \forall \varphi^f \in \mathcal{V}^f. \tag{2.13b}$$

Due to the linearity of a we immediately get:

Lemma 2.5. The variational formulation (2.13) is well defined and admits a one-to-one correspondence to (2.7).

Now, let $\mathcal{T} : \mathcal{V}^{f*} \to \mathcal{V}^f$ denote the solution operator associated with (2.13b). More precisely, $\mathcal{T}$ satisfies

$$a(\mathcal{T} f, \varphi^f) = \left\langle f, \varphi^f \right\rangle \quad \forall\, \varphi^f \in \mathcal{V}^f, \forall\, f \in \mathcal{V}^{f*}. \tag{2.14}$$

With

$$\mathcal{L} \,:\, \mathcal{V} \to \mathcal{V}^*, \quad \mathcal{L}\,\varphi = a(\varphi, .) \tag{2.15}$$

equation (2.13a) can be restated as

$$a(u^H, \varphi^H) + a(\mathcal{T}(f - \mathcal{L}\,u^H), \varphi^H) = \left\langle f, \varphi^H \right\rangle \quad \varphi^H \in V^H. \tag{2.16}$$

This *statically condensed* equation is a typical starting point for variational multiscale formulations. Furthermore, it allows for a precise definition of the notion of multiscale character (in the context of VMM).

Definition 2.6. A problem is of *multiscale character* (in the sense of VMM) if the fine-scale contribution $a(\mathcal{T}(f - \mathcal{L}\,u^H), \varphi^H)$ is of the same order as $a(u^H, \varphi^H)$ for every computationally feasible V^H.

The idea of the VMM approach is to find a localized, finite-dimensional subspace formulation

$$V^f = \sum_i V_i^f \subset \mathcal{V}^f \tag{2.17}$$

such that with the solution operators $\mathcal{T}_i : V_i^{f*} \to V_i^f$ it holds true that

$$\sum_i a\big(\mathcal{T}_i(f - \mathcal{L}\,u^H), \varphi^H\big) \approx a\big(\mathcal{T}(f - \mathcal{L}\,u^H), \varphi^H\big). \tag{2.18}$$

However, it is a delicate matter to construct a good subspace approximation (2.17) that *also* allows for an efficient approximation of $a(\mathscr{T}(f-\mathscr{L}u^H,\varphi^H)$. One has to keep in mind that the goal is to construct an efficient multiscale method that should allow for good parallelization of assembly and solution processes, and should *in total* be cheaper than a straightforward numerical treatment with a full resolution of the original problem.

We examine two reconstruction approaches in the following.

2.1.2 Residual-Free Bubbles and Element Green's Functions

The traditional approach developed to approximate the fine-scale contribution is via so-called *Residual-Free Bubbles* or *Element Green's Functions*. Let $\mathbb{T}_H$ be a mesh covering Ω and let V^H be the corresponding ansatz space with linear finite elements. Define a space of element-wise *bubbles*[29]

$$\mathscr{V}_B^f \coloneqq \prod_{K\in\mathbb{T}_H} H_0^1(K). \tag{2.19}$$

Remark 2.7. Above choice (2.19) for a fine-scale approximation is consistent and, in case of a linear space V^H, the sum of macro- and microscale spaces is direct, i. e. $V^H \oplus \mathscr{V}_B^f \subset \mathscr{V}$. For a higher polynomial degree for the finite-element ansatz V^H additional constraints are necessary to ensure direct summability.

Equation (2.13) decouples (2.13b) into localized fine-scale contributions

$$u^f = \sum_{K\in\mathbb{T}_H} u_K^f, \quad u_K^f \in H_0^1(K), \tag{2.20a}$$

$$a(u_K^f,\varphi_K^f) = \left\langle f-\mathscr{L}u^H,\varphi_K^f\right\rangle \quad \forall \varphi_K^f \in H_0^1(K). \tag{2.20b}$$

Let $\mathscr{T}_K : H^{-1}(K) \to H_0^1(K)$ denote the solution operator associated with (2.20b). Inserted into (2.13a) the coarse-scale equation now reads

$$a(u^H,\varphi^H) + \sum_{K\in\mathbb{T}_H} a\big(\mathscr{T}_K(f-\mathscr{L}u^H),\varphi^H\big) = \left\langle f,\varphi^H\right\rangle. \tag{2.21}$$

Remark 2.8. The name *Residual-Free Bubbles* comes from the following fact: Given a fine-scale discretization to solve the fine-scale problems approximately

$$V_B^f \coloneqq \prod_{K\in\mathbb{T}_H} V^h(K) \subset \mathscr{V}_B^f = \prod_{K\in\mathbb{T}_H} H_0^1(K), \tag{2.22}$$

and a *localizing property* for the variational form a (e. g. in form of problem (2.10)), $a(\varphi,\psi)=(A\nabla\varphi,\nabla\psi)$, relatively few macroscale degrees of freedom couple into the fine-scale problem. Consequently, the coarse-scale equation (2.16) can be rewritten by a *Schur's complement approach*:

$$\sum_{K,\operatorname{supp}\varphi_i\cap K\neq\emptyset}\int_K A\nabla\varphi_j^H\cdot\nabla\varphi_i^H\,\mathrm{d}x-\int_K A\nabla(\mathscr{T}_K\varphi_j^H)\cdot\nabla\varphi_i^H\,\mathrm{d}x$$
$$=\sum_{K,\operatorname{supp}\varphi_i\cap K\neq\emptyset}\int_K f\varphi_i^H\,\mathrm{d}x-\int_K A\nabla(\mathscr{T}_K f)\cdot\nabla\varphi_i^H\,\mathrm{d}x. \quad (2.23)$$

This allows for an efficient parallel assembly of the macroscale problem where no macroscale residual enters the fine-scale feedback computations explicitly, hence *residual free*[60].

Remark 2.9. The motivation behind ansatz (2.19) is owed to the restricted computational possibilities of the time the method was first formulated. Spending some degrees of freedom for a local subspace reconstruction of (2.20b) with a priori considerations can result in a good *stabilization* of the subscale behavior. Given the nature of Equation (2.20b), a better approximation of the fine-scale response is likely reachable by a bubble space approximation than by pure enrichment of the macroscale finite-element ansatz.

A similar approach, originally formulated by Hughes[60] arises from the use of an *Element Green's Function* formulation by expressing all terms coming from the fine-scale contribution by the Green's function associated with (2.20b):

Definition 2.10. Let g_K denote the classical Green's function associated with the solution of (2.20b), i. e. for a given right hand side f and with $\mathscr{L}_K$ denoting the variational operator of (2.20b) in strong form, the solution of $\mathscr{L}_K u\equiv f$ can be expressed as

$$u=\int g_K(x,y)f(y)\mathrm{d}y. \quad (2.24)$$

With the help of g_K the fine-scale contribution $a\big(\mathscr{T}_K(f-\mathscr{L}u^H),\varphi^H\big)$ in the condensed equation (2.16) can be expressed as

$$a\big(\mathscr{T}_K(f-\mathscr{L}u^H),\varphi^H\big)=\big\langle\mathscr{L}\varphi^H,\mathscr{T}_K(f-\mathscr{L}u^H)\big\rangle$$
$$=\int_K\int_K g_K(x,y)\mathscr{L}\varphi^H(x)(f-\mathscr{L}u^H)(y)\,\mathrm{d}y\,\mathrm{d}x. \quad (2.25)$$

Remark 2.11. The motivation behind a description of the fine-scale contribution in form of Equation (2.25) lies in the possibility of a pure a priori treatment of (2.25). The behavior of Green's functions is a well established field of research.

Both approaches, Residual-Free Bubbles and Element Green's Functions, lead to similar descriptions depending on the concrete approximation chosen (for the fine-scale problems, or the Green's function, respectively). It is important to note that the basic idea behind both approaches is to enable a relatively cheap *stabilization technique* for multiscale problems. It is possible to allow only for a few degrees of freedom (even just one degree of freedom) for the fine-scale reconstruction, for example by choosing a one-dimensional bubble element

$$\mathcal{V}_B^f \coloneqq \left\{ \varphi \in P^4(K) \,:\, \varphi(x)\big|_K \equiv 0 \right\}, \tag{2.26}$$

or by approximating the Green's function g_K by

$$g_K \approx \tau \delta(x-y) \tag{2.27}$$

with a scalar $\tau \in \mathbb{R}$ suitably chosen. The correspondence between both approaches is examined in depth by Brezzi et al.[30]

Remark 2.12. Various stabilization techniques, that were developed and formulated independently of the VMM ansatz, can be recast as a derivation based on (2.25) in context of the VMM ansatz[60]. We state an example by Hughes[60] in the context of above abstract model problem (2.7). Consider an approximation of $g_K(x,y)$ by the Dirac distribution,

$$g_K(x,y) \approx \tau \delta(x-y), \tag{2.28}$$

with a degree of freedom $\tau \in \mathbb{R}$ chosen such hat

$$\int_K \int_K g_K(x,y)\,\mathrm{d}x\,\mathrm{d}y = \tau \int_K \int_K \delta(x-y)\,\mathrm{d}x\,\mathrm{d}y. \tag{2.29}$$

Inserting (2.28) and (2.29) into (2.25) leads to

$$a(u^H,\varphi^H) + \sum_{K\in\mathbb{T}_H} \tau(K) \int_K (f - \mathscr{L}u^H)\mathscr{L}\varphi^H \,\mathrm{d}x = \left\langle f, \varphi^H \right\rangle, \tag{2.30a}$$

with a cell-wise stabilization parameter

$$\tau(K) \coloneqq \frac{1}{|K|} \int_K \int_K g_K(x,y)\,\mathrm{d}x\,\mathrm{d}y. \tag{2.30b}$$

This is a classical *Galerkin least-squares* stabilization, except for the choice of scaling parameter τ that is not determined by a classical heuristic approach (bound to the local mesh size) but by a refined choice determined by (2.29) that takes fine-scale influences into account. For a thorough investigation as well as an overview of more involved schemes we refer to Hughes et al.[60,62,63].

2.1.3 Localization techniques by partition of unities

The strict confinement to function spaces with support inside a cell in the bubble-space ansatz

$$V^f \subset \mathcal{V}_B^f = \prod_{K \in \mathbb{T}_H} H_0^1(K) \tag{2.31}$$

has the severe drawback that it only allows for fine-scale responses *localized within* a macro-cell K. The a priori assumption that no inter-cell fine-scale responses occur does not have to be satisfied, meaning that a VMM approach with bubbles might result in an unstable approximation with

$$\left| \sum_i a\big(\mathcal{T}_i(f - \mathcal{L}u^H), \varphi^H\big) \right| \ll \left| a\big(\mathcal{T}(f - \mathcal{L}u^H), \varphi^H\big) \right|. \tag{2.32}$$

This phenomenon, i. e. the error due to non-matching periodicities of the sampling (in this case due to strict homogeneous Dirichlet conditions on macro-cell boundaries) with the inherent periodicity of the micro-structure is known as *resonance error*[59].

It is possible to augment $\mathcal{V}^f$ to take missing (inter macro-cell) fine-scale features into account. The tradeoff is a higher computational complexity. In the following we present a modified localization technique proposed and analyzed in depth by Larson and Målqvist[67,68,81] that addresses this problem. It is based on a (nodal-oriented) partition of unity where the size of the individual fine-scale-reconstruction *patches*, more precisely the position of artificial boundary conditions of the fine-scale reconstructions, can be chosen at will—ideally controlled by a priori or a posteriori techniques (as will be discussed later).

Let $\{\psi_i\}_{i=1}^n \subset H^1(\Omega)$ be a partition of unity, e. g. the nodal basis of a linear finite-element space V^H. Split the fine-scale equation as follows:

$$u^f = \sum_{i=1}^n u_i^f, \quad u_i^f \in \mathcal{V}^f, \tag{2.33a}$$

$$a(u_i^f, \varphi^f) = \left\langle f - \mathcal{L}u^H, \psi_i \varphi^f \right\rangle \quad \forall \varphi^f \in \mathcal{V}^f,\ i = 1, \ldots, n. \tag{2.33b}$$

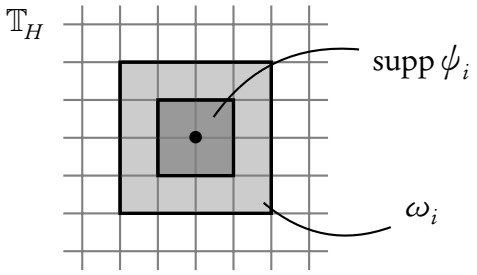


Figure 2.3. A patch $\omega_i \subset \mathbb{T}_H$ for a nodal basis function ψ_i.

The fine-scale equations are now localized by imposing artificial boundary conditions. For each nodal basis function ψ_i fix a simply-connected patch $\omega_i \subset \mathbb{T}_H$ with $\operatorname{supp}\psi_i \subset \omega_i$ (see Figure 2.3) and define a space

$$\mathcal{V}^f_{\omega_i} \coloneqq \Big\{ \psi \in \mathcal{V}^f \,:\, \operatorname{supp}\psi \subset \omega_i \Big\}. \tag{2.34}$$

This leads to

$$a(u^H, \varphi^H) + a(u^f, \varphi^H) = \langle f, \varphi^H \rangle \quad \forall \varphi^H \in V^H, \tag{2.35a}$$

$$a(u^f_i, \varphi^f) = \big\langle f - \mathscr{L} u^H, \psi_i \varphi^f \big\rangle \quad \forall \varphi^f \in \mathcal{V}^f_{\omega_i}, \forall\, i = 1, \ldots, n. \tag{2.35b}$$

By construction, the splitting (2.35) is again a conforming ansatz, i. e.

$$V^H \oplus \sum_i \mathcal{V}^f_{\omega_i} \,\subset\, \mathcal{V}. \tag{2.36}$$

The difference lies in the possibility to increase the sampling region as necessary.

Remark 2.13. It is noteworthy to mention that (2.35) is by construction always "oversampling", i. e. even for the smallest non-degenerate choice of ω_i every macro cell is sampled for every nodal basis function with intersecting support in ω_i. This makes an a posteriori method for controlling the local patch size of ω_i highly necessary. On the other hand, the splitting (2.35) is more natural in the sense that it suppresses the influence of artificial boundary conditions "isotropically" (for every nodal point). In contrast, the bubble-space ansatz forces strict confinement to

$$\big\{ \varphi \in \mathcal{V}^f : \varphi\big|_{\partial K} \equiv 0 \quad \forall K \in \mathbb{T}_H \big\}, \tag{2.37}$$

which completely removes any fine-scale modes not identically vanishing on individual element boundaries.

2.1.4 On the choice of subspaces

Besides the influence of artificial boundary equations another crucial aspect of a VMM approach is the choice of "orthogonality" for the space decomposition $V^H \oplus \mathcal{V}^f$. By construction, the bubble-space ansatz is always direct for the choice of a linear finite-element space V^H. For higher polynomial degrees, or for the partition-of-unity approach, the fine-scale space $\mathcal{V}^f$ has to be restricted. From an implementational standpoint a natural choice is the *hierarchical basis method*, where $V^h(\omega_i)$ is chosen to be part of a fine, global finite-element ansatz V^h on a refined mesh $\mathbb{T}_h$ of $\mathbb{T}_H$ divided by V^H to ensure strict summability[81]:

Definition 2.14. Let $\mathbb{T}_H(\Omega)$ be a mesh covering Ω, and $\mathbb{T}_h(\Omega) \supset \mathbb{T}_H(\Omega)$ be a refined mesh. With the nodal interpolant $I_H : V^h(\Omega) \to V^H(\Omega)$ define

$$V^h(\omega_i) \coloneqq \left\{ \varphi \in V^h(\Omega) : \operatorname{supp} \varphi \subset \bar{\omega}_i \text{ and } I_H \varphi \equiv 0 \right\}. \tag{2.38}$$

A more expensive approach than the hierarchical basis method is to enforce L^2-, H^1-, a-orthogonality for the subspaces by orthogonalizing $V^h(\omega_i)$ such that

$$\left(\varphi^H, \varphi^f\right)_{L^2(\omega_i)} = 0 \quad \forall\, \varphi^H \in V^H,\ \forall\, \varphi^f \in V^h(\omega_i), \text{ or} \tag{2.39}$$

$$\left(\nabla\varphi^H, \nabla\varphi^f\right)_{L^2(\omega_i)^d} = 0 \quad \forall\, \varphi^H \in V^H,\ \forall\, \varphi^f \in V^h(\omega_i), \text{ or} \tag{2.40}$$

$$a\left(\varphi^H, \varphi^f\right) = 0 \quad \forall\, \varphi^H \in V^H,\ \forall\, \varphi^f \in V^h(\omega_i), \tag{2.41}$$

respectively. This leads to considerably improved localization of fine-scale problems, and consequently less influence of artificial boundary conditions. We refer to Hughes and Sangalli[61] for a detailed theoretical and numerical analysis for different projector choices (L^2- and H^1-orthogonal), as well as to a publication by Målqvist and Peterseim[82] that shows results for the localizing property of an $a(.,.)$-orthogonal splitting. However, constructing spaces with high orthogonality requirements is very expensive[81].

2.1.5 Implementational aspects and scale separability

Given the nodal basis $\{\psi_i^H\}_{i \in \mathcal{N}}$ of V^H the Equation (2.35) can be rewritten as a linear system of equations $AU = F$,

$$A_{ij} \coloneqq a(\varphi_j^H, \varphi_i^H) - \sum_{k \in \mathcal{N}} a\left(\mathcal{T}_k^h(\psi_k^H \varphi_j^H), \varphi_i^H\right), \tag{2.42}$$

$$F_i \coloneqq (f, \varphi_i^H) - \sum_{k \in \mathcal{N}} a\left(\mathcal{T}_k^h(\psi_k^H f), \varphi_i^H\right), \tag{2.43}$$

with the solution operator $\mathcal{T}_k^h : H^{-1}(\Omega) \to V^h(\omega_k)$ given by (2.35b). In summary, we have established Algorithm 1.

Algorithm 1: General VMM algorithm.

– Choose a macroscale discretization V^H of Ω with nodal basis $\{\psi_i^H\}_{i\in\mathcal{N}}$.

for *each* $i \in \mathcal{N}$ **do**
  – Choose a patch ω_i with a finite-element discretization $V^h(\omega_i)$.

while *stopping criterion not reached* **do**
  for $k \in \mathcal{N}$ **do**
    – Compute $\mathcal{T}_k(\psi_k^H f)$ according to (2.35b).
    for *every nodal basis* $\varphi_i^H \in \{\psi_\nu^H\}_\nu$ *with* $\operatorname{supp}(\varphi_i^H) \cap \operatorname{supp}(\psi_k^H) \neq \emptyset$ **do**
      – Compute $a(\mathcal{T}_k(\psi_k^H f), \varphi_i^H)$ and add contribution to right hand side F.
    for *every* $\varphi_j^H \in \{\psi_\nu^H\}_\nu$ *with* $\operatorname{supp}(\varphi_j^H) \cap \operatorname{supp}(\psi_k^H) \neq \emptyset$ **do**
      – Compute $\mathcal{T}_k(\psi_k^H \varphi_j^H)$ according to (2.35b).
      for *every* $\varphi_i^H \in \{\psi_\nu^H\}_\nu$ *with* $\operatorname{supp}(\varphi_i^H) \cap \operatorname{supp}(\psi_k^H) \neq \emptyset$ **do**
        – Compute $a(\mathcal{T}_k(\psi_k^H \varphi_j^H), \varphi_i^H)$ and add contribution to system matrix A.
  – Solve macroscale problem $AU = F$.
  – Adapt V^H and/or $\{V^h(\omega_i)\}_{i\in\mathcal{N}}$ with a suitable strategy.

Remark 2.15. Special care must be taken when numerically computing (2.42) or (2.43). Even with φ_i^H, $\varphi_j^H \in V^H(\Omega)$ being functions from a coarse scale, $a(\varphi_j^H, \varphi_i^H)$ still exhibits fine-scale fluctuations. This makes a summed quadrature rule necessary in order to evaluate (2.42) and (2.43) correctly.

The above VMM approach allows for a straight forward parallelization of the assembly process, where for a given entry A_{ij} or F_i the fine-scale contribution (2.43) and coarse scale par (2.42) can be computed individually on different threads of execution (in case of thread parallelization).

In light of Remark 2.15 it is best to do *all* assembly work on the fine-scale grid $\mathbb{T}_h(\omega_i)$ associated with the patch-wise fine-scale problems (and especially use it for all numerical quadrature). Such an assembly of the matrix A is essentially a Schur's complement approach, where the resulting linear equation $AU = F$ is much smaller than the full system and can be solved in a single thread of execution (or computing node). If a local fine-scale reconstruction is desired, it is possible to store the linear system of equation defined by (2.13b) in a suitable manner and reconstruct the contribution U_i^f with it.

Remark 2.16. Given the fact that for a each individual microscale region the fine-scale response has to be computed for each nodal basis with non-vanishing support in the microscale region, direct solvers are preferred for the fine-scale problems. For example, for a linear finite-element ansatz it is already necessary to compute nine values for each of the nodal basis functions and one additional inhomogeneity, see Equation (2.43).

Taking both remarks into account, a modified, detailed version of Algorithm 1 can be formulated, see Algorithm 2 on the next page.

Remark 2.17. The maximal resolution τ_{res} for the above (essentially) two-scale VMM is roughly given by

$$\tau_{\text{res}} = 2^{-H_{\max} - h_{\max}}. \tag{2.44}$$

This implies a natural limit in the degree of scale separation that can be captured by a two-scale VMM formulation. For a higher degree of scale separation, it is necessary to introduce intermediate scales, or to use modeling assumptions in order to cut the fine-scale problems down in size.

2.1.6 Related methods

The localization principle employed in the VMM ansatz is closely related to a number of different numerical multiscale methods. Notable examples are the

Algorithm 2: Detailed VMM algorithm with direct solver.

- Start with an initial macroscale discretization $\mathbb{T}_H(\Omega)$ of Ω with nodal basis $\{\psi_i^H\}_{i\in\mathcal{N}}$, patches $\{\omega_i\}_{i\in\mathcal{N}}$ with discretization $\mathbb{T}_h(\omega_i)$.

while *stopping criterion not reached* **do**

  for ψ_k^H, $k\in\mathcal{N}$ **do**

    - Set up $\mathbb{T}_h(\omega_k)$ and assemble A_k^F:
    $\left(A_k^F\right)_{\nu\mu} = (\varphi_\mu^f, \varphi_\nu^f)\ \forall\, \varphi_\mu^f, \varphi_\nu^f \in V^h(\omega_k)$.

    - Compute matrix decomposition of A_k^F: $\quad LR = A_k^F$.

    - Solve for inhomogeneity $\psi_k^H f$: $\left(B_k^F\right)_\nu := (\psi_k^H f, \varphi_\nu^f)$,
    $U_k^F = R^{-1}(L^{-1}B_k^F)$.

    for *every nodal basis* $\varphi_i^H \in \{\psi_\nu^H\}_\nu$ *with* $\operatorname{supp}(\varphi_j^H)\cap\operatorname{supp}(\psi_k^H)\neq\emptyset$ **do**

      - Let Φ_i^F be the interpolation of φ_i^H onto $V^h(\omega_k)$.

      - Add contribution $B_k^F\cdot\Phi_i^F$ to right hand side F_i.

      - Subtract contribution $(A_K^F U_K^F)\cdot\Phi_i^F$ from right hand side F_i.

      - Solve for $\psi_k^H\varphi_i^H$: $\left(B_{k,i}^F\right)_\nu := (\psi_k^H\varphi_i^H, \varphi_\nu^f)$, $U_{k,i}^F = R^{-1}(L^{-1}B_{k,i}^F)$.
      for *every nodal basis* $\varphi_j^H \in \{\psi_\nu^H\}_\nu$ *with*
      $\operatorname{supp}(\varphi_i^H)\cap\operatorname{supp}(\psi_k^H)\neq\emptyset$ **do**

        - Let Φ_j^F be the interpolation of φ_j^H onto $V^h(\omega_k)$.

        - Add contribution $B_{k,i}^F\cdot\Phi_j^F$ to system matrix A_{ji}.

        - Subtract $(A_K^F U_{K,i}^F)\cdot\Phi_j^F$ from right hand side A_{ji}.

  - Solve macroscale problem $AU = F$.

  - Modify $\mathbb{T}_H(\Omega)$ and/or $\{\mathbb{T}_h(\omega_i)\}_{i\in\mathcal{N}}$ with a suitable strategy.

Multiscale Finite-Element Method (MsFEM) introduced by Hou and Wu[59] and Efendiev and Hou[47], the *Mixed Multiscale Methods* by Arbogast[11,12] and Chen and Hou[32], the *Two-Scale* or *Generalized Finite-Element Method* by Matache and Schwab[75,76] and variants thereof.

These models have different modeling assumptions but can be recast in the same variational context (2.13) as the VMM. All have the same fundamental idea in common, to augment or replace an insufficient coarse-scale space $V^H(\Omega)$ by a suitable approximation with (ideally) the same magnitude of degrees of freedom (on the coarse-scale). This can happen by a variety of approaches.

Generalized-basis methods

The underlying idea of *generalized-basis methods* is to replace $V^H(\Omega)$ by a space of *generalized* ansatz functions that are able to cover the essential features of the fine scale—an idea that can be traced back to Babuška and Osborn[14].

In the context of multiscale problems this ansatz was introduced by Hou and Wu[59] and Efendiev and Hou[47,48] as the *Multiscale Finite-Element Method* (MsFEM). The construction of generalized basis functions is done with the help of a "*multiscale mapping*"[48] that connects a coarse-scale (finite-element) function φ^H with its fine-scale counterpart φ^h (see Definition 2.18). The reconstruction of φ^h leads (in some sense) to localized sampling problems. For the sake of simplicity we present a prototypical example (with a straightforward multiscale mapping) which can be regarded as a simplification of the full abstract framework[47,59].

Definition 2.18 (MsFEM ansatz[47,59])**.** Let $\mathbb{T}_H(\Omega)$ be a coarse-scale mesh. Further, let $V^H(\Omega)$ be an associated finite-element space and $V^h(\Omega) \supset V^H(\Omega)$ a fine-scale space with sufficient resolution. For $K \in \mathbb{T}_H(\Omega)$ define

$$V^h(K) = \left\{\varphi \in V^h(\Omega) \,:\, \operatorname{supp}\varphi \in \overline{K}\right\}, \tag{2.45}$$

i. e. $V^h(K)$ shall have homogeneous Dirichlet boundary conditions. Now, for $K \in \mathbb{T}_H(\Omega)$ and $\varphi_i^H \in V^H(\Omega)$ with $\operatorname{supp}(\varphi_i^H) \cap K \neq \emptyset$ define $\varphi_{K,i}^h \in \varphi_i^H + V^h(\Omega)$ by

$$\left(A^\varepsilon \nabla \varphi_{K,i}^h, \nabla \psi^h\right)_K = 0 \quad \forall\, \psi^h \in V^h(K), \tag{2.46}$$

and set

$$\tilde{V}^h(\Omega) \coloneqq \left\{ \sum_K \varphi_{K,i}^h \,:\, \varphi_i^H \in V^H \right\}. \tag{2.47}$$

This is a conforming ansatz space. The MsFEM problem formulation (as proposed by Hou and Wu[59]) reads: Find $u^h \in \tilde{V}^h(\Omega)$ s. t.

$$\left(A^\varepsilon \nabla u^h, \nabla \varphi^h\right) = \left\langle f, \varphi^h \right\rangle \quad \forall\, \varphi^h \in \tilde{V}^h. \tag{2.48}$$

Proposition 2.19 (Equivalence of MsFEM and VMM)**.** The conforming MsFEM formulation (2.48) can be embedded into the VMM ansatz defined by (2.16).

Proof. Let $i_H : \tilde{V}^h(\Omega) \to V^H(\Omega)$ denote the nodal interpolation onto $V^H(\Omega)$ and consider the split exact sequence

$$0 \to V^H(\Omega) \to \tilde{V}^h(\Omega) \to \tilde{V}^h(\Omega)/V^H(\Omega) \to 0 \tag{2.49}$$

defined by the canonical embedding and the projection $\varphi^h \mapsto \varphi^h - i_H \varphi^h$. With $\varphi^f := \varphi^h - i_H \varphi^h$ and $\varphi^H := i_H \varphi^h$ Equation (2.48) takes the form

$$\begin{aligned}\left(A^\varepsilon \nabla(u^H + u^f), \nabla(\varphi^H + \varphi^f)\right) &= \left\langle f, \varphi^H + \varphi^f \right\rangle \\ &\forall\, \varphi^H \in V^H(\Omega),\ \forall\, \varphi^f \in \tilde{V}^h(\Omega)/V^H(\Omega).\end{aligned} \tag{2.50}$$

By linearity this is equivalent to (2.13) and therefore to (2.16). □

Remark 2.20. The ansatz (2.48) suffers from the very same type of resonance error as the bubble-space ansatz (2.19). Similar mitigation techniques can be applied as formulated in Subsection 2.1.3 for the VMM ansatz.

Remark 2.21. The MsFEM approach is by no means limited to a conforming ansatz or the specific multiscale mapping on macro cells $K \in \mathbb{T}_H$ as employed in Definition 2.18. Different boundary conditions have been studied that can lead to better approximation properties than the restriction to linear boundary conditions in (2.46)[47,59].

2.2 Multiscale methods based on homogenization principles

Classical homogenization theory is a mathematical field of research that examines the "approximability" of a series of solutions u^ε of *microscale models*

$$\mathcal{L}^\varepsilon u^\varepsilon = f^\varepsilon. \tag{2.51}$$

Here, $\mathcal{L}^\varepsilon$ and f^ε exhibit fine-scale features associated with a scaling factor $\varepsilon > 0$. Under appropriate assumptions on the sequence of operators $\mathcal{L}^\varepsilon$ and data f^ε (that

will be formulated in the following), an effective solution u^0 of a *homogenized problem*

$$\mathscr{L}^0 u^0 = f^0 \tag{2.52}$$

can be constructed. The individual fine-scale solutions u^ε then converge,

$$u^\varepsilon \to u^0, \quad \text{for } \varepsilon \to 0 \tag{2.53}$$

in a suitable norm.

A vast body of literature is available on general and specific homogenization processes. For a general overview we refer to Bensoussan et al.[22], Sanchez-Palencia[92], Allaire[7], Cioranescu and Donato[34], and Tartar[97].

Remark 2.22. Mathematical homogenization theory has its roots in the field of *asymptotic analysis* in whose context formal expansions

$$u^\varepsilon(x) = u^0(x, \tfrac{x}{\varepsilon}) + \varepsilon u^1(x, \tfrac{x}{\varepsilon}) + \varepsilon^2 u^2(x, \tfrac{x}{\varepsilon}) + \dots, \tag{2.54}$$

$$\mathscr{L}^\varepsilon = \mathscr{L}^0 + \varepsilon \mathscr{L}^1 + \varepsilon^2 \mathscr{L}^2 + \dots \tag{2.55}$$

are used to match components in powers of ε to construct effective models[7,66,92]. The field of asymptotic analysis is a heuristic one; there exist examples in literature that do *not* admit an expansion of the form (2.54) or (2.55)[46]. In contrast, mathematical homogenization theory is rigorous in the sense that precise convergence results (in specific norms and orders of ε) can be proved.

2.2.1 Homogenization of the elliptic model problem

We assume that the model problem (2.7) has the following (local) periodicity property; namely that $\mathscr{L}^\varepsilon$ admits a representation of the form

$$\mathscr{L}^\varepsilon \varphi^\varepsilon = \big(A^\varepsilon(x)\nabla\varphi^\varepsilon, \nabla\,.\big) \tag{2.56}$$

with coefficients $A^\varepsilon \in L^\infty(\Omega)^{d\times d}$ that can be expressed as

$$A^\varepsilon(x) = A(x, \tfrac{x}{\varepsilon}), \text{ with } A \in C^{0,1}\big(\Omega, L^\infty_{\mathrm{per}}(Y)^{d\times d}\big), \tag{2.57}$$

where $Y \coloneqq [0,1]^d$ is the so-called *unit cell*, and $C^{0,1}(\Omega, X)$ denotes the space of Lipschitz continuous functions defined on Ω with values in a Banach space X. The precise definition of $L^\infty_{\mathrm{per}}(Y)$ and other periodic function spaces is given below.

Remark 2.23. The introduction of an additional degree of freedom y in $A(x,y)$ can be justified by a *scale-separation* assumption: Assume that $\varepsilon \ll 1$. Then, a small variation δx of the location x on the microscale is so minuscule on the coarse scale that it can be effectively regarded as being independent. Therefore, it is justified to assume that every observable f with (spatial) multiscale features can be described in terms of two variables, x for the coarse scale and $y = \frac{x}{\varepsilon}$ on the fine scale, hence $f(x,y)$. The variable y can be physically interpreted as a *hidden variable* describing fine-scale oscillations not resolvable on the coarse scale[46].

Remark 2.24. Only the *local* behavior of an observable f with multiscale character is expressed with the finescale variable y—global behavior is already expressed with the macroscopic variable x. This justifies the introduction of a periodicity assumption on the observable with respect to y[66].

We introduce the following function spaces:

Definition 2.25 (Periodic function spaces)**.** Let Y be the unit cell $Y = [0,1]^d$ and fix the notation Y^δ for a scaling of Y with $\delta > 0$, $Y^\delta := \delta Y$. Define an equivalence relation on $\mathbb{R}^d$ by

$$x \equiv y \mod Y^\delta \quad \Longleftrightarrow \quad \frac{x-y}{\delta} \in \mathbb{Z}^d. \tag{2.58}$$

Furthermore, define $x \mod Y^\delta$ to be equal to the uniquely determined $y \in Y^\delta$ with $x \equiv y \mod Y^\delta$ and $y_i < \delta$, $i = 1, \dots, d$.

A function $\varphi \in C^k(\mathbb{R}^d)$ is said to be *δ-periodic* if $\varphi(x + \delta z) = \varphi(x)$ $\forall x \in \mathbb{R}^d$, $z \in \mathbb{Z}^d$. We denote the space of all k-times differentiable δ-periodic functions on $\mathbb{R}^d$ by $C^k_{\text{per}}(Y^\delta)$. This allows to add a notion of periodicity to the usual closure of Lebesgue- and Sobolev-spaces: Let $L^p_{\text{per}}(Y^\delta)$ be the closure of $C^\infty_{\text{per}}(Y^\delta)$ in $L^p(Y^\delta)$ and let $W^{m,p}_{\text{per}}(Y^\delta)$ be the closure of $C^\infty_{\text{per}}(Y^\delta)$ in $W^{m,p}(Y^\delta)$.

For the choice $p = 2$ we use the usual notation H for Hilbert spaces, i. e.

$$H^m_{\text{per}}(Y^\delta) := W^{m,2}_{\text{per}}(Y^\delta). \tag{2.59}$$

Furthermore, let $\tilde{H}^m_{\text{per}}(Y^\delta) := \left\{\varphi \in H^m_{\text{per}}(Y^\delta) : \fint_Y \varphi \, \mathrm{d}x = 0\right\}$ be equipped with the homogeneous scalar product $(\nabla \cdot, \nabla \cdot)$. Here, $\fint$ denotes the mean value,

$$\fint_X \dots = \frac{1}{|X|} \int_X \dots \quad \text{for measurable } X \subset \Omega. \tag{2.60}$$

By abuse of notation, we allow functions $\varphi \in W^{m,p}_{\text{per}}(Y^\delta)$, or $\varphi \in L^p_{\text{per}}(Y^\delta)$, to be interpreted as functions in $W^{m,p}_{\text{loc}}(\mathbb{R}^d)$, or $L^p_{\text{loc}}(\mathbb{R}^d)$ respectively, by setting $\varphi(x) =$

$\varphi(x \mod Y^\delta)$ for $x \notin Y^\delta$. Finally, we fix the notation Y_x^δ for a translation of the rescaled unit cell Y^δ to a new midpoint x,

$$Y_x^\delta = x - (0.5, \ldots, 0.5)^T + Y^\delta. \tag{2.61}$$

All periodic spaces can also be defined for Y_x^δ in a straightforward manner.

Consider the canonical elliptic model problem (2.10) to find $u^\varepsilon \in H_0^1(\Omega)$ s. t.

$$\left(A^\varepsilon \nabla u^\varepsilon, \nabla \varphi\right) = (f, \varphi) \quad \forall \varphi \in H_0^1(\Omega). \tag{2.62}$$

In the remainder of this subsection we present a homogenization result for the elliptic multiscale problem (2.62), as well as rescaling results that will serve as a starting point for the Heterogeneous Multiscale Method (HMM). The homogenized problem corresponding to (2.62) is:

Definition 2.26 (Homogenized Problem)**.** Find $u^0 \in H_0^1(\Omega)$ s. t.

$$\left(A^0 \nabla u^0, \nabla \varphi\right) = (f, \varphi) \quad \forall \varphi \in H_0^1(\Omega), \text{ where} \tag{2.63}$$

$$A_{ij}^0(x) := \int_Y A(x,y)\left[\nabla_y \omega_i(x,y) + \boldsymbol{e}_i\right] \cdot \left[\nabla_y \omega_j(x,y) + \boldsymbol{e}_j\right] \mathrm{d}y, \tag{2.64}$$

and the $\omega_i \in \tilde{H}^1_{\mathrm{per}}(Y)$ are solutions of so-called *cell problems*

$$\int_Y A(x,y)\left[\nabla_y \omega_i(x,y) + \boldsymbol{e}_i\right] \cdot \nabla \varphi \, \mathrm{d}y = 0 \quad \forall \varphi \in \tilde{H}^1_{\mathrm{per}}(Y). \tag{2.65}$$

Remark 2.27. The inherent difficulty to construct an effective equation and solution of (2.62) lies in the pairing of two weak convergence processes: The control of u^ε in $H_0^1(\Omega)$ is as expected: u^ε is uniformly bounded in $H_0^1(\Omega)$ with respect to ε under the uniform ellipticity condition (2.9). This implies that there exists a $u^0 \in H_0^1(\Omega)$ such that up to a subsequence

$$u^\varepsilon \rightharpoonup u^0 \quad \text{weakly in } H_0^1(\Omega). \tag{2.66}$$

However, the coefficients $A^\varepsilon(x)$ do not admit sufficient control. It is generally only possible to show[34] the following:

$$A^\varepsilon(x) \rightharpoonup \bar{A}(x) := \frac{1}{|Y|} \int_Y A(x,y) \, \mathrm{d}y \quad \text{weak-}* \text{ in } L^\infty(\Omega). \tag{2.67}$$

This convergence is not strong enough to take the simultaneous limit of the expression $A^\varepsilon \nabla u^\varepsilon$ [97, Ch. 4], and in fact, it generally holds that $A^\varepsilon \nabla u^\varepsilon \not\to \bar{A} \nabla u^0$.

The principle idea to construct an effective equation for u^0 emerging from (2.62) is to exploit the periodicity assumption in context of a generalized concept of convergence. Established methods are, e. g., the *G-* and Γ*-convergence* by Spagnolo[95,96] and De Giorgi[39], *H-convergence* coined by Murat and Tartar[80,97], or the concept of *two-scale convergence* introduced by Allaire[7–10]. All these approaches have different advantages and are in principle also applicable to non-periodic cases. We concentrate on the concept of two-scale convergence in the following. The classical convergence result reads:

Proposition 2.28 (Classical homogenization result[7,22])**.** Problems (2.62) and (2.63) are well defined and admit a unique solution. In particular, all cell problems (2.65) are well defined and the resulting A^0 defined by (2.64) is elliptic, symmetric and positive definite. The sequence of solutions u^ε converges weakly in $H_0^1(\Omega)$ to u^0,

$$u^\varepsilon \rightharpoonup u^0 \quad \text{in } H_0^1(\Omega). \tag{2.68}$$

Proof. For a detailed proof of the statements we refer to Allaire[7], Bensoussan et al.[22] and Cioranescu and Donato[34]. □

Remark 2.29. It is also possible to derive the effective problem by means of asymptotic expansions and matching in the field of asymptotic analysis[22,46].

The qualitative convergence of the last proposition can be improved by the following corrector result:

Proposition 2.30 (A corrector result by Allaire[7] and Hoang and Schwab[58])**.** u^0 is also part of the solution of the so-called *two-scale homogenized problem*: Find $u^0 \in H_0^1(\Omega)$, $u^1 \in L^2\big(\Omega, \tilde{H}^1_{\mathrm{per}}(Y)\big)$ s. t.

$$\begin{aligned}\int_\Omega \int_Y A(x,y)\big[\nabla_x u^0(x) + \nabla_y u^1(x,y)\big] \cdot \big[\nabla_x \varphi^0(x) + \nabla_y \varphi^1(x,y)\big] \,\mathrm{d}y \,\mathrm{d}x \\ = \int_\Omega f(x)\varphi^0(x)\,\mathrm{d}x \quad \forall \varphi^0 \in H_0^1(\Omega),\, \varphi^1 \in L^2\big(\Omega, \tilde{H}^1_{\mathrm{per}}(Y)\big).\end{aligned} \tag{2.69}$$

It holds that

$$u^1(x,y) = \sum_i \nabla_i u^0(x)\omega_i(x,y), \tag{2.70}$$

where ω_i are the solutions of the cell problems as defined in (2.65). $u^\varepsilon(x)$ *two-scale converges* to $\nabla u^0(x) + \nabla_y u^1(x,y)$, i. e.

$$\lim_{\varepsilon\to 0}\int_\Omega \nabla u^\varepsilon(x)\cdot\psi\Big(x,\frac{x}{\varepsilon}\Big)\,\mathrm{d}x = \int_\Omega\int_Y \big[\nabla u^0(x)+\nabla_y u^1(x,y)\big]\cdot\psi(x,y)\,\mathrm{d}y\,\mathrm{d}x \qquad \forall\psi\in C_c^\infty\big(\Omega, C^\infty_{\mathrm{per}}(Y)\big)^d. \quad (2.71)$$

In addition, if A admits the regularity $A\in C^{0,1}\big(\Omega, C^{0,1}(Y)\big)$ the following quantitative convergence results are available

$$\big\|u^\varepsilon - u^0\big\|_{L^2(\Omega)} = \mathcal{O}\big(\varepsilon\big), \quad (2.72)$$

$$\big\|\nabla u^\varepsilon - \nabla\big(u^0+\varepsilon u^1(x,\tfrac{x}{\varepsilon})\big)\big\|_{L^2(\Omega)^d} = \mathcal{O}\big(\sqrt{\varepsilon}\big). \quad (2.73)$$

Proof. The correspondence of (2.69) to the two-scale problem was established by Allaire[7]. The quantitative convergence rate in terms of ε was formulated by Hoang and Schwab[58]. □

To embed modern numerical HMM formulations into this homogenization context, a rescaled variant formulated by Ohlberger[88] is helpful.

Proposition 2.31 (Reformulation by Ohlberger[88])**.** It is possible to restate the homogenized equation (2.63) with the homogenized coefficient A^0 given by (2.64) as follows: Find $u^0\in H^1_0(\Omega)$ s. t.

$$\int_\Omega \fint_{Y^\delta} A(x,\tfrac{y}{\delta})\big[\nabla_x u^0+\nabla_y\mathscr{R}(u^0)(x,y)\big]\cdot\big[\nabla_x\varphi^0+\nabla_y\mathscr{R}(\varphi^0)(x,y)\big]\,\mathrm{d}y\,\mathrm{d}x = \int_\Omega f(x)\varphi^0(x)\,\mathrm{d}x \quad \forall\varphi^0\in H^1_0(\Omega), \quad (2.74)$$

and the solution $\mathscr{R}(\varphi)(x,y)\in L^2\big(\Omega,\tilde{H}^1_{\mathrm{per}}(Y^\delta)\big)$ of the rescaled cell problem

$$\int_{Y^\delta} A(x,\tfrac{y}{\delta})\big[\nabla_x\varphi(x)+\nabla_y\mathscr{R}(\varphi)(x,y)\big]\cdot\nabla\psi\,\mathrm{d}y = 0 \quad \forall\psi\in\tilde{H}^1_{\mathrm{per}}(Y^\delta). \quad (2.75)$$

Proof. Follows immediately by elementary rescaling[88]. □

2.2.2 Heterogeneous Multiscale Method (HMM)

We present a derivation of the Finite-Element Heterogeneous Multiscale Method (HMM) in the context of mathematical homogenization theory based on work

by Ohlberger[88] and Abdulle[2]. The traditional derivation of the HMM can be found in the original publications by E and Engquist[42,43]. For a given macroscale discretization $\mathbb{T}_H(\Omega)$ a standard finite-element approximation of (2.10) with (for simplicity) Q^1 discretization, a linear mapping and a given quadrature rule $\{(\hat{x}_i, q_i)\}_i$ for a given cell K takes the form: Find $U \in V^H$ s. t.

$$\sum_{K\in\mathbb{T}_H} |K| \sum_i q_i A^\varepsilon(x_i)\nabla U(x_i)\cdot\nabla\varphi(x_i) = \sum_{K\in\mathbb{T}_H} |K| \sum_i q_i f(x_i)\varphi(x_i) \quad \forall \varphi \in V^H. \quad (2.76)$$

The principle idea of the HMM is to use the homogenized equation (2.74) as a starting point for an improved approximation. In order to deal with the unknown hidden variable $\frac{y}{\delta}$ in $A(x, \frac{y}{\delta})$ an approximation $A(x, \frac{y}{\delta}) \approx A^\varepsilon(x)$ is necessary. Then, discretizing (2.74) and (2.75) leads to: Find $U \in V^H$ s. t.

$$\sum_{K\in\mathbb{T}_H} |K| \sum_i q_i \fint_{Y^\delta_{K,i}} A^\varepsilon(x)\nabla\big[U(x) + \mathscr{R}^h_{K,i}(U)(x)\big]\cdot\nabla\big[\varphi(x) + \mathscr{R}^h_{K,i}(\varphi)(x)\big]\mathrm{d}x = \sum_{K\in\mathbb{T}_H} |K| \sum_i q_i f(x_i)\varphi(x_i) \quad \forall \varphi \in V^H, \quad (2.77)$$

with a discretized fine-scale reconstruction $\mathscr{R}^h_{K,i} : H^1(\Omega) \to V^h(Y^\delta_{K,i})$ associated with a finite-element discretization $V^h(Y^\delta_{K,i}) \subset \tilde{H}^1_{\mathrm{per}}(Y^\delta_{K,i})$ and defined by

$$\fint_{Y^\delta_{K,i}} A^\varepsilon(x)\nabla\big[\mathscr{R}^h_{K,i}(\varphi)(x) + \varphi(x)\big]\cdot\nabla\psi(x)\,\mathrm{d}x = 0 \quad \forall \psi \in V^h(Y^\delta_{K,i}). \quad (2.78)$$

Remark 2.32. It is insightful to note that the HMM discretization (2.78) is up to quadrature algebraically equivalent to a direct numerical treatment of (2.63): An approximate assembly of A^0 defined by (2.64) with a finite-element discretization for the cell problems (2.65) in combination with a standard finite-element discretization of the macroscale problem (2.63) yields the same algebraic system.

Remark 2.33. Under the strong periodicity assumption on A—as required by the model problem—and under scale separation, i. e. $\varepsilon \ll 1$, the HMM allows for both, an efficient parallel assembly of the fine-scale reconstructions (2.78) as well as for a *net* saving of total computational cost because δ can be chosen as a small multiple of ε resulting in a small total area with fine-scale reconstruction, see Figure 2.4. However, strong periodicity is a very restrictive assumption and cannot always be assumed to hold in practice.

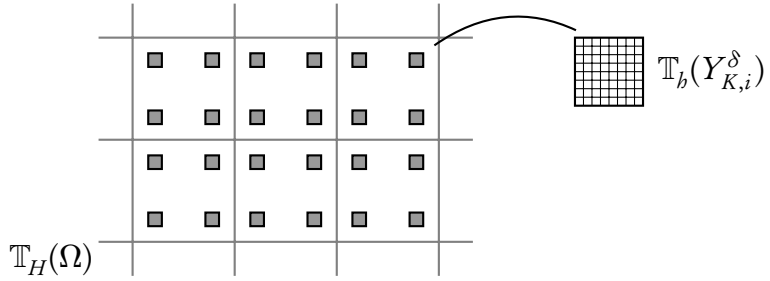


Figure 2.4. Macroscale discretization $\mathbb{T}_H$ with fine-scale sampling regions $Y^\delta_{K,i}$ (in gray) at quadrature points

2.2.3 Implementational aspects

From an implementational point of view, HMM approaches are closely related to VMM. Both rely on a coarse macroscale description that can be treated by traditional numerical methods. Due to the fact that fine-scale responses are modeled by localized problems that couple to the macroscale it is possible to employ the same Schur's complement approach as in the case of VMM. Let $\{\Psi^H_i\}_{i\in\mathcal{N}}$ denote the nodal basis of $V^H(\Omega)$. Then, Equation (2.77) can be written as a linear system of equations, $AU = F$, where the matrix A and right hand side F are defined as

$$A_{\mu\nu} = \sum_K |K| \sum_i q_i \fint_{Y^\delta_{K,i}} A^\varepsilon(x) \nabla\big[\varphi_\nu + \mathscr{R}^h_{K,i}(\varphi_\nu)\big] \cdot \nabla\big[\varphi_\mu + \mathscr{R}^h_{K,i}(\varphi_\mu)\big] \mathrm{d}x, \quad (2.79)$$

$$F_\nu = \sum_K |K| \sum_i q_i f(x_i) \varphi_\nu(x_i), \text{ with} \quad (2.80)$$

$$\fint_{Y^\delta_{K,i}} A^\varepsilon(x) \nabla\big[\mathscr{R}^h_{K,i}(\varphi)(x) + \varphi(x)\big] \cdot \nabla\psi(x)\,\mathrm{d}x = 0 \quad \forall \psi \in V^h(Y^\delta_{K,i}). \quad (2.81)$$

Remark 2.34. In contrast to the VMM approach, the HMM has the advantage that all functions already "live" on the correct scale, i. e. in contrast to VMM (see Remark 2.15) every sampling/quadrature process that encounters fine-scale features is already on $Y^\delta_{K,i}$ with an appropriate discretization $\mathbb{T}_h(Y^\delta_{K,i})$.

Following the same considerations as expressed in Subsection 2.1.5 for the VMM, it is best to compute (2.81) by a direct method. Furthermore, the same considerations with respect to (thread) parallelization apply. In summary, we end up with Algorithm 3 for the HMM (that has remarkable similarities to Algorithm 2).

Algorithm 3: General HMM approach

– Start with an initial macroscale discretization $\mathbb{T}_H(\Omega)$ and a quadrature rule $(\hat{x}_i, q_i)_i$ as well as for each cell K and quadrature point $\hat{x}_i$ a sampling region $Y^\delta_{K,i}$ and discretization $\mathbb{T}_h(Y^\delta_{K,i})$.

while *stopping criterion not reached* **do**

  for *each* $K \in \mathbb{T}_H(\Omega)$ **do**

    for *each quadrature point* $(\hat{x}_i, q_i)$ **do**

      – Set up $\mathbb{T}_h(Y^\delta_{K,i})$ and assemble $A^F_{K,i}$: $\left(A^F_{K,i}\right)_{\nu\mu} = a(\varphi^f_\mu, \varphi^f_\nu)$.

      – Compute matrix decomposition of $A^F_{K,i}$: $\quad LR = A^F_{K,i}$.

      for *every nodal basis* $\varphi^H_\mu \in \{\psi^H_k\}_k$ *with* $\operatorname{supp}(\varphi^H_\mu) \cap Y^\delta_{K,i} \neq \emptyset$ **do**

        – Solve for φ^H_μ:
        $\left(B^F_{K,i,\mu}\right)_k := -\fint_{Y^\delta_{K,i}} A^\varepsilon \nabla \varphi^H_\mu \cdot \nabla \varphi^f_k \,\mathrm{d}x$, $U^F_{K,i,\mu} = R^{-1}(L^{-1}B^F_{k,i,\mu})$

        – Store $U_{K,i,\mu}$.

        – Add contribution $(f, \varphi^H_\mu)_{Y^\delta_{K,i}}$ to right hand side F_μ.

      for *every nodal basis* $\varphi^H_\mu \in \{\psi^H_k\}_k$ *with* $\operatorname{supp}(\varphi^H_\mu) \cap Y^\delta_{K,i} \neq \emptyset$ **do**

        for *every nodal basis* $\varphi^H_\nu \in \{\psi^H_k\}_k$ *with* $\operatorname{supp}(\varphi^H_\nu) \cap Y^\delta_{K,i} \neq \emptyset$ **do**

          – Add contribution $q_i\,|K|/|Y^\delta_{K,i}|\left(A^F_{K,i}\,U^F_{K,\mu}\right) \cdot U^F_{K,\nu}$ to system matrix $A_{\mu\nu}$.

      – Solve macroscale problem $AU = F$.

      – Modify macroscale discretization $\mathbb{T}_H$ and/or microscale discretization $\mathbb{T}_h(Y^\delta_{K,i})$.

2.2.4 On the question of periodicity and scale separation

The underlying regularity assumption of the discussion in the previous subsection, namely the strong periodicity assumption $A^\varepsilon(x) = A(x, \frac{x}{\varepsilon})$ is critical. On the one hand it allows for a thorough theoretical treatment—and a good "saving" of total computational time (see Remark 2.33 and Figure 2.4)—on the other hand strong periodicity conditions are a severe limitation.

Especially the implied exact knowledge of ε is a problem. If the sampling size δ is not an integral multiple of ε the model error $u^\delta - u^\varepsilon$ might be of order 1. This phenomenon is known as *resonance error*[44,59]. Thus, a practical implementation is usually forced to choose local sampling sizes δ quite large compared to the assumed periodicity of the roughness ε.

Furthermore, for the HMM method as defined above, a refinement in the macroscale also forces a (significant) increase of individual sampling regions—and will inevitably lead to drastic redundant sampling when the fine-scale regions start to overlap. This is due to the fact that macro- and microscale refinements are not entirely decoupled from each other.

2.2.5 Generalizations

The basic HMM presented so far has been generalized to a big class of nonlinear and time-dependent problems. We refer to Henning and Ohlberger[52–55] and Abdulle and Nonnenmacher[2–5,83]. In the following, a generalization to an advection-diffusion problem is presented.

Generalized advection-diffusion problems

The model problem (2.62) is a special case of a generalized class of time-dependent advection-diffusion problems for which a (mathematical) homogenization theory is available[74]. We give a brief overview and point to relevant literature.

Definition 2.35 (Multiscale advection-diffusion-reaction-problem[52,53])**.**
Let $I = [0, T]$ be a time interval and $A : I \times Y \to \mathbb{R}^{d\times d}$ be coefficients and $b : I \times Y \to \mathbb{R}^d$ be an advection field with multiscale character. Let $c : I \times Y \to \mathbb{R}$ be a reaction rate. We define the following *advection-diffusion-reaction* problem: Find $u^\varepsilon : I \times \Omega \to \mathbb{R}$ s. t.

$$\partial_t u^\varepsilon - \nabla \cdot \left(A^\varepsilon(t;x)\nabla u^\varepsilon\right) + \frac{1}{\varepsilon} b^\varepsilon(t;x) \cdot \nabla u^\varepsilon + \frac{1}{\varepsilon^2} c^\varepsilon(t;x) u^\varepsilon = 0, \quad (2.82)$$

with the initial condition $u^\varepsilon(0;.) = u_0$ in Ω for a given $u_0 \in L^2(\Omega)$, and the definitions $A^\varepsilon(t;x) = A(t, \frac{x}{\varepsilon})$, $b^\varepsilon(t;x) = b(t, \frac{x}{\varepsilon})$, $c^\varepsilon(t;x) = c(t, \frac{x}{\varepsilon})$.

For a discussion of well-posedness and unique solvability of the advection-diffusion-reaction problem (2.82) we refer to Henning and Ohlberger[52,53].

Assume that A is symmetric and uniformly elliptic in the sense of (2.8) and (2.9) and admits a representation in $W^{1,\infty}\big(I, W^{1,\infty}_{\mathrm{per}}(Y)^d\big)$. Let the advection field b be of class $W^{1,\infty}\big(I, W^{1,\infty}_{\mathrm{per}}(Y)^d\big)$ and divergence-free, i. e.

$$\nabla_x \cdot b = 0, \quad \nabla_y \cdot b = 0, \quad \text{a. e. on } I \times Y, \tag{2.83}$$

and let $c \in W^{1,\infty}\big(I, L^{\infty}_{\mathrm{per}}(Y)^d\big)$. In correspondence to the two-scale homogenized formulation (2.69) the following convergence result holds true:

Proposition 2.36 (Homogenization result for the adv.-diff. prob.[52,53,74])**.** Let $u^0 \in L^2\big(I; H^1(\Omega)\big)$, $u^1 \in L^2\big(I; \tilde{H}^1_{\mathrm{per}}(Y)\big)$ be the solution of the *two-scale homogenized equation with drift*[53]:

$$\begin{aligned} -\int_I (u^0, \partial_t \varphi^0)_{L^2(\Omega)} \mathrm{d}t + \int_I \int_\Omega \int_Y A(t,y)\big[\nabla u^0 + \nabla_y u^1\big] \cdot \big[\nabla \varphi^0 + \nabla_y \varphi^1\big] \mathrm{d}y \, \mathrm{d}x \, \mathrm{d}t \\ + \int_I \int_\Omega \int_Y b(t,y) \cdot \big[\nabla u^0 + \nabla_y u^1\big] \varphi^1(y) \, \mathrm{d}y \, \mathrm{d}x \, \mathrm{d}t \\ - \int_I \int_\Omega \int_Y b(t,y) \cdot \nabla \varphi^0 \, u^1(y) \, \mathrm{d}y \, \mathrm{d}x \, \mathrm{d}t = \big(u_0, \varphi^0(0,.)\big)_{L^2(\Omega)}, \end{aligned} \tag{2.84}$$

with the average $\bar{b}(t) \coloneqq \int_Y (b(t,y) \, \mathrm{d}y$. Then,

$$\begin{aligned} u^\varepsilon &\to u^0 && \text{two-scale with drift,} \\ \nabla u^\varepsilon &\to \nabla_x u^0 + \nabla_y u^1 && \text{two-scale with drift.} \end{aligned} \tag{2.85}$$

Two-scale convergence with drift is defined as[10,52,74]:

Definition 2.37 (Two-scale convergence with drift[10,52,74])**.** Let

$$B(t) \coloneqq \int_0^t \int_Y b(s,y) \, \mathrm{d}y \, \mathrm{d}s. \tag{2.86}$$

In analogy of (2.71) a sequence $(v^\varepsilon) \in L^2(I \times Y)$ *two-scale converges with drift* $B(t)$ to a limit $v^0 \in L^2(I \times \Omega \times Y)$ if and only if

$$\begin{aligned} \lim_{\varepsilon \to 0} \int_I \int_\Omega v^\varepsilon(t,x) \cdot \psi\Big(t, x - \frac{B(t)}{\varepsilon}, \frac{x}{\varepsilon}\Big) \mathrm{d}x \, \mathrm{d}t \\ = \int_I \int_\Omega \int_Y v^0(t,x,y) \psi(t,x,y) \, \mathrm{d}y \, \mathrm{d}x \, \mathrm{d}t \\ \forall \psi \in C^\infty\big(I, C^\infty_c\big(\Omega, C^\infty_{\mathrm{per}}(Y)\big)^d\big). \end{aligned} \tag{2.87}$$

Corollary 2.38. u^0 is the solution of the homogenized equation

$$\big(\partial_t u^0, \varphi\big) + \big(A^0 \nabla u^0, \nabla \varphi\big) = 0 \quad \forall\, \varphi \in H_0^1(\Omega), \tag{2.88a}$$

$$u^0(0, .) = u_0 \tag{2.88b}$$

with

$$A^0_{ij}(t) \coloneqq \int_Y A(t,y)\big[\nabla_y \omega_i(t,y) + \boldsymbol{e}_i\big] \cdot \big[\nabla_y \omega_j(t,y) + \boldsymbol{e}_j\big] \mathrm{d}y, \tag{2.89}$$

and the $\omega_i(t,.) \in \tilde{H}^1_{\mathrm{per}}(Y)$ are solutions of the cell problems

$$\int_Y A(t,y)\big[\nabla_y \omega_i(t,y) + \boldsymbol{e}_i\big] \cdot \nabla_y \varphi \,\mathrm{d}y + \int_Y b(t,y) \cdot \big[\nabla_y \omega_i(t,y) + \boldsymbol{e}_i\big] \varphi \,\mathrm{d}y = 0 \quad \forall \varphi \in \tilde{H}^1_{\mathrm{per}}(Y). \tag{2.90}$$

Proof of Proposition 2.36 and Corollary 2.38. A proof of a more general variant of the statements is given by Henning[52,53]. The proof is based on the general framework of two-scale convergence and the notion of "two-scale convergence with drift" developed by Allaire[7,10] and Marušić-Paloka and Piatnitski[74]. □

Now, an HMM can be derived by means of a *Rothe method*. First, Equation (2.88) is discretized in time with a suitable time-stepping scheme. After that, (2.88) and (2.90) are simultaneously discretized in space with the HMM discretization developed in Subsection 2.2.2.

In analogy of (2.77), the discretized version of the diffusive term $\big(A^0 \nabla u^0, \varphi^0\big)$ takes the form:

$$\sum_{K \in \mathbb{T}_H} |K| \sum_i q_i \fint_{Y^\delta_{K,i}} A^\varepsilon(x) \nabla\big[U(x) + \mathscr{R}^h_{K,i}(U)(x)\big] \cdot \nabla\big[\varphi(x) + \mathscr{R}^h_{K,i}(\varphi)(x)\big] \mathrm{d}x, \tag{2.91}$$

where a discretized fine-scale reconstruction

$$\mathscr{R}^h_{K,i} : H^1(\Omega) \to V^h(Y^\delta_{K,i}) \tag{2.92}$$

associated with a finite-element discretization $V^h(Y^\delta_{K,i}) \subset \tilde{H}^1_{\mathrm{per}}(Y^\delta_{K,i})$ is defined by:

$$\fint_{Y^\delta_{K,i}} A^\varepsilon(x) \nabla\big[\mathscr{R}^h_{K,i}(\varphi)(x) + \varphi(x)\big] \cdot \nabla \psi(x) \,\mathrm{d}x + \fint_{Y^\delta_{K,i}} b^\varepsilon(t,y) \cdot \big[\nabla_y \omega_i(t,y) + \boldsymbol{e}_i\big] \varphi \,\mathrm{d}y = 0 \quad \forall \psi \in V^h(Y^\delta_{K,i}). \tag{2.93}$$

2.3 Multiscale methods based on physical upscaling principles

In contrast to the mathematical derivation of multiscale schemes that was presented in the last two subsections, the emphasis within *upscaling* and *computational homogenization* techniques lies on a *physically sound* derivation of multiscale descriptions. In general, this involves the formulation of (possibly independent) macroscale and microscale models coupled by physical *first principles*[49] (see Figure 2.1 on page 8).

Generally, *scale separation* is assumed in the sense that it is possible to describe the fine scale in terms of a *hidden variable* ("hidden" because not visible on the macroscale, see Remark 2.23 on page 25). This justifies the introduction of localized fine-scale problems for each macroscopic point $x \in \Omega$ —the so-called *representative volume element* (RVE). Similarly to cell problems in mathematical homogenization theory, these fine-scale problems are localized by artificial boundary values. However, in contrast to mathematical homogenization the justification for the introduction of localized fine-scale problems is fundamentally different. The former proves rigorous convergence results under a strong periodicity assumption (or similar a priori assumptions) in a mathematical sense. The latter justifies the introduction purely by a scale-separation hypothesis in combination with upscaling principles: The RVE shall on average be "*structurally typical*"[56] of the material, and the sampling region large enough so that the error introduced by artificial boundary conditions is controlled[56].

2.3.1 An example from linear elasticity

To exemplify the approach we give a modeling example from linear elasticity. The following discussion is based on the outline and methodology given by Miehe and Bayreuther[17,78].

Given a heterogeneous material in $\Omega \in \mathbb{R}^d$ described by a macroscopic deformation

$$\bar{\boldsymbol{u}} : \Omega \to \mathbb{R}^d, \tag{2.94}$$

assume that the material follows the *constitutive* equation of linear elasticity, $\bar{\boldsymbol{u}}$ is the solution of an optimization problem defined over the space of all *admissible* deformations:

$$\inf_{\bar{\boldsymbol{u}}\ \text{adm.}} \frac{1}{2}\big(\bar{\boldsymbol{\sigma}}, \bar{\boldsymbol{\varepsilon}}\big)_{L^2(\Omega)^{d\times d}} - \big(\bar{\boldsymbol{f}}, \bar{\boldsymbol{u}}\big)_{L^2(\Omega)^d}, \tag{2.95}$$

with the *strain tensor*

$$\bar{\boldsymbol{\varepsilon}} = \frac{1}{2}\left(\nabla \bar{\boldsymbol{u}} + \nabla \bar{\boldsymbol{u}}^T\right) \in L^2(\Omega)^{d\times d} \tag{2.96}$$

and the *Cauchy-stress tensor* $\bar{\boldsymbol{\sigma}} \in L^2(\Omega)^{d\times d}$. Equivalently, the optimization problem reads in weak form:

$$\begin{aligned} \left(\bar{\boldsymbol{\sigma}}, \nabla\varphi\right)_{L^2(\Omega)^{d\times d}} &= \left(\bar{\boldsymbol{f}}, \varphi\right)_{L^2(\Omega)^d} && \forall \varphi \in H^1(\Omega)^d, \\ \boldsymbol{n}\cdot\bar{\boldsymbol{\sigma}} &= \tau && \text{on } \partial\Omega_\tau, \\ \bar{\boldsymbol{u}} &= \bar{\boldsymbol{u}}_d && \text{on } \partial\Omega_d, \end{aligned} \tag{2.97}$$

Remark 2.39. For a material without microstructure a relation of the form

$$\bar{\boldsymbol{\sigma}} = \overline{C} : \bar{\boldsymbol{\varepsilon}}, \tag{2.98}$$

with the fourth order elasticity tensor $\overline{C} : \Omega \to \mathbb{R}^{d\times d\times d\times d}$ is known and the combined system of equations (2.97) and (2.98) is closed.

We assume that the material is constituted by a fine scale (not visible on the macroscale) so that a closed form of the macroscopic elasticity tensor $\overline{C}$ is not available. The fine scale shall also be governed by a model of linear elasticity: For every macro-point $x \in \Omega$ there exists a microscopic displacement field $u^F(x,y) : \mathrm{RVE}(x) \to \mathbb{R}^d$

$$\left(\boldsymbol{\sigma}^F, \nabla\varphi\right)_{L^2(\mathrm{RVE}(x))^{d\times d}} = 0 \quad \forall \varphi \in H^1(\mathrm{RVE}(x))^d, \tag{2.99}$$

$$\boldsymbol{\sigma}^F = C^F : \boldsymbol{\varepsilon}^F, \tag{2.100}$$

with a known, microscopic fourth order elasticity tensor C^F. In this form, the microscopic model (2.99) is still ill-defined due to unknown boundary conditions, and consequently needs a *closure assumption* that will define the appropriate reconstruction and compression operations.

A famous principle dating back to Germain[50] is the *principle of virtual power*. It states that given

> *a system S* [...] *in equilibrium with respect to a given Galilean frame; then in any virtual motion, the virtual power of all the "internal forces" and "external forces" acting on S is null*[50].

It was introduced by Hill[56] and Mandel in the context of multiscale modeling, where it is now known as the *Hill-Mandel homogeneity condition* that enforces

that the virtual power of virtual motions of macro- and microscale with respect to a simultaneous displacement on macro- and microscale is equal[57].

With respect to the given elasticity problem this implies

$$\bar{\boldsymbol{\sigma}} : \bar{\boldsymbol{\varepsilon}}(x) = \fint_{\mathrm{RVE}(x)} \boldsymbol{\sigma}^F : \boldsymbol{\varepsilon}^F(y)\,\mathrm{d}y, \tag{2.101}$$

where the binary operator : denotes the full contraction of both tensors. From the defining properties of a RVE the following averaging principle is directly deducible[17]:

$$\bar{\boldsymbol{\sigma}}(x) = \fint_{\mathrm{RVE}(x)} \boldsymbol{\sigma}^F(y)\,\mathrm{d}y. \tag{2.102}$$

Proof. In the state of an equilibrium, the total energy of a displacement on macro- and microscale is $\bar{\boldsymbol{\sigma}} : \bar{\boldsymbol{\varepsilon}}$ and $\fint_{\mathrm{RVE}(x)} \boldsymbol{\sigma}^F : \boldsymbol{\varepsilon}^F \mathrm{d}y$, respectively. Hence, by virtue of the Hill-Mandel condition[17] Equation (2.101) holds true. Furthermore, it must hold that for any displacement $\left(\delta\boldsymbol{\varepsilon}^F, \delta\bar{\boldsymbol{\varepsilon}} = \fint \boldsymbol{\varepsilon}^F \mathrm{d}y\right)$ the change on total energy on macro- and microscale is equal. This must especially hold true for a homogeneous displacement $\delta\bar{\boldsymbol{\varepsilon}} = \delta\boldsymbol{\varepsilon}^F = \boldsymbol{e}_{ij}$, with a unit tensor $\boldsymbol{e}_{ij}$. Consequently, Equation (2.102) is fulfilled. □

A sufficient condition for (2.101) to be fulfilled is to assume that the difference between the microscopic displacement $\boldsymbol{u}^F(x,y)$ on RVE(x) and the linear extension $\nabla\bar{\boldsymbol{u}}(x)\cdot\boldsymbol{y}$ of the macroscopic displacement is periodic[17].

$$\boldsymbol{u}^F(x,y) - \nabla_x\bar{\boldsymbol{u}}(x)\cdot\boldsymbol{y} \in H^1_{\mathrm{per}}\big(\mathrm{RVE}(x)\big). \tag{2.103}$$

Remark 2.40. This coupling is similar to a *first-order* homogenization approach known from computational homogenization for general nonlinear mechanical problems[49].

Remark 2.41. Other choices of boundary conditions in order to fulfill the Hill-Mandel condition (2.101) are possible. For a detailed overview we refer to a discussion given by Miehe and Bayreuther[17,78].

In summary, choosing periodic boundary conditions, we have established the following *compression* and *reconstruction process* (in the sense of Figure 2.1 on page 8):

- The macroscale-stress is given by

$$\bar{\boldsymbol{\sigma}}(x) = \fint_{\mathrm{RVE}(x)} \boldsymbol{\sigma}^F(y)\,\mathrm{d}y. \tag{2.104}$$

– For a given stress $\bar{\boldsymbol{\sigma}}$ on the macroscale the microscale displacement $\boldsymbol{\sigma}^F$ is given by the solution of (2.99) with boundary conditions

$$\boldsymbol{u}^F \in H^1_{\text{per}}\big(\text{RVE}(x)\big). \tag{2.105}$$

In order to convert this computational homogenization approach to a computable numerical scheme, both, the macroscale (2.97) and microscale (2.99) equations are discretized. Similarly to the HMM, a corresponding micro-problem is solved on $\text{RVE}(x)$ for every quadrature point x. The resulting scheme is known as *FE*^{2} *computational homogenization*[102].

2.3.2 Relation to HMM and MsFEM

Computational homogenization schemes are closely related to HMM (cf. Section 2.2) and MsFEM (cf. Subsection 2.2.5) approaches. For the choice of structurally identical macro- and microscale models all three methods are in fact equivalent up to quadrature and discretization: The RVE of the computational homogenization schemes either corresponds to local cell, or sampling problems in case of HMM. Or, the microscale problems defined on $\text{RVE}(x)$ together with the reconstruction and compression operations correspond to the multiscale mapping defined in context of MsFEM.

We exemplify this claim by proving the correspondence of a scalar version of the computational homogenization scheme (2.97) and (2.99) with periodic boundary conditions (2.103) to the HMM formulation that was outlined in Subsection 2.2.2.

Definition 2.42 (Scalar computational homogenization problem)**.** A scalar pendant of the above computational homogenization scheme (2.97) and (2.99) with homogeneous Dirichlet boundary conditions reads

$$(\bar{\boldsymbol{\sigma}}, \nabla\varphi)_{L^2(\Omega)^3} = (\bar{f}, \varphi) \quad \forall\, \varphi \in H^1_0(\Omega), \tag{2.106a}$$

$$(A^\varepsilon \nabla u^F, \nabla\varphi)_{L^2(\Omega)^3} = (f^F, \varphi) \quad \forall\, \varphi \in H^1_{\text{per}}(\text{RVE}(x)), \tag{2.106b}$$

with A^ε denoting, for the sake of consistency with the model problem (2.10), the known relation $\boldsymbol{\sigma}^F = A^\varepsilon \nabla u^F$ on the fine scale.

Let A^0 denote the unknown (and to be determined) pendant of the elasticity

tensor on the macroscale. Then, it follows by virtue of (2.102) that

$$\begin{aligned} A^0(x)\nabla\bar{u} &= \fint_{\mathrm{RVE}(x)} A^\varepsilon(y)\nabla u^F(x,y)\,\mathrm{d}y \\ &= \fint_{\mathrm{RVE}(x)} A^\varepsilon(y)\big[\nabla_x\bar{u}+\nabla_y\hat{u}^F(x,y)\big]\mathrm{d}y, \end{aligned} \tag{2.107}$$

with the definition $\hat{u}^F(x,y) \coloneqq u^F(x,y) - \nabla_x\bar{u}\cdot y$ according to (2.103). Furthermore, expressing (2.99) with $\hat{u}^F(x,y)$ leads to

$$\fint_{\mathrm{RVE}(x)} A^\varepsilon(y)\big[\nabla_y\hat{u}^F(x,y)+\nabla_x\bar{u}\big]\cdot\nabla_y\hat{\varphi}^F\,\mathrm{d}y = 0 \quad \forall\hat{\varphi}^F\in H^1_{\mathrm{per}}\big(\mathrm{RVE}(x)\big). \tag{2.108}$$

This is exactly the cell problem (2.62). Finally, observe that with (2.107) and (2.108) it follows:

$$\begin{aligned} &\big(A^0(x)\nabla\bar{u},\nabla\bar{\varphi}\big)_{L^2(\Omega)^d} \\ &\qquad= \fint_{\mathrm{RVE}(x)} A^\varepsilon(y)\big[\nabla_x\bar{u}+\nabla_y\hat{u}^F(x,y)\big]\cdot\nabla_x\bar{\varphi}\,\mathrm{d}y\,\mathrm{d}x \\ &\qquad= \fint_{\mathrm{RVE}(x)} A^\varepsilon(y)\big[\nabla_x\bar{u}+\nabla_y\hat{u}^F(x,y)\big]\cdot\big[\nabla_x\bar{\varphi}+\nabla_y\hat{\varphi}^F(x,y)\big]\mathrm{d}y\,\mathrm{d}x, \end{aligned} \tag{2.109}$$

which is equivalent to Equation (2.74). In summary, we have established the following Proposition:

Proposition 2.43 (Equivalence of computational homogenization and HMM)**.** The scalar simplification (2.106) of the original computational homogenization scheme (2.97) and (2.99) is equivalent to the homogenized problem (2.62); and therefore after discretization (and in combination with the discussion given in Remark 2.32) equivalent to the HMM method (2.77).

2.3.3 Averaging schemes

An even simpler upscaling technique that determines effective parameters solely by means of averaging on the finescale—without solving finescale problems, and without a coupling "reconstruction" process—is very popular in groundwater flow simulations (and related fields of research). Here, the heterogeneous coefficient is given by a fluctuating permeability $K^\varepsilon(x)$:

$$A^\varepsilon(x) = K^\varepsilon(x)\,\mathrm{Id}, \tag{2.110}$$

with the identity matrix Id. The general idea is to generate piecewise constant, effective coefficients,

$$A^\delta(x) = K^\delta(x)\mathrm{Id} : \Omega \to \mathbb{R}^{d\times d} \tag{2.111}$$

that are (in contrast to A^ε) resolvable by a coarse discretization and reproduce the effective macroscale behavior. More precisely, one could seek an effective coefficient A^δ such that the solution $u^\delta \in H_0^1(\Omega)$ of the variational problem

$$\left(A^\delta \nabla u^\delta, \nabla\varphi\right) = \left(f,\varphi\right) \quad \forall\,\varphi \in H_0^1(\Omega) \tag{2.112}$$

is a good approximation of u^ε in the $L^2(\Omega)$-norm (but not necessarily in the $H^1(\Omega)$-norm). Very simple averaging schemes involve the geometric or harmonic mean value taken over small sampling regions $\mathrm{RVE}(x)$ (in the notation of the previous section):

$$\log K^\delta(x) = \fint_{\mathrm{RVE}(x)} \log K^\varepsilon(x)\,\mathrm{d}x. \tag{2.113}$$

or the harmonic mean value

$$\log K^\delta(x) = \fint_{\mathrm{RVE}(x)} K^\delta(x)^{-1} = \fint_{\mathrm{RVE}(x)} K^\varepsilon(x)^{-1}\,\mathrm{d}x. \tag{2.114}$$

Remark 2.44. It is a well known fact that the arithmetic average

$$K^\delta(x) = \fint_{\mathrm{RVE}(x)} K^\varepsilon(x)\,\mathrm{d}x, \tag{2.115}$$

is not appropriate in practice. In case of classical homogenization theory it can be shown, e. g., that in 1D the homogenization limit A^0 is given by the harmonic mean value and that in higher space dimension the correctors are non-vanishing[13,34], see the summary of classical homogenization results given in Section 2.2.1.

It was shown by Cardwell and Parsons[31] that for a large class of randomly distributed permeabilities a good effective value is expected to lie between the arithmetic and harmonic mean value,

$$\left(\fint_{Y_K^\delta} \frac{1}{K^\varepsilon}\,\mathrm{d}y\right)^{-1} \le K^0 \le \fint_{Y_K^\delta} K^\varepsilon(y)\,\mathrm{d}y. \tag{2.116}$$

The geometric mean value was further examined extensively for log-normally distributed permeabilities, i. e. permeabilities where the logarithm of the permeability values is pointwise normally distributed with a specific correlation structure. First numerical tests were performed by Warren and Price[100].

Remark 2.45. In context of linear elasticity it can also be shown that effective coefficients can be expected to lie between the harmonic and arithmetic mean value, see Dederichs and Zeller[40] and Kröner[65].

Simple averaging strategies such as (2.113) and (2.114) can only lead to satisfying approximation results under the assumption of strong statistical properties to hold true, e. g. the underlying microstructure must be *isotropic*, i. e. it must have the same characteristics for each spatial dimension. A multitude of refined, heuristic averaging schemes have been proposed to resolve those shortcomings of simple averages in case of more complex random permeabilities. This involves cases of layered permeabilities, and with permeabilities with specific correlation properties, possibly different for individual spatial dimensions. Usually, a sophisticated *heuristic* postprocessing is used that involves the adaptation of simple (spatial) averages with the help of local statistical properties. For a detailed description of such a method as well as a general overview we refer to Li et al.[71]

Remark 2.46. Nevertheless, the simple averaging strategies (2.113) and (2.114) will also be considered as computationally inexpensive reconstruction strategies in the abstract framework introduced in Chapter 3. This is justifiable, because *the quality* of the resulting method is controlled by a posteriori techniques in an adaptive algorithm tuned for a quantity of interest. In this case the goal is to achieve a significant computational saving (of several orders of magnitude) by means of adaptive discretization and model control, see Chapters 5 and 6.

3 An abstract multiscale scheme for model adaptation

As evidenced in the previous chapter, multiscale schemes have many intrinsic refinement and quality parameters. For the methods presented in Chapter 2 there are choices for macroscale discretization (denoted by H), microscale discretizations (denoted by h) and *modeling parameters*. The latter involves the location of sampling regions or cell problems, the choice of sampling size and placement of artificial boundary conditions, and orthogonality relations (as discussed in subsection 2.1.4). Such model parameters (that are not related to macroscale or microscale discretization parameters) shall subsequently be denoted by δ.

A good choice for all discretization and model parameters is necessary in order to have an efficient multiscale method. However, in the methods presented so far the choice of modeling parameters is directly coupled to a macroscale discretization. For example, by prescribing a sampling or cell problem in each quadrature point (in case of HMM and computational homogenization), or by a cell-wise fine-scale reconstruction (in case of VMM). This leads to the effect that refinement in the macro region induces a simultaneous adaptation of the effective model.

There are situations in which such a strict coupling of macroscale resolution and model parameters is not ideal. Consider, e. g., the case of a singularity only present on the macroscale (for example a corner singularity). This makes a local grid refinement on the macroscale necessary but has no influence on the quality of the sampling process on the microscale. A local refinement with fixed sampling in each quadrature point, however, leads to severe "oversampling" due to increasingly many, closely located sampling regions. But, for closely located sampling regions effective parameters do not vary much.

Further, all approaches examined in Chapter 2 have an underlying scale separation assumption: in case of HMM the assumption $\varepsilon \ll 1$ (and a periodicity assumption), and in case of VMM the assumption that the influence of artificial boundary conditions is small (which implies smallness of microscale features). Such a scale-separation assumption might not always be fulfilled—especially

when the *modeling error*, i. e., $u^\varepsilon - u^0$ (in the notation of Section 2.2), is not homogeneously small for all $x \in \Omega$. In such cases a *model adaptation* strategy is desirable.

In light of this discussion, this chapter presents an *abstract multiscale scheme for model adaptation* that explicitly decouples all discretization and modeling parameters. It is suitable as a *model framework* for the development of discretization and model adaptation approaches. Furthermore, it can be regarded as a generalization of the HMM and shares some ideas with the model adaptation approaches by Oden and Vemaganti[84–87,91] and Braack and Ern[26] (see Section 5.1 for details). The novelty lies in the explicit decoupling of the sampling processes from the macroscopic discretization.

3.1 On the choice of the underlying multiscale formulation

A fundamental question for formulating a multiscale framework is on which of the mathematical multiscale formulations, that were introduced in Chapter 2, it should be based on. This choice has to be done while keeping in mind that the goal of the discussion is to use the framework for goal oriented a posteriori error estimation and model adaptation (cf. Chapter 4). Here, the variational ansatz pursued in the VMM formulation suffers from an intrinsic problem: Consider the DWR ansatz[19–21] (that will be introduced in detail in Chapter 4): Let $j \in H^{-1}(\Omega)$ be a functional. For the quantity of interest $\langle j, u^\varepsilon \rangle$ define the dual problem to find $z \in \mathcal{V}$

$$\left(A^\varepsilon \nabla \varphi, \nabla z\right) = \langle j, \varphi \rangle \quad \forall\, \varphi \in \mathcal{V}. \tag{3.1}$$

By using the split $z = z^H + z^f$ given by $\mathcal{V} = V^H \oplus \mathcal{V}^f$ we derive the following error identity for a VMM approximation $U = U^H + U^f \in V^H \oplus V^f$:

$$\begin{aligned} \langle j, u^\varepsilon - U \rangle &= (f, z) - (A^\varepsilon \nabla U, \nabla z) \\ &= (f, z) - (A^\varepsilon \nabla U^f, \nabla z) - (A^\varepsilon \nabla U^H, \nabla z) \\ &= (f, z^f) - (A^\varepsilon \nabla U^f, \nabla z^f) - (A^\varepsilon \nabla U^H, \nabla z^f). \end{aligned} \tag{3.2}$$

The last equation holds true due to the variational split (2.13) of U:

$$(f, z^H) - (A^\varepsilon \nabla U^f, \nabla z^H) - (A^\varepsilon \nabla U^H, \nabla z^H) = 0. \tag{3.3}$$

Thus, the error identity reduces entirely to a *residual on the fine-scale space* $\mathcal{V}^f$:

$$\langle j, u^\varepsilon - U\rangle = (f, z^f) - (A^\varepsilon \nabla U^f, \nabla z^f) - (A^\varepsilon \nabla U^H, \nabla z^f). \tag{3.4}$$

No information about the "quality" of the macroscale discretization can be reconstructed. Given the modeling ansatz of the VMM approach this comes at no surprise—after all, the underlying idea is to augment the space V^H by an "orthogonal" correction. But, from an a posteriori adaptation perspective, this is unfortunate as it possibly forces resolving macroscale effects with fine-scale reconstructions. The HMM formulation on the other hand does not suffer from this kind of defect (see Chapter 4).

Another point that has to be taken into consideration is the question in which form *model adaptation* is actually possible within a given approach. We note that the VMM ansatz (2.13) uses the same equations for coarse and fine scale—the *modeling aspect* lies solely in the choice of split $V^H \oplus \mathcal{V}^f$ and approximation of $\mathcal{V}^f$. More precisely, model adaptivity in context of VMM formulations is reduced to the choice of approximations of the fine-scale space (2.17):

$$V^f = \sum_i V_i^f \subset \mathcal{V}^f. \tag{3.5}$$

This boils down to the sole question of choosing sufficient *discretization parameters*. In contrast, the HMM approach of constructing an effective operator $\mathcal{L}^0$ (2.52) is not only closer to the physical approach of upscaling and effective parameters (see Section 2.3), but also allows for model adaptivity by *combining different upscaling principles*, e. g., by using effective parameters computed by HMM cell problems in combination with simple averaging strategies (see Sections 2.2.2 and 2.3.3). Thus, we will base the framework on an abstraction of those approaches.

Remark 3.1. An extensive study for goal-oriented a posteriori control of discretization parameters (involving microscale discretization and patch size) for the VMM ansatz is given by Larson and Målqvist [67–69,81].

3.2 An abstract multiscale scheme

The discussion in this Subsection follows in parts a publication by the author[72]. Consider the model problem (2.10) to find $u^\varepsilon \in H_0^1(\Omega)$ s. t.

$$\left(A^\varepsilon \nabla u^\varepsilon, \nabla \varphi\right) = (f, \varphi) \quad \forall \varphi \in H_0^1(\Omega), \tag{3.6}$$

with coefficients $A^\varepsilon \in L^\infty(\Omega)^{d\times d}$ that employ multiscale behavior in the sense of Definition 2.2.

Remark 3.2. For the formulation of the method, we will not make any further assumptions on the coefficient A^ε other than (2.8) and (2.9) that guarantees well-posedness of (3.6). For the a priori convergence results, however, mild regularity assumptions have to be made.

3.2.1 Effective problem

In order to be able to adapt *model parameters* separately from the discretization it is necessary to introduce an abstract notion of a model. In context of the mathematical homogenization theory for the elliptic problem (2.10) a natural starting point is the homogenized equation (2.62): Find $u^0 \in H_0^1(\Omega)$ s. t.

$$\left(A^0 \nabla u^0, \nabla \varphi\right) = \left(f, \varphi\right) \quad \forall\, \varphi \in H_0^1(\Omega). \tag{3.7}$$

In case of a heterogeneous problem, the coefficient $A^0 \in L^\infty(\Omega)^{d\times d}$ is a function depending on x. This space dependency has to be discretized for a numerical scheme. A possible choice (as employed by the HMM) is to start with a finite-element discretization of (3.7) and define a sampling problem for every quadrature point in Ω for which $A^0(x)$ has to be evaluated.

However, this is exactly the type of coupling between (macroscale) discretization and sampling process that we try to avoid. Consequently, in order to decouple the sampling process from the coarse-scale discretization, it is necessary to abstract this choice and incorporate it into an *effective model*.

Definition 3.3 (Effective model)**.** Define an *effective model* to be a pair

$$\left(\mathbb{T}_\delta(\Omega), A^\delta\right) \tag{3.8}$$

consisting of a discretization $\mathbb{T}_\delta(\Omega)$ of Ω called a *sampling mesh* together with a function

$$\bar{A}^\delta : \mathbb{T}_\delta(\Omega) \to \mathbb{R}^{d\times d} \tag{3.9}$$

with region-wise constant values (see Figure 3.1). We assume that $\bar{A}^\delta$ fulfills (2.8) and (2.9). The discretization parameter δ shall denote the typical length scale of $\mathbb{T}_\delta(\Omega)$.

The effective model $\left(\mathbb{T}_\delta(\Omega), \bar{A}^\delta\right)$ can be constructed by different means. In spirit of the classical HMM, it is possible to define a reconstruction process by solving a local cell problem on a sampling region:

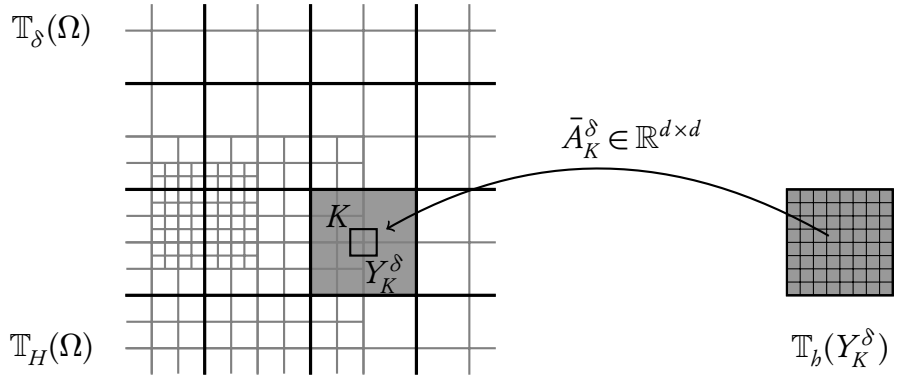


Figure 3.1. The computational domain Ω together with the *sampling mesh* $\Pi_\delta(\Omega)$ consisting of sampling regions $K \in \Pi_\delta$ that are in turn discretized by a fine-scale mesh $\Pi_h(K)$. The coarse mesh $\Pi_H(\Omega)$ used for the final finite-element discretization is a refinement of the sampling mesh Π_δ.

Definition 3.4 (Homogenization sampling strategy)**.** For $K \in \mathbb{T}_\delta(\Omega)$ let Y_K^δ be a sampling region associated with a sampling cell K, where every Y_K^δ is sought to be a simple translation and rescaling of the unit cell Y. Define $\bar{A}^\delta(K)$ as

$$\bar{A}_{ij}^\delta(K) \coloneqq \fint_{Y_K^\delta} A^\varepsilon(x)\big(\nabla_x \omega_i(x) + \boldsymbol{e}_i\big)\cdot\big(\nabla_x \omega_j(x) + \boldsymbol{e}_j\big)\,\mathrm{d}x, \tag{3.10}$$

where the $\omega_i \in \tilde{H}^1_{\mathrm{per}}(Y_K^\delta)$ are solutions of

$$\int_{Y_K^\delta} A^\varepsilon(x)\big(\nabla_x \omega_i(x) + \boldsymbol{e}_i\big)\cdot\nabla\varphi = 0 \quad \forall \varphi \in \tilde{H}^1_{\mathrm{per}}(Y_K^\delta). \tag{3.11}$$

Other choices of sampling strategies are possible. As discussed in Section 2.3.3 simple averaging schemes can be defined in order to avoid solving local sampling problems. Transferred to the heterogeneous coefficient this leads to an averaging sampling strategy:

Definition 3.5 (Averaging sampling strategies)**.** Let Y_K^δ, $K \in \mathbb{T}_\delta(\Omega)$ be the set of sampling regions as defined in Definition 3.4. A simple averaging process is given by the arithmetic mean value

$$\bar{A}_{ij}^\delta(K) \coloneqq \fint_{Y_K^\delta} A_{ij}^\varepsilon(y)\,\mathrm{d}y, \tag{3.12}$$

by the *geometric mean value*

$$\log \bar{A}^{\delta}_{ij}(K) \coloneqq \fint_{Y^{\delta}_K} \log A^{\varepsilon}_{ij}(y)\,\mathrm{d}y, \quad \text{for } A^{\varepsilon}_{ij}(K) \neq 0, \quad \bar{A}^{\delta}_{ij} = 0, \text{ otherwise,} \tag{3.13}$$

or by the *harmonic mean value*

$$\bar{A}^{\delta}_{ij}(K)^{-1} \coloneqq \fint_{Y^{\delta}_K} \frac{1}{A^{\varepsilon}_{ij}(y)}\,\mathrm{d}y, \quad \text{for } A^{\varepsilon}_{ij}(K) \neq 0, \quad \bar{A}^{\delta}_{ij} = 0, \text{ otherwise.} \tag{3.14}$$

Remark 3.6. In general, the (arithmetic) average (3.12) does not lead to the correct homogenization limit A^0 (see the discussion in Remark 2.27 and Section 2.3.3). However, in context of linear elasticity it can be shown that good, effective coefficients lie between the harmonic (3.14) and arithmetic (3.12) mean value (cf. Dederichs and Zeller[40] and Kröner[65]); expressed in terms of a scalar coefficient:

$$\Big(\fint_{Y^{\delta}_K} \frac{1}{A^{\varepsilon}(y)}\,\mathrm{d}y\Big)^{-1} \leq A^0 \leq \fint_{Y^{\delta}_K} A^{\varepsilon}(y)\,\mathrm{d}y. \tag{3.15}$$

As already discussed in Section 2.3.3, the same observation holds true for groundwater flow simulations with $A^{\varepsilon}(x) = K^{\varepsilon}(x)\mathrm{Id}$ and an effective permeability $K^{\delta}(x)$.

Remark 3.7. The geometric mean value (3.13) is usually a reasonable choice for a large class of log-normally distributed permeabilities[100], i. e., permeabilities A^{ε} where the (component-wise) logarithm $\log A^{\varepsilon}$ is normally distributed with a specific correlation structure. Consequently, the correct choice of averaging strategy is highly micro structure and problem dependent.

In order to avoid a reduction in regularity the cell-wise *constant* (globally discontinuous) coefficients $\bar{A}^{\delta}$ may be further post-processed to yield a globally *continuous* function:

Definition 3.8 (Post processing)**.** Let $V^{\delta}(\Omega)$ denote the space of linear finite-elements associated with $\mathbb{T}_{\delta}(\Omega)$. Define $A^{\delta} \in V^{\delta}(\Omega)^{d\times d}$ with the help of an interpolation of *Clément-type* (cf. Clément[35] and Scott and Zhang[94]): Let x_i be a nodal point of $\mathbb{T}_{\delta}(\Omega)$ and let $\mathscr{K}_i$ be the set of all cells $K \in \mathbb{T}_{\delta}(\Omega)$ with $\bar{K} \cap x_i \neq \emptyset$. Define

$$A^{\delta}(x_i) = \frac{\sum_{K\in\mathscr{K}_i} \bar{A}^{\delta}(K)\,|K_i|}{\sum_{K\in\mathscr{K}_i} |K_i|}. \tag{3.16}$$

With the help of the effective model $\left(\mathbb{T}_\delta(\Omega), A^\delta\right)$ we define an *effective problem*.

Definition 3.9 (Effective problem)**.** Let $\left(\mathbb{T}_\delta(\Omega), \bar{A}^\delta\right)$ be an effective model in the sense of Definition 3.3. The *effective problem* reads: Find $u^\delta \in H_0^1(\Omega)$ s. t.

$$\left(A^\delta \nabla u^\delta, \nabla \varphi\right) = (f, \varphi) \quad \forall\, \varphi \in H_0^1(\Omega). \tag{3.17}$$

Here, A^δ shall either denote the post-processed coefficient according to (3.16) or $\bar{A}^\delta$ defined by Definition 3.5.

Remark 3.10. The well-posedness and a priori convergence of the numerical homogenization scheme (3.17), as well as the well-posedness of the sampling problems (3.11), will be established in Section 3.4.

3.2.2 Semi-discretized problem

A numerical evaluation of the sampling processes (3.11), (3.12), or (3.13) introduces quadrature and discretization errors on the fine (sampling) scale that have to be taken into account. Therefore, we denote such a *numerically computed, effective model* with a global discretization parameter h and a sequence $\{h_K\}_{K \in \mathbb{T}_\delta(\Omega)}$ of local discretization parameters for each sampling region Y_K^δ. Here, h_K denotes the refinement level of a fine-scale mesh $\mathbb{T}_h(Y_K^\delta)$ of Y_K^δ, and h is the maximum of the h_K. With the help of the meshes $\mathbb{T}_h$ a summed quadrature rule $Q_{h,K}$ on Y_K^δ is defined:

Definition 3.11 (Summed quadrature rule)**.** Given a base quadrature rule $\hat{Q}$ on Y with support points $\left\{\hat{x}_i\right\}_i$ and weights $\left\{q_i\right\}_i$ and let $\mathcal{T}_{\tilde{K}}$ denote the transformation $\mathcal{T}_{\tilde{K}} : Y \to \tilde{K}$ for $\tilde{K} \in \mathbb{T}_h(Y_K^\delta)$. Define a *summed quadrature rule*

$$\begin{aligned} Q_{h,K}(f) &:= \sum_{\tilde{K} \in \mathbb{T}_h(Y_K^\delta)} |\tilde{K}|\, \hat{Q}\left(f \circ \mathcal{T}_{\tilde{K}}\right) \\ &= \sum_{\tilde{K} \in \mathbb{T}_h(Y_K^\delta)} |\tilde{K}| \sum_i q_i\, f \circ \mathcal{T}_{\tilde{K}}(\hat{x}_i). \end{aligned} \tag{3.18}$$

Define, in correspondence to (3.12), (3.13), and (3.14) the *numerically computed,*

effective model $\big(\mathbb{T}_\delta(\Omega), \bar{A}^{\delta,h}\big)$ by

$$\bar{A}^{\delta,h}_{ij}(K) \coloneqq \frac{1}{|Y^\delta_K|} Q_{h,K}\big(A^\varepsilon_{ij}(y)\big), \text{ or} \tag{3.19}$$

$$\big(\bar{A}^{\delta,h}_{ij}(K)\big)^{-1} \coloneqq \frac{1}{|Y^\delta_K|} Q_{h,K}\Big(\frac{1}{A^\varepsilon_{ij}(y)}\Big), \text{ or} \tag{3.20}$$

$$\log \bar{A}^{\delta,h}_{ij}(K) \coloneqq \frac{1}{|Y^\delta_K|} Q_{h,K}\big(\log A^\varepsilon_{ij}(y)\big)\,\mathrm{d}y, \text{ respectively.} \tag{3.21}$$

In case of the homogenization scheme (see Definition 3.4), an additional fine-scale discretization has to be taken into account. Namely, the choice of a discretized fine-scale space $V^h(Y^\delta_K) \subset \tilde{H}^1_{\mathrm{per}}(Y^\delta_K)$ associated with a mesh $\mathbb{T}_h(Y^\delta_K)$ of Y^δ_K. With such a choice at hand we set

$$\bar{A}^{\delta,h}_{ij}(x) \coloneqq \frac{1}{|Y^\delta_K|} Q_{h,K}\Big(A^\varepsilon(x)\big(\nabla_x \omega^h_i(x) + \boldsymbol{e}_i\big)\cdot\big(\nabla_x \omega^h_j(x) + \boldsymbol{e}_j\big)\Big), \text{ where} \tag{3.22}$$

$$Q_{h,K}\Big(A^\varepsilon(x)\big(\nabla_x \omega_i(x) + \boldsymbol{e}_i\big)\cdot\nabla\varphi\Big) = 0 \quad \forall \varphi \in V^h(Y^\delta_K). \tag{3.23}$$

In summary, introducing a fine-scale discretization for the sampling process leads to another auxiliary problem:

Definition 3.12 (effective, semi-discretized problem)**.**
For a numerically computed, effective model $\big(\mathbb{T}_\delta(\Omega), \bar{A}^{\delta,h}\big)$ let $A^{\delta,h}$ be its post-processed variant and consider the variational problem: Find $u^{\delta,h} \in H^1_0(\Omega)$ s. t.

$$\big(A^{\delta,h}\nabla u^{\delta,h}, \nabla\varphi\big) = (f,\varphi) \quad \forall\, \varphi \in H^1_0(\Omega). \tag{3.24}$$

Remark 3.13. Well-posedness and a priori convergence of the semi-discretized problem (3.24) will be established in Section 3.4.

3.2.3 Fully discretized problem

As a last step introduce a coarse grid $\mathbb{T}_H(\Omega)$ for numerically approximating the variational equation (3.24):

Definition 3.14 (Fully discretized problem)**.** Let $\mathbb{T}_H$ be a mesh covering Ω, and let $V_H(\Omega) \subset H^1_0(\Omega)$ be a finite-element ansatz space. The fully discrete problem reads: Find $U \in V_H(\Omega)$ s. t.

$$\big(A^{\delta,h}\nabla U, \nabla\varphi^H\big) = \big(f,\varphi^H\big) \quad \forall\, \varphi^H \in V_H(\Omega). \tag{3.25}$$

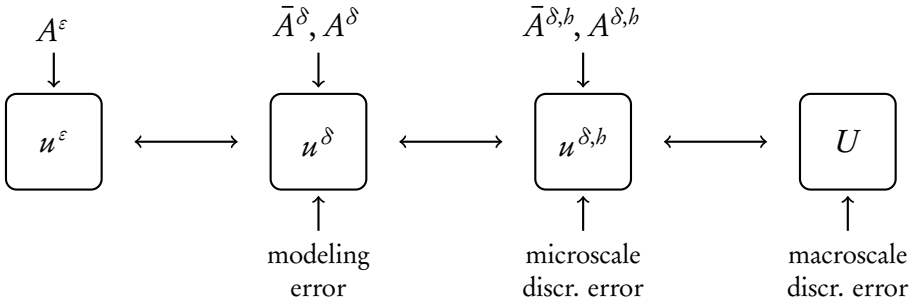


Figure 3.2. The interplay of the different sources of error ranging from the solution u^ε of the full model problem (3.6), over the solutions u^δ and $u^{\delta,h}$ of the auxiliary problems (3.17), (3.24) to the solution U of the fully discretized problem (3.25).

In summary, we have established a framework that explicitly decouples all discretization parameters and respective sources of error, see Figure 3.2.

Remark 3.15. Oden and Vemaganti, and Braack and Ern considered a similar abstract model framework[26,84,86] that can be regarded as a simplification of above scheme by neglecting fine-scale discretization errors (discussed in Section 3.2.2) and assuming $u^\delta \equiv u^{\delta,h}$, as well as removing the choice of a different sampling discretization by setting $\mathbb{T}_\delta = \mathbb{T}_H$.

3.3 A note on the generality of the framework

The model framework is general in the sense that it allows the presented VMM and HMM, as well as the computational homogenization scheme to be formulated within its boundaries.

In case of the HMM scheme create a sampling discretization $\mathbb{T}_\delta(\Omega)$ such that every quadrature point of the macroscale discretization is located in a separate cell $K \in \mathbb{T}_\delta(\Omega)$. Further, fix the midpoint of the sampling regions Y_K^δ to coincide with the respective quadrature point. Neglecting quadrature on the microscale discretization for simplicity, it remains to show that after introducing a macroscale discretization the system

$$\sum_{K\in\mathbb{T}_H(\Omega)} |K| \sum_i q_i A_K^\delta \nabla U(x_i)\cdot\nabla\varphi^H(x_i) = \sum_{K\in\mathbb{T}_H(\Omega)} |K| \sum_i q_i A_K^\delta f(x_i)\varphi^H(x_i) \quad \forall \varphi^H \in V^H(\Omega) \quad (3.26)$$

is equivalent to

$$\sum_{K\in\mathbb{T}_H(\Omega)} |K| \sum_i q_i \fint_{Y^\delta_{K,i}} A^\varepsilon \nabla \big[U(x_i) + \mathscr{R}^h_{K,i}(U)(x) \big] \cdot \big[\varphi^H(x_i) + \mathscr{R}^h_{K,i}(\varphi^H)(x) \big] = \sum_{K\in\mathbb{T}_H(\Omega)} |K| \sum_i q_i A^\delta_K f(x_i) \varphi^H(x_i) \quad \forall \varphi^H \in V^H(\Omega). \tag{3.27}$$

Such an equivalence follows fully analogous to the continuous case examined in Proposition 2.31. In light of the discussion in Section 2.3.2 the same holds true for the computational homogenization scheme defined in Subsection 2.3.1. Hence, both, HMM and the computational homogenization scheme can be embedded into the given framework.

In contrast to the former approaches, the VMM ansatz is based on a split of the ansatz space. In order to show equivalence to the given framework is is necessary to express this split in terms of an effective parameter:

Proof. Applying the VMM ansatz (2.16) to the model problem and a subsequent discretization with a finite-element space $V^H(\Omega)$ and a quadrature rule (q_i, x_i) results in

$$\begin{aligned} \sum_{K\in\mathbb{T}_H(\Omega)} |K| \sum_i q_i A^\varepsilon \nabla u^H(x_i) \cdot \nabla \varphi^H(x_i) \\ + \sum_{K\in\mathbb{T}_H(\Omega)} |K| \sum_i q_i A^\varepsilon \nabla (\mathscr{T}(f - \mathscr{L} u^H)(x_i) \cdot \nabla \varphi^H(x_i) \\ = \sum_{K\in\mathbb{T}_H(\Omega)} |K| \sum_i q_i A^\delta_K f(x_i) \varphi^H(x_i) \quad \forall \varphi^H \in V^H(\Omega). \end{aligned} \tag{3.28}$$

In order to express this equation in terms of (3.25) it must hold true that

$$A^\delta_K \nabla u^H(x_i) = A^\varepsilon(x_i) \nabla u^H(x_i) + A^\varepsilon(x_i) \nabla u^f(x_i) \quad \forall u^H(x_i). \tag{3.29}$$

It is important to note that the functional dependency $\nabla u^f(x_i)\big(\nabla u^H(x_i)\big)$ is effectively a linear dependency. Thus, it can be expressed by a *reconstruction* matrix $\mathscr{F}$ with

$$\nabla u^f(x_i) = \mathscr{F} \nabla u^H(x_i). \tag{3.30}$$

Therefore,

$$A^\delta_K(x_i) = A^\varepsilon(x_i) + A^\varepsilon(x_i) \mathscr{F} \tag{3.31}$$

is a solution. □

Remark 3.16. The developed methodology and all model adaptation strategies that will be formulated in the following are in principle applicable to any method that fulfills the following conditions:

- There exists a complete (microscopic) model of the phenomenon in question such that the problem can be stated as: Find $u^\varepsilon \in \mathcal{V}^\varepsilon$ s. t.

$$\mathcal{L}^\varepsilon u^\varepsilon = f^\varepsilon, \tag{3.32}$$

  i. e. formulated as an abstract model problem in the sense of (2.7).

- There exists an effective model

$$\mathcal{L}^\delta u^\delta = f^\delta, \tag{3.33}$$

  that has a *localization property* for its parameters (that are denoted by δ), i. e. $\mathcal{L}^\delta$ can be iteratively refined to a sequence of improved, effective model $\mathcal{L}^{d_i}$ and $\mathcal{L}^{d_i} \to \mathcal{L}^\varepsilon$ in the operator norm.

- Model specific residuals of the form $(\mathcal{L}^\delta_K - \mathcal{L}^\varepsilon_K)$ can be estimated.

3.4 Well-posedness and a priori error analysis

In this section an existence and uniqueness result for the developed multiscale framework is established and a priori error estimates in the homogeneous H^1- and L^2-norm are derived. This is done for both types of sampling processes—the averaging schemes given in Definition 3.5 and the homogenization process (cf. Definition 3.4). The following assumptions are made:

(A1) $A^\varepsilon \in L^\infty(\Omega)^{d\times d}$ is Lipschitz continuous with an ε-dependent coefficient,

$$\left\|A^\varepsilon(y) - A^\varepsilon(x)\right\| \leq C(\varepsilon)\|y - x\|. \tag{3.34}$$

(A2) Ω is a polygonal/polyhedral domain and all meshes involved in the multiscale method, $\mathbb{T}_H(\Omega)$, $\mathbb{T}_\delta(\Omega)$, $\{\mathbb{T}_h(Y^\delta_K), K \in \mathbb{T}_\delta(\Omega)\}$ shall be *quasi-uniform*, i. e., we assume that the following properties are fulfilled (see Ciarlet[33] or Brenner and Scott[28]):

  - *Structural regularity*: $\bigcup \overline{K} = \overline{\Omega}$, and every intersection $\overline{K}_i \cap \overline{K}_j$, $i \neq j$ is a corner, an edge, or a face.

- *Uniform shape regularity*: There exists $C \geq 0$, such that $\left\| \det \nabla \mathcal{T}_K \right\| + \left\| \det \nabla \mathcal{T}_K^{-1} \right\| \leq C$ uniformly for all cells K of all families of meshes. Here, $\mathcal{T}_K : Y \to K$ denotes a d-linear transformation of the unit cell Y onto K. This allows for a decomposition of $\overline{\Omega}$ consisting of (closed and convex) quadrilaterals/hexahedra.

(A3) Additional structural requirements:

- *Interior-angle condition* on the sampling mesh $\mathbb{T}_\delta(\Omega)$: There exist numbers $\alpha_1, \alpha_2 > 0$ such that $\alpha_1 \leq \alpha \leq \alpha_2$ for all interior angles α of $K \in \mathbb{T}_\delta(\Omega)$ of the sampling mesh $\mathbb{T}_\delta(\Omega)$.
- $Y_K^\delta \subset K$ for all sampling regions Y_K^δ associated with a cell K for the family of sampling meshes $\left\{ \mathbb{T}_\delta(\Omega) \right\}_{\delta > 0}$.

Remark 3.17. There is no requirement about *size regularity* of the meshes in order to allow for local mesh adaptation. To further ease local mesh refinement the regularity assumption (A2) can be relaxed to allow for *hanging nodes* (see Becker and Braack[18]). This leads to the following relaxed conditions:

- Lift the structural regularity by introducing *hanging nodes*: Allow for a difference of one degree of refinement between neighboring cells and introduce an additional linear constraint.

- Only require *polynomial size regularity*: There exist $c_1, c_2 > 0$ and $\varkappa > 0$ s. t. $c_1 \nu^\varkappa \leq \nu_K \leq c_2 \nu$ for all meshes $\mathbb{T}_\nu$ and $\nu = H, h, \delta$.

Remark 3.18. The main focus of the thesis is an a posteriori treatment of the abstract model framework. Thus, the discussion given below will only derive convergence results for uniform refinement to ensure well-posedness and (principal) a priori convergence. The *efficiency* of the framework will be solely based on a posteriori techniques, not strong a priori assumptions.

3.4.1 Convergence of the averaging schemes

In order to derive a priori error estimates and convergence results for the modeling error $u^\delta - u^\varepsilon$ it is necessary to assume some control over the quality of effective models. Therefore, we make the following *a priori assumption* on the effective model $\left(\mathbb{T}_\delta(\Omega), A^\delta \right)$ that takes the choice of model (denoted by δ) as well as the finescale scaling ε into account:

Definition 3.19 (Applicability)**.** Let $\{A^\varepsilon\}_{\varepsilon>0}$ be a family of heterogeneities and let $\{(\mathbb{T}_\delta(\Omega), A^\delta)\}_{\delta>0}$ be a family of effective models. The family of effective models is said to be *applicable for* A^ε if there exists a possibly ε-dependent $C(\varepsilon)$ s. t.

$$\sup_{K\in\mathbb{T}_\delta(\Omega)} \left\|A^\varepsilon - A^\delta\right\|_{L^\infty(K)} \leq C(\varepsilon)\,\delta. \tag{3.35}$$

Remark 3.20. The a priori assumption (3.35) is used for the averaging schemes (3.12) and (3.13) to show convergence in H^1. Given its ε-dependence expressed as $C(\varepsilon)$ the convergence results *can be arbitrarily ill-posed for the limit* $\varepsilon \to 0$. Thus, the a priori assumption (3.34) can be interpreted as the requirement for a *typical length scale of fluctuations* such that, with a sufficient refinement of $\mathbb{T}_\delta(\Omega)$, it is possible to resolve the fine scale completely. In order to control the limit $\varepsilon \to 0$, *additional* regularity assumptions on A^ε are necessary. A possibility is the requirement of a local periodic structure $A^\varepsilon(x) = A(x, \frac{x}{\varepsilon})$ with $A(x,y)$ being Y-periodic in y as done in the context of the mathematical homogenization theory (cf. Section 2.2.1). This periodicity assumption will be used for the homogenization sampling approach (Definition 3.4) for an improved convergence result, see Section 3.4.2.

The question of a priori convergence of the solution u^δ of the effective equation (3.17) to the solution u^ε of the model problem (3.6) is subtle. The inherent difficulty lies in the fact that δ is *not only* a discretization parameter *but also* denotes the choice of effective parameters for an effective model. In its extremes, one has the option between a convergence result

$$u^\delta \to u^\varepsilon \quad \text{for (discretization parameter) } \delta \to 0 \tag{3.36}$$

for a pure *discretization-adaptation strategy* (with fixed effective model derivation) and a pure *model adaptation* by improving the values $\bar{A}^\delta$ while maintaining a fixed sampling discretization $\mathbb{T}_\delta(\Omega)$.

It has to be noted that at first sight an asymptotic refinement of the discretization $\mathbb{T}_\delta(\Omega)$ seems to be contrary to the very philosophy of multiscale methods that try to *decrease* computational costs by maintaining a *coarse* sampling grid $\mathbb{T}_\delta(\Omega)$. However, this is only true if the microscale discretization associated with an individual sampling cell cannot be simultaneously coarsened in order to maintain an equal amount of total microscale resolution. In Chapter 5, a sampling-adaptation strategy will be presented that is based on this pure discretization adaptation approach.

The definition of applicability given in Definition 3.19 is fulfilled by every class of coefficients A^ε that is at least Lipschitz continuous as defined in (A1):

Lemma 3.21. Let $A^\varepsilon \in L^\infty(\Omega)^{d\times d}$ be Lipschitz continuous as defined in (3.34). Then, every sequence of effective models with sampling discretization fulfilling

$$\sup_{K\in\mathbb{T}_\delta(\Omega)} (\operatorname{diam} K) \leq C\,\delta \tag{3.37}$$

is applicable.

Proof. In case of the arithmetic average it holds true that

$$\begin{aligned}\sup_{K\in\mathbb{T}_\delta(\Omega)} \left\|A^\varepsilon - \bar{A}^\delta\right\|_{L^\infty(K)} &= \sup_{K\in\mathbb{T}_\delta(\Omega)} \left\|A^\varepsilon - \fint_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y\right\|_{L^\infty(K)} \\ &\leq \sup_{K\in\mathbb{T}_\delta(\Omega)} \left[\left\|A^\varepsilon - A^\varepsilon(\hat{x})\right\|_{L^\infty(K)} + \left\|A^\varepsilon(\hat{x}) - \fint_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y\right\|_{\mathbb{R}^{d\times d}}\right] \end{aligned} \tag{3.38}$$

for some $\hat{x} \in Y_K^\delta \subset K$. Furthermore, the second term can be bounded by

$$\begin{aligned}\left\||Y_K^\delta| A^\varepsilon(\hat{x}) - \int_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y\right\|_{\mathbb{R}^{d\times d}} &= \left\| \int_{Y_K^\delta} A^\varepsilon(\hat{x}) - A^\varepsilon(y)\,\mathrm{d}y\right\|_{\mathbb{R}^{d\times d}} \\ &\leq |Y_K^\delta| \left\|A^\varepsilon(\hat{x}) - A^\varepsilon(y)\right\|_{L^\infty(Y_K^\delta)}. \end{aligned} \tag{3.39}$$

Inserting (3.39) into (3.38):

$$\begin{aligned}\sup_{K\in\mathbb{T}_\delta(\Omega)} \left\|A^\varepsilon - \bar{A}^\delta\right\|_{L^\infty(K)} &\leq \sup_{K\in\mathbb{T}_\delta(\Omega)} C(\varepsilon)\left\{(\operatorname{diam} Y_K^\delta) + (\operatorname{diam} K)\right\} \\ &\leq C(\varepsilon)\,\delta. \end{aligned} \tag{3.40}$$

In case of the post-processed coefficients it also holds that

$$\sup_{K\in\mathbb{T}_\delta(\Omega)} \left\|A^\delta - \bar{A}^\delta\right\|_{L^\infty(K)} \leq C\,\delta \tag{3.41}$$

by virtue of the Clément-type interpolation (see Definition 3.8). A similar estimate yields the result for the geometric and harmonic mean value. Observe for example that

$$\begin{aligned}&\left\|A_{ij}^\varepsilon(\hat{x}) - \exp\Big(\fint_{Y_K^\delta} \log A_{ij}^\varepsilon \,\mathrm{d}y\Big)\right\|_{\mathbb{R}^{d\times d}} \\ &\leq \left|A_{ij}^\varepsilon(\hat{x})\right| \left[1 - \exp\Big(\fint_{Y_K^\delta} \log A_{ij}^\varepsilon \,\mathrm{d}y - \log A_{ij}^\varepsilon(\hat{x})\Big)\right] \\ &\leq \left|A_{ij}^\varepsilon(\hat{x})\right| \left[1 - \exp\Big(\left\|\log A_{ij}^\varepsilon - \log A_{ij}^\varepsilon(\hat{x})\right\|_{L^\infty(Y_K^\delta)}\Big)\right] \end{aligned}$$

$$\leq \left|A_{ij}^{\varepsilon}(\hat{x})\right|\left[1-\exp\left(\mathcal{O}\left(\log(C(\varepsilon)\operatorname{diam} Y_K^{\delta})\right)\right)\right]$$
$$\leq \left|A_{ij}^{\varepsilon}(\hat{x})\right|\left[\mathcal{O}\left(C(\varepsilon)\operatorname{diam} Y_K^{\delta}\right)\right]. \tag{3.42}$$

□

Effective problem In the following we assume (A1) and (A2) to be fulfilled. Alternatively, the Lipschitz continuity assumed in (A1) can be replaced by

(A*) $\left\{(\mathbb{T}_\delta(\Omega), A^\delta)\right\}_{\delta>0}$ is applicable for $\left\{A^\varepsilon\right\}_{\varepsilon>0}$ in the sense of Definition 3.19.

Under these assumptions, the effective problem (3.17) is well-posed:

Lemma 3.22. $\bar{A}^\delta$ as defined in (3.12), (3.13), or (3.14), and its post-processed counterpart A^δ are symmetric and elliptic with the same constants as A^ε stated in (2.8) and (2.9). Therefore, the effective problem of Definition 3.9 to find $u^\delta \in H^1(\Omega)$ s. t.

$$\left(A^\delta \nabla u^\delta, \nabla\varphi\right) = (f,\varphi) \quad \forall\, \varphi \in H_0^1(\Omega). \tag{3.43}$$

admits a unique solution.

Proof. The statement follows immediately from the fact that the corresponding properties are fulfilled point-wise a. e. and that the Clément-type interpolation preserves symmetry and ellipticity. □

Lemma 3.23. It holds true that

$$||\nabla u^\delta - \nabla u^\varepsilon||_{L^2(\Omega)} \leq \frac{1}{\alpha} \max_{K\in\mathbb{T}_\delta(\Omega)} \left\|A^\delta - A^\varepsilon\right\|_{L^\infty(K)} \left\|\nabla u^\varepsilon\right\|_{L^2(\Omega)}. \tag{3.44}$$

Proof. By (2.10) and (3.17) it follows that

$$\left(A^\delta \nabla u^\delta, \nabla\varphi\right) = \left(A^\varepsilon \nabla u^\varepsilon, \nabla\varphi\right). \tag{3.45}$$

Subtracting $A^\delta \nabla u^\varepsilon$ from (3.17), testing with $u^\delta - u^\varepsilon$ and utilizing (3.45):

$$\left(A^\delta \nabla u^\delta - A^\delta \nabla u^\varepsilon, \nabla u^\delta - \nabla u^\varepsilon\right) = \left(A^\varepsilon \nabla u^\varepsilon - A^\delta \nabla u^\varepsilon, \nabla u^\delta - \nabla u^\varepsilon\right). \tag{3.46}$$

Therefore,

$$\alpha||\nabla u^\delta - \nabla u^\varepsilon||_{L^2(\Omega)}^2 \leq \left((A^\varepsilon - A^\delta)\nabla u^\varepsilon, \nabla u^\delta - \nabla u^\varepsilon\right)_{L^2(\Omega)}$$
$$\leq \max_K ||A^\delta - A^\varepsilon||_{L^\infty(K)} ||\nabla u^\varepsilon|| \, ||\nabla u^\delta - \nabla u^\varepsilon||. \tag{3.47}$$

□

Corollary 3.24. The corresponding estimate for the H^1-norm,

$$\|u^\delta - u^\varepsilon\|_{H^1(\Omega)} \leq C \frac{1}{\alpha} \max_{K \in \mathbb{T}_\delta(\Omega)} \left\|A^\delta - A^\varepsilon\right\|_{L^\infty(K)} \left\|\nabla u^\varepsilon\right\|_{L^2(\Omega)}, \tag{3.48}$$

follows by virtue of Poincaré's inequality.

With this prerequisites at hand we have established:

Proposition 3.25 (Convergence)**.** Under assumptions (A1) and (A2) the effective solution u^δ converges to u^ε with the same order in H^1,

$$\left\|u^\varepsilon - u^\delta\right\|_{H^1(\Omega)} \leq C(\varepsilon)\,\delta. \tag{3.49}$$

Proof. The result follows immediately from Lemma 3.23 together with the a priori estimate

$$\|\nabla u^\varepsilon\|_{L^2(\Omega)^d} \leq C \|f\|_{L^2(\Omega)}, \tag{3.50}$$

with C independent of ε due to the uniform ellipticity assumption (2.9). □

Remark 3.26. Improved convergence rates can be shown for the $L^2(\Omega)$ norm, provided the oscillatory coefficient A^ε has enough regularity. Such results follow by the same reasoning as the duality arguments given in the a posteriori error analysis in Chapter 4.

Semi-discretized problem Next, we examine the numerically computed, effective model $\left(\mathbb{T}_\delta(\Omega), A^{\delta,h}\right)$ that takes quadrature and discretization errors on the fine (sampling) scale into account. Let $u^{\delta,h} \in H^1_0(\Omega)$ be the solution of the numerically computed, effective problem (3.24):

$$\left(A^{\delta,h} \nabla u^{\delta,h}, \nabla \varphi\right) = (f, \varphi) \quad \forall\, \varphi \in H^1_0(\Omega). \tag{3.51}$$

The requirement for Lipschitz continuity made in (A1) is sufficient to establish a convergence result for an approximation with a summed quadrature rule $Q_{h,K}$ of A^ε.

Remark 3.27. Alternative approaches can be made by requiring statistical properties to hold true[90]. We restrict the discussion to the non-statistical case.

Lemma 3.28. Assume that (A1) is fulfilled, in particular that A^ε is Lipschitz continuous in the sense of (3.34). Let Q_h be a summed quadrature rule as defined in Definition 3.11 and let $\hat{Q}$ be its base quadrature defined on the unit cell, with support points $\{\hat{x}_i\}_i$ and weights $\{q_i\}_i$. Then, there exists $\delta_0 > 0$ s. t.

$$\left\| \frac{1}{|Y_K^\delta|} Q_{h,K}\big(A^\varepsilon\big) - \fint_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y \right\| \leq C(\varepsilon)\, h \tag{3.52}$$

uniformly in $K \in \mathbb{T}_\delta(\Omega)$, $\delta \leq \delta_0$.

Proof. By virtue of (3.18):

$$\begin{aligned}
\left| Q_{h,K}\big(A^\varepsilon\big) - \int_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y \right| &= \left| \sum_{\tilde{K} \in \mathbb{T}_h(Y_K^\delta)} |\tilde{K}| \sum_i q_i\, A^\varepsilon \circ \mathscr{T}_{\tilde{K}}(\hat{x}_i) - \int_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y \right| \\
&\leq \sum_{\tilde{K} \in \mathbb{T}_h(Y_K^\delta)} |\tilde{K}| \left| \sum_i q_i A^\varepsilon\big(\mathscr{T}_{\tilde{K}}(\hat{x}_i)\big) - \int_{\tilde{K}} A^\varepsilon \,\mathrm{d}y \right| \\
&\leq \sum_{\tilde{K} \in \mathbb{T}_h(Y_K^\delta)} |\tilde{K}| \left| \sum_i q_i \big\| A^\varepsilon\big(\mathscr{T}_{\tilde{K}}(\hat{x}_i)\big) - A^\varepsilon \big\|_{L^\infty(\tilde{K})} \right| \\
&\leq |Y_K^\delta|\, C(\varepsilon) \max_{\tilde{K} \in \mathbb{T}_h(Y_K^\delta)} \operatorname{diam}(\tilde{K}).
\end{aligned} \tag{3.53}$$

Due to (A2) and (A3) this result is uniform in $K \in \mathbb{T}_\delta(\Omega)$, $\delta \leq \delta_0$. □

Remark 3.29. A corresponding result can be shown for the geometric and harmonic mean value (3.13).

Proposition 3.30. With assumptions (A1) and (A2), it holds true that

$$\left\| u^{\delta,h} - u^\delta \right\|_{H^1(\Omega)} \leq C(\varepsilon)\, h \tag{3.54}$$

uniformly in $\delta \leq \delta_0$.

Proof. With Lemma 3.28 and the same argumentation as in Lemma 3.23 it follows that

$$\begin{aligned}
||\nabla u^{\delta,h} - \nabla u^\delta|| &\leq \frac{1}{\alpha^*} \max_{K \in \mathbb{T}_\delta(\Omega)} ||A^\delta - A^{\delta,h}||_{L^\infty(K)} ||\nabla u^\delta||_\Omega \\
&\leq \frac{1}{\alpha^*} (1 + C\,\delta) \max_{K \in \mathbb{T}_\delta(\Omega)} ||\bar{A}^\delta - \bar{A}^{\delta,h}||_{L^\infty(K)} ||\nabla u^\delta|| \\
&\leq \frac{1}{\alpha^*} (1 + C\,\delta) \max_{K \in \mathbb{T}_\delta(\Omega)} \left| \frac{1}{|Y_K^\delta|} Q_{h,K}\big(A^\varepsilon\big) - \fint_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y \right| ||\nabla u^\delta||_{L^2(\Omega)} \\
&\leq \frac{1}{\alpha^*} (1 + C\,\delta) \max_{K \in \mathbb{T}_\delta(\Omega)} C(\varepsilon)\, h_K ||\nabla u^\delta||_{L^2(\Omega)}.
\end{aligned} \tag{3.55}$$

The first transformation follows from

$$\begin{aligned}||A^\delta - A^{\delta,h}|| &\leq ||(A^\delta - A^{\delta,h}) - (\bar{A}^\delta - \bar{A}^{\delta,h})|| + ||\bar{A}^\delta - \bar{A}^{\delta,h}|| \\ &\leq \big(1 + C\,\delta\big)||\bar{A}^\delta - \bar{A}^{\delta,h}||.\end{aligned}$$

Applying an a priori bound on $||\nabla u^\delta||$ by virtue of Lemma 3.22 gives the desired result. The corresponding estimate in the L^2-norm follows with the help of Poincaré's inequality. □

Remark 3.31. This convergence result can be substantially improved in case of a smooth coefficient $A^\varepsilon \in C^k(\Omega)^{d\times d}$ for some small k. In this case (3.55) can be combined with a standard interpolation estimate:

$$\begin{aligned}&||\nabla u^{\delta,h} - \nabla u^\delta|| \\ &\quad\leq \frac{1}{\alpha^*} \max_{K\in\mathbb{T}_\delta(\Omega)} ||A^\delta - A^{\delta,h}||_{L^\infty(K)} ||\nabla u^\delta||_\Omega \\ &\quad\leq \frac{1}{\alpha^*}(1 + C\,\delta) \max_{K\in\mathbb{T}_\delta(\Omega)} \frac{1}{|Y_K^\delta|} \left| Q_{h,K}\big(A^\varepsilon\big) - \int_{Y_K^\delta} A^\varepsilon \,\mathrm{d}y \right| ||\nabla u^\delta||_{L^2(\Omega)} \\ &\quad\leq \frac{1}{\alpha^*}(1 + C\,\delta) \max_{K\in\mathbb{T}_\delta(\Omega)} \frac{1}{|Y_K^\delta|} \sum_{Q\in\mathbb{T}_h(Y_K^\delta)} \left| Q_h\big(A^\varepsilon\big) - \int_Q A^\varepsilon \,\mathrm{d}y \right| ||\nabla u^\delta||_{L^2(\Omega)} \\ &\quad\leq \frac{1}{\alpha^*}(1 + C\,\delta) \max_{K\in\mathbb{T}_\delta(\Omega)} \frac{1}{|Y_K^\delta|} \sum_{Q\in\mathbb{T}_h(K)} C\,h^k\,|Q| \\ &\quad\leq C\,\frac{1}{\alpha^*}\,h^k. \end{aligned} \tag{3.56}$$

In regard of the fully discretized problem that will be discussed in the next paragraph, a stability result for $\big\|\nabla^2 u^{\delta,h}\big\|_{L^2(\Omega)^{d\times d}}$ is necessary:

Lemma 3.32. In case of the post-processed, continuous $A^{\delta,h}$, it holds true that

$$\big\|A^{\delta,h}\big\|_{W^{1,\infty}(\Omega)} \leq C(\delta) \leq C(\varepsilon), \tag{3.57}$$

where $C(\delta)$ describes the same functional dependency as the coefficient abstractly introduced in (3.35). Consequently, $u^{\delta,h} \in H^2(\Omega)$ with

$$||\nabla^2 u^{\delta,h}||_{L^2(\Omega)^{d\times d}} \leq C(\delta). \tag{3.58}$$

Proof. $A^{\delta,h}$ is a piecewise d-linear function and therefore of class $W^{1,\infty}(\Omega)$. Further, the uniform bound $||A^{\delta,h}||_{L^\infty(\Omega)^{d\times d}} \leq C$ holds true by a priori assumptions.

The derivatives of $A^{\delta,h}$ are solely determined by the variation of $\bar{A}^{\delta,h}$ and the sampling-cell sizes of $\mathbb{T}_\delta(\Omega)$. Hence, the constant $C(\delta)$ occurs, where $C(\delta)$ generally behaves like $1/\delta$ and describes the same functional dependency as $C(\varepsilon)$. □

Fully discretized problem Let $U \in V_H(\Omega)$ be the solution of the fully discretized problem (3.25),

$$\left(A^{\delta,h}\nabla U, \nabla\varphi^H\right) = \left(f, \varphi^H\right) \quad \forall\, \varphi^H \in V_H(\Omega). \tag{3.59}$$

In preparation for a convergence result of the fully discretized problem a regularity result for the semi-discrete problem is needed.

Proposition 3.33 (Regularity of the semi-discrete problem)**.** Let $u^{\delta,h} \in H^1(\Omega)$ be the solution of (3.24) with post-processed, continuous $A^{\delta,h}$. Then, $A^{\delta,h}$ is of class $W^{1,\infty}(\Omega)$ and, assuming sufficient smoothness of the boundary, it is $u^{\delta,h} \in H^2(\Omega)$ and the well known a priori error estimate

$$\left\|u^{\delta,h} - U\right\|_{L^2(\Omega)} + H\left\|\nabla u^{\delta,h} - \nabla U\right\|_{L^2(\Omega)^d} \leq C(\delta)H^2\left\|\nabla^2 u^{\delta,h}\right\|_{L^2(\Omega)^{d\times d}} \tag{3.60}$$

holds true.

Proof. The statement follows immediately by standard a priori error analysis (see Ciarlet[33]). Further, due to the fact that $A^{\delta,h}$ is a continuous and patch-wise d-linear interpolation the bound $||A^{\delta,h}||_{W^{1,\infty}(\Omega)} \leq C(\delta)$ is available, where $C(\delta)$ is the abstract functional dependency described in Lemma 3.32. □

In case of *discontinuous coefficients* $A^{\delta,h} = \bar{A}^{\delta,h}$ a problem arises from the fact that A^δ is discontinuous over the boundaries of the sampling regions. The solution $u^{\delta,h}$ is generally only of class $H^1(\Omega)$ and a full regularity $H^2(\Omega)$ (required by standard finite-element estimates) cannot be expected. Even in individual subregions $K \in \mathbb{T}_\delta(\Omega)$ there holds in general that $u^{\delta,h}\big|_K \notin H^2(K)$. The reason for this lies in the fact that a discontinuous permeability $A^{\delta,h} : \mathbb{T}_\delta(\Omega) \to \mathbb{R}^{d\times d}$ produces the same type of singularities as "reentrant corners" in the case of polygonal or polyhedral domains[24,25]. Given the fact that an approximation with piecewise constant $\bar{A}^{\delta,h}$ is numerically desirable, we go into detail and state the corresponding convergence results.

The question of regularity of solutions (and especially the question of the correct notion of regularity) has been studied extensively for different special cases in two or three dimensions. We refer to work by Blumenfeld[24,25], who gives an explicit decomposition into singular functions for the case of two dimensions, and cite the following result.

Proposition 3.34 (Regularity of the semi-discrete problem in 2D[24,25]).
Let $u^{\delta,h} \in H^1(\Omega)$ be the solution of (3.24) for the case $d = 2$. Then, given the set $\mathbb{P}_\delta(\Omega)$ of all vertices of $\mathbb{T}_\delta(\Omega)$, the solution $u^{\delta,h}$ can be decomposed into

$$u^{\delta,h}(x) = \sum_{k \in \mathbb{P}_\delta(\Omega)} \sum_{j \le J^k} \sigma_j^k s_j^k(x) + \bar{u}(x), \tag{3.61}$$

with $\bar{u} \in \prod_{K \in \mathbb{T}_\delta(\Omega)} H^2(K)$ and $\sigma_j^k \in \mathbb{R}$. Here, $\{\sigma_j^k, j \le J^k\}$ denotes the singular-function expansion associated with vertex k: The s_j^k have the form

$$s_j^k(x) = |x|^{\lambda_j^k} \varphi_j^k\big(\arg(x)\big) \tag{3.62}$$

with pairs of eigenfunctions $\varphi_j^k(\vartheta)$ and eigenvalues $0 \le \lambda_j^k \le 1$ coming from a one-dimensional eigenvalue problem associated with the vertex $k \in \mathbb{P}_\delta(\Omega)$. It holds,

$$\sum_{k \in \mathbb{P}_\delta(\Omega)} \sum_{j \le J^k} (\sigma_j^k)^2 + \sum_{K \in \mathbb{T}_\delta(\Omega)} \|\bar{u}\|^2_{H^2(K)} \le C(\delta) \|f\|^2_{L^2(\Omega)}, \tag{3.63}$$

with the same functional dependency $C(\delta)$ as found in Lemma 3.32.

Proof. The statement was proved by Blumenfeld[25, Th. 3.2,Th. 5.1]. The dependency on $C(\delta)$ follows by a similar argument as given in Lemma 3.32. □

Corollary 3.35. The result copies verbatim to the solution u^δ of the numerically homogenized problem (3.17).

Remark 3.36. In case of reentrant corners the functional dependency of

$$(\lambda_j^k, \varphi_j^k)_{j \le J^k} \tag{3.64}$$

on the interior angle ω^k of the reentrant corner is well known[64]:

$$\text{if } \pi < \omega^k < \tfrac{3}{2}\pi: \quad J^k = 1 \text{ with } \lambda_1^k = \frac{\pi}{\omega^k}, \tag{3.65}$$

$$\text{if } \pi \le \omega^k < 2\pi: \quad J^k = 2 \text{ with } \lambda_1^k = \frac{\pi}{\omega^k} \text{ and } \lambda_2^k = \frac{2\pi}{\omega^k}. \tag{3.66}$$

The case of internal interfaces is more complicated because multiple interior angles have to be considered simultaneously in an eigenvalue problem[25]: Find $\varphi \in H^1_{\text{per}}(I)$ on $I = [0, 2\pi]$

$$\int_I a \partial_\vartheta \varphi \partial_\vartheta \psi \, \mathrm{d}\vartheta = \lambda^2 \int_I a \varphi \psi \, \mathrm{d}\vartheta \quad \forall\, \psi \in H^1_{\text{per}}(I), \tag{3.67}$$

with a sector-wise constant a defined in therms of $A^{\delta,h}$.

The singular expansion (3.61) is used to establish the following a priori result:

Proposition 3.37. Assume (A2) – (A3), as well as

$$\mathbb{T}_H(\Omega) \supset \mathbb{T}_\delta(\Omega) \tag{3.68}$$

to hold true. Let $V^H \subset H_0^1(\Omega)$ be a space of linear finite elements associated with $\mathbb{T}_H(\Omega)$ and $U \in V^H$ a solution of (3.25). Then,

$$\begin{aligned}\left\|u^{\delta,h} - U\right\|_{L^2(\Omega)} + H^\gamma \left\|\nabla u^{\delta,h} - \nabla U\right\|_{L^2(\Omega)^d} \\ \leq C(\delta) H^{2\gamma}\left(\left\|\nabla^2 \bar{u}\right\|_{L^2(\Omega)^{d\times d}} + \sqrt{\sum_{k\in\mathbb{P}_\delta(\Omega)}\sum_{j\leq J^k}(\sigma_j^k)^2}\right),\end{aligned} \tag{3.69}$$

with the decomposition of $u^{\delta,h}$ according to Proposition 3.34 and a constant γ,

$$\min_{k\in\mathbb{P}_\delta(\Omega)} \frac{\pi}{\overline{\omega}^k} \leq \gamma \leq 1, \tag{3.70}$$

that depends on the explicit form of the expansion (3.61). Here, $\overline{\omega}^k$ denotes the smallest interior angle for a given vertex k.

Proof. The explicit expansion (3.61) of $u^{\delta,h}$ admits a corresponding asymptotic error expansion

$$\begin{aligned}u^{\delta,h}(x) - U(x) \\ = \sum_{k\in\mathbb{P}_\delta(\Omega)}\sum_{j\leq J^k} \sigma_j^k(H)(\varphi_j^k)'(\arg(x)) H^{2\lambda_i^k} + \mathcal{O}\left(||\nabla^2 \bar{u}||\right) H^2 |\log H|.\end{aligned} \tag{3.71}$$

This expansion follows by applying results given by Dobrowolski[41], and Blum and Rannacher[23, Th. 4]. By virtue of (A2) it holds $\sigma_j^k(H) = \sigma_j^k + \mathcal{O}(1)$[23, Th. 2]. Due to the quasi uniformity required in (A2) and the interior angle condition assumed in (A3) the derivatives $(\varphi_j^k)'$ are uniformly bounded[24, Th. 3.5]. □

Remark 3.38. $\min_{k\in\mathbb{P}_\delta(\Omega)} \frac{\pi}{\overline{\omega}^k}$ is uniformly bounded from below by virtue of (A3).

3.4.2 Convergence of the homogenization scheme

The multiscale method introduced in Sections 3.2.1 to 3.2.3 can be readily applied in the usual homogenization context. Here, we present the corresponding a priori estimates and existence results for the homogenization sampling strategy.

Effective problem The effective problem (3.17) is well defined for the homogenization strategy given in Definition 3.4 (cf. Corollary 3.43 on page 66). Furthermore, under suitable a priori assumptions the convergence of the numerically homogenized solution u^δ to u^ε (for $\delta, \varepsilon \to 0$) can be established (cf. Proposition 3.44 and Proposition 3.48 on pages 66 and 70, respectively). The results in this section are only proved for the case of piecewise constant coefficients $\bar{A}^\delta$, $\bar{A}^{\delta,h}$. The transfer of the results to the case of interpolated coefficients A^δ and $A^{\delta,h}$, respectively, can be done in full analogy to the sampling strategies discussed above. For brevity it is omitted.

Some preparations are necessary in order to establish the convergence results.

Lemma 3.39. For the cell problems (3.11)

$$\int_{Y_K^\delta} A^\varepsilon(x)\big(\nabla_x \omega_i(x) + \boldsymbol{e}_i\big)\cdot\nabla\varphi = 0 \quad \forall \varphi \in \tilde{H}^1_{\mathrm{per}}(Y_K^\delta) \tag{3.72}$$

it holds:

a) They are uniquely solvable.

b) If A^ε is constant over Y_K^δ, then $\omega_i \equiv 0$.

c) $\|\nabla\omega_i\|_{Y_K^\delta} \leq \frac{1}{\alpha}\sqrt{|Y_K^\delta|}\,\Big\| \fint_{Y_K^\delta} A^\varepsilon(y)\,\mathrm{d}y - A^\varepsilon \Big\|_{L^\infty(Y_K^\delta)}$.

Proof. The statement given in (a) is obvious. To prove (b) use partial integration for the right-hand side of (3.11):

$$\int_{Y_K^\delta} A^\varepsilon e_i\cdot\nabla\varphi\,\mathrm{d}y = 0 - \int_{Y_K^\delta} (\nabla\cdot(A^\varepsilon e_i)\varphi\,\mathrm{d}y = 0 \quad \forall\varphi\in\tilde{H}^1_{\mathrm{per}}(Y_K^\delta), \tag{3.73}$$

and $\omega_i \equiv 0$ is due to (a) the only solution for vanishing right hand side. Finally, for (c) we observe that

$$\begin{aligned}\int_{Y_K^\delta} A^\varepsilon\nabla\omega_i\cdot\nabla\omega_i\,\mathrm{d}y &= -\int_{Y_K^\delta} A^\varepsilon e_i\cdot\nabla\omega_i\,\mathrm{d}y \\ &= \int_{Y_K^\delta}\Big(\fint_{Y_K^\delta} A^\varepsilon\,\mathrm{d}y - A^\varepsilon\Big)e_i\cdot\nabla\omega_i\,\mathrm{d}y.\end{aligned} \tag{3.74}$$

Utilizing the uniform ellipticity (2.9) leads to:

$$\alpha\|\nabla\omega_i\|^2 \leq \sqrt{|Y_K^\delta|}\,\Big\|\fint_{Y_K^\delta} A^\varepsilon\,\mathrm{d}y - A^\varepsilon\Big\|_{L^\infty(Y_K^\delta)}\|\nabla\omega_i\|_{L^2(Y_K^\delta)}. \tag{3.75}$$

This concludes the proof. □

Lemma 3.40. Then, $\bar{A}^{\delta}$ defined by (3.11) is symmetric and elliptic.

Proof. We follow closely an argument by Cioranescu and Donato[34]. Symmetry follows from the fact that A^{ε} allows the representation:

$$\bar{A}^{\delta}_{ij} = \fint_{Y^{\delta}_K} A^{\varepsilon}_{ij}\,\mathrm{d}y - \fint_{Y^{\delta}_K} A^{\varepsilon}\nabla\omega_i \cdot \nabla\omega_j\,\mathrm{d}y. \tag{3.76}$$

For ellipticity observe that

$$\bar{A}^{\delta}_{K,ij}\xi_i\xi_j = \fint_{Y^{\delta}_K} A^{\varepsilon}_{kl}\partial_k(\omega_i + x_i)\cdot\partial_l(\omega_j + x_j)\,\mathrm{d}y \geq \alpha \fint_{Y^{\delta}_K} |\nabla\zeta|^2\,\mathrm{d}y \quad \forall \xi \in \mathbb{R}^d, \tag{3.77}$$

with $\zeta = \xi_i(\omega_i + x_i)$. It holds true that $\fint_{Y^{\delta}_K} ||\nabla\zeta||^2 > 0$, otherwise $\xi_i\omega_i(y) = -\xi_i x_i$ contradicts $\omega_i \in H^1_{\mathrm{per}}(K)$. □

For a uniform lower bound on the ellipticity of $\bar{A}^{\delta}$ in case of the homogenization strategy a further regularity assumption on A^{ε} is necessary. It is sufficient to assume (A1) to hold true, i. e., to assume that A^{ε} is Lipschitz continuous. Alternatively, in spirit of classical homogenization theory (cf. Section 2.2.1) it is possible to replace (A1) with:

(A4) There exists $A \in C^{0,1}\big(\Omega, C^{0,1}(Y)\big)^{d\times d}$ such that $A^{\varepsilon} = A(x, \frac{x}{\varepsilon})$ a. e. on Ω.

Here, $C^{0,1}\big(\Omega, C^{0,1}(Y)\big)^{d\times d}$ denotes the space of Lipschitz continuous functions defined on Ω with values in the space $C^{0,1}(Y)$. Additionally, we assume

(A5) $\delta_K = \delta\ \forall K \in \mathbb{T}_{\delta}(\Omega)$ and δ is an integral multiple of ε.

Remark 3.41. (A4) is an example of a heterogeneity with a "bad" behavior in ε in the sense of assumption (A1): It only holds true that

$$\big\|A^{\varepsilon}(x) - A^{\varepsilon}(y)\big\| \leq C\left(1 + \frac{1}{\varepsilon}\right)||y - x||, \tag{3.78}$$

i. e., A^{ε} is not uniformly Lipschitz continuous.

With this prerequisites the following Lemma can be shown:

Lemma 3.42. In case of (A1) it holds true that

$$\bar{A}^{\delta}_{ij}\xi_i\xi_j \geq \big(\alpha - C(\varepsilon)\,\delta^2\big)|\xi|^2, \tag{3.79}$$

with the ellipticity constant α defined in (2.9). The result is also true if the assumptions (A4) and (A5) are fulfilled instead of (A1).

Proof. The statement follows directly by virtue of (3.76), the definition of the cell problems (3.11) and Lemma 3.39:

$$\begin{aligned}
\bar{A}^{\delta}_{K,ij}\xi_i\xi_j &= \fint_{Y^{\delta}_K} A^{\varepsilon}_{ij}\xi_i\xi_j \,\mathrm{d}y - \fint_{Y^{\delta}_K} A^{\varepsilon}\nabla\omega_i\cdot\nabla\omega_j\,\xi_i\xi_j \,\mathrm{d}y \\
&= \fint_{Y^{\delta}_K} A^{\varepsilon}_{ij}\xi_i\xi_j \,\mathrm{d}y + \fint_{Y^{\delta}_K} A^{\varepsilon} e_i\cdot\nabla\omega_j\,\xi_i\xi_j \,\mathrm{d}y \\
&= \fint_{Y^{\delta}_K} A^{\varepsilon}_{ij}\xi_i\xi_j \,\mathrm{d}y - \frac{1}{|Y^{\delta}_K|}\int_{Y^{\delta}_K}\Big[\fint_{Y^{\delta}_K} A^{\varepsilon}\,\mathrm{d}y - A^{\varepsilon}\Big] e_i\cdot\nabla\omega_j\,\xi_i\xi_j \,\mathrm{d}y \\
&\geq \alpha|\xi|^2 - \frac{1}{|Y^{\delta}_K|}\Big\|\fint_{Y^{\delta}_K} A^{\varepsilon}\,\mathrm{d}y - A^{\varepsilon}\Big\|_{L^{\infty}(Y^{\delta}_K)}\sqrt{|Y^{\delta}_K|}\,\|\nabla\omega_j\|_{L^2(Y^{\delta}_K)}\,\xi_i\xi_j \\
&\geq \alpha|\xi|^2 - \frac{1}{\alpha}\Big\|\fint_{Y^{\delta}_K} A^{\varepsilon}\,\mathrm{d}y - A^{\varepsilon}\Big\|^2_{L^{\infty}(Y^{\delta}_K)}\xi_i\xi_j. \qquad (3.80)
\end{aligned}$$

Now, apply the result from Lemma 3.21. The case of (A4) and (A5) will be proved together with Proposition 3.45 on page 67. □

With this result well-posedness follows immediately:

Corollary 3.43 (Well-posedness)**.** Assume either (A1), or alternatively (A4) and (A5), to hold true. The numerically homogenized problem (3.17) is well defined for the homogenization strategy given in Definition 3.4, i. e. it admits a unique solution.

In the following we prove a convergence result for both cases of assumptions, (A1) or (A4).

Proposition 3.44. Let (A1) – (A3) be fulfilled. Then, the solution u^{δ} of the effective problem (3.17) in case of the homogenization strategy (cf. Definition 3.4) converges against the solution u^{ε} of the model problem (2.10):

$$\big\|\nabla u^{\delta} - \nabla u^{\varepsilon}\big\|_{L^2(\Omega)^{d\times d}} \leq C(\varepsilon)\,\delta. \qquad (3.81)$$

Proof. By virtue of (3.76) in combination with Lemma 3.39 and following the same strategy as in the proof of Lemma 3.42:

$$\begin{aligned}
\big\|\bar{A}^{\delta}_{ij} - A^{\varepsilon}_{ij}\big\|_{L^{\infty}(K)^{d\times d}} &= \Big\|\fint_{Y^{\delta}_K} A^{\varepsilon}_{ij}\,\mathrm{d}y - \fint_{Y^{\delta}_K} A^{\varepsilon}\nabla\omega_i\cdot\nabla\omega_j\,\mathrm{d}y - A^{\varepsilon}\Big\|_{L^{\infty}(K)^{d\times d}} \\
&\leq \Big\|\fint_{Y^{\delta}_K} A^{\varepsilon}_{ij}\,\mathrm{d}y - A^{\varepsilon}\Big\|_{L^{\infty}(K)^{d\times d}} + \frac{1}{\alpha}\Big\|\fint_{Y^{\delta}_K} A^{\varepsilon}_{ij}\,\mathrm{d}y - A^{\varepsilon}\Big\|^2_{L^{\infty}(K)^{d\times d}}. \qquad (3.82)
\end{aligned}$$

The statement follows with Lemma 3.21. □

Assumptions (A4) and (A5) require a different approach. In preparation for the convergence result stated in Proposition 3.48, we prove the following proposition.

Proposition 3.45. Let (A2) – (A5) be fulfilled, in particular assume that $A^\varepsilon(x) = A(x, \frac{x}{\varepsilon})$ a. e. with $A \in C^{0,1}\big(\Omega, L^\infty_{\mathrm{per}}(Y)^{d\times d}\big)$. Then, for δ and ε suitably small it holds

$$\big\|A^0 - \bar{A}^\delta\big\|_{L^\infty(K)} \leq C(\delta + \varepsilon), \tag{3.83}$$

with C only depending on a priori data. Here, A^0 denotes the homogenized matrix (see Section 2.2.1), i. e., it is defined by (2.64),

$$A^0_{ij}(x) = \fint_Y A(x,y)\big(\nabla\tilde{\omega}_i(x,y) + e_i\big)\cdot\big(\nabla\tilde{\omega}_j(x,y) + e_j\big)\,\mathrm{d}y, \tag{3.84}$$

with the solution $\tilde{\omega}_i \in \tilde{H}^1_{\mathrm{per}}(Y)$ of the local cell problem (2.65),

$$\fint_Y A(x,y)\big(\nabla\tilde{\omega}_i(x,y) + e_i\big)\cdot\nabla\varphi(y)\,\mathrm{d}y = 0 \quad \forall\varphi \in \tilde{H}^1_{\mathrm{per}}(Y). \tag{3.85}$$

Proof. As a first step, apply the pull-back defined by the transformation

$$y \mapsto x_0 + \varepsilon y \tag{3.86}$$

on the integral (3.10) defining $\bar{A}^\delta$:

$$\begin{aligned}
\bar{A}^\delta_{ij}(x) &= \fint_{Y_K^\delta} A^\varepsilon_{ij}(y)\,\mathrm{d}y - \fint_{Y_K^\delta} A^\varepsilon(y)\nabla\omega_i(y)\cdot\nabla\omega_j(y)\,\mathrm{d}y \\
&= \big(\frac{\varepsilon}{\delta}\big)^d \int_{\frac{\delta}{\varepsilon}Y} A_{ij}\big(x_0 + \varepsilon y, \frac{x_0}{\varepsilon} + y\big)\,\mathrm{d}y \\
&\quad - \big(\frac{\varepsilon}{\delta}\big)^d \int_{\frac{\delta}{\varepsilon}Y} A\big(x_0 + \varepsilon y, \frac{x_0}{\varepsilon} + y\big)\frac{1}{\varepsilon}\nabla_y\,\omega_i(x_0 + \varepsilon y)\cdot\frac{1}{\varepsilon}\nabla_y\,\omega_j(x_0 + \varepsilon y)\,\mathrm{d}y,
\end{aligned} \tag{3.87}$$

where x_0 denotes the lower left point of the sampling region K with $x \in K$. By virtue of periodicity, shift the integrals in the definition of A^0 by x_0/ε and expand the integration area by the factor δ/ε:

$$\begin{aligned}
A^0_{ij}(x) &= \big(\frac{\varepsilon}{\delta}\big)^d \int_{\frac{\delta}{\varepsilon}Y} A_{ij}\big(x, \frac{x_0}{\varepsilon} + y\big)\,\mathrm{d}y \\
&\quad - \big(\frac{\varepsilon}{\delta}\big)^d \int_{\frac{\delta}{\varepsilon}Y} A\big(x, \frac{x_0}{\varepsilon} + y\big)\nabla_y\,\tilde{\omega}_i\big(\frac{x_0}{\varepsilon} + y\big)\cdot\nabla_y\,\tilde{\omega}_j\big(\frac{x_0}{\varepsilon} + y\big)\,\mathrm{d}y.
\end{aligned} \tag{3.88}$$

Combining these equations leads to

$$\begin{aligned}\bar{A}^{\delta}_{ij}(x)-A^{0}_{ij}(x) &= \Big(\frac{\varepsilon}{\delta}\Big)^{d}\int_{\frac{\delta}{\varepsilon}Y} A_{ij}\Big(x_0+\varepsilon y,\frac{x_0}{\varepsilon}+y\Big)-A_{ij}\Big(x,\frac{x_0}{\varepsilon}+y\Big)\,\mathrm{d}y\\ &+\Big(\frac{\varepsilon}{\delta}\Big)^{d}\cdot\int_{\frac{\delta}{\varepsilon}Y}\Big[A\Big(x,\frac{x_0}{\varepsilon}+y\Big)-A\Big(x_0+\varepsilon y,\frac{x_0}{\varepsilon}+y\Big)\Big]\frac{1}{\varepsilon}\nabla_y\,\omega_i\cdot\frac{1}{\varepsilon}\nabla_y\,\omega_j\,\mathrm{d}y\\ &+\Big(\frac{\varepsilon}{\delta}\Big)^{d}\int_{\frac{\delta}{\varepsilon}Y}A\Big(x,\frac{x_0}{\varepsilon}+y\Big)\Big[\nabla_y\,\tilde{\omega}_i\Big(\frac{x_0}{\varepsilon}+y\Big)\cdot\nabla_y\,\tilde{\omega}_j\Big(\frac{x_0}{\varepsilon}+y\Big)\\ &\qquad-\frac{1}{\varepsilon}\nabla_y\,\omega_i(x_0+\varepsilon y)\cdot\frac{1}{\varepsilon}\nabla_y\,\omega_j(x_0+\varepsilon y)\Big]\mathrm{d}y. \quad (3.89)\end{aligned}$$

The first integral term can be directly bounded with the help of the Lipschitz continuity and Lemma 3.39:

$$\begin{aligned}\Big(\frac{\varepsilon}{\delta}\Big)^{d}&\Big\|A\Big(x_0+\varepsilon y,\frac{x_0}{\varepsilon}+y\Big)-A\Big(x,\frac{x_0}{\varepsilon}+y\Big)\Big\|_{L^\infty\left(\frac{\delta}{\varepsilon}Y\right)}\\ &\le C\,\|x_0+\varepsilon y-x\|\le C\,(\delta+\varepsilon), \quad (3.90)\end{aligned}$$

because $\frac{\varepsilon}{\delta}\le 1$. For the second term observe that

$$\begin{aligned}&\Big(\frac{\varepsilon}{\delta}\Big)^{d}\int_{\frac{\delta}{\varepsilon}Y}\Big[A\Big(x,\frac{x_0}{\varepsilon}+y\Big)-A\Big(x_0+\varepsilon y,\frac{x_0}{\varepsilon}+y\Big)\Big]\frac{1}{\varepsilon}\nabla_y\,\omega_i\cdot\frac{1}{\varepsilon}\nabla_y\,\omega_j\,\mathrm{d}y\\ &\le\Big\|A\Big(x_0+\varepsilon y,\frac{x_0}{\varepsilon}+y\Big)-A\Big(x,\frac{x_0}{\varepsilon}+y\Big)\Big\|_{L^\infty\left(\frac{\delta}{\varepsilon}Y\right)}\\ &\qquad\times\Big(\frac{\varepsilon}{\delta}\Big)^{d}\Big[\Big\|\frac{1}{\varepsilon}\nabla_y\,\omega_i(x_0+\varepsilon y)\Big\|^2_{L^2\left(\frac{\delta}{\varepsilon}\right)}+\Big\|\frac{1}{\varepsilon}\nabla_y\,\omega_j(x_0+\varepsilon y)\Big\|^2_{L^2\left(\frac{\delta}{\varepsilon}\right)}\Big]. \quad (3.91)\end{aligned}$$

The gradient of ω_i arising in the second term can be uniformly bounded. Transforming back and employing Lemma 3.39 yields

$$\begin{aligned}\Big(\frac{\varepsilon}{\delta}\Big)^{d}\int_{\frac{\delta}{\varepsilon}Y}\Big|\frac{1}{\varepsilon}\nabla_y\,\omega_i(x_0+\varepsilon y)\Big|^2\,\mathrm{d}y &= \int_{Y_K^\delta}\big|\nabla_y\omega_i(y)\big|^2\,\mathrm{d}y\\ &\le C\,\frac{1}{\alpha^2}\,|Y_K^\delta|\,\|A\|_{L^\infty(\Omega\times Y)}\le C. \quad (3.92)\end{aligned}$$

For the last integral term, add and subtract

$$A\Big(x,\frac{x_0}{\varepsilon}+y\Big)\nabla_y\,\tilde{\omega}_i\Big(\frac{x_0}{\varepsilon}+y\Big)\cdot\frac{1}{\varepsilon}\nabla_y\,\omega_j(x_0+\varepsilon y) \quad (3.93)$$

and utilize symmetry (2.8) to derive an inequality

$$\begin{aligned}
&\Big|\big(\tfrac{\varepsilon}{\delta}\big)^d \int_{\frac{\delta}{\varepsilon}Y} A\big(x,\tfrac{x_0}{\varepsilon}+y\big)\Big[\nabla_y \tilde{\omega}_i\big(\tfrac{x_0}{\varepsilon}+y\big)\cdot\nabla_y \tilde{\omega}_j\big(\tfrac{x_0}{\varepsilon}+y\big) \\
&\qquad\qquad -\frac{1}{\varepsilon}\nabla_y \omega_i(x_0+\varepsilon y)\cdot\frac{1}{\varepsilon}\nabla_y \omega_j(x_0+\varepsilon y)\Big]\mathrm{d}y\Big| \\
&\leq \big(\tfrac{\varepsilon}{\delta}\big)^d \|A\|_{L^\infty}\Big[\big\|\nabla_y\tilde{\omega}_i\big\|_{L^2(\frac{\delta}{\varepsilon}Y)}\big\|\nabla_y\tilde{\omega}_j\big(\tfrac{x_0}{\varepsilon}+y\big)-\frac{1}{\varepsilon}\nabla_y\omega_j\big(x_0+\varepsilon y\big)\big\|_{L^2(\frac{\delta}{\varepsilon}Y)} \\
&\qquad +\big\|\nabla_y\tilde{\omega}_j\big\|_{L^2(\frac{\delta}{\varepsilon}Y)}\big\|\nabla_y\tilde{\omega}_i\big(\tfrac{x_0}{\varepsilon}+y\big)-\frac{1}{\varepsilon}\nabla_y\omega_i\big(x_0+\varepsilon y\big)\big\|_{L^2(\frac{\delta}{\varepsilon}Y)}\Big]. \qquad (3.94)
\end{aligned}$$

It remains to show that

$$\big(\tfrac{\varepsilon}{\delta}\big)^{d/2}\Big\|\nabla_y\tilde{\omega}_i\big(\tfrac{x_0}{\varepsilon}+y\big)-\frac{1}{\varepsilon}\nabla_y\omega_i\big(x_0+\varepsilon y\big)\Big\|_{L^2(\frac{\delta}{\varepsilon}Y)}=\mathcal{O}\big(\delta+\varepsilon\big) \qquad (3.95)$$

and that the following stability estimate holds true:

$$\big(\tfrac{\varepsilon}{\delta}\big)^{d/2}\Big[\big\|\frac{1}{\varepsilon}\nabla\omega_i\big\|_{L^2\left(\frac{\delta}{\varepsilon}Y\right)}+\big\|\nabla\omega_i\big\|_{L^2\left(\frac{\delta}{\varepsilon}Y\right)}\Big]\leq\frac{1}{\alpha}\big(\frac{\delta}{\varepsilon}\big)^{d/2}\|A^\varepsilon\|_{L^\infty\left(\frac{\delta}{\varepsilon}Y\right)}. \qquad (3.96)$$

Rescaling the corresponding cell problems (3.11) and (2.65) in the same manner as done in (3.87) and (3.88) and subtracting the result yields

$$\begin{aligned}
&\int_{\frac{\delta}{\varepsilon}Y} A\big(x_0+\varepsilon y,\tfrac{x_0}{\varepsilon}+y\big)\frac{1}{\varepsilon}\nabla_y\omega_i(x_0+\varepsilon y)\cdot\nabla\varphi\,\mathrm{d}y \\
&\qquad=\int_{\frac{\delta}{\varepsilon}Y} A\big(x,\tfrac{x_0}{\varepsilon}+y\big)\nabla_y\tilde{\omega}_i\big(\tfrac{x_0}{\varepsilon}+y\big)\cdot\nabla\varphi\,\mathrm{d}y \\
&\qquad\qquad+\int_{\frac{\delta}{\varepsilon}Y}\Big[A\big(x,\tfrac{x_0}{\varepsilon}+y\big)-A\big(x_0+\varepsilon y,\tfrac{x_0}{\varepsilon}+y\big)\Big]e_i\cdot\nabla\varphi\,\mathrm{d}y. \qquad (3.97)
\end{aligned}$$

Following the same strategy as used in the proof of Lemma 3.23 by testing with $\frac{1}{\varepsilon}\omega_i-\tilde{\omega}_i$ leads to

$$\begin{aligned}
&\Big\|\nabla_y\tilde{\omega}_i\big(\tfrac{x_0}{\varepsilon}+y\big)-\frac{1}{\varepsilon}\nabla_y\omega_i\big(x_0+\varepsilon y\big)\Big\|_{L^2(\frac{\delta}{\varepsilon}Y)} \\
&\qquad\leq\frac{1}{\alpha}\big\|A\big(x_0+\varepsilon y,\tfrac{x_0}{\varepsilon}+y\big)-A\big(x,\tfrac{x_0}{\varepsilon}+y\big)\big\|_{L^\infty\left(\frac{\delta}{\varepsilon}Y\right)} \\
&\qquad\qquad\times\Big(\big(\frac{\delta}{\varepsilon}\big)^{d/2}+\|\nabla\tilde{\omega}\|_{L^2(\frac{\delta}{\varepsilon}Y)}\Big). \qquad (3.98)
\end{aligned}$$

Finally, Equation (3.96) follows by a similar argument as used in the proof of Lemma 3.39. □

Remark 3.46. Only the fact that δ is an integral multiple of ε was used in the proof, the factor δ/ε does not have to be bounded. In fact, assumption (A5) could be weakened to

(A6) δ_K is a (not necessarily uniform) integral multiple of ε for $K \in \mathbb{T}_\delta$, $\delta > 0$.

Remark 3.47. If the requirement of δ being an integral multiple of ε is lifted, Equation (3.88) does not hold anymore. Instead, it is necessary to estimate

$$\left|A_{ij}^{\delta}(x) - A_{ij}^{0}(x)\right| \leq \left|A_{ij}^{\delta}(x) - A_{ij}^{\tilde{\delta}}(x)\right| + \left|A_{ij}^{\tilde{\delta}}(x) - A_{ij}^{0}(x)\right|, \tag{3.99}$$

with $\tilde{\delta}$ being defined as the next integral multiple of ε smaller than δ. The term $\left|A_{ij}^{\delta}(x) - A_{ij}^{\tilde{\delta}}(x)\right|$ exhibits a *resonance error*[59] that only allows for an a priori estimate

$$\left|A_{ij}^{\delta}(x) - A_{ij}^{\tilde{\delta}}(x)\right| \leq C\left[\delta + \frac{\varepsilon}{\delta}\right]. \tag{3.100}$$

In summary,

$$\left\|A^0 - A^\delta\right\|_{L^\infty(K)} \leq C\left[\delta + \varepsilon + \frac{\varepsilon}{\delta}\right]. \tag{3.101}$$

Proof. With the help of the two auxiliary estimates

$$\begin{aligned}\left|\fint_{Y_K^\delta} A(x, \tfrac{x}{\varepsilon})\,\mathrm{d}y - \fint_{Y_K^{\tilde{\delta}}} A(x, \tfrac{x}{\varepsilon})\,\mathrm{d}y\right| &\leq \frac{1}{|Y_K^\delta|}\left|\int_{Y_K^\delta \setminus Y_K^{\tilde{\delta}}} A(x, \tfrac{x}{\varepsilon})\,\mathrm{d}y\right| \\ &\leq \frac{1}{\delta^d}\,\delta^{d-1}\,\varepsilon\,\|A\|_{L^\infty(\Omega\times Y)} \leq C\frac{\varepsilon}{\delta}\end{aligned} \tag{3.102}$$

and

$$\|\nabla\omega_i\|^2_{L^2\left(Y_K^\delta\right)^d} + \|\nabla\hat{\omega}_i^\delta\|^2_{L^2\left(Y_K^\delta\right)^d} \leq C \tag{3.103}$$

the statement follows analogous to Proposition 3.48. □

By virtue of Proposition 3.45 and with the help of classical homogenization results convergence against the homogenized solution u^0 follows directly:

Proposition 3.48. Under the same assumptions (A2) – (A5) as used in Proposition 3.45, the solution u^δ converges to u^0 in H^1,

$$\left\|u^0 - u^\delta\right\|_{H^1(\Omega)} \leq C\big(\delta + \varepsilon\big), \tag{3.104}$$

with a constant C independent of δ and ε, and for δ, ε suitably small.

Proof. The results follow immediately from Lemma 3.23 and Proposition 3.45 together with the a priori estimate

$$\|u^\varepsilon\|_{H^1(\Omega)} \leq C\|f\|_{L^2(\Omega)}, \tag{3.105}$$

with C independent of ε, resulting from the uniform ellipticity assumption (2.9). □

Finally, combining the result of Proposition 3.48 with the a priori convergence of u^ε against u^0, which was established in Proposition 2.30, leads to a convergence result:

Proposition 3.49 (Convergence)**.** Let $A \in C^{0,1}\big(C^{0,1}(Y)\big)$. Then,

$$\big\|u^\varepsilon - u^\delta\big\| \leq C\,\varepsilon, \tag{3.106}$$

$$\big\|\nabla u^\varepsilon - \varepsilon \nabla_y u^1(x, \frac{x}{\varepsilon}) - \nabla u^\delta\big\| \leq C\,\varepsilon^{1/2}. \tag{3.107}$$

Unfortunately u^ε only converges to u^0 weakly in H^1, for strong convergence in H^1 the corrector u^1 is needed. At least we can formulate the following corollary:

Corollary 3.50. Under the assumption stated in Proposition 3.48, the difference u^ε and u^δ remains uniformly bounded with respect to δ and ε, i. e.

$$\big\|u^\varepsilon - u^\delta\big\|_{H^1(\Omega)} \leq C \quad \forall\, \varepsilon \leq \varepsilon_0\, \forall\, \delta \leq \delta_0, \tag{3.108}$$

where ε_0 and δ_0 only depend on a priori data.

Semi-discretized problem Similarly to the averaging strategy, we establish a convergence result for the weaker case of a periodic coefficient A^ε as defined in (A4). In order to estimate quadrature errors and discretization errors separately, introduce an auxiliary problem. Given a finite-element space $V^h\big(Y_K^\delta\big)$, let $\hat{A}_{ij}^{\delta,h}$ be defined as

$$\hat{A}_{ij}^{\delta,h}(K) \coloneqq \fint_{Y_K^\delta} A^\varepsilon(x)\big(\nabla_x \hat{\omega}_i^h(x) + \boldsymbol{e}_i\big) \cdot \big(\nabla_x \hat{\omega}_j^h(x) + \boldsymbol{e}_j\big)\,\mathrm{d}x, \tag{3.109}$$

where the $\hat{\omega}_i^h \in V^h\big(Y_K^\delta\big) \subset \tilde{H}^1_{\mathrm{per}}(Y_K^\delta)$ are solutions of

$$\int_{Y_K^\delta} A^\varepsilon(x)\big(\nabla_x \hat{\omega}_i^h(x) + \boldsymbol{e}_i\big) \cdot \nabla\varphi = 0 \quad \forall \varphi \in V^h\big(Y_K^\delta\big). \tag{3.110}$$

Lemma 3.51. Let $A_{ij}^{\delta,h}$ be defined according to (3.22) and (3.23) with a summed quadrature rule $Q_{h,K}$ associated with $\mathbb{T}_h(Y_K^\delta)$ as defined in (3.18). Then, under assumptions (A2) – (A5) it holds true that

$$\left|A_{ij}^{\delta,h}(x) - \hat{A}_{ij}^{\delta,h}(x)\right| \leq C\,\frac{h}{\varepsilon}. \tag{3.111}$$

Proof. Utilizing (3.109) and (3.22), the error $\left|A_{ij}^{\delta,h}(x) - \hat{A}_{ij}^{\delta,h}(x)\right|$ can be reduced to a quadrature error for approximating A^ε:

$$\begin{aligned}
&\left|A_{ij}^{\delta,h}(x) - \hat{A}_{ij}^{\delta,h}(x)\right| \\
&\quad\leq \left|\frac{1}{|Y_K^\delta|}Q_{h,K}\big(A_{ij}^\varepsilon\big) - \fint_{Y_K^\delta} A_{ij}^\varepsilon \,\mathrm{d}y\right| \\
&\qquad + \left\|\frac{1}{|Y_K^\delta|}Q_{h,K}\big(A_{ij}^\varepsilon\big) - \fint_{Y_K^\delta} A_{ij}^\varepsilon \,\mathrm{d}y\right\| \left\|\nabla\omega_i^h\right\|_{L^2\left(Y_K^\delta\right)^d} \left\|\nabla\omega_j^h\right\|_{L^2\left(Y_K^\delta\right)^d} \\
&\qquad + \left\|\fint_{Y_K^\delta} A_{ij}^\varepsilon \,\mathrm{d}y\right\| \left\|\nabla\omega_i^h - \nabla\hat{\omega}_i^h\right\|_{L^2\left(Y_K^\delta\right)^d} \left\|\nabla\omega_j^h - \nabla\hat{\omega}_i^h\right\|_{L^2\left(Y_K^\delta\right)^d},
\end{aligned} \tag{3.112}$$

where, shown by similar techniques as used in the proofs of Proposition 3.45 and Lemma 3.28,

$$\begin{aligned}
\left\|\nabla\omega_i^h - \nabla\hat{\omega}_i^h\right\|_{L^2\left(Y_K^\delta\right)^d} &\leq \frac{1}{\alpha}\Big[\sup_{K\in\mathbb{T}_h\left(Y_K^\delta\right)}\sup_{x,y\in K}\left\|A^\varepsilon(x) - A^\varepsilon(y)\right\|\Big]\left\|\nabla\hat{\omega}_i^h\right\| \\
&\leq \frac{1}{\alpha}\,\frac{h}{\varepsilon}\left\|\nabla\hat{\omega}_i^h\right\|.
\end{aligned} \tag{3.113}$$

In analogy of Lemma 3.28 it can be shown that

$$\left\|\frac{1}{|Y_K^\delta|}Q_{h,K}\big(A_{ij}^\varepsilon\big) - \fint_{Y_K^\delta} A_{ij}^\varepsilon \,\mathrm{d}y\right\| \leq C\,\frac{h}{\varepsilon}. \tag{3.114}$$

□

Lemma 3.52. Assume (A2) – (A5) to be fulfilled. Let $\hat{\omega}_i^h \in V^h\big(Y_K^\delta\big)$ be the solution of (3.110) and $\omega_i \in \tilde{H}^1_{\mathrm{per}}(Y_K^\delta)$ be the solution of (3.11). Then, by virtue of the Lipschitz continuity of A^ε, the ω_i are already of class $H^2(Y_K^\delta)$ and the following a priori estimate holds true:

$$\left\|\nabla\omega_i - \nabla\hat{\omega}_i^h\right\|_{L^2(Y_K^\delta)^d} \leq C\,\sqrt{|Y_K^\delta|}\,\frac{h}{\delta}, \tag{3.115}$$

with a constant C only depends on a priori data independent of δ and h.

Proof. For the regularity result $\omega_i \in H^2(Y_K^\delta)$ in case of Lipschitz continuous coefficients A^ε on the rescaled unit square Y_K^δ we refer to Grisvard[51, Th. 3.2.1.2]; the corresponding regularity result by Grisvard can be immediately adapted to the case of periodic boundary conditions and right hand side $(A^\varepsilon \boldsymbol{e}_i \nabla .)$. Then, by standard finite-element approximation theory:

$$\left\|\nabla\omega_i - \nabla\hat{\omega}_i^h\right\|_{L^2(Y_K^\delta)^d} \leq C\, h \left\|\nabla^2\omega_i\right\|_{L^2(Y_K^\delta)^{d\times d}}. \tag{3.116}$$

A rescaling argument now shows that

$$\left\|\nabla^2\omega_i\right\|_{L^2(Y_K^\delta)^{d\times d}} \sim \sqrt{|Y_K^\delta|}\frac{1}{\delta}. \tag{3.117}$$

□

Lemma 3.53. Let A_{ij}^δ be defined according to (3.10) and (3.11). Then, under the same assumptions as stated in Lemma 3.51, it holds

$$\left|A_{ij}^\delta(x) - \hat{A}_{ij}^{\delta,h}(x)\right| \leq C\,\frac{h}{\delta}. \tag{3.118}$$

Proof. Utilizing the definition of A^δ and $A^{\delta,h}$ leads to

$$\begin{aligned}\left|A_{ij}^\delta(K) - A_{ij}^{\delta,h}(K)\right| &= \left|\fint_{Y_K^\delta} A^\varepsilon \boldsymbol{e}_i \cdot \left(\nabla\omega_j - \nabla\hat{\omega}_j^h\right)\mathrm{d}y\right| \\ &\leq C\,\frac{1}{|Y_K^\delta|}\left\|A^\varepsilon\right\|_{L^\infty\left(Y_K^\delta\right)}\sqrt{|Y_K^\delta|}\left\|\nabla\omega_i - \nabla\hat{\omega}_i^h\right\|_{L^2\left(Y_K^\delta\right)^d}\end{aligned} \tag{3.119}$$

The statement now follows by virtue of Lemma 3.52. □

With this prerequisites at hand a convergence result can be established.

Proposition 3.54. In case of the homogenization strategy and under the same assumptions as in Lemma 3.51:

$$\left\|A^{\delta,h} - A^\delta\right\|_{L^\infty(\Omega)} \leq C\left(\frac{h}{\varepsilon} + \frac{h}{\delta}\right). \tag{3.120}$$

This estimate leads to the error estimate (with C only depending on a priori data):

$$\left\|u^{\delta,h} - u^\delta\right\|_{H^1(\Omega)} \leq C\left(\frac{h}{\varepsilon} + \frac{h}{\delta}\right). \tag{3.121}$$

Proof. The first statement is a consequence of Lemma 3.51 and 3.53. The second statement follows with the same technique as used in Lemma 3.23 and the previous propositions. □

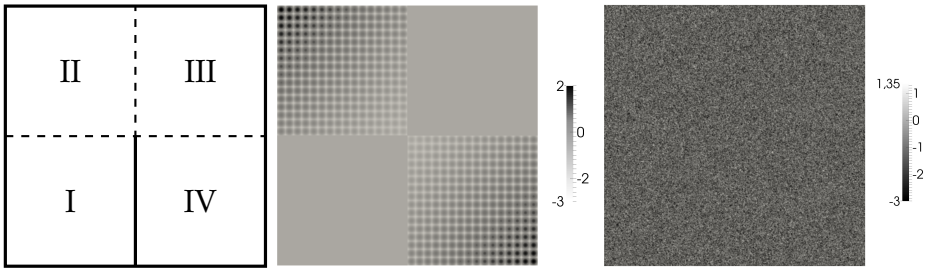


Figure 3.3. The slit domain $\Omega = [0,1]^2 \setminus s$ together with the 4 quadrants (a). Computations are done for an artificial periodic coefficient (b) and a log-normally distributed permeability (c). Both coefficients are shown in logarithmic scale with respect to base 10.

3.5 Numerical validation

The a priori convergence results shall be examined with a series of small numerical tests. Those are done with a C++[1] program that utilizes the finite-element library DEAL.II[15,16] as a general framework, the QUANTIM[99] library for generating stochastic coefficients as well as the UMFPACK solver from the SUITESPARSE[37,38] collection as a direct solver for cell problems.

Two different kinds of heterogeneous coefficients are examined. The first one is an artificial, periodic structure with quadrant-wise different character (cf. Figure 3.3):

$$A^\varepsilon(x) = I_d\,\gamma \begin{cases} \exp(3) & \text{in I and III,} \\ \exp\big(6(1-x_1)\,x_2\,[\cos(\pi\hat{x}_1)+\cos(\pi\hat{x}_2)]\big) & \text{in II,} \\ \exp\big(6(1-x_2)\,x_1\,[\cos(\pi\hat{x}_1)+\cos(\pi\hat{x}_2)]\big) & \text{in IV.} \end{cases} \tag{3.122}$$

Here, $\gamma = 0.001$ and the rescaling $\hat{x}_i$ is defined as $\hat{x}_i := \lfloor x_i/\varepsilon \rfloor - 1/2$ with the notation $\lfloor . \rfloor$ for the *floor function* that selects the largest previous integral value. For the periodic structure a mild scale separation with $\varepsilon = 2^{-5}$ is chosen. The second example is given by a log-normally distributed random field with Gaussian correlation: $A^\varepsilon(x) = I_d \times \gamma \times \exp(10 \times g(x)/255)$, where $g(x)$ is an 8 bit grayscale picture (with integral values between 0 and 255) with 1024×1024 pixels resolution (see Figure 3.3). A Gaussian correlation with a strong scale separation due to a small correlation length of $r = 0.0025$ is chosen.

Due to the exponential scaling the microscale fluctuations in the coefficients exhibit a strong influence on the macroscale in both cases. This manifests in the fact that a high resolution that resolves the finescale completely is necessary for a good approximation. The periodic structure is an example that fulfills the periodicity assumption of the modified HMM scheme. It exhibits a quadrant-wise different behavior with a heterogeneous character. Both coefficients are tested on the slit domain $\Omega = [0,1]^2 \setminus s$ (see Figure 3.3).

3.5.1 Uniform refinement

In a first numerical test the convergence rate in the $L^2(\Omega)$- and $H^1(\Omega)$-norm for a standard finite-element discretization under uniform refinement is examined. For this, a reference solution with 1.67×10^7 cells is computed against which intermediate results are compared. The results are given in Table 3.1. The right column for each norm is the logarithmic reduction rate, i. e., $\log_2(||e_k||/||e_{k-1}||)$. A reduced convergence rate can be observed. Compared to the corresponding Laplace problem with smooth coefficient, a drastically increased resolution of several orders of magnitude is necessary in order to reach the same relative error. Further, even for a very fine discretization of over 1×10^6 cells, a relative error of less than 5 % in the $H^1(\Omega)$-norm cannot be achieved. In conclusion, a clear two-scale behavior with a significant influence of the microscopic permeability on the macroscale can be observed.

3.5.2 Model error of averaging schemes and HMM scheme

Another point of interest is the typical numerical range of the model error. To get an estimate for it, a computation with high resolution on macroscale and microscale discretization is performed. 2.56×10^6 cells on macroscale and 1.0×10^6 cells on microscale are chosen (in case of HMM 4.0×10^6). The sampling mesh $\mathbb{T}_\delta(\Omega)$ is chosen to be a uniform refinement of Ω. For simplicity, the individual sampling regions are fixed to the full, respective sampling cell, $Y_K^\delta = K$. Further, $\delta = \varepsilon = 2^{-5}$ is used in the case of the periodic structure and $\delta = 2^{-3}$, as well as 2^{-5}, in the case of the randomly distributed permeability is chosen. The results are given in Table 3.2. It can be seen that the effective macroscale behavior expressed in the $L^2(\Omega)$-norm can be reproduced by the HMM scheme and by the geometric averaging approach. In contrast to this, the effective coefficients based on arithmetic and harmonic mean values result in very large relative errors of around 50 %. The relative error with respect to the $H^1(\Omega)$-norm (which also captures local oscillations) remains large throughout.

Table 3.1. Error development in L^2- and H^1-norm under uniform refinement for both types of permeabilities (given in absolute and relative error, as well as the logarithmic reduction rate with respect to base 2).

	periodic coefficient				random coefficient			
# macro	$L^2(\Omega)$		$H^1(\Omega)$		$L^2(\Omega)$		$H^1(\Omega)$	
256	2.9e-1 (35%)	—	3.5e-0 (59%)	—	7.0e-2 (41%)	—	8.7e-1 (68%)	—
1024	2.6e-1 (32%)	0.1	3.3e-0 (55%)	0.1	6.8e-2 (40%)	0.0	8.6e-1 (67%)	0.0
4096	2.6e-1 (31%)	0.0	3.3e-0 (55%)	0.0	6.4e-2 (37%)	0.1	8.3e-1 (65%)	0.1
16384	1.2e-1 (15%)	1.0	2.3e-0 (39%)	0.5	5.1e-2 (30%)	0.3	7.4e-1 (58%)	0.2
65536	4.1e-2 (4.9%)	1.6	1.4e-0 (23%)	0.8	2.7e-2 (16%)	0.9	5.3e-1 (41%)	0.5
262144	1.2e-2 (1.5%)	1.8	7.8e-1 (13%)	0.8	7.3e-3 (4.3%)	1.9	2.7e-1 (21%)	1.0
1048576	3.7e-3 (0.5%)	1.7	4.6e-1 (7.7%)	0.8	7.3e-4 (0.4%)	3.3	8.4e-2 (6.5%)	1.7

Table 3.2. Model error in L^2- and H^1-norm for the proposed sampling schemes (with arithmetic, geometric and harmonic mean value) and the HMM scheme for (a) the periodic coefficients and (b) the random permeability.

strategy	$L^2(\Omega)$	$H^1(\Omega)$
arithmetic averaging	2.6e-1 (31%)	3.2e-0 (54%)
geometric averaging	5.2e-2 (6.4%)	2.7e-0 (45%)
harmonic averaging	4.4e-1 (54%)	4.0e-0 (67%)
HMM scheme	1.6e-2 (2.0%)	2.7e-0 (45%)

(a) periodic coefficient

	$\delta = 2^{-3}$		$\delta = 2^{-5}$	
strategy	$L^2(\Omega)$	$H^1(\Omega)$	$L^2(\Omega)$	$H^1(\Omega)$
arithmetic av.	6.9e-2 (40%)	8.6e-1 (67%)	6.8e-2 (40%)	8.5e-1 (66%)
geometric av.	3.7e-3 (2.2%)	7.5e-1 (59%)	2.6e-3 (1.5%)	7.5e-1 (59%)
harmonic av.	1.2e-1 (71%)	1.1e-0 (86%)	1.2e-1 (67%)	1.0e-0 (78%)
HMM scheme	2.9e-3 (1.7%)	7.5e-1 (59%)	2.2e-3 (1.3%)	7.5e-1 (59%)

(b) random coefficient

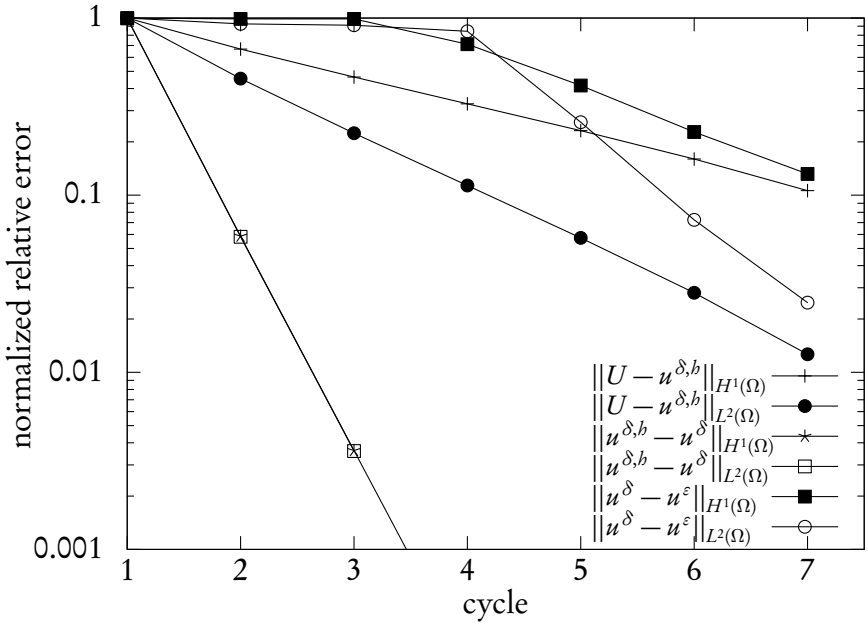
3.5.3 Convergence rates for uniform refinement

Next, we examine the convergence rates achieved under uniform refinement and compare them to the theoretical results given in Propositions 3.25, 3.30, 3.33, and 3.37. This is done for the simple averaging strategy. In light of the numerical results of the first two tests the geometric average is chosen.

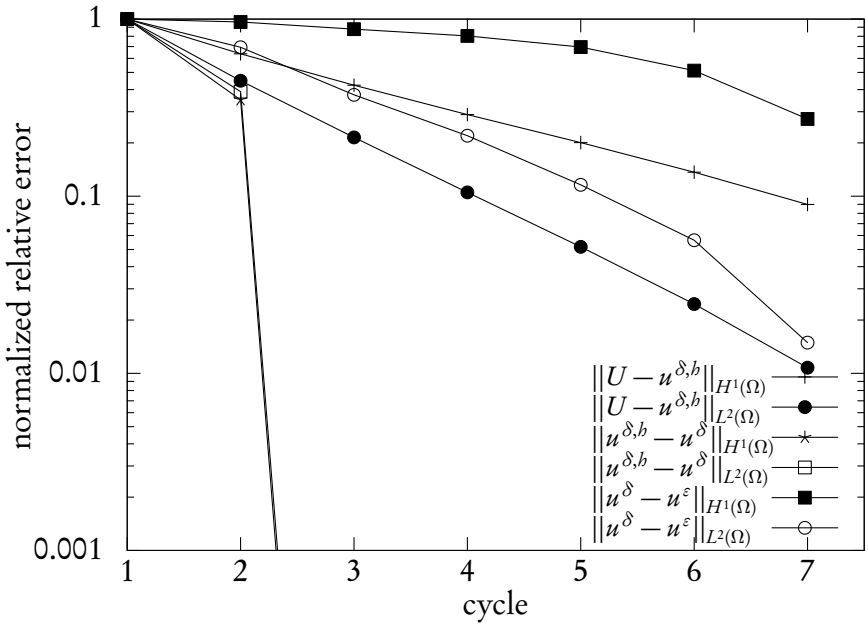
In order to verify the a priori results, additional approximations of u^δ and $u^{\delta,h}$ are computed with 4.0×10^6 cells on the macroscale; with a fine resolution on the microscale ($h = 2^{-12}$) for u^δ and a sequence of decreasing $h = 2^{-4}, \ldots, 2^{-10}$ for $u^{\delta,h}$. Finally, the error $u^\varepsilon - u^\delta$ (as a function of δ), $u^\delta - u^{\delta,h}$ (as a function of h), and $u^\varepsilon - u^\delta$ (as a function of δ) is determined in both, $L^2(\Omega)$- and $H^1(\Omega)$-norm.

The results are depicted in Figure 3.4 in form of a semi-logarithmic plot of the relative error normalized to an initial error of 1. Due to the macroscopic singularity coming from the reentrant corner and the discontinuous coefficient $A^{\delta,h}$ only a reduced convergence rate of $2/3$ in the $H^1(\Omega)$- and $4/3$ in the $L^2(\Omega)$-norm is expected (cf. Proposition 3.37 and Remark 3.36). This can be verified for both types of microstructures.

In contrast, in case of the microscale error $u^\delta - u^{\delta,h}$ under refinement of $\mathbb{T}_h(K)$, a similar behavior for both global norms, i. e., in $L^2(\Omega)$ and $H^1(\Omega)$, is expected—and this is indeed the case. The convergence rate in h is identical for both norms and both types of microstructure. Here, a high convergence order of 4 was achieved in case of the periodic permeability (for a 4^{th} order quadrature rule that was used). In case of the random coefficients, a roughly linear convergence rate can be observed for the first steps. The sudden drop of the microscale error to 10^{-10} is due to the way how the random permeability is implemented, namely, as a piecewise constant coefficient on a 1024×1024 background mesh. A soon as the summed quadrature rule resolves the background mesh, the error drops abruptly. The observed convergence rates under sampling-grid refinement (in δ) need some explanation. For the periodic coefficients (Figure 3.4), at first, no convergence in $H^1(\Omega)$ or $L^2(\Omega)$ is observable. This is due to the fact that up to cycle 3 the parameter δ is still an exact integral multiple of ε. Later, the model error $||\nabla(u^\varepsilon - u^\delta)||$ shows an asymptotic linear convergence order for the periodic coefficients. In case of the random coefficients a reduced convergence order can be observed that increases from ~ 0.1 to linear convergence. This has to be expected as the microstructure is not continuous, but piecewise constant. In both cases, the model error in the $L^2(\Omega)$ norm shows a roughly doubled convergence order as the $H^1(\Omega)$ case: 1.8 instead of 0.9 for the periodic coefficients, linear instead of 0.5 for the random permeability—in agreement with Remark 3.26 and Proposition 3.49.



(a) Periodic coefficients



(b) Random coefficients

Figure 3.4. Convergence order of macroscale, microscale and discretization for (a) periodic and (b) random coefficients. The error is plotted on a semi-logarithmic scale as relative error normalized to an initial error of 1.

4 A posteriori error estimation

In this section we derive *a posteriori* error estimates for the multiscale scheme. This is done in the framework of the *Dual Weighted Residual* (DWR) method introduced by Becker and Rannacher[19–21]. A so-called *dual problem* is solved that is defined in terms of the coefficients A^ε and a *quantity of interest* represented by a linear functional j. The goal of this discussion is to derive independent a posteriori error estimates and local indicators for all sources of error—for the discretization errors on the macro- and microscale, as well as for the model error. The error indicators will be used as the basis of different adaptation strategies formulated in Chapters 5 and 6.

A fundamental difficulty that has to be taken care of arises from the fact that computing the solution of the dual problem (in case of the elliptic model problem) is of the same complexity as the primal problem, but a good approximation of the dual solution is needed for certain types of model adaptation strategies.

Thus, in order to make a numerical evaluation of the dual problem computationally feasible, different strategies to reconstruct the dual solution will be examined. Such reconstruction strategies range from coarse macroscale approximations of the dual solution (with high quadrature) to *local enhancement* by solving localized reconstruction problems.

Remark 4.1. Comparable techniques by using the DWR method in a model-adaptive context have been developed by Oden and Vemaganti[84–87,91] and Braack and Ern[26] (see Section 5.1), albeit with a different treatment of localization and a posteriori control.

4.1 Duality-based error identity

Let $j \in H^{-1}(\Omega)$ be a linear and continuous functional and suppose that a *quantity of interest* is given by the value $\langle j, u^\varepsilon \rangle$. Define a *dual problem*:

Definition 4.2 (Dual problem)**.** Find $z^\varepsilon \in H^1(\Omega)$ s. t.

$$\left(A^\varepsilon \nabla \varphi, \nabla z^\varepsilon\right) = \langle j, \varphi \rangle \quad \forall\, \varphi \in H^1(\Omega). \tag{4.1}$$

The dual problem is well-posed and admits a unique solution.

At this point the usual approach is to transform the error expressed in the quantity of interest, $\langle j, u^\varepsilon\rangle - \langle j, U\rangle$, into a residual tested with the dual solution[19].

$$\langle j, u^\varepsilon\rangle - \langle j, U\rangle = (f, z^\varepsilon) - \langle j, U\rangle = (f, z^\varepsilon) - \left(A^\varepsilon \nabla U, \nabla z^\varepsilon\right). \tag{4.2}$$

By introducing an effective model and a microscale discretization, the situation becomes more complicated. Several different splitting strategies can be used. In principle, it is possible to use the error identity independently between all intermediate problems leading to an identity of the form

$$\begin{aligned}
&\langle j, u^\varepsilon\rangle - \langle j, U\rangle \\
&\quad = \langle j, u^\varepsilon\rangle - \langle j, u^\delta\rangle + \langle j, u^\delta\rangle - \langle j, u^{\delta,h}\rangle + \langle j, u^{\delta,h}\rangle - \langle j, U\rangle \\
&\quad = \left([A^\delta - A^\varepsilon]\nabla u^\delta, \nabla z^\varepsilon\right) + \left([A^{\delta,h} - A^\delta]\nabla u^{\delta,h}, \nabla z^\delta\right) \\
&\qquad + (f, z^{\delta,h}) - \left(A^{\delta,h}\nabla U, \nabla z^{\delta,h}\right).
\end{aligned} \tag{4.3}$$

Here, A^δ, $A^{\delta,h}$ denote either the post-processed coefficients according to Definition 3.8, or its piecewise constant counterpart $\bar{A}^\delta$, $\bar{A}^{\delta,h}$ (depending on which is actually used). z^δ, $z^{\delta,h}$ are solutions of corresponding intermediate dual problems that will be defined below. This form of splitting has the disadvantage of introducing all primal and dual intermediate solutions in the equation. Especially the generally unavailable intermediate solutions u^δ and $u^{\delta,h}$ pose a problem. An alternative approach is to split only between discretization and model contribution:

$$\begin{aligned}
&\langle j, u^\varepsilon\rangle - \langle j, u^\delta\rangle + \langle j, u^\delta\rangle - \langle j, U\rangle \\
&\quad = (f, z^\varepsilon) - \left(A^\varepsilon \nabla u^\delta, \nabla z^\varepsilon\right) + (f, z^\delta) - \left(A^\delta \nabla U, \nabla z^\delta\right) \\
&\quad = \left([A^\delta - A^\varepsilon]\nabla u^\delta, \nabla z^\varepsilon\right) + \left([A^{\delta,h} - A^\delta]\nabla U, \nabla z^\delta\right) \\
&\qquad + (f, z^\delta) - \left(A^\delta \nabla U, \nabla z^\delta\right),
\end{aligned} \tag{4.4}$$

with the solution $z^\delta \in H^1_0(\Omega)$ of an effective, dual solution

$$\left(A^\delta \nabla \varphi, \nabla z^\delta\right) = \langle j, \varphi\rangle \quad \forall\, \varphi \in H^1_0(\Omega). \tag{4.5}$$

This splitting has the advantage that at least both terms that relate to discretization errors entirely consist of evaluable expressions. The question how to efficiently approximate the model error will be examined in detail in Section 4.4. In summary, we have established the following result.

Proposition 4.3 (Error identity)**.** It holds true that

$$\begin{aligned}\langle j, u^\varepsilon\rangle - \langle j, U\rangle &= \underbrace{\big(f, z^\delta\big) - \big(A^{\delta,h}\nabla U, \nabla z^\delta\big)}_{=:\theta^H} + \underbrace{\big(A^{\delta,h}\nabla U, \nabla z^\delta\big) - \big(A^\delta \nabla U, \nabla z^\delta\big)}_{=:\theta^h} \\ &\quad + \underbrace{\big(A^\delta \nabla u^\delta, \nabla z^\varepsilon\big) - \big(A^\varepsilon \nabla u^\delta, \nabla z^\varepsilon\big)}_{=:\theta^\delta}, \end{aligned} \tag{4.6}$$

with the following *error estimators*: θ^H, which is a *residual on the macroscale*, θ^h that takes the form of a *residual on the microscale*, and θ^δ that estimates the model error.

The macroscale and microscale error estimators can be rewritten in the usual form of a residual tested with the dual solution. By employing Galerkin orthogonality and partial integration (in case of the macroscale error estimator) the error estimators can be localized and split into local error indicators.

Proposition 4.4. The macroscale error estimator θ^H permits the representation

$$\theta^H = \sum_{K\in\mathbb{T}_H(\Omega)} \eta_K^H, \tag{4.7}$$

$$\begin{aligned}\eta_K^H &:= \big(f + \nabla\cdot A^{\delta,h}\nabla U, z^\delta - \varphi^H\big)_{L^2(K)} \\ &\quad - \frac{1}{2}\big(\big[\boldsymbol{n}\cdot A^{\delta,h}\nabla U\big]_{\partial Q}, z^\delta - \varphi^H\big)_{L^2(\partial K)^d}\end{aligned} \tag{4.8}$$

for arbitrary $\varphi^H \in V_H(\Omega)$. Here, $[\,.\,]_{\partial K}$ denotes the jump over ∂K and $\boldsymbol{n}$ is the outward unit normal. The microscale error indicator is given by

$$\theta^h = \sum_{K\in\mathbb{T}_\delta(\Omega)} \eta_K^h, \tag{4.9}$$

$$\eta_K^h := \big(\{A^\delta - A^{\delta,h}\}\nabla U, \nabla z^\delta\big)_K. \tag{4.10}$$

Finally, based on the definition of θ^δ, also define local *model-error* indicators

$$\theta^\delta = \sum_{K\in\mathbb{T}_\delta(\Omega)} \eta_K^\delta, \tag{4.11}$$

$$\eta_K^\delta := \big(\{A^\varepsilon - A^\delta\}\nabla u^\delta, \nabla z^\varepsilon\big)_K. \tag{4.12}$$

Remark 4.5. In case of the homogenization sampling strategy (3.10), the microscale error estimator can be expressed in terms of microscale residuals (neglecting quadrature):

$$\eta_K^h \coloneqq \sum_{ij} \int_K \nabla_i U \nabla_j z^\delta \,\mathrm{d}x \, \frac{1}{|Y_K^\delta|} \sum_{Q \in \mathbb{T}_h(Y_K^\delta)} \int_Q A^\varepsilon \big(-\nabla \omega_j^h - \boldsymbol{e}_j\big) \cdot \big(\nabla \omega_i - \nabla \psi_K^h\big) \,\mathrm{d}y, \quad (4.13)$$

with a choice $\psi_K^h \in V^h(Y_K^\delta)$ due to Galerkin orthogonality.

Proof of proposition and remark. Equation (4.7) immediately follows by virtue of Galerkin orthogonality and by partial integration,

$$\begin{aligned}\theta^H &= \sum_{K \in \mathbb{T}_H(\Omega)} \big(f, z^\delta - \pi_H z^\delta\big)_{L^2(K)} - \big(A^{\delta,h} \nabla U, \nabla z^\delta - \nabla \pi_H z^\delta\big)_{L^2(K)^d} \\ &= \sum_{K \in \mathbb{T}_H(\Omega)} \big(f + \nabla \cdot A^{\delta,h} \nabla U, z^\delta - \pi_H z^\delta\big)_{L^2(K)} \\ &\qquad - \frac{1}{2} \Big(\big[\boldsymbol{n} \cdot A^{\delta,h} \nabla U\big]_{\partial K}, z^\delta - \nabla \pi_H z^\delta \Big)_{L^2(\partial K)^d}. \quad (4.14)\end{aligned}$$

Equation (4.13) is a consequence of a straightforward calculation with the identity

$$⨍_{Y_K^\delta} A^\varepsilon \nabla \omega_i^h \cdot \nabla \omega_j^h \,\mathrm{d}y = ⨍_{Y_K^\delta} A^\varepsilon \nabla \omega_i^h \cdot \nabla \omega_j \,\mathrm{d}y \quad (4.15)$$

and the symmetric characterization (3.76) of ω_i and ω_i^h:

$$\begin{aligned}\theta^h &= \sum_{K \in \mathbb{T}_\delta(\Omega)} \int_K \big(A^{\delta,h} - A^\delta\big) \nabla U \cdot \nabla z^\delta \,\mathrm{d}x \\ &= \sum_{K \in \mathbb{T}_\delta(\Omega)} \sum_{ij} \int_K ⨍_{Y_K^\delta} A^\varepsilon(y) \nabla_y \omega_i \cdot \nabla_y \omega_j - A^\varepsilon(y) \nabla_y \omega_i^h \cdot \nabla_y \omega_j^h \,\mathrm{d}y \nabla_i U \nabla_j z^\delta \,\mathrm{d}x \\ &= \sum_{K \in \mathbb{T}_\delta(\Omega)} \sum_{ij} \int_K \nabla_i U \nabla_j z^\delta \,\mathrm{d}x ⨍_{Y_K^\delta} -A^\varepsilon(y) \boldsymbol{e}_i \cdot \nabla_y \omega_j - A^\varepsilon(y) \nabla_y \omega_i^h \cdot \nabla_y \omega_j \,\mathrm{d}y \\ &= \sum_{K \in \mathbb{T}_\delta(\Omega)} \sum_{ij} \int_K \nabla_i U \nabla_j z^\delta \,\mathrm{d}x \, \frac{1}{|Y_K^\delta|} \\ &\qquad \times \sum_{Q \in \mathbb{T}_h\left(Y_K^\delta\right)} \int_Q A^\varepsilon(y) \big[-\boldsymbol{e}_i - \nabla_y \omega_i^h\big] \cdot \big[\nabla_y \omega_j - \nabla \psi_K^h\big] \mathrm{d}y. \quad (4.16)\end{aligned}$$

□

Remark 4.6. Residual based a posteriori results for the discretization errors for the underlying HMM were already presented by Ohlberger[88], Henning and Ohlberger[53], Henning et al.[55], and Abdulle[3]. Corresponding *goal-oriented* error estimates for discretization errors on macro- and microscale were formulated by Abdulle and Nonnenmacher[4]. Those are similar to the local error estimators η_K^H and η_K^h developed for the abstract framework. Furthermore, goal-oriented a posteriori results for the VMM ansatz that use the same methodology were derived by Larson and Målqvist[67–69], see also Section 3.1.

4.2 Efficiency of the error estimators

The residual-type error indicators θ^H and θ^h are well behaved, i. e. they are uniformly bounded in powers of H and h, respectively. Let $\pi_H : H_0^1(\Omega) \to V^H(\Omega)$ be the Scott-Zhang interpolant[94] defined on $H_0^1(\Omega)$. Or alternatively, let π_H be any other choice of interpolation that fulfills the local estimate

$$\begin{aligned}\left\|\varphi - \pi_H\varphi\right\|_{L^2(K)} + H_K^{1/2}\left\|\varphi - \pi_H\varphi\right\|_{L^2(\partial K)} + H_K\left\|\nabla\varphi - \nabla\pi_H\varphi\right\|_{L^2(K)^d} \\ \leq C\, H_K \left\|\nabla\varphi\right\|_{L^2(\Omega)^d} \quad \forall \varphi \in H_0^1(\Omega). \quad (4.17)\end{aligned}$$

Proposition 4.7. Let $A^{\delta,h}$ be the post-processed, continuous coefficient from Definition 3.8 and let assumptions (A1) – (A3), or alternatively (A2) – (A5) in case of the homogenization strategy, hold true. Then, the macroscale error estimator and indicators converge linearly in H (with a constant depending on δ as described in Lemma 3.32),

$$|\theta^H| \leq \sum_{K \in \mathbb{T}_H(\Omega)} |\eta_K^H| \leq C(\delta) H. \quad (4.18)$$

Proof. The macroscale error indicator reads

$$\begin{aligned}\eta_K^H = \left(f + \nabla\cdot A^{\delta,h}\nabla U, z^\delta - \pi_H z^\delta\right)_{L^2(K)} \\ - \frac{1}{2}\left(\left[\boldsymbol{n}\cdot A^{\delta,h}\nabla U\right]_{\partial K}, z^\delta - \pi_H z^\delta\right)_{L^2(\partial K)}. \quad (4.19)\end{aligned}$$

Estimating the first term of (4.19) yields

$$\begin{aligned}\left|\left(f + \nabla\cdot A^{\delta,h}\nabla U, z^\varepsilon - \pi_H z^\varepsilon\right)_{L^2(K)}\right| \\ \leq \left\|f + \nabla\cdot A^{\delta,h}\nabla U\right\|_{L^2(K)} \left\|z^\varepsilon - \pi_H z^\varepsilon\right\|_{L^2(K)} \\ \leq \left\|f + \nabla\cdot A^{\delta,h}\nabla U\right\|_{L^2(K)} C\, H_K. \quad (4.20)\end{aligned}$$

To employ the same strategy on the second term needs a variant of a well known trace theorem[6]. The function $u^{\delta,h} \in H_0^1(\Omega)$ admits a representation

$$\nabla \cdot A^{\delta,h} \nabla u^{\delta,h} \equiv f \in L^2(\Omega). \tag{4.21}$$

Thus, for arbitrary $K \in \mathbb{T}_H(\Omega)$, the trace $\boldsymbol{n} \cdot A^{\delta,h} \nabla u^{\delta,h} \in L^2(\partial K)$ exists and the following estimate holds by virtue of a rescaling argument:

$$\begin{aligned} &\left\| \boldsymbol{n} \cdot A^{\delta,h} \nabla (u^{\delta,h} - U) \right\|^2_{L^2(\partial K)} \\ &\quad \leq C \Big(H_K \left\| f + \nabla \cdot A^{\delta,h} \nabla U \right\|^2_{L^2(K)} + H_K^{-1} \left\| A^{\delta,h} \nabla (u^{\delta,h} - U) \right\|^2_{L^2(K)^d} \Big), \end{aligned} \tag{4.22}$$

with a constant C independent of H and H_K denoting the local cell size of K. Hence,

$$\begin{aligned} &H_K^{-1} \left\| \left[\boldsymbol{n} \cdot A^{\delta,h} \nabla U \right]_{\partial K} \right\|^2_{L^2(\partial K)^d} \\ &\quad = H_K^{-1} \left\| \left[\boldsymbol{n} \cdot A^{\delta,h} \left(\nabla U - \nabla u^{\delta,h} \right) \right]_{\partial K} \right\|^2_{L^2(\partial K)^d} \\ &\quad \leq C \left\| f + \nabla \cdot A^{\delta,h} \nabla U \right\|^2_{L^2(K)} + C\, H_K^{-2} \left\| A^{\delta,h} (\nabla U - \nabla u^{\delta,h}) \right\|^2_{L^2(K)^d}. \end{aligned} \tag{4.23}$$

This enables us to estimate the second term:

$$\begin{aligned} &\left| \left([\boldsymbol{n} \cdot A^{\delta,h} \nabla U]_{\partial K}, z^\delta - \pi_H z^\delta \right)_{L^2(\partial K)} \right| \\ &\quad \leq C \Big(\left\| f + \nabla \cdot A^{\delta,h} \nabla U \right\|^2_{L^2(K)} + H_K^{-2} \left\| A^{\delta,h} (\nabla U - \nabla u^{\delta,h}) \right\|^2_{L^2(K)^d} \Big) H_K. \end{aligned} \tag{4.24}$$

As a last ingredient a local stability estimate is needed:

$$\left\| f + \nabla \cdot A^{\delta,h} \nabla U \right\|^2_{L^2(K)} + H_K^{-2} \left\| A^{\delta,h} (\nabla U - \nabla u^{\delta,h}) \right\|^2_{L^2(K)^d} \leq C(\delta). \tag{4.25}$$

For the post-processed variant and thus uniformly Lipschitz continuous coefficient $A^{\delta,h}$ this is indeed the case[6], see Lemma 3.32 and Proposition 3.33. □

Remark 4.8. In case of the piecewise constant coefficients $\bar{A}^\delta$ and $\bar{A}^{\delta,h}$ only a reduced stability estimate

$$\left\| f + \nabla \cdot A^{\delta,h} \nabla U \right\|^2_{L^2(K)} + H_K^{-2} \left\| A^{\delta,h} (\nabla U - \nabla u^{\delta,h}) \right\|^2_{L^2(K)^d} \leq C(\delta) H_K^{-2+2\gamma} \tag{4.26}$$

holds true (cf. Proposition 3.37). This also leads to a reduced convergence order for the macroscale error estimator and local indicators:

$$\left| \theta^H \right| \leq \sum_{K \in \mathbb{T}_H(\Omega)} \left| \eta_K^H \right| \leq C(\delta) H^\gamma. \tag{4.27}$$

Remark 4.9. For a large class of quantities of interest and associated dual solutions an improved interpolation estimate as the one given in (4.17) holds true and a higher convergence rate can be shown: Suppose z^ε admits the estimate

$$\left\|z^\varepsilon - \pi_H z^\varepsilon\right\|_{L^2(K)} + H_K^{1/2}\left\|z^\varepsilon - \pi_H z^\varepsilon\right\|_{\partial L^2(K)} \leq C\, H^{1+\varkappa}, \tag{4.28}$$

then, in case of the post-processed coefficient $A^{\delta,h}$, it holds true that

$$\left|\theta^H\right| \leq \sum_{K\in\mathbb{T}_H(\Omega)} \left|\eta_K^H\right| \leq C(\delta) H^{1+\varkappa}. \tag{4.29}$$

Similarly, in case of the piecewise constant coefficient $\bar{A}^{\delta,h}$

$$\left|\theta^H\right| \leq \sum_{K\in\mathbb{T}_H(\Omega)} \left|\eta_K^H\right| \leq C(\delta) H^{\gamma+\varkappa}. \tag{4.30}$$

The required regularity assumptions for such an improved interpolation estimate are fulfilled for a big class of quantities of interests j. Elliptic equations with very weak regularity on the coefficient A^ε have been studied extensively in the literature[79,93,98]. As an example, consider j to be of class $L^2(\Omega)$. Then, the dual solution z^ε already admits a Hölder-continuous representation[79, Ch. 5], i. e.,

$$\operatorname*{ess\,sup}_{\Omega} \frac{\left|z^\varepsilon(y) - z^\varepsilon(x)\right|}{||y-x||^\varkappa} < \infty, \tag{4.31}$$

for some Hölder-exponent $0 < \varkappa \leq 1$. This result only requires a Lebesgue-integrable function A^ε, which is the case for $A^\varepsilon \in L^\infty(\Omega)^{d\times d}$ on a bounded domain Ω. However, the Hölder-continuity is only uniform in ε in case of stronger assumptions such as (A1). Otherwise, no uniform bound with respect to ε can be established.

In contrast to the macroscopic discretization indicator θ^H, the microscopic pendant θ^h depends only on the regularity of the solutions of the local sampling problems and the approximation order of the quadrature rule used.

Proposition 4.10. In case of the averaging strategy (3.12) and if (A1) – (A3) hold true, the microscale indicator θ^h admits an estimate

$$|\theta^h| \leq \sum_{K\in\mathbb{T}_\delta(\Omega)} \sum_{Q\in\mathbb{T}_h(Y_K^\delta)} |\eta_Q^h| \leq C(\varepsilon) h. \tag{4.32}$$

In case of the homogenization sampling strategy and with (A2) – (A5) being fulfilled a slightly different result follows:

$$|\theta^h| \leq \sum_{K\in\mathbb{T}_\delta(\Omega)} \sum_{Q\in\mathbb{T}_h(Y_K^\delta)} |\eta_Q^h| \leq C\left(\frac{h}{\varepsilon}\right). \tag{4.33}$$

Proof. The statements are a direct consequence of Lemma 3.28 as well as Proposition 3.54, respectively, where it is shown that $||A^{\delta,h} - A^{\delta}||_{L^\infty(\Omega)}$ is bounded with the respective convergence order. □

The model-error indicator θ^{δ} is of fundamentally different character as the discretization indicators θ^{H} and θ^{h}. Only for the considerably stronger assumption of applicability (cf. Definition 3.19) a convergence with respect to a discretization parameter for a family of effective models $\left\{(\mathbb{T}_\delta(\Omega), A^{\delta})\right\}_{\delta>0}$ can be shown.

Proposition 4.11. Assume that a family of effective models $\left\{(\mathbb{T}_\delta(\Omega), A^{\delta})\right\}_{\delta>0}$ is applicable for A^{ε}. Then,

$$|\theta^{\delta}| \leq \sum_{K\in\mathbb{T}_\delta(\Omega)} |\eta_K^{\delta}| \leq \sum_{K\in\mathbb{T}_\delta(\Omega)} C(\varepsilon)||\nabla z^{\varepsilon}||_K\, \delta \leq C(\varepsilon)\,\delta. \tag{4.34}$$

Proof. This is an immediate consequence of (3.35). □

Remark 4.12. The preceding proposition is too pessimistic in practice. The *key point* of the whole a posteriori approach is the fact that the model error indicator (4.11), $\eta_K^{\delta} := \left(\{A^{\varepsilon} - A^{\delta}\}\nabla u^{\delta}, \nabla z^{\varepsilon}\right)_K$, *contains the dual solution* as a *weighting factor*. Thus, for a localized functional, where z^{ε} is of Green's function type, the bad error constant $C(\varepsilon)$ is only present in a small region that needs to be resolved in full. We refer to the numerical results given in Section 5.3.

Under the weaker assumptions (A2) – (A5) alone (in the case of the homogenization sampling strategy) a general a priori result for the convergence of the model-error estimator is not possible. Additional assumptions are required. The reason for this lies in the fact that a relation between $\langle j, u^{\varepsilon}\rangle$ and $\langle j, u^{0}\rangle$ in the homogenization limit has to hold true.

Remark 4.13. This observation is one of the reasons that motivate alternative model-adaptation strategies that do not rely on asymptotic behavior with respect to a discretization parameter δ, see Chapter 6.

For the sake of completeness we formulate the following convergence result—but stress the fact that due to the occurrence of all discretization parameters its usefulness is rather limited.

Proposition 4.14. Let the homogenization sampling strategy given in Definition 3.4 be applicable for the given quantity of interest j in the sense that

$$\left|\langle j, u^{0}\rangle - \langle j, u^{\varepsilon}\rangle\right| \leq C\,\varepsilon. \tag{4.35}$$

Then, the model-error estimator θ^δ permits the estimate

$$|\theta^\delta| \leq C\left(H + \frac{h}{\varepsilon} + \delta + \varepsilon\right). \tag{4.36}$$

Remark 4.15. The additional assumption (4.35) is analogous to the assumption

$$||u^0 - u^\varepsilon|| \leq C\,\varepsilon \tag{4.37}$$

made in the previous a priori analysis where a convergence $||u^\delta - u^0|| \leq C\,\delta$ was shown.

Proof of the proposition. Starting with the definition of θ^δ:

$$\begin{aligned}\theta^\delta &= \left((A^\varepsilon - A^\delta)\nabla U, \nabla z^\varepsilon\right)_{L^2(\Omega)^d} \\ &= \left((A^0 - A^\delta)\nabla U, \nabla z^\varepsilon\right)_{L^2(\Omega)^d} + \left((A^\varepsilon - A^0)\nabla[U - u^0], \nabla z^\varepsilon\right)_{L^2(\Omega)^d} \\ &\qquad + \left((A^\varepsilon - A^0)\nabla u^0, \nabla z^\varepsilon\right)_{L^2(\Omega)^d}.\end{aligned} \tag{4.38}$$

The last term can be further transformed into

$$\begin{aligned}\left(A^\varepsilon \nabla u^0 - A^0 \nabla u^0, \nabla z^\varepsilon\right)_{L^2(\Omega)^d} &= \left(A^\varepsilon \nabla u^0 - A^\varepsilon \nabla u^\varepsilon, \nabla z^\varepsilon\right)_{L^2(\Omega)^d} \\ &= \langle j, u^0\rangle - \langle j, u^\varepsilon\rangle.\end{aligned} \tag{4.39}$$

Hence,

$$\begin{aligned}|\theta^\delta| \leq{}& \left\|A^0 - A^\delta\right\|_{L^\infty(\Omega)^{d\times d}} ||\nabla U||_{L^2(\Omega)^d}\, ||\nabla z^\varepsilon||_{L^2(\Omega)^d} \\ &+ \left(||A^\varepsilon||_{L^\infty(\Omega)^{d\times d}} + ||A^0||_{L^\infty(\Omega)^{d\times d}}\right) \left\|\nabla U - \nabla u^0\right\|_{L^2(\Omega)^d} ||\nabla z^\varepsilon||_{L^2(\Omega)^d} \\ &+ \left|\langle j, u^0\rangle - \langle j, u^\varepsilon\rangle\right|.\end{aligned} \tag{4.40}$$

The result now follows by virtue of Propositions 3.37, 3.45, 3.48, and 3.54. □

Remark 4.16. The last proposition illustrates the fundamental problem one is faced with in context of model adaptation. In light of Proposition 4.7 and 4.10 it can be safely assessed that θ^δ is (at least for H and h suitably small) an estimator for the model error. So it is possible to detect a large model error—but the possibility to adapt the model locally in order to improve the situation is not necessarily given. As an illustration consider the case of vanishing H, h and δ. Then, $||\theta^\delta|| \leq C\,\varepsilon$ by virtue of above result. However, within the confinement of one specific model derivation (in this case HMM) no further model adaptation to improve the situation can be done; ε is a fixed problem parameter.

4.3 Localization strategies for the dual problem

A fundamental difficulty arises from the fact that computing the solution of the dual problem is (in case of the elliptic model problem) of the same complexity as the primal problem itself. A *global* fine-scale approximation of z^ε has to be considered infeasible. Thus, a strategy to approximate the dual problem with low computational overhead is needed. This section discusses a number of possible approaches to address this problem are discussed. The chapter concludes with a numerical validation of the different reconstruction principles.

As a first step, we make the observation that the model-error indicator η_K^δ is of fundamentally different character than the residual-type indicators η_K^H and η_K^h. This can be seen by the following heuristic reasoning: η_K^H, and similarly η_K^h, correspond to *moments of first order* with respect to microscale fluctuations when defined with the full dual solution z^ε. For example, the microscale indicator reads

$$\eta_K^h = \big((A^\delta - A^{\delta,h})\nabla U, \underbrace{\nabla z^\varepsilon}_{\text{fluct.}}\big)_K. \tag{4.41}$$

In case of the error splitting given in Proposition 4.3 the macro- and microscale error indicators actually contain no fluctuations at all. Whereas η_K^δ always corresponds to a *moment of second order*,

$$\eta_K^\delta = \big(\underbrace{(A^\varepsilon - A^\delta)}_{\text{fluct.}}\nabla u^\delta, \underbrace{\nabla z^\varepsilon}_{\text{fluct.}}\big). \tag{4.42}$$

Due to the fact that z^ε influences η_K^h mainly by its mean value $\int_K z^\varepsilon \,\mathrm{d}y$ and not by its variation $\int_K \|z^\varepsilon\| \,\mathrm{d}y$, a replacement of ∇z^ε in η_K^H and η_K^h by an averaged approximation can be justified. As already introduced, a canonical candidate is the solution $z^\delta \in H_0^1(\Omega)$ of a corresponding *effective dual problem*

$$\big(A^\delta \nabla\varphi, \nabla z^\delta\big) = \langle j, \varphi\rangle \quad \forall\, \varphi \in H_0^1(\Omega) \tag{4.43}$$

defined with in terms of the effective model A^δ. The idea is that an effective approximation of z^δ that, for example, *globally* approximates z^ε in $L^2(\Omega)$ also produces suitable averages (for a sampling region K large enough) with

$$\int_K \nabla z^\delta \,\mathrm{d}y \approx \int_K \nabla z^\varepsilon \,\mathrm{d}y. \tag{4.44}$$

For the model-error indicator, however, such a (heuristic) argument cannot be applied because replacing z^ε by a mean value z^δ destroys the coupling of

fluctuations in ∇z^ε with those in $A^\delta - A^\varepsilon$, and thus the property of being a moment of second order.

Different strategies for effectively approximating the dual solution have been discussed in the literature:

- Braack and Ern[26] used the homogenized dual solution z^δ directly instead of z^ε in order to estimate model errors. This approach can be justified, if the difference between A^ε and A^δ is suitably small, i. e. if A^δ is an effective model already very close to the full model A^ε, or if only a *qualitative—not a quantitative*—error estimate is needed, see Chapter 5.

- Oden and Vemaganti[84–86] also used the homogenized dual solution z^δ but with an additional cell-wise local inverse of the form

$$I_d - (A^\varepsilon)^{-1} A^\delta \tag{4.45}$$

  in the estimation of the model-error indicator. This results in an estimate of the form[84, Th. 3.1] (adapted to the modified HMM scheme)

$$\left|\langle j, u^\delta\rangle - \langle j, u^\varepsilon\rangle\right|^2 \leq C \sum_{K\in\mathbb{T}_\delta(\Omega)} \left|\left(\nabla u^\delta, \left[I_d - \left(A^\varepsilon\right)^{-1} A^\delta\right] z^\delta\right)_K\right|^2 + \mathscr{R}. \tag{4.46}$$

  with a (not directly computable) residual-type remainder $\mathscr{R}$.

- A different strategy was explored by Romkes and Moody[91]. They replaced the global dual problem (4.1) by a local problem solely defined on the local sampling region: Find $\hat{z}^\varepsilon \in H^1(K)$, s. t.

$$\begin{aligned} \left(A^\varepsilon \nabla\varphi, \nabla\hat{z}^\varepsilon\right) &= \langle j, \varphi\rangle, \quad && \forall\, \varphi \in H^1(K) \\ \boldsymbol{n} \cdot A^\varepsilon \nabla\hat{z}^\varepsilon &= 0 && \text{on } \partial K. \end{aligned} \tag{4.47}$$

  This approach is computationally rather expensive and does not capture global information about the error distribution that is part of z^ε (or even the reduced version z^δ).

All of this strategies were derived in a context of model adaptation strategies that only require a *qualitative* error estimate (see Sections 5.1 and 5.2). In particular, a correct balancing of discretization estimators θ^H, θ^h and model error indicator θ^δ was not needed and therefore not examined.

In this thesis, two slightly different reconstruction approaches are used. The first one is based on the idea to approximate z^ε directly on a coarse discretization

but with high (summed) quadrature. The idea is that if the coarse discretization is chosen to be a (sufficient) refinement of $\mathbb{T}_H(\Omega)$ a compromise between computational complexity and quantitative information can be found. This requires the dual discretization to be still coarse enough to have a computable number of degrees of freedom, but refined enough to capture essential oscillations in the finite-element ansatz space of the dual problem.

Definition 4.17 (Coarse dual approximation)**.** Let $\mathbb{T}_{H/2}(\Omega)$ be the triangulation resulting from a one time refinement of $\mathbb{T}_H(\Omega)$ and let Q_h be a fine, summed quadrature rule. Define $\tilde{Z} \in V^{H/2}(\Omega)$ as the solution of

$$Q_h\big(A^\varepsilon \nabla \tilde{Z} \cdot \nabla \varphi^{H/2}\big) = Q_h\big(j\,\varphi^{H/2}\big) \quad \forall \varphi^{H/2} \in V^{H/2}(\Omega). \tag{4.48}$$

Remark 4.18. This reconstruction is of a similar nature as the arithmetic averaging sampling strategy. $\nabla \tilde{Z} \otimes \nabla \varphi^{H/2}$ is approximately constant for the summed quadrature rule Q_h such that

$$Q_{h,K}\big(A^\varepsilon \nabla \tilde{Z} \cdot \nabla \varphi^{H/2}\big) \approx Q_{h,K}\big(A^\varepsilon\big)\overline{\nabla \tilde{Z} \cdot \nabla \varphi^{H/2}} \tag{4.49}$$

on a macro cell K. Consequently, this approximation suffers from the same problem of probably bad global approximation in $L^2(\Omega)$, see the numerical results of Subsection 3.5.2. But, as evidenced by standard DWR techniques, the approximation of the dual solution z^ε can be relatively coarse and might still contain enough qualitative information about error distribution to enable a stable adaptation process.

It is clear that the approximation property of (4.48) will degrade with increasing scale separation. For such cases a more elaborate approximation strategy is needed. Thus as a second reconstruction principle we propose a strategy that combines the usage of a global, effective approximation of z^ε with a local enhancement. The enhancement is given by localized reconstruction problems in spirit of the VMM ansatz.

Definition 4.19 (Local enhancement)**.** Let z^δ be the solution of (4.43),

$$\big(A^\delta \nabla \varphi, \nabla z^\delta\big) = \langle j, \varphi \rangle \quad \forall\, \varphi \in H_0^1(\Omega), \tag{4.50}$$

and let $\big\{\omega_K \,:\, K \in \mathbb{T}_\delta(\Omega)\big\}$ be a set of *reconstruction patches* fulfilling $\omega_K \supset K$. Define a patch-wise reconstruction $z_K^\delta \in H_0^1(\omega_K)$ by

$$\big(A^\varepsilon \nabla \varphi, \nabla (z^\delta + z_K^\delta)\big) = \langle j, \varphi \rangle \quad \forall \varphi \in H_0^1(\omega_K). \tag{4.51}$$

Remark 4.20. The reconstruction patches $\omega(K)$ can usually be chosen as small as $\omega(K) = K$ (and the majority of numerical results in this thesis are computed with this choice). If a very precise error estimate is necessary, however, a slightly increased patch-size must be used. Further, for technical reasons faced with in the model optimization framework in Chapter 6 the reconstruction processes is sometimes used on a patch $\omega(K)$ around K with patch-depth 1.

Remark 4.21. It is also possible to use other boundary conditions than homogeneous Dirichlet conditions—homogeneous Neumann, natural, or periodic boundary conditions are a possible. This results in a non-conforming ansatz that unconditionally requires to choose $\omega(K)$ slightly larger than K in order to control the effect of the imposed boundary conditions.

With the choice $\omega(K) = K$, the locally reconstructed dual solution leads to a conforming ansatz $z^\delta + \sum_{K\in\mathbb{T}_\delta} z_K^\delta \in H^1(\Omega)$. In this case above local enhancement strategy can be regarded as a variant of the VMM formulation that only has a *reconstruction* coupling and omits the opposite coupling via *compression* (cf. Figure 2.1).

Remark 4.22. More precisely, the error identity (6.3) lifts the question of suitable approximations in terms of a quantity of interest (for the primal problem) to the question of suitable approximation properties of the localization technique for the dual problem. The catch here is that the latter is typically measured in the L^2-norm of the gradient of the error of the dual approximation, for which—depending on the localization approach—strong approximation properties are available.

In general, no strict projection property onto $V^H(\Omega) \oplus \sum_{K\in\mathbb{T}_\delta} H_0^1(K)$ with respect to the functional j holds true, i. e., $((A^\delta - A^\varepsilon)\nabla U, \nabla z^\delta + z_K^\delta)_K$ does not necessarily have to be a better approximation of η_K^δ than the reduced variant $((A^\delta - A^\varepsilon)\nabla U, \nabla z^\delta)_K$.

Remark 4.23. A heuristic argument can be given for the case of homogeneous Dirichlet boundary conditions in case of a global approximation property to hold true: The difference between η_K^δ and its approximation $\tilde{\eta}_K^\delta = \big((A^\varepsilon - A^\delta)\nabla u^\delta, \nabla z^\delta + \nabla z_K^\delta\big)_K$ can be expressed as

$$\begin{aligned}\eta_K^\delta - \tilde{\eta}_K^\delta &= \big((A^\varepsilon - A^\delta)\nabla u^\delta, \nabla z^\varepsilon - \nabla(z^\delta + z_K^\delta)\big)_K \\ &= \big((A^\varepsilon - A^\delta)\nabla(u^\delta - \varphi), \nabla(z^\varepsilon - z^\delta - z_K^\delta)\big)_K \\ &\qquad + \big(A^\delta \nabla\varphi, \nabla(z^\varepsilon - z^\delta - z_K^\delta)\big)_K \qquad (4.52)\end{aligned}$$

for arbitrary $\varphi \in H_0^1(K)$. Now, choose φ to be equal to the projection of the function $u^\delta \in H^2(\Omega) \cap H_0^1(\Omega)$ onto $H_0^1(K)$, $\varphi = \pi_0 u^\delta$, and estimate the last term as follows:

$$\begin{aligned}
&\left|\left(A^\delta \nabla\varphi, \nabla(z^\varepsilon - z^\delta - z_K^\delta)\right)_K\right| \\
&\quad \leq \left| \int_{\partial K} \boldsymbol{n} \cdot A^\delta \nabla \pi_0 u^\delta \, (z^\varepsilon - z^\delta) \, \mathrm{d}o_x \right| + \left|\left(\nabla \cdot A^\delta \nabla \pi_0 u^\delta, z^\varepsilon - z^\delta - z_K^\delta\right)_K\right| \\
&\quad = \left|\left(\nabla \cdot A^\delta \nabla \pi_0 u^\delta, z^\varepsilon - z^\delta - z_K^\delta\right)_K\right| \\
&\quad \leq \left\|\nabla \cdot A^\delta \nabla \pi_0 u^\delta\right\|_K \left\|z^\varepsilon - z^\delta - z_K^\delta\right\|_K . \qquad (4.53)
\end{aligned}$$

Given the fact that $\pi_0 u^\delta$ deviates from φ only close to the boundary, it can be heuristically justified to approximate the first term by

$$\left|\left((A^\varepsilon - A^\delta)\nabla(u^\delta - \pi_0 u^\delta), \nabla(z^\varepsilon - z^\delta - z_K^\delta)\right)_K\right| \sim \left| \int_{\partial K} z^\varepsilon - z^\delta - z_K^\delta \, \mathrm{d}o_x \right|. \quad (4.54)$$

One conludes that the approximate error estimator $\tilde{\eta}_K^\delta$ is indeed a sufficient approximation on η_K^δ provided that the heuristic properties

$$\left\|z^\varepsilon - z^\delta - z_K^\delta\right\|_K \ll 1, \qquad (4.55)$$

$$\left| \int_{\partial K} z^\varepsilon - z^\delta - z_K^\delta \, \mathrm{d}o_x \right| \ll 1 \qquad (4.56)$$

hold true. The critical property is (4.56) which can be highly influenced by the artificial boundary conditions imposed on z_K^δ.

A detailed numerical study about qualitative and quantitative behavior of the coarse approximation and local enhancement strategy is given in Section 4.5.

4.4 Evaluation of the estimators and indicators

In contrast to pure residual-type estimators that can be evaluated in a simple post-processing step, the practical evaluation of the error estimators require the approximation of an additional, intermediate dual solution and effective coefficients. The evaluation of the discretization error indicators η_K^H and η_K^h according to (4.7) and (4.9) involve the a priori unknown effective coefficients A^δ and dual solution z^δ. Both have to be approximated with higher order than the

corresponding solution of the primal problem U, or the numerically computed, effective coefficients $A^{\delta,h}$. Thus, we define an approximation of the dual solution with appropriately chosen refinement parameters for both localization strategies that were given by Definitions 4.17 and 4.19.

Definition 4.24 (Coarse dual approximation)**.** As already introduced in Definition 4.17, let $\tilde{Z} \in V^{H/2}(\Omega)$ be the solution of

$$\left(A^{\varepsilon}\nabla\varphi^{H/2}, \nabla\tilde{Z}\right) = \langle j, \varphi^{H/2}\rangle \quad \forall \varphi^{H} \in V^{H/2}(\Omega). \tag{4.57}$$

Definition 4.25 (Local higher order approximation)**.** Let the coefficient $A^{\delta,h/2}$ denote the numerically computed, effective model with respect a one time refined microscale discretization $\left\{\mathbb{T}_{h/2}(Y_K^{\delta}) : K \in \mathbb{T}_{\delta}(\Omega)\right\}$. Define $\tilde{Z} \in V^{H}(\Omega)$ to be the solution of

$$\left(A^{\delta,h/2}\nabla\varphi^{H}, \nabla\tilde{Z}\right) = \langle j, \varphi^{H}\rangle \quad \forall \varphi^{H} \in V^{H}(\Omega). \tag{4.58}$$

Let $\pi_{2H}^{(2)}$ denote the patch-wise interpolation to a d-quadratic finite-element space defined on a one times coarsened mesh $\mathbb{T}_{2H}(\Omega)$. Define the approximate error indicators

$$\tilde{\eta}_K^H \coloneqq \left(f, \pi_{2H}^{(2)}\tilde{Z} - \tilde{Z}\right)_K - \left(A^{\delta,h}\nabla U, \nabla(\pi_{2H}^{(2)}\tilde{Z} - \tilde{Z})\right)_K, \tag{4.59}$$

$$\tilde{\eta}_K^h \coloneqq \left(\{A^{\delta,h/2} - A^{\delta,h}\}\nabla U, \nabla z^{\delta}\right)_K. \tag{4.60}$$

The replacement of $z^{\delta} - \pi_H z^{\delta}$ by $\pi_{2H}^{(2)}\tilde{Z} - \tilde{Z}$ is a well-known post-processing technique[21]. It allows to use the same ansatz space $V^H(\Omega)$ for the dual solution that is used for the primal problem where, normally, due to Galerkin orthogonality an approximation of the dual solution in the same ansatz space would be unsuitable. Its usage leads to a slight reduction in effectivity (i. e. the quantitative prediction of the estimator).

Remark 4.26. The usage of $\tilde{Z}$ instead of z^{δ} in the evaluation of η_K^h is uncritical because z^{δ} only takes the role of a weighting factor outside of the residual such that no Galerkin orthogonality occurs. Recalling (4.13) for the HMM method:

$$\eta_K^h = \sum_{ij}\int_K \nabla_i U \nabla_j z^{\delta}\,\mathrm{d}x \frac{1}{\left|Y_K^{\delta}\right|} \sum_{Q \in \mathbb{T}_h(Y_K^{\delta})} \textit{residual on micro scale}. \tag{4.61}$$

In order to define an approximation of the model-error indicator

$$\eta_K^{\delta} = \left((A^{\varepsilon} - A^{\delta})\nabla u^{\delta}, \nabla z^{\varepsilon}\right) \tag{4.62}$$

the additional unknown u^δ occurring in the definition of η_K^δ has to be taken care of. It would be possible to approximate u^δ by a higher order approximation $\tilde{U}$ solving—similarly to (4.58)—the equation

$$\left(\bar{A}^{\delta,h/2}\nabla\tilde{U},\nabla\varphi^H\right)=\langle j,\varphi^H\rangle \quad \forall\varphi^H\in V^H(\Omega). \tag{4.63}$$

An alternative approach is to avoid this higher order approximation of the primal problem entirely by ensuring that the macroscale is always approximated with a higher accuracy, e. g., by ensuring that $\mathbb{T}_H$ is always sufficiently refined such that θ^H is an order of magnitude smaller than θ^δ. This approach is justified by the fact that the computational cost for a higher accuracy on the macroscale is negligible compared to the necessary work to improve the model error. We thus define:

Definition 4.27 (Approximate model-error indicator)**.** Let $A^{\delta,h/2}$ be the higher order effective model of Definition 4.25 and let $\tilde{Z}$ be the corresponding (discrete) dual solution of problem (4.58). We define the approximative model-error indicator

$$\tilde{\eta}_K^\delta \coloneqq \left((A^{\delta,h/2}-A^\varepsilon)\nabla U,\nabla\tilde{Z}\right)_K. \tag{4.64}$$

The local problems of the proposed local enhancement strategy are numerically discretized in a straight forward manner.

Definition 4.28 (Approximate local enhancement)**.** Per sampling region $K\in\mathbb{T}_\delta(\Omega)$ define a local correction $\tilde{\eta}_K^\delta+\tilde{\eta}_{K,\mathrm{rec}}^\delta$ of the error indicator $\tilde{\eta}_K^\delta$ by

$$\tilde{\eta}_{K,\mathrm{rec}}^\delta \coloneqq \left((A^{\delta,h/2}-A^\varepsilon)\nabla U,\nabla\tilde{Z}_K\right)_K, \tag{4.65}$$

with a finite-element approximation $\tilde{Z}_k$ solving the local reconstruction problem: Find $\tilde{Z}_k\in V^h(K)$ s. t.

$$\left(A^\varepsilon\nabla\varphi,\nabla\tilde{Z}+\nabla\tilde{Z}_K\right)_K=\langle j,\varphi\rangle \quad \forall\,\varphi\in V^h(K). \tag{4.66}$$

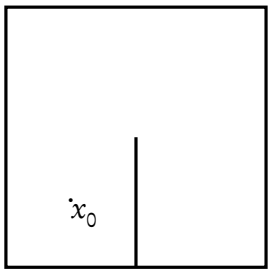


Figure 4.1. Evaluation point $x_0 = (0.25, 0.25)$ for the point functionals inside the slit domain Ω.

4.5 Numerical validation of the error estimators

The error splitting and practical evaluation of the error indicators developed in this chapter deserve a thorough numerical validation. Consider again both microstructures defined on the domain Ω introduced in Section 3.5. In addition, three quantities of interest will be examined; a non-local average given by

$$\langle j_1, \varphi \rangle = \int_\Omega \varphi \, \mathrm{d}x, \tag{4.67}$$

and two localized point functionals

$$\langle j_2, \varphi \rangle = \varphi(x_0), \tag{4.68}$$

$$\langle j_3, \varphi \rangle = \partial_y \varphi(x_0), \tag{4.69}$$

evaluated at point $x_0 = (0.25, 0.25)$, see Figure 4.1. The functionals are chosen in such a way that the microstructure has an increasingly pronounced influence on the quantity of interest—ranging from a purely macroscopic dependency of j_1 to the high influence of the micro structure on j_3.

In the following, several different aspects of the a posteriori error analysis are examined. In particular, the convergence results of Section 4.2 are verified and the different reduction and reconstruction approaches introduced in Section 4.3 are examined for both, qualitative and quantitative behavior.

4.5.1 Behavior under uniform refinement

In a first numerical test the behavior of the three error estimators θ^H, θ^h and θ^δ given by (4.6) under uniform refinement is examined. The numerical test serves two purposes. Firstly, the idea is to verify that the predicted convergence rates

derived in Section 4.2 can be reproduced. Secondly—and more important—the question arises whether the error splitting into $\theta^H + \theta^h + \theta^\delta$ is well behaved. Meaning, whether the individual estimators (and indicators) are *separated*. For example, pure refinement in the macroscale should only influence the estimator θ^H, but θ^h and θ^δ should remain approximately constant, etc. Such a separation property is important for the model- and discretization-adaptive strategies that will be developed in the next chapter. Otherwise, a simultaneous treatment of different sources with above error splitting cannot be justified.

All numerical results are compared with a reference computation of 1.67×10^7 cells. Instead of the reduced dual approximations (4.57) or (4.58), a fully resolved dual approximation of z^ε given by a discretization of (4.1) with 1.67×10^7 macro cells is used. Otherwise, the individual error estimators are computed as described in Section 4.4 (see Equations 4.59, 4.60, and 4.64).

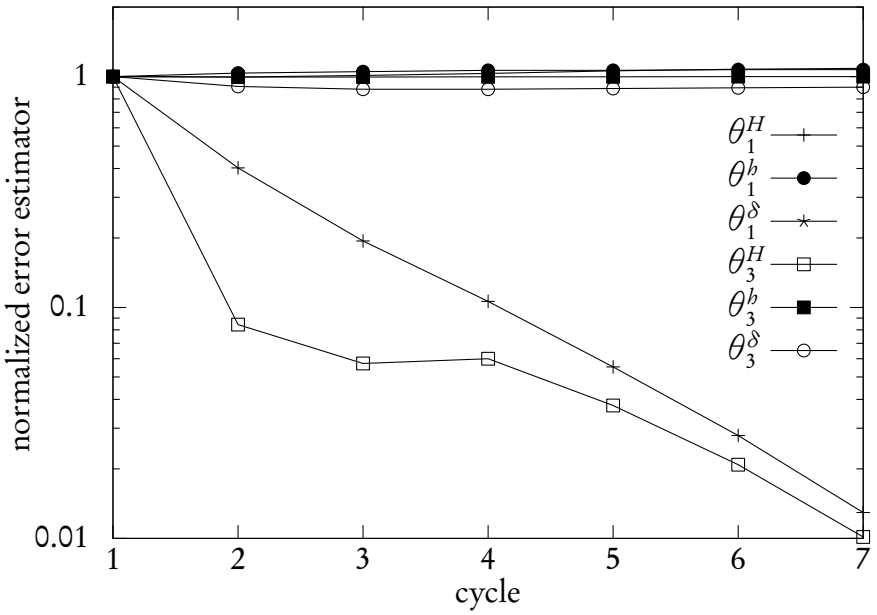
Remark 4.29. The behavior of the error indicators and estimators for different reconstructions of the dual problem will be examined in the next numerical test.

A base resolution of $H = 2^{-9}$, $h = 2^{-11}$ and $\delta = 2^{-3}$ is chosen for two series of computations: One is performed for a successive macroscale refinement with $H = 2^{-3}, \ldots, 2^{-9}$; another for a successive microscale refinement given by the sequence $h = 2^{-6}, \ldots, 2^{-12}$. Figures 4.2 and 4.3 show the results in form of a semi-logarithmic plot of the error estimators over refinement cycles normalized to an initial value of 1.

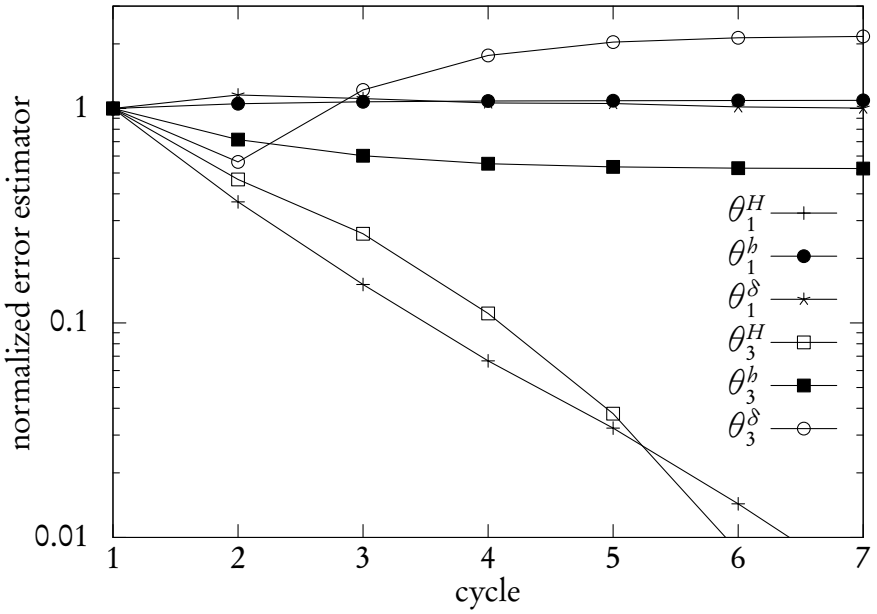
Under uniform refinement in H, the estimators θ^h and θ^δ remain close to their respective initial values: within 10 % for the periodic coefficients and within a factor of 2 for the random permeability. Whereas the macroscale estimator θ^H decreases with roughly linear order for all cases.

Remark 4.30. An increased convergence rate with respect to the regular functional j_1 (or j_2) cannot be observed because of a lack of regularity—for all computations in this chapter the piecewise constant variant $\bar{A}^{\delta,h}$ is used instead of the post-processed and therefore continuous variant given in Definition 3.8.

A similar behavior for uniform refinement in the microscale discretization $\mathbb{T}_h(K)$ can be observed, see Figure 4.3. The estimators θ^H and θ^δ remain close to their initial value with the exception of θ_3^H for the localized functional j_3 on the random permeability that changes by around one order of magnitude. Again, in analogy to the numerical a priori results of Section 3.5.3, a high convergence rate of order 4 occurs for θ^h. For the random permeability the microscale indicators show the same behavior of a sudden drop to vanishing values as was observed for uniform refinement in the $L^2(\Omega)$- and $H^1(\Omega)$-norms (cf. Section 3.5.3).

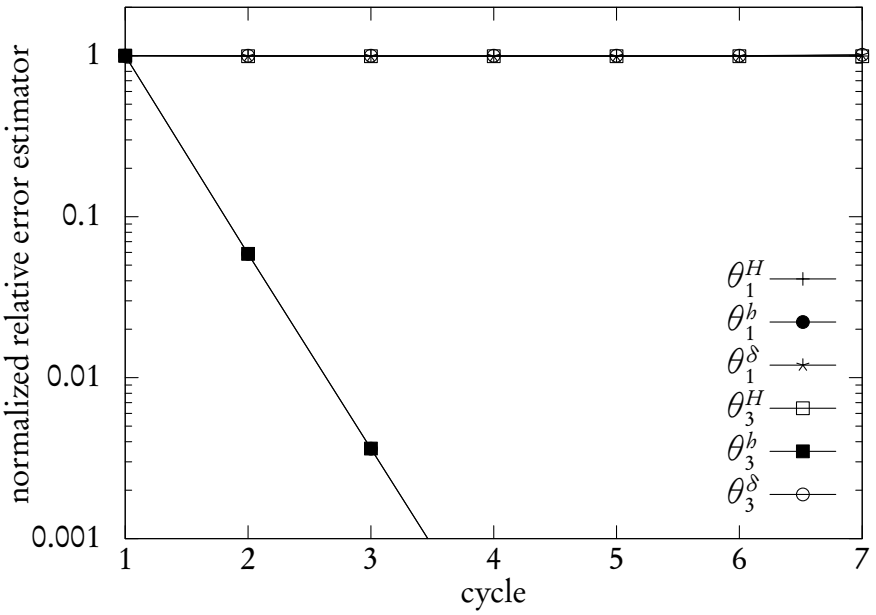


(a) Periodic coefficients, macroscale refinement (H)

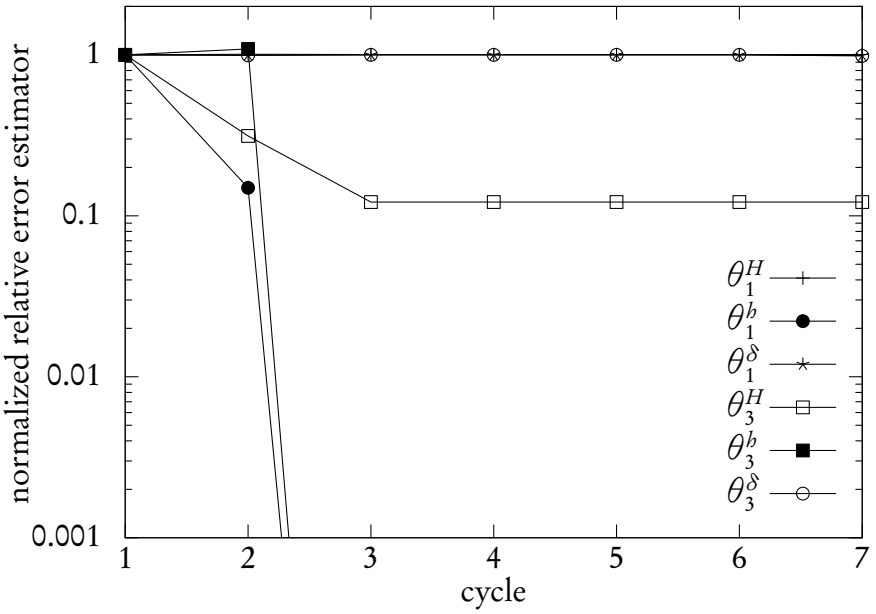


(b) Random coefficients, macroscale refinement (H)

Figure 4.2. Behavior of macroscale (θ^H), microscale (θ^h), and sampling (θ^δ) error estimates under uniform refinement of $\mathbb{T}_H(\Omega)$ with $H = 2^{-3}, \ldots, 2^{-8}$ for the global functional j_1 and the highly localized variant j_3.



(a) Periodic coefficients, microscale refinement (h)



(b) Random coefficients, microscale refinement (h)

Figure 4.3. Behavior of macroscale (θ^H), microscale (θ^h), and sampling (θ^δ) error estimates under uniform refinement of $\mathbb{T}_H(\Omega)$ with $h = 2^{-6}, \ldots, 2^{-11}$ for the global functional j_1 and the highly localized variant j_3.

4.5.2 Qualitative and quantitative behavior

A second numerical test is concerned with the reconstruction strategies introduced in Definitions 4.17 and 4.19. As we are primarily interested in the quantitative estimation of the model error, a fine macro- and fine microscale mesh is chosen with 6.5×10^4 cells on the macro level and a resolution of $h = 2^{-11}$ for the sampling patches. The error estimators θ^H, θ^h, and θ^δ are computed with 4 different dual solutions: a reference dual solution with 4.1×10^6 degrees of freedom ("full") with the help of (4.1), the coarse variant computed on a macroscale mesh consisting of 6.5×10^4 cells and with a summed quadrature rule of high order ("coarse"), and a reduced version (4.43) utilizing the effective coefficients without any reconstruction ("reduced"). Finally, a reduced variant with local reconstruction ("enhanced") is used that is given by (4.51) with the choice $\omega(K) = K$. Again, the geometric mean value is chosen as a simple averaging strategy. The results are given in Tables 4.1 and 4.2, as well as, in Figures 4.4 and 4.5 on page 101 et seqq.

Here, the performance of the different reconstruction approaches are measured with two different quantities. The *effectivity index* I_{eff} defined as

$$I_{\text{eff}} := \frac{\tilde{\theta}^H + \tilde{\theta}^h + \tilde{\theta}^\delta}{\langle j, u^\varepsilon - U \rangle}, \tag{4.70}$$

expresses the *quantitative behavior* of the error estimation. The closer it is to 1, the better. The index I_{loc} defined as

$$I_{\text{loc}} := \frac{\sum_{K \in \mathbb{T}_\delta(\Omega)} |\tilde{\theta}^H| + |\tilde{\theta}^h| + |\tilde{\theta}^\delta|}{|\tilde{\theta}^H + \tilde{\theta}^h + \tilde{\theta}^\delta|} \tag{4.71}$$

is a measure for the oscillatory behavior of the error indicators.

For all quantities of interest the effectivity of the (almost) fully resolved dual solution is around 0.97 in case of the periodic coefficients and remains between 1.0 – 1.4 for the random microstructure. The "coarse" approximation shows also a reasonable quantitative behavior for the chosen resolution with an effectivity that ranges roughly between 0.4 – 4. Further, the "reduced" reconstruction strategy generally leads to false quantitative estimates $\tilde{\theta}^\delta$ for the model error that can be significantly improved by the local reconstruction strategy. The macroscale and microscale resolutions are already chosen relatively fine. Nevertheless, the macroscale and microscale discretization indicators θ^H and θ^h remain well approximated for all different approximation strategies of the dual solution and all examined quantities of interest.

In general, the results for the periodic test case are similar for all three functionals. The localization I_{loc} remains close to 1 for all reconstruction approaches. The very bad quantitative estimate $I_{\text{eff}} \approx -6$ for the reduced variant is improved to a very good effectivity of $I_{\text{eff}} \approx 0.7$ by means of local enhancement.

The observations made for the periodic test case generally also hold true for the microstructure with random coefficients. The higher values in I_{loc} indicate a pronounced oscillative behavior that is also evidenced by generally worse efficiency values $I_{\text{eff}} \approx 4$. Obviously, the concrete behavior examined is owed to the random nature of the coefficients and can change slightly with different samples chosen.

Remark 4.31. The same numerical test performed with the arithmetic mean value leads to a strong *underestimation* with $I_{\text{eff}} \approx 0.001$ in case of the reduced reconstruction strategy. The effectivity of $I_{\text{eff}} \approx -6.00$ in case of the periodic coefficients comes from the fact that the geometric mean value was used.

For the reconstruction strategies to be usable for model adaptation not just a quantitative behavior $\tilde{\theta}^\delta \approx \theta^\delta$ of the *error estimators* is important, but also a mostly quantitative approximation $\tilde{\eta}_K^\delta \approx \eta_K^\delta$ of the *local error indicators*. Thus, this property shall be discussed exemplarily for the case of periodic coefficients and local functional j_3, $\langle j_3, \varphi \rangle = \partial_y \varphi(x_0)$, and for the case of random coefficients and global functional j_1.

In Figures 4.4, 4.5, and 4.6 on page 103 et seqq. the local model-error indicator $\tilde{\eta}_K^\delta$ is depicted for the different reconstruction processes. It can be seen that the local Dirichlet reconstruction leads to a significantly improved quantitative behavior. Furthermore, the numerical results also indicate (Figure 4.6) that the *qualitative behavior* of the *reduced* approximation can still be adequate when used for *marking strategies* in case of *localized functionals* (cf. Section 5.2), but not if quantitative information is actually needed (cf. Chapter 6).

Remark 4.32. It has to be noted that this numerical test of trying to achieve a *quantitative estimate* for the elliptic model problem is a kind of worst case scenario. Normally, when applied in *context of an adaptation cycle*—as will be formulated in the next chapter—the successive improvement of the effective model $\bar{A}^\delta$ will also improve the quantitative character of $z^\delta + \sum z_K^\delta$. We refer to the numerical results presented in Chapter 5.

Table 4.1. Error estimators θ^H, θ^h, θ^δ for the different quantities of interest in case of the periodic coefficients computed with different reconstruction strategies for the dual solution: *full*, *reduced* and *coarse* variants following (4.1) and (4.43), as well as the reduced and locally reconstructed variant (*enhanced*) given by (4.51).

	$\langle j_1, U\rangle$	$\langle j_1, u-U\rangle$	θ^H	θ^h	θ^δ	I_{eff}	I_{loc}
full	5.90e-1	5.08e-2	1.61e-3	-1.50e-9	4.78e-2	0.97	1.05
coarse	5.90e-1	5.08e-2	7.66e-4	-1.37e-9	1.99e-2	0.39	1.22
reduced	5.90e-1	5.08e-2	3.83e-4	-1.29e-9	-3.23e-1	-6.36	1.01
enhanced	5.90e-1	5.08e-2	3.83e-4	-1.29e-9	3.32e-2	0.65	1.06

(a) Periodic coefficients, global functional j_1

	$\langle j_2, U\rangle$	$\langle j_2, u-U\rangle$	θ^H	θ^h	θ^δ	I_{eff}	I_{loc}
full	1.15e+0	2.22e-2	4.57e-4	-6.59e-10	2.10e-2	0.97	1.22
coarse	1.15e+0	2.22e-2	2.40e-4	-6.14e-10	9.39e-3	0.42	1.37
reduced	1.15e+0	2.22e-2	1.73e-4	-5.95e-10	-1.37e-1	-6.15	1.02
enhanced	1.15e+0	2.22e-2	1.73e-4	-5.95e-10	1.57e-2	0.71	1.14

(b) Periodic coefficients, local functional j_2

	$\langle j_3, U\rangle$	$\langle j_3, u-U\rangle$	θ^H	θ^h	θ^δ	I_{eff}	I_{loc}
full	1.55e+0	1.5032e-01	2.92e-3	-4.50e-9	1.43e-1	0.97	3.21
coarse	1.55e+0	1.5032e-01	1.36e-3	-4.20e-9	6.41e-2	0.43	1.37
reduced	1.55e+0	1.5032e-01	9.10e-4	-4.07e-9	-9.32e-1	-6.20	1.02
enhanced	1.55e+0	1.5032e-01	9.10e-4	-4.07e-9	1.07e-1	0.71	1.14

(c) Periodic coefficients, local functional j_3

Table 4.2. Error estimators θ^H, θ^h, θ^δ for the different quantities of interest in case of the random coefficients computed with different reconstruction strategies for the dual solution: *full*, *reduced* and *coarse* variants following (4.1) and (4.43), as well as the reduced and locally reconstructed variant (*enhanced*) given by (4.51).

	$\langle j_1, U\rangle$	$\langle j_1, u-U\rangle$	θ^H	θ^h	θ^δ	I_{eff}	I_{loc}
full	1.48e-1	-1.82e-3	1.02e-4	-1.01e-14	-2.64e-3	1.39	4.95
coarse	1.48e-1	-1.82e-3	6.35e-5	-9.65e-15	-8.11e-3	4.42	2.08
reduced	1.48e-1	-1.82e-3	8.67e-5	-1.01e-14	-1.02e-1	56.1	1.00
enhanced	1.48e-1	-1.82e-3	8.67e-5	-1.01e-14	-6.75e-3	3.66	1.08

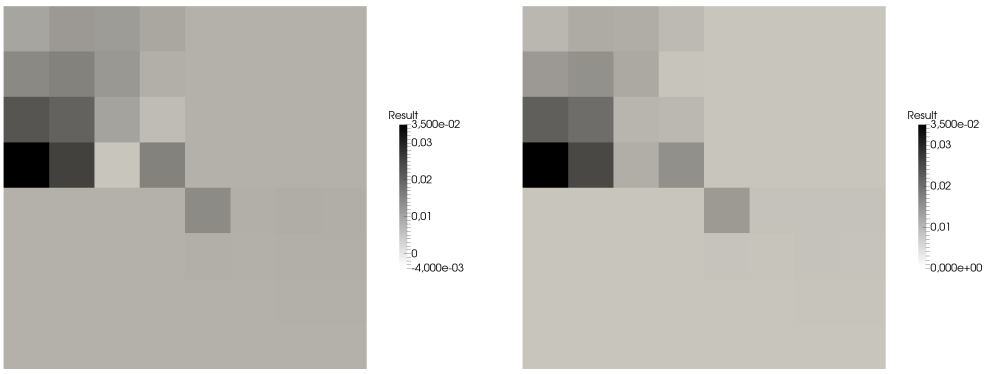
(a) Random coefficients, global functional j_1

	$\langle j_2, U\rangle$	$\langle j_2, u-U\rangle$	θ^H	θ^h	θ^δ	I_{eff}	I_{loc}
full	1.92e-1	-2.14e-3	2.52e-5	1.18e-15	-3.12e-3	1.45	7.67
coarse	1.92e-1	-2.14e-3	3.43e-5	9.89e-16	-1.04e-2	4.84	2.44
reduced	1.92e-1	-2.14e-3	4.83e-5	8.33e-16	-1.29e-1	60.3	1.01
enhanced	1.92e-1	-2.14e-3	4.83e-5	8.33e-16	-8.92e-3	4.15	1.07

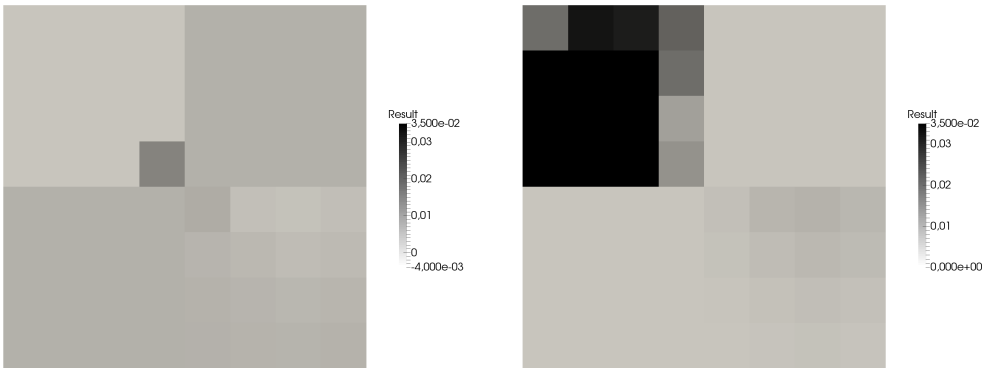
(b) Random coefficients, local functional j_2

	$\langle j_3, U\rangle$	$\langle j_3, u-U\rangle$	θ^H	θ^h	θ^δ	I_{eff}	I_{loc}
full	3.76e-1	7.93e-1	2.74e-3	-8.30e-13	7.34e-1	0.94	2.21
coarse	3.76e-1	7.93e-1	3.41e-3	-5.84e-13	4.77e-1	0.61	2.34
reduced	3.76e-1	7.93e-1	3.01e-4	-1.89e-13	2.38e-1	0.30	4.48
enhanced	3.76e-1	7.93e-1	3.01e-4	-1.89e-13	3.03e-1	0.38	1.27

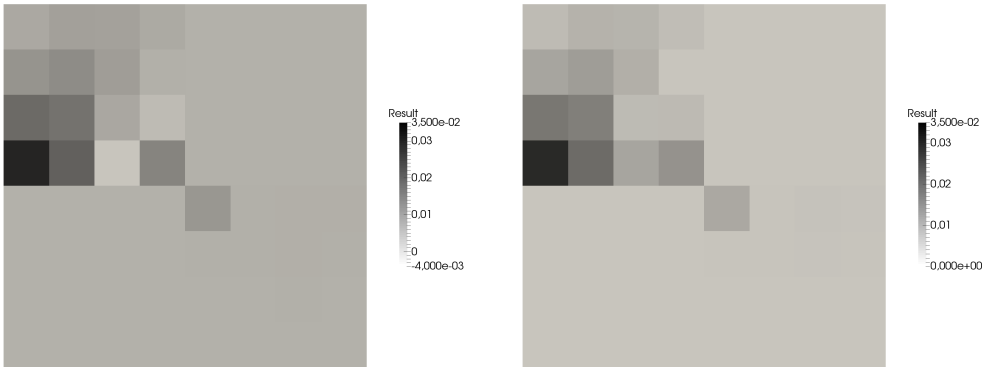
(c) Random coefficients, local functional j_3



(a) η_K^δ (left) and $|\eta_K^\delta|$ (right) for fully resolved dual solution

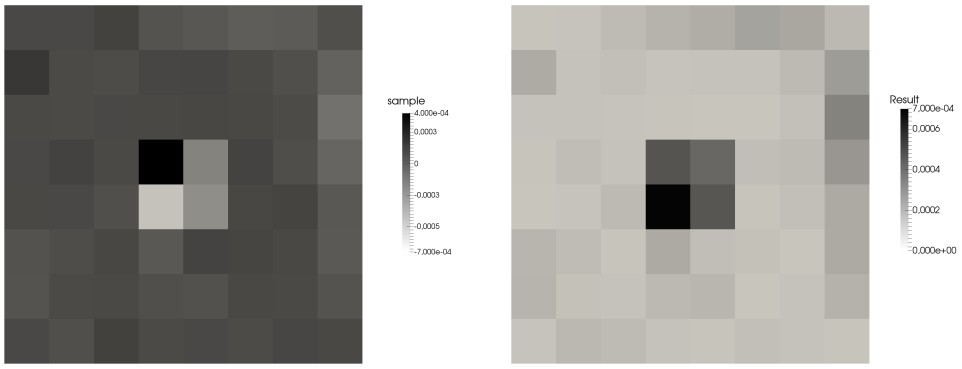


(b) η_K^δ (left) and $|\eta_K^\delta|$ (right) for reduced dual solution

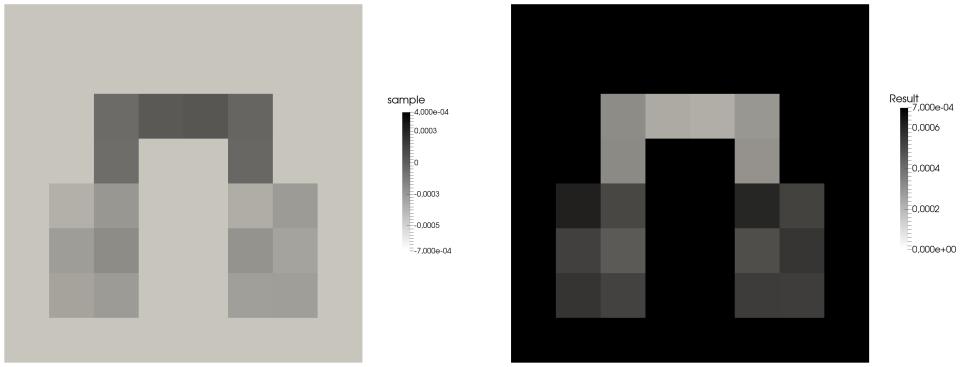


(c) η_K^δ (left) and $|\eta_K^\delta|$ (right) for the local enhancement strategy

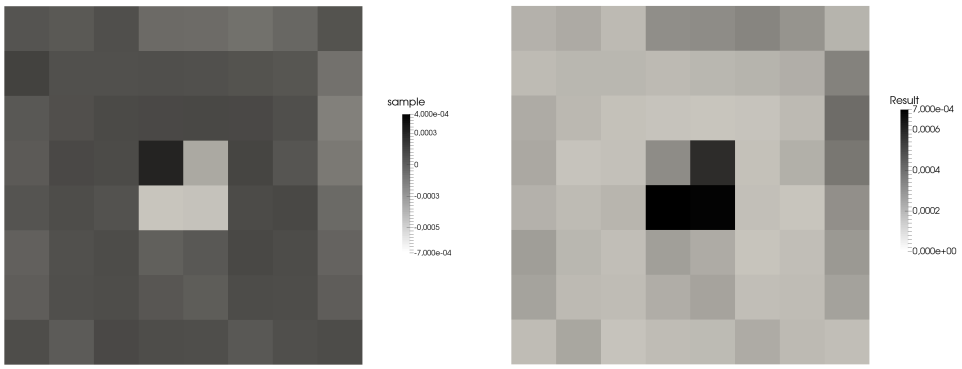
Figure 4.4. Local error indicators η_K^δ (left column) and its absolute values $|\eta_K^\delta|$ (right column) in the case of periodic coefficients and functional j_3 computed with different reconstruction strategies: *full* (a), *reduced* (b), and *enhanced* (c). The color ranges for left and right colum are fixed, respectively.



(a) η_K^δ (left) and $|\eta_K^\delta|$ (right) for fully resolved dual solution

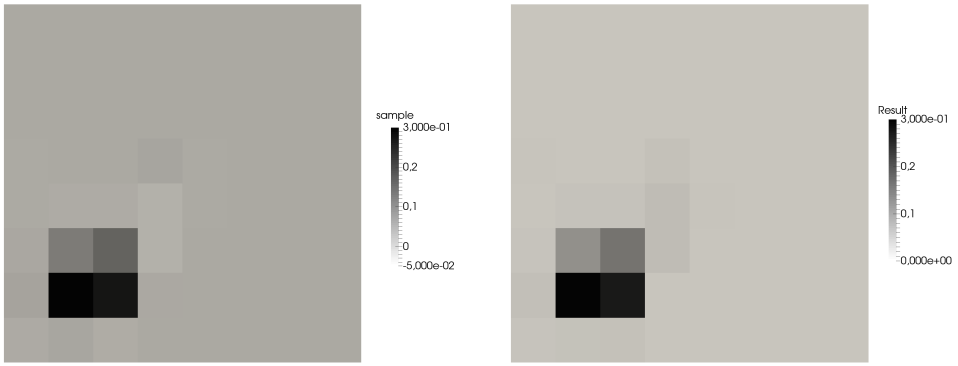


(b) η_K^δ (left) and $|\eta_K^\delta|$ (right) for reduced dual solution

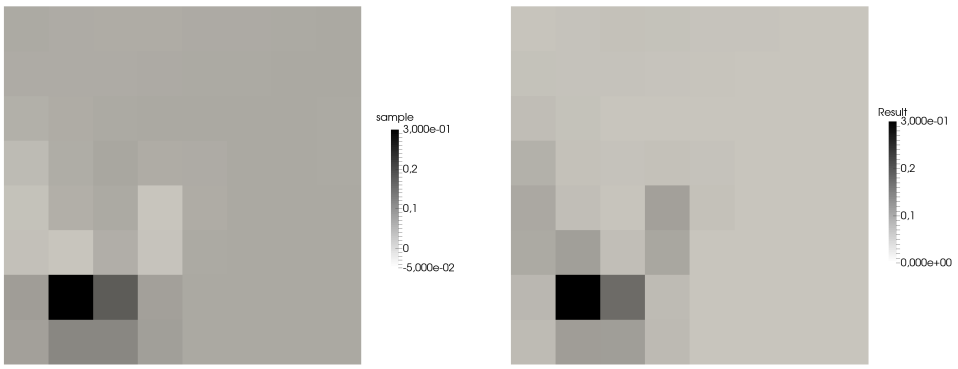


(c) η_K^δ (left) and $|\eta_K^\delta|$ (right) for the local enhancement strategy

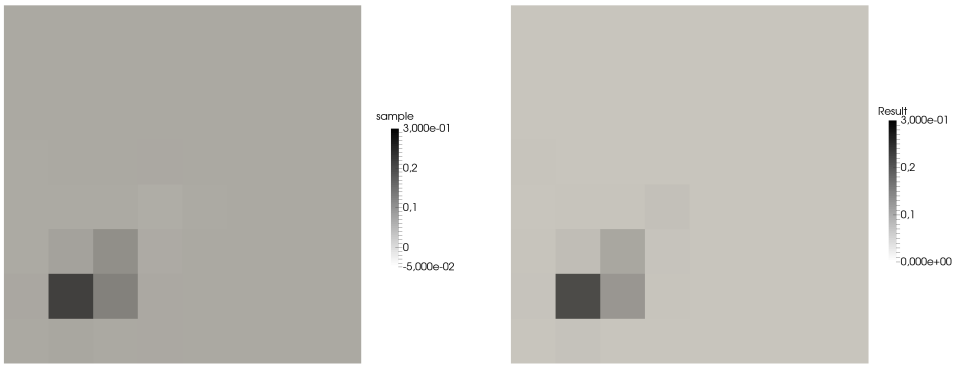
Figure 4.5. Local error indicators η_K^δ (left column) and its absolute values $|\eta_K^\delta|$ (right column) in the case of random coefficients and functional j_1 computed with different reconstruction strategies: *full* (a), *reduced* (b), and *enhanced* (c). The color ranges in the left and right colum are fixed.



(a) η_K^δ (left) and $|\eta_K^\delta|$ (right) for fully resolved dual solution



(b) η_K^δ (left) and $|\eta_K^\delta|$ (right) for reduced dual solution



(c) η_K^δ (left) and $|\eta_K^\delta|$ (right) for the local enhancement strategy

Figure 4.6. Local error indicators η_K^δ (left column) and its absolute values $|\eta_K^\delta|$ (right column) in the case of random coefficients and functional j_3 computed with different reconstruction strategies: *full* (a), *reduced* (b), and *enhanced* (c). The color ranges in the left and right colum are fixed.

5 Model-adaptation strategies

The abstract model framework with the a posteriori error estimates developed in the preceding chapters can be used for different model adaptation strategies. Based on the concept of an effective model

$$\bar{A}^\delta : \mathbb{T}_\delta(\Omega) \to \mathbb{R}^{d\times d}, \tag{5.1}$$

which was introduced in the abstract model framework (see Definition 3.3), two fundamentally different approaches for model adaptivity are possible.

The first is based on the refinement of the sampling mesh $\mathbb{T}_\delta(\Omega)$ and associated sampling regions $\{Y_K^\delta : K \in \mathbb{T}_\delta(\Omega)\}$ while keeping the same reconstruction process for all sampling regions. It is thus comparable to classical discretization adaptation—and is indeed the model-adaptation strategy for which a priori convergence results were shown in Chapter 3. The second strategy consists of switching the effective model used for the reconstruction process[26,84–87,91]. This is done by locally selecting a more expensive but also more precise sampling strategy from an a priori chosen list of effective models. Typically, the same fixed sampling discretization is used throughout the process.

This chapter examines model-adaptation strategies based on these two approaches. In particular, a novel *sampling-adaptation strategy* is derived that employs model adaptivity by means of locally refining the sampling mesh $\mathbb{T}_\delta(\Omega)$. A balancing strategy is introduced for the sampling-adaptation process that simultaneously controls macro- and microscale discretization in addition to the sampling process.

Remark 5.1. A completely different approach for model adaptation (but still in the spirit of the second strategy) can be derived by starting with the *error identity* (4.6),

$$\langle j, u^\varepsilon\rangle - \langle j, U\rangle = \theta^H + \theta^h + \theta^\delta, \quad \theta^\delta = \big((A^\delta - A^\varepsilon)\nabla u^\delta, \nabla z^\varepsilon\big), \tag{5.2}$$

and *interpreting model adaptivity* as an *optimization problem*

$$\arg\inf_{A^\delta} \sum_{K\in\mathbb{T}_\delta(\Omega)} \Big[\big|\big((A^\delta - A^\varepsilon)\nabla u^\delta, \nabla z^\varepsilon\big)\big|^2 + \text{regularization}\Big]. \tag{5.3}$$

This approach is examined in detail in Chapter 6.

5.1 Model switching

In this section, adaptation strategies based on the local switching of effective parameters from an a priori chosen set of models are briefly discussed. A refined strategy based on local sampling-mesh adaptation is discussed in detail in the next section. In the following the sampling discretization $\mathbb{T}_\delta(\Omega)$ shall be a priori chosen and remains fixed.

5.1.1 General model switching

One of the first strategies for model adaptation that comes to mind are *model switching* strategies. Here, depending on the local model-error estimate η_K^δ, a more expensive, but also more precise sampling strategy is *locally* chosen from a fixed list

$$\mathscr{A} = \left\{\bar{A}^{\delta,0}, \bar{A}^{\delta,1}, \ldots, \bar{A}^{\delta,n}\right\} \tag{5.4}$$

of effective models. Obviously, this makes a priori knowledge about the quality of the individual effective models necessary. At least, $\eta_K^\delta \approx 0$ for $K \in \mathbb{T}_\delta(\Omega)$ must hold true for the most precise model $\bar{A}^{\delta,n}$ that is available. Otherwise, the adaptation limit, where $\bar{A}^{\delta,n}$ is chosen throughout the domain Ω, is ill-posed.

Combining this adaptation approach with standard mesh refinement for $\mathbb{T}_H(\Omega)$ and $\mathbb{T}_h(\Omega)$ results in three independent adaptation processes being available: a local adaptation of the macroscale discretization $\mathbb{T}_H(\Omega)$ controlled by η_K^H, as well as the microscale discretizations $\{h_K : K \in \mathbb{T}_\delta(\Omega)\}$ controlled by η_K^δ, and a local switch of the model by using $\bar{A}_K^{\delta,i+1}$ instead of $\bar{A}_K^{\delta,i}$ as long as $i < n$. The individual selection of macro-, micro-, and sampling cell for adaptation can be carried out by a number of different marking strategies.

Due to the fact that this approach is—with the exception of the model adaptation step—similar to the *sampling-adaptation strategy* discussed in the next section, we refer for any further details, such as a *balancing strategy* for the individual error sources, to Section 5.2. A reduced model-adaptation algorithm that utilizes the model switching approach that does not deal with discretization errors is given in Algorithm 4.

5.1.2 Binary switching

In context of the elliptic model problem (2.10) one is usually limited to *one* effective model (e. g., to an HMM approach, or to a sampling with an appropriate

Algorithm 4: General model-switching algorithm.

Choose fixed discretizations $\mathbb{T}_H(\Omega)$ for macroscale, $\mathbb{T}_\delta(\Omega)$ for sampling, and $\left\{\mathbb{T}_h(Y_K^\delta), K \in \mathbb{T}_\delta(\Omega)\right\}$ for the microscale.

while *stopping criterion not reached* **do**

- Compute U and $\tilde{Z}$ with the help of (3.25) and (4.58), respectively.
- Compute the model-error estimator and local indicators

$$\tilde{\theta}^\delta = \sum_{K \in \mathbb{T}_\delta(\Omega)} \tilde{\eta}_K^\delta.$$

according to (4.64).

- Optionally, determine a local enhancement $\tilde{\theta}^\delta_{\text{rec}} = \sum_{K \in \mathbb{T}_\delta(\Omega)} \tilde{\eta}^\delta_{K,\text{rec}}$ following Definition 4.19.
- If $|\tilde{\theta}^\delta| \leq \text{TOL}$, then stop. Otherwise continue.
- Based on the local error indicators $|\eta_K^\delta|$ select sampling cells K for model adaptation with the help of a suitable marking strategy (e. g. select a fixed fraction of the cells ordered in decreasing magnitude of the absolute indicator values $|\eta_K^\delta|$).
- *Model adaptation*. For every marked sampling cell K, switch the model from $\bar{A}_K^{\delta,i}$ to $\bar{A}_K^{\delta,i+1}$ if $i < n$.

averaging strategy), because from an a priori standpoint it is often clear what the best available model is. Deliberately choosing a sequence of sampling strategies (increasing in quality) that all exhibit the same computational workload does not make very much sense. So, the only more precise model left is to actually resolve the microscale in full. This consideration gives rise to a *binary switching* approach where either the effective value $\bar{A}_K^\delta$ is used for a given sampling cell, or the microstructure is resolved in full by not using the constant effective value but $A^\varepsilon(x)$ directly.

This approach was first described in context of the DWR method by Oden and Vemaganti[84–86] as a form of post-processing technique. The explicit formulation by locally switching the effective model is due to Braack and Ern[26], who also gave a generalization to nonlinear problems. In detail, Oden and Vemaganti

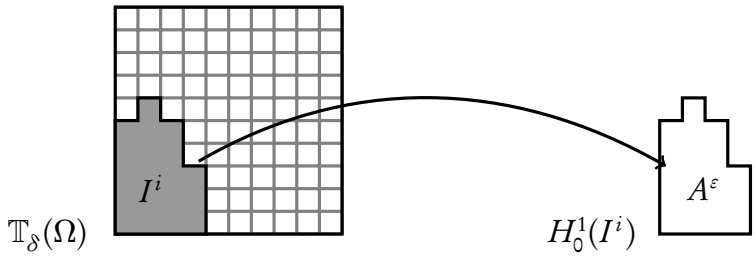


Figure 5.1. The (iteratively increased) *region of influence*[84–86] I^i for a *local* quantity of interest j. A correction with homogeneous Dirichlet boundary conditions and full coefficients $A^\varepsilon(x)$ is computed on it.

proposed to use approximate model-error indicators $\tilde{\eta}_K^\delta$, e. g., as defined in (4.46), to iteratively define a *region of influence*[85] I^n (see Figure 5.1):

Definition 5.2 (Region of influence[84–86])**.** For a *local* quantity of interest j define an initial region of influence $I^0 := \{K \in \mathbb{T}_\delta(\Omega) : K \cap \operatorname{supp} j \neq \emptyset\}$, where $\operatorname{supp} j$ is defined as $\operatorname{supp} f \cup \operatorname{supp} g$ for a representation $\langle j, . \rangle = (f, .) + (g, \nabla .)$. Now, iteratively increase I^i by selecting all *neighboring* sampling regions K of $\mathbb{T}_\delta(\Omega)$ with high absolute indicator values $|\tilde{\eta}_K^\delta|$,

$$I^{i+1} := I^i \cup \left\{K \in \mathbb{T}_\delta(\Omega) : \overline{K} \cap \overline{I}^i \neq \emptyset,\ |\tilde{\eta}_K^\delta| \text{ is high}\right\}, \tag{5.5}$$

with a suitable marking strategy[84–86].

On the region of influence I^i a correction is defined: Find $u^{\delta,i} \in H_0^1(I^i)$ s. t.

$$\left(A^\varepsilon \nabla(u^{\delta,i} + u^\delta), \nabla\varphi\right) = (f, \varphi) \quad \forall \varphi \in H_0^1(I^i). \tag{5.6}$$

This results in an iterative *post-processing strategy* where I^i is subsequently increased by selecting more and more sampling regions until the value $\langle j, u^\delta + u^{\delta,i} \rangle$ is a good enough estimate for the quantity of interest. An improved reformulation of above post-processing strategy was introduced by Braack and Ern[26] that actually switches from an effective model to a full variational formulation within an adaptation cycle similar to Algorithm 4.

The binary switching approach suffers from an intrinsic deficiency in case of bad effective models. As an illustration, consider the case of an effective model $\bar{A}^\delta$ that produces entirely false effective values. Then, the full model has to be used on a macroscopic part of the domain Ω in order to reach a certain level of accuracy, which is of the same computational complexity as a full resolution.

Consequently, even the coarse effective model $\bar{A}^\delta$ must in general fulfill a certain level of "applicability", e. g., sufficient approximation properties in the $L^2(\Omega)$-norm, to allow for local model adaptation.

Remark 5.3. One of the biggest drawbacks of a binary switching approach is the fact that a sudden change to full resolution is quite "sharp". It not only comes with a high computational penalty (full resolution) but also contains sharp transitions from $A^\varepsilon(x)$ to constant effective values. A *gradual transition* from $\bar{A}^\delta$ to full resolution might actually be preferred instead, and is thus the root concept of the *sampling-adaptation strategy* introduced in the next section.

5.2 A sampling-adaptation strategy

In this section, a *sampling-adaptation* strategy is derived that solely uses local mesh refinement of $\mathbb{T}_\delta(\Omega)$ for model adaptation. It falls into the first category of adaptation strategies outlined in the introduction of this chapter. The following discussion is based on a publication by the author[72].

5.2.1 Model adaptation by means of local refinement

The basic idea of the sampling-adaptation strategy is to use the local model-error indicators η_K^δ to adaptively refine $\mathbb{T}_\delta(\Omega)$. This approach is based on the result that for suitable regularity in the coefficient A^ε a convergence of $u^\delta \to u^\varepsilon$ for $\delta \to 0$ is available (see Proposition 3.25).

For this approach to result in an *efficient approximation*, i. e., one that actually allows for localized refinement in the sampling mesh $\mathbb{T}_\delta(\Omega)$, the chosen effective model A^δ must have some *macroscale approximation property*. For example, a good approximation when measured in $L^2(\Omega)$ might hold true. Otherwise, an adaptation process resulting in a *homogeneously* highly refined sampling mesh $\mathbb{T}_\delta(\Omega)$ would exhibit no significant computational saving compared to a straightforward numerical treatment of the model problem.

Consequently, the primary application for the proposed sampling-adaptation approach is the *local improvement* of an otherwise applicable effective model for a badly approximated, *localized* quantity of interest (such as the point functional j_3 introduced in Section 4.5). An example of such a multiscale problem was introduced in Sections 3.5 and 4.5. The sampling with the *geometric average* provides a good effective model for log-normally distributed random coefficients when measured in $L^2(\Omega)$. However, the point value $\langle j_3, u^\varepsilon \rangle$ is inadequately approximated by $\langle j_3, u^\delta \rangle$.

Remark 5.4. An insightful counterexample where a refinement in $\mathbb{T}_\delta(\Omega)$ cannot be localized is also given by the very same log-normally distributed coefficients A^ε as above, but with the *arithmetic average* chosen as sampling strategy instead.

The adaptation strategy outlined in this section is not fundamentally different from the previously examined model switching strategy. For the case of a *binary* model switching, it can be interpreted as a refined strategy instead. Consider the example of a binary model-switching strategy between a reduced model A^δ and a full resolution by using A^ε directly (with sufficient macroscale resolution). The key point is the observation that in the refinement limit of $\mathbb{T}_\delta(\Omega)$ and with a microstructure fulfilling (A1) the sampling strategy becomes equivalent to a quadrature rule. Thus, a subsequent refinement of $\mathbb{T}_\delta(\Omega)$ can be regarded as an *improved strategy* of a binary model switching in the sense that *intermediate resolutions* between the reduced model and full resolution are also considered.

Remark 5.5. This claim also holds true for homogenization strategies with local cell problems. It can be shown that the corrector defined by the local cell problem vanishes asymptotically for $\delta \ll \varepsilon$ and under assumption (A1), i. e., also homogenization processes degrade in this sense to mere quadrature rules.

A fundamental problem, one is faced with in context of model adaptation, is the fact that the model-error indicator η_K^δ only provides a scalar value as an indicator for the (local) quality of the effective model. This is less of a problem for a model switching strategy where an ordering in quality of different models is available and thus switching to the next, more expensive model is a natural strategy. In contrast, adaptation of the sampling grid $\mathbb{T}_\delta(\Omega)$ and the sampling region Y_K^δ involves a number of different parameters such that a natural choice is not obvious. In particular, the ideal position and size of Y_K^δ is an open problem.

In light of the a priori analysis of the averaging scheme given in Section 3.4 a simple strategy is chosen: The sampling regions Y_K^δ are fixed to be located in the center of each sampling region K and further scaled to a width that is a fixed fraction of the local mesh size δ_K.

Remark 5.6. Some first results for an a posteriori treatment of the sampling size of Y_K^δ were given by Larson and Målqvist [67,68,81] in the context of a VMM localization technique based on a partition of unity and presented in Subsection 2.1.3, as well as an a posteriori error analysis based on it (cf. Section 3.1). An artificial split in the error identity (3.2) is used to patch-wise increase the sampling regions ω_i [67,68]. But, only under strong conditions on the splitting (such as H^1-, or $(A^\varepsilon \nabla \cdot, \nabla \cdot)$-orthogonality) a priori convergence results of this approach are available [61,82].

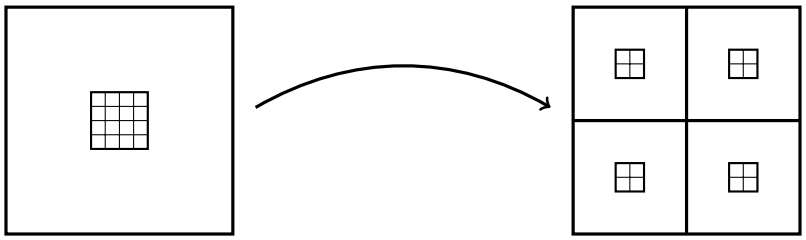


Figure 5.2. Refinement of a sampling cell K and the associated sampling region Y_K^δ. The degree of microlevel resolution is preserved.

5.2.2 The adaptation algorithm in detail

For the sake of simplicity, let the individual meshes $\mathbb{T}_h(Y_K^\delta)$ of sampling regions Y_K^δ be homogeneous and let h_K denote the local (uniform) refinement parameter of this finescale discretization. No further restriction on the discretization is made, i. e., the macroscale and sampling discretization $\mathbb{T}_H(\Omega)$ and $\mathbb{T}_\delta(\Omega)$, respectively, are allowed to be a family of locally refined meshes according to Remark 3.17 in Chapter 3.

In order to control the microscale discretization when adapting the sampling region, the local discretization $\mathbb{T}_h(K)$ of a sampling region $K \in \mathbb{T}_\delta(\Omega)$ involved has to be kept at the *same level of resolution*. More precisely, the following approach is chosen:

- If η_K^δ is large for some $K \in \mathbb{T}_\delta(\Omega)$, split K into 2^d sampling regions K_i, correspondingly shrink the sampling region Y_K^δ by a factor of 2^{-d} and set $h_{K_i} = h_K$. This amounts to a coarsening of one level if expressed in terms of number of grid refinements (see Figure 5.2).

Combining this refinement procedure with standard mesh refinement for $\mathbb{T}_H(\Omega)$ and $\mathbb{T}_h(Y_K^\delta)$ by splitting individual cells results in three independent refinement processes being available: a local adaptation of the macroscale discretization $\mathbb{T}_H(\Omega)$ controlled by η_K^H, as well as the microscale discretizations $\{h_K : K \in \mathbb{T}_\delta(\Omega)\}$ controlled by η_K^h, and finally the sampling $\mathbb{T}_\delta(\Omega)$ itself controlled by η_K^δ. The individual local mesh refinement itself can be carried out by a number of different *marking strategies*. A straightforward method is to select a fixed fraction of the cells ordered in decreasing magnitude of the absolute indicator values $|\eta_K^\nu|$ ($\nu = H, h, \delta$). More sophisticated marking strategies have been derived that take approximation orders into account, see Braack and Richter[27].

The last part missing is a *balancing strategy* for those independent refinement strategies in order to maintain a balanced error contribution. Thus, not all cells

that are originally selected on the individual discretization are refined but only the fraction

$$\frac{\alpha_\nu |\tilde{\theta}^\nu|}{\alpha_H |\tilde{\theta}^H| + \alpha_\delta |\tilde{\theta}^\delta| + \alpha_h |\tilde{\theta}^h|} \tag{5.7}$$

for $\nu = H, h, \delta$, where α_ν are a priori chosen scaling parameters. Now, let the goal be to reach a certain error tolerance

$$\left|\theta^H + \theta^h + \theta^\delta\right| \leq \mathrm{TOL}. \tag{5.8}$$

To achieve this, a finite number of repetitions of above adaptation cycle is carried out, see Algorithm 5.

Remark 5.7. A local enhancement $\tilde{\theta}^\delta_{\mathrm{rec}} = \sum_{K \in \mathbb{T}_\delta(\Omega)} \tilde{\eta}^\delta_{K,\mathrm{rec}}$, as proposed in Definition 4.19, is usually not necessary for this adaptation strategy to be efficient. The reason for this lies in the fact that the reduced variant of error indicators still gives a reasonable qualitative localization, see Section 4.5.2. In order to compensate for the loss of quantitative information, however, the scaling parameters α_ν ($\nu = H, h, \delta$) have to be chosen appropriately. The numerical results of Section 4.5.2 indicate that a slight initial overestimation of the absolute value of the model error by a factor of 6 has to be expected, hence the microscale discretization should be given slightly more weight by approximately the same factor: $\alpha_h \approx 10$, $\alpha_\delta \approx 1$. The macroscale discretization is usually less critical because it is implicitly coupled by the technical requirement of $\mathbb{T}_H(\Omega)$ to be a refined discretization of $\mathbb{T}_\delta(\Omega)$. Thus, depending on emphasis, a choice of $\alpha_H = 1 - 10$ (compared to $\alpha_\delta = 1$) is adequate.

5.3 Numerical results for the sampling-adaptation strategy

The proposed algorithm is tested on both microstructures introduced in Section 3.5. The resulting mesh refinement is compared with uniform and local refinement as well as the HMM. The geometric averaging strategy is chosen as a base sampling strategy for the adaptive sampling algorithm, starting from a coarse macroscale and microscale discretization of 256 and 1 024 cells, respectively, with 64 distinct sampling regions. As quantity of interest the highly localized third functional j_3, as given in (4.69), is chosen,

$$\langle j_3, \varphi \rangle = \partial_y \varphi(x_0). \tag{5.10}$$

Algorithm 5: Sampling-adaptation algorithm.

Start with initial discretization meshes $\mathbb{T}_H(\Omega)$, $\mathbb{T}_\delta(\Omega)$, $\{\mathbb{T}_h(Y_K^\delta), K \in \mathbb{T}_\delta(\Omega)\}$ and choose scaling parameters α_H, α_h, and α_δ.

while *stopping criterion not reached* **do**

– Compute $A^{\delta,h}$ and $\bar{A}^{\delta,h/2}$ according to one of the strategies in Definition 3.11 and compute U and $\tilde{Z}$ with the help of (3.25) and (4.58), respectively.

– Compute the error estimators and local indicators

$$\tilde{\theta}^H = \sum_{K\in\mathbb{T}_H(\Omega)} \tilde{\eta}_K^H, \qquad \tilde{\theta}^h = \sum_{K\in\mathbb{T}_\delta(\Omega)} \tilde{\eta}_K^h, \qquad \tilde{\theta}^\delta = \sum_{K\in\mathbb{T}_\delta(\Omega)} \tilde{\eta}_K^\delta.$$

according to (4.59), (4.60) and (4.64). Optionally, determine a local enhancement $\tilde{\theta}^\delta_{\mathrm{rec}} = \sum_{K\in\mathbb{T}_\delta(\Omega)} \tilde{\eta}^\delta_{K,\mathrm{rec}}$ following Definition 4.19.

– If $|\tilde{\theta}^H + \tilde{\theta}^h + \tilde{\theta}^\delta| \leq \mathrm{TOL}$, then stop. Otherwise continue.

– Based on the local error indicators $|\eta_K^\nu|$ select, for each source of error independently, cells for refinement with the help of a suitable marking strategy.

– *Balancing*. In order to balance the adaptation, not all of the selected cells are used but only the fraction

$$\frac{\alpha_\nu |\tilde{\theta}^\nu|}{\alpha_H |\tilde{\theta}^H| + \alpha_\delta |\tilde{\theta}^\delta| + \alpha_h |\tilde{\theta}^h|}, \tag{5.9}$$

where α_ν are fixed scaling parameters.

– *Microlevel refinement*. Set $h_K \leftarrow h_K/2$ for all selected cells $K \in \mathbb{T}_\delta(\Omega)$.

– *Sampling refinement*. In order to maintain $\mathbb{T}_H(\Omega) \supset \mathbb{T}_\delta(\Omega)$, mark every macrocell $K \in \mathbb{T}_H(\Omega)$ for refinement that is equal to a selected sampling cell K of $\mathbb{T}_\delta(\Omega)$. Split each selected sampling cell K into a finite number of children cells Q, associate new sampling regions Y_Q^δ with half edge length and discretizations $\mathbb{T}_\nu(Y_Q^\delta)$ with $\nu = h_K$, cf. Figure 5.2.

– *Macrolevel refinement*. Split each selected macrolevel cell $K \in \mathbb{T}_H(\Omega)$ into a finite number of child cells (while ensuring that $\mathbb{T}_H(\Omega) \supset \mathbb{T}_\delta(\Omega)$).

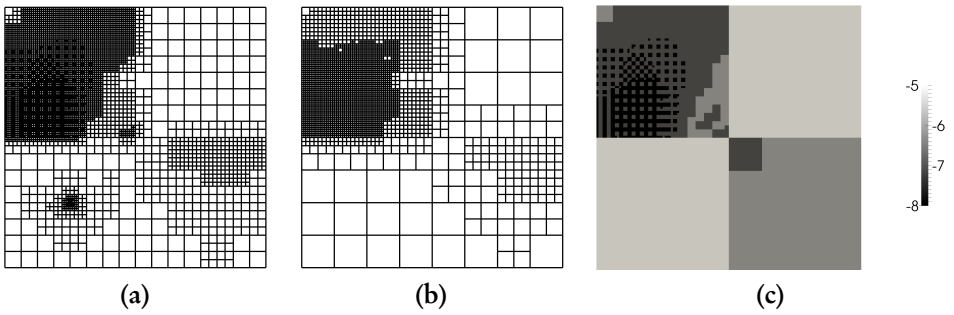


Figure 5.3. Refinement on cycle 6 (with 7 174 macro cells) for the periodic coefficients on (a) the macroscale $\mathbb{T}_H(\Omega)$, (b) the sampling discretization $\mathbb{T}_\delta(\Omega)$, and (c) on the microscale discretization $\{h_K\}$ (logarithmically scaled).

5.3.1 Periodic coefficients

Following Remark 5.7, the scaling parameters in the balancing step of the adaptive algorithm are set to $\alpha_H = \alpha_\delta = 1$ and $\alpha_h = 10$. This enforces slightly more accuracy on the microscale. The initial microscale resolution is chosen in such a way that the microstructures are coarsely resolved. The results of the adaptation process are shown in Table 5.1 together with the intermediate values for the different error estimators. A slight initial overestimation, $|I_{\text{eff}}| > 1$, mainly of the model error, changes into a pronounced underestimation in the refinement limit. The qualitative character of the error indicators is still well preserved.

Figure 5.3 shows the adapted meshes for cycle 6 (which corresponds to 5% relative error): The macroscale discretization (Figure 5.3a) is locally refined at the point $(0.25, 0.25)$ in quadrant I (due to the choice of quantity of interest) as well as where the sampling discretization enforces a local refinement. The sampling discretization is adapted in quadrants II and IV, with significantly more refinement in quadrant II than in quadrant IV. The same observation holds true for the microscale discretization. The adaptive sampling algorithm leads to a significant improvement in terms of necessary refinement (on macro- and microscale) compared to a direct finite-element discretization and HMM (cf. Sections 3.5.1 and 3.5.2). The values for the HMM are determined with the help of the adaptive algorithm and a fixed sampling mesh in order to provide a more sensible comparison than uniform refinement.

From Table 5.2 it can be seen that the adaptive sampling is consistently better than uniform and local refinement. This is mainly due to the fact that the geometric average allows for a higher localization in the mesh adaptation than

the arithmetic average, which is automatically used in the classical finite-element discretization. Compared to the adaptive HMM the adaptive algorithm needs significantly more cells for macroscale and sampling (compared to the fixed 1 024 sampling regions of the HMM). However, the 3.0×10^5 cells on the microscale correspond to 1 024 full sampling problems in case of the HMM, whereas for the averaging scheme only a significantly less expensive *averaging procedure* with 9.0×10^4 cells is necessary.

Table 5.1. Refinement history, error in the quantity of interest and error estimator for the sampling strategy with geometric averaging.

#macro	#sampl.	#micro	$\lvert\langle j_3, u^\varepsilon - U\rangle\rvert$	$\lvert\theta^H\rvert$	$\lvert\theta^b\rvert$	$\lvert\theta^\delta\rvert$	$\lvert I_{\text{eff}}\rvert$
256	64	1 024	9.04e-2 (11%)	3.61e-2	1.04e-1	9.99e-1	11.811
262	91	1 456	2.63e-1 (15%)	1.48e-1	1.80e-2	9.00e-1	2.929
385	217	1 696	2.13e-1 (13%)	9.87e-2	8.87e-3	9.05e-1	3.830
886	709	2 032	1.87e-1 (11%)	7.31e-2	5.76e-3	9.06e-1	4.489
2 803	2 620	4 180	1.61e-1 (9.5%)	4.83e-2	9.84e-4	3.36e-1	1.794
7 357	7 102	8 374	6.58e-2 (3.9%)	9.60e-3	2.23e-4	1.17e-1	1.633
17 749	17 407	18 355	3.12e-2 (1.8%)	5.29e-3	6.74e-5	5.25e-2	1.515
39 736	39 262	39 775	2.63e-2 (1.5%)	1.34e-2	1.46e-5	2.15e-2	0.310
89 911	88 675	89 056	1.11e-2 (0.7%)	5.48e-3	3.02e-6	9.06e-3	0.323

Table 5.2. Refinement levels in number of cells on macro, micro and sampling discretization to achieve at least (a) 5% and (b) 1% error.

strategy	#macro	#micro	#sampl.	$\lvert\langle j, u^\varepsilon - U\rangle\rvert$
uniform	65 536	—	—	8.62e-2 (5.1%)
local	35 377	—	—	1.06e-1 (6.2%)
ad. sampl.	7 174	6 679	8 149	7.13e-2 (4.1%)

(a)

strategy	#macro	#micro	#sampl.	$\lvert\langle j, u^\varepsilon - U\rangle\rvert$
uniform	1 048 576	—	—	6.64e-3 (0.4%)
local	568 201	—	—	1.19e-2 (0.7%)
HMM	19 330	302 896	1 024	1.55e-2 (0.9%)
ad. sampl.	93 245	89 017	89 500	1.12e-2 (0.7%)

(b)

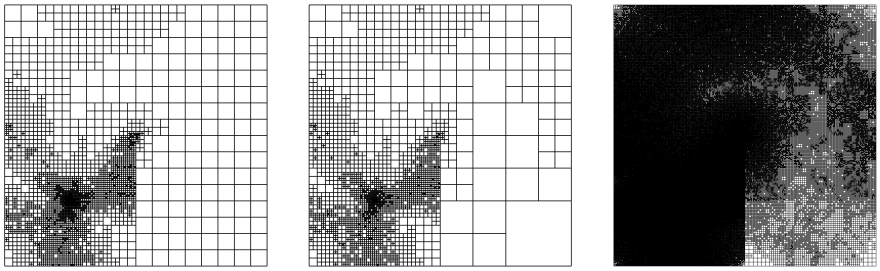


Figure 5.4. Refinement on cycle 10 (that reaches 2% accuracy) on (a) the macroscale $\mathbb{T}_H(\Omega)$ and (b) the sampling discretization $\mathbb{T}_\delta(\Omega)$. Panel (c) shows a corresponding locally refined mesh to reach 2% accuracy with a direct finite-element discretization.

5.3.2 Random coefficients

As a last numerical test, we perform a full adaptation process for the random permeability with coarse initial macroscale discretization and a coarse sampling mesh. In contrast to the first numerical example, a high initial resolution of the microscale with about 10^6 cells is chosen (resulting in $h = 2^{-11}$ respecting quadrature). This is done in order to avoid unnecessary refinement on macroscale and sampling, which turns out to happen if the microscale resolution is not sufficiently good. This is due to the bad convergence behavior with respect to the microscale discretization parameter h as evidenced in Sections 3.5 and 4.5. Given the fact that in case of averaging schemes a high microscale resolution does only introduce a high quadrature, such a choice is still computationally acceptable.

A slightly different scaling is chosen as in the first numerical example. We set $\alpha_\delta = \alpha_h = 1$, because the microscale is already resolved by the initial choice of h, and enforce a higher accuracy on the macroscale discretization by setting $\alpha_H = 20$. The results of the adaptation process are given in Table 5.3. Compared to uniform and local refinement with a standard finite-element discretization, a significant saving can be observed for macroscale and sampling discretization with a difference of almost two orders of magnitude in refinement (see Figure 5.4 and Figure 5.5). The huge saving is due to the fact that the geometric average allows for a very localized refinement process in order to improve the accuracy in the quantity of interest. With this, a relative error of around 2% can be reached with very little coarse scale and sampling refinement. After that, convergence stagnates and the adaptive sampling algorithm "degenerates" to local refinement.

Table 5.3. Refinement history: error in the quantity of interest and error estimators for the adaptive sampling strategy with geometric averaging.

	#macro	#sampl.	#micro	$\lvert\langle j, u^\varepsilon - U\rangle\rvert$	$\lvert\theta^H\rvert$	$\lvert\theta^b\rvert$	$\lvert\theta^\delta\rvert$	I_{eff}
1	256	16	$\approx 10^6$	8.09e-1 (70%)	1.07e-2	1.96e-13	2.14e-1	0.25
3	388	37	$\approx 10^6$	7.94e-1 (68%)	7.56e-3	6.25e-15	3.14e-1	0.38
5	652	166	$\approx 10^6$	7.75e-1 (66%)	6.08e-3	2.98e-16	9.84e-2	0.13
7	1 285	760	$\approx 10^6$	5.22e-1 (45%)	1.22e-2	6.29e-15	2.12e-1	0.43
9	3 442	1 981	$\approx 10^6$	1.78e-1 (15%)	2.13e-2	7.47e-14	5.26e-1	2.83
11	11 788	8 581	$\approx 10^6$	1.77e-2 (1.5%)	1.14e-2	5.76e-13	2.59e-1	14.05
13	50 836	26 641	$\approx 10^6$	2.10e-2 (1.8%)	1.37e-2	5.80e-13	1.55e-1	6.73
15	191 368	62 290	$\approx 10^6$	1.61e-2 (1.4%)	9.66e-3	5.84e-13	9.24e-2	5.14
17	658 525	129 910	$\approx 10^6$	1.25e-2 (1.1%)	5.29e-3	6.00e-13	4.68e-2	4.88

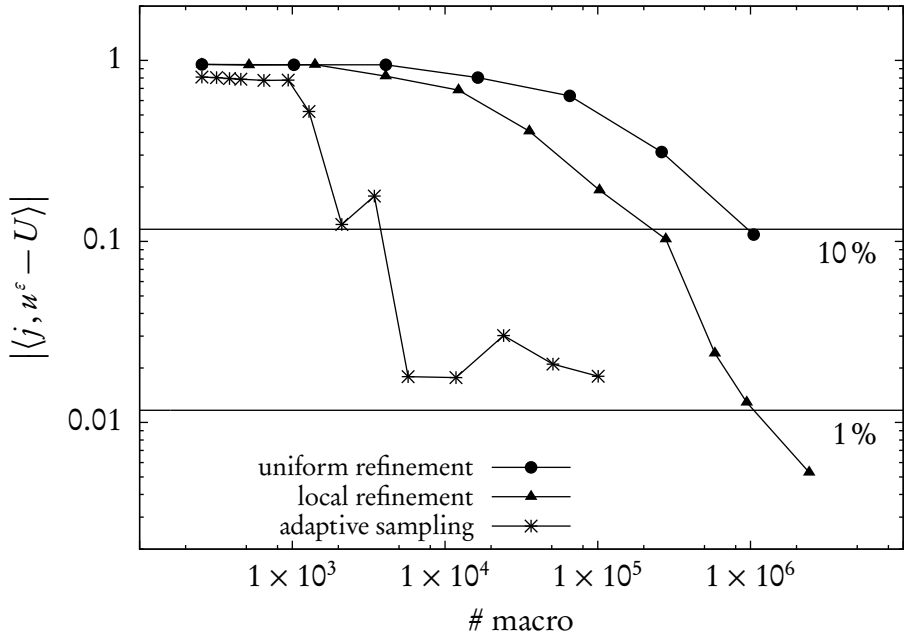


Figure 5.5. Performance plot showing the error in quantity of interest over macroscale discretization for uniform and local refinement with a standard finite element discretization and the adaptive sampling strategy with geometric averaging

6 Optimization strategies for model adaptivity

The model-adaptation strategies discussed in the preceding chapter are all based on the availability of a priori knowledge about the quality of effective models. The efficiency of the model-adaptive process itself is solely based on the selection of an improved model, or alternatively, a refinement of the sampling discretization. In this sense, only *qualitative information* of the model-error indicators η_K^δ enter the adaptation process—its *stability* is solely based on a priori knowledge.

On the other hand, the numerical results of Chapter 4 indicate that the model-error estimators and indicators based on a coarse approximation of the full dual solution z^ε, or its reduced, locally enhanced variant $z^\delta + \sum z_K^\delta$, already exhibit nearly quantitative behavior. Given the fact that a concise choice of improved models might not be available (see Section 5.1.2) the question arises whether the almost quantitative error information can be directly used for model adaptation.

This chapter presents a novel approach for model adaptivity that expresses the adaptation process as a *minimization problem* of the error estimator θ^δ: Given the *error identity* (4.6),

$$\langle j, u^\varepsilon\rangle - \langle j, U\rangle = \theta^H + \theta^h + \theta^\delta, \quad \theta^\delta = \big((A^\delta - A^\varepsilon)\nabla u^\delta, \nabla z^\varepsilon\big), \tag{6.1}$$

model adaptivity is interpreted as an *optimization problem*

$$\arg\inf_{A^\delta} \sum_{K\in\mathbb{T}_\delta(\Omega)} \Big[\big|\big((A^\delta - A^\varepsilon)\nabla u^\delta, \nabla z^\varepsilon\big)\big|^2 + \text{regularization}\Big]. \tag{6.2}$$

This approach can be used as a *model-optimization framework* to locally select optimal coefficients from a set of available models (see Section 6.2), as well as in situations where an effective model is not known and, thus, an efficient post-processing strategy is needed to construct one (see Section 6.3). It has the advantage that no a priori knowledge about effective models and reconstruction principles has to be available. The optimization problem itself is used to select the optimal model.

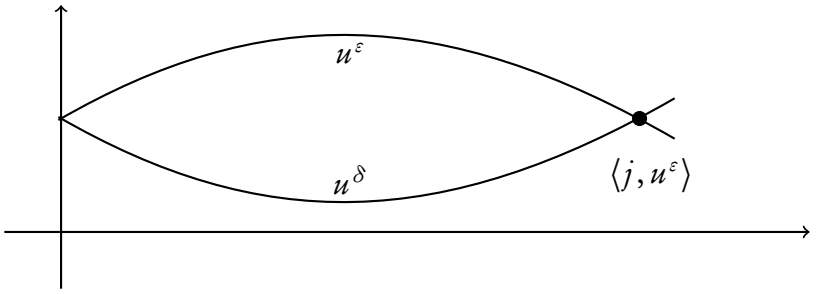


Figure 6.1. Ill-posedness of the naive optimization problem (6.5): For a point functional j an effective solution u^δ might result in a good approximation $\langle j, u^\delta \rangle \approx \langle j, u^\varepsilon \rangle$ while not reproducing the global behavior of u^ε at all.

6.1 Model-optimization framework

In context of the a posteriori error analysis that was derived for the abstract model framework the *quality* of an effective model A^δ *with respect to a quantity of interest* j can be measured with the help of the error identity (4.6)

$$\langle j, u^\varepsilon \rangle - \langle j, u^\delta \rangle = \big((A^\delta - A^\varepsilon)\nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\big)_{L^2(\Omega)^d}. \tag{6.3}$$

Now, given an a priori chosen sampling discretization $\mathbb{T}_\delta(\Omega)$, define a set of *admissible coefficients* consisting of symmetric and elliptic coefficients (as defined in Chapter 2) for this fixed sampling discretization,

$$\mathscr{A}^\delta := \big\{\bar{A}^\delta : \mathbb{T}_\delta(\Omega) \to \mathbb{R}^{d\times d} : \bar{A}^\delta \text{ fulfills (2.8) and (2.9)}\big\}. \tag{6.4}$$

The error identity (6.3) encourages to define an optimization problem over above set $\mathscr{A}^\delta$ of admissible effective models:

$$\operatorname*{arg\,inf}_{\bar{A}^\delta \in \mathscr{A}^\delta} \big|\big((\bar{A}^\delta - A^\varepsilon)\nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\big)_{L^2(\Omega)}\big|. \tag{6.5}$$

However, in this form the fact whether the optimization problem is well-behaved highly depends on the locality of the functional j in question (see Remark 6.7). As a counterexample, consider a point functional (such as j_2 or j_3 defined in Section 4.5). For such a choice, an infinite number of effective values $\bar{A}^\delta$ are possible that all fulfill the error identity (6.3). A schematic illustration of this is given in Figure 6.1. Hence, above optimization problem is ill-posed.

From a physical point of view such a behavior is also not desirable. As examined in Section 2.3, upscaling strategies are usually based on physical averaging

principles motivated by *first principles*. Thus, those upscaling principles should in general be fulfilled by a numerical approximation as well—even if they fail to reproduce localized quantities of interest correctly.

A solution for this dilemma is to add a penalization term to the optimization problem that controls the deviation of the effective values $\bar{A}^\delta$ from a physical model $\bar{A}^{\delta,0}$ (or any other choice of initial model). In this sense, the resulting optimization problem can be restated as a *local optimization of a given initial model* $\bar{A}^{\delta,0}$:

Definition 6.1 (Model-optimization problem)**.** Let $\bar{A}^{\delta,0}$ be an initial effective model and let $\{\alpha_K\}_{K\in\mathbb{T}_\delta(\Omega)}$, $\alpha_K \in \mathbb{R}^+$, be a set of (local) regularization parameters. Then, an optimal model $\bar{A}^{\delta,\mathrm{opt}}$ is defined to be a solution of

$$\underset{\bar{A}^\delta\in\mathscr{A}^\delta}{\arg\inf} \sum_{K\in\mathbb{T}_\delta(\Omega)} \left|\left((\bar{A}^\delta - A^\varepsilon)\nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\right)_K\right|^2 + \alpha_K \left\|\bar{A}^\delta_K - \bar{A}^{\delta,0}_K\right\|^2_{\mathbb{R}^{d\times d}}, \tag{6.6}$$

subject to the side condition

$$\left(\bar{A}^\delta \nabla u^\delta(\bar{A}^\delta), \nabla\varphi\right) = (f,\varphi) \quad \forall\varphi \in H^1_0(\Omega). \tag{6.7}$$

Remark 6.2. Given the fact that the ellipticity (2.9) is impractical to enforce in practice, because the correct lower-bound α is usually not known, the ellipticity constraint present in $\mathscr{A}^\delta$ is dropped in the concrete numerical computations to avoid the hard side-condition. The regularization term together with a factor α_K appropriately chosen is enough to ensure sensible coefficients $\bar{A}^\delta$.

The regularization parameter α_K is best fixed to a uniform value $\alpha_K = \alpha_0$ on all sampling regions. Here, α_0 is chosen to be roughly $0.01 - 1$ times the typical size of $|\tilde{\theta}^\delta|^2/|\bar{A}^\delta_K|^2$.

Remark 6.3. While such a uniform choice is often sufficient, improved strategies are also possible. In spirit of above discussion of *a local optimization* of a given initial model $\bar{A}^{\delta,0}$, the penalization can, for example, be further localized with the help of the dual solution z^ε. Assuming z^ε is the corresponding dual solution of a local functional such as j_2, or j_3 (defined in Section 4.5), then z^ε—interpreted as the propagator of error—is of pronounced Green's function character. In such a case it is often desirable to localize the model adaptation to the region with high values in ∇z^ε. Hence, an alternative choice

$$\alpha_K \sim \frac{1}{\|\nabla z^\varepsilon\|_{L^2(\Omega)^d}} \tag{6.8}$$

can be made instead of a uniformly chosen parameter α_K.

Proposition 6.4. The optimization problem is well-posed.

Proof. The functional dependency $u^\delta(\bar{A}^\delta)$ described by (6.7) with respect to $\bar{A}^\delta \in \mathscr{A}^\delta$ is well-posed—i. e. (6.7) is always uniquely solvable—and continuous. Further,

$$||\nabla u^\delta(\bar{A}^\delta)|| \leq \frac{1}{\alpha}||f||, \tag{6.9}$$

by definition of $\mathscr{A}^\delta$. Hence, the function

$$\mathscr{F}(\bar{A}^\delta) := \sum_{K\in\mathbb{T}_\delta(\Omega)} \left|\left((\bar{A}^\delta - A^\varepsilon)\nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\right)_K\right|^2 + \alpha_K \left\|\bar{A}_K^\delta - A_K^{\delta,0}\right\|^2_{\mathbb{R}^{d\times d}} \tag{6.10}$$

is well-defined, continuous, and coercitive, i. e., it holds

$$\mathscr{F}(\bar{A}^\delta) \to \infty \text{ for } ||\bar{A}^\delta|| \to \infty. \tag{6.11}$$

The optimization problem is thus well-posed. □

Remark 6.5. The functional dependency $u^\delta(\bar{A}^\delta)$ given by the side-condition (6.7) is highly nonlinear. In fact, $||\nabla u^\delta||_{L^2(K)} \to 0$ has to be expected for the limit $||\bar{A}_K^\delta|| \to \infty$. Consequently, $\left|\left((\bar{A}^\delta - A^\varepsilon)\nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\right)_K\right|^2$ is in general not convex. The optimization problem is therefore not necessarily uniquely solvable.

In preparation for the numerical treatment of the optimization problem (6.6), we formulate the following regularity result for the cost functional $\mathscr{F}$ given in (6.10):

Proposition 6.6. The functional dependency $\mathscr{F}(\bar{A}^\delta)$ is *Gâteaux-differentiable* and its derivative $\mathrm{D}\mathscr{F}(\bar{A}^\delta)[\delta\bar{A}^\delta]$ in direction $\delta\bar{A}^\delta$ is given by

$$\begin{aligned}\mathrm{D}\mathscr{F}(\bar{A}^\delta)[\delta\bar{A}^\delta] = \sum_{K\in\mathbb{T}_\delta(\Omega)} 2\eta_K^\delta \Big\{ &\left(\delta\bar{A}_K^\delta \nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\right)_K \\ &+ \left((\bar{A}^\delta - A^\varepsilon)\nabla \mathrm{D}u^\delta(\bar{A}^\delta)[\delta\bar{A}^\delta], \nabla z^\varepsilon\right)_K \Big\} \\ &+ \sum_{K\in\mathbb{T}_\delta(\Omega)} 2\alpha_K (\bar{A}_K^\delta - \bar{A}_K^{\delta,0}) : \delta\bar{A}_K^\delta,\end{aligned} \tag{6.12}$$

with the solution $\mathrm{D}u^\delta(\bar{A}^\delta)[\delta\bar{A}^\delta]$ of the equation

$$\left(\bar{A}^\delta \nabla \mathrm{D}u^\delta(\bar{A}^\delta)[\delta\bar{A}^\delta], \nabla\varphi\right) + \left(\delta\bar{A}^\delta \nabla u^\delta(\bar{A}^\delta), \nabla\varphi\right) = 0 \quad \forall\varphi \in H^1(\Omega). \tag{6.13}$$

Proof. The crucial part is to assert that the side condition (6.7) interpreted as a functional dependency $u^\delta(\bar{A}^\delta)$ is Gâteaux-differentiable and its derivative is given by (6.13). The rest of the statement follows in a straightforward manner.

Due to the fact that $\mathscr{A}^\delta$ is finite dimensional it suffices to show that the limit

$$\lim_{s \searrow 0} \mathrm{D}_s u^\delta, \quad \mathrm{D}_s u^\delta := \frac{1}{s}\big(u^\delta(\bar{A}^\delta + \delta\bar{A}^\delta) - u^\delta(\bar{A}^\delta)\big) \tag{6.14}$$

is well defined for arbitrary $\delta\bar{A}^\delta$. For this we note that the difference $\mathrm{D}_s u^\delta(\bar{A}^\delta)$ is, for s sufficiently small, given by

$$\big((\bar{A}^\delta + s\,\delta\bar{A}^\delta)\nabla(u^\delta(\bar{A}^\delta) + s\,\mathrm{D}_s u^\delta), \nabla\varphi\big) = (f, \varphi) \quad \forall \varphi \in H^1(\Omega). \tag{6.15}$$

Equivalently,

$$\big(\bar{A}^\delta \nabla \mathrm{D}_s u^\delta, \nabla\varphi\big) + s\,\big(\delta\bar{A}^\delta \nabla \mathrm{D}_s u^\delta, \nabla\varphi\big) + \big(\delta\bar{A}^\delta u^\delta(\bar{A}^\delta), \nabla\varphi\big) = 0. \tag{6.16}$$

By continuity, it follows that the limit of (6.16) for $s \to 0$ is well-defined and indeed given by (6.13). □

Remark 6.7 (Large model deviations)**.** The necessity for the regularization term in Definition 6.1 limits its application to a local optimization of a given initial model $\bar{A}^{\delta,0}$. Consequently, this imposes some conditions on the quality of the base model. Or expressed the opposite way, the worse the quality of $\bar{A}^{\delta,0}$ is the smaller the regularization parameters $\{\alpha_K\}$ have to be chosen to allow for a large enough deviation. The regularization term is introduced to primarily ensure robustness in case of localized functionals—and, in regard of Remark 6.2, to provide an alternative for a strong ellipticity side condition. In case of a *global* quantity of interest the optimization problem itself is in general already stable without the regularization term. As an example, consider the global quantity of interest

$$\langle j, u^\varepsilon \rangle = \int_\Omega u^\varepsilon \,\mathrm{d}x. \tag{6.17}$$

A heuristic argument is

$$\sum_{K \in \mathbb{T}_\delta(\Omega)} \big|\big((\bar{A}^\delta - A^\varepsilon)\nabla u^\delta(\bar{A}^\delta), \nabla z^\varepsilon\big)_K\big|^2 \approx \sum_{K \in \mathbb{T}_\delta(\Omega)} \Big|\int_K u^\varepsilon - u^\delta(\bar{A}^\delta)\,\mathrm{d}x\Big|^2. \tag{6.18}$$

Now, utilizing $u^\varepsilon(x) > 0$, $u^\delta(x) > 0$ a. e. on Ω, as well as the functional dependency of $u^\delta(\bar{A}^\delta)$, one concludes that

$$\left| \int_K u^\varepsilon - u^\delta(\bar{A}^\delta)\,\mathrm{d}x \right| \to \left| \int_K u^\varepsilon\,\mathrm{d}x \right| \quad \text{for} \quad \|\bar{A}^\delta\| \to \infty. \tag{6.19}$$

Together with the observation that there is always a choice for the local coefficient A_K^δ for which $\left| \int_K u^\varepsilon - u^\delta(\bar{A}^\delta)\,\mathrm{d}x \right| < \left| \int_K u^\varepsilon\,\mathrm{d}x \right|$, it follows that the optimization problem (6.6) with $\alpha_K = 0$ is already well-posed.

Thus, for model adaptation with large deviation from the base model $\bar{A}^{\delta,0}$, global functionals with such stabilizing character should be preferred.

6.2 Model switching revisited

The optimization problem (6.6) can also be used for a reinterpretation of the model-switching strategy discussed in Section 5.1. Let $\mathscr{A}^0$ be a finite set of $n+1$ different admissible models $\bar{A}^{\delta,i}$ that fulfill (2.8) and (2.9),

$$\mathscr{A}^0 = \left\{ \bar{A}^{\delta,0}, \ldots, \bar{A}^{\delta,n} \right\}, \tag{6.20}$$

and let the task be to locally switch to an appropriate linear combination of those candidates. Therefore, restrict $\mathscr{A}^\delta$ to the affine hull of $\mathscr{A}^0$:

$$\operatorname{aff} \mathscr{A}^0 := \Big\{ \bar{A}^\delta : \ \forall K \in \mathbb{T}_\delta \ \exists \gamma \in \mathbb{R}^{n+1},\ \gamma \geq 0,\ \sum_i \gamma_i = 1 \text{ s. t. } \bar{A}_K^\delta = \sum_i \gamma_i \bar{A}_K^{\delta,i} \Big\}. \tag{6.21}$$

Lemma 6.8. The modified optimization problem (6.6) corresponding to a minimization over the closed, affine subset $\operatorname{aff} \mathscr{A}^0 \subset \mathscr{A}^\delta$,

$$\operatorname*{arg\,inf}_{\gamma \geq 0,\ \sum_i \gamma_i = 1} \sum_{K \in \mathbb{T}_\delta(\Omega)} \left| \left((\bar{A}^\delta(\gamma) - A^\varepsilon) \nabla u^\delta(\bar{A}^\delta(\gamma)), \nabla z^\varepsilon \right)_K \right|^2, \tag{6.22}$$

with the side condition $u^\delta(\bar{A}^\delta)$ given in (6.7), is already well-posed without a regularization term, i. e., with the choice $\alpha_K = 0$.

Proof. Due to $\operatorname{aff} \mathscr{A}^0 \subset \mathscr{A}^\delta$, the side condition (6.7) remains well-posed. Further, $\operatorname{aff} \mathscr{A}^0$ is compact and the function $\mathscr{F} : \operatorname{aff} \mathscr{A}^0 \mapsto \mathbb{R}^+$ that was defined in the proof of Proposition 6.4 is continuous. □

In contrast to the model switching discussed in Section 5.1, this optimization approach has the advantage that no a priori ordering of the effective models in terms of *quality* has to be known. The optimization problem itself locally selects the optimal effective coefficient, or (to be more precise) an affine combination of coefficients. This additional information comes at the prize that it requires to compute all effective coefficients $\bar{A}^{\delta,i} \in \mathscr{A}^0$ beforehand.

Consequently, the primary purpose of this modified optimization problem (6.22) is *not* to serve as a *post-processing step* in order to reduce the computational cost of using the most expensive model throughout. Instead, it is used to *find an optimal configuration* of effective coefficients in situations where the best choice is not obvious a priori. This applies, e. g., to situation with strongly heterogeneous behavior in different parts of the computational domain Ω.

This information can then be further used, for example, for model validation, or as *offline data* for a subsequent computation (e. g., tune the model with a stationary computation and run a time-dependent computation afterwards).

Remark 6.9. If an affine-linear combination of coefficients is not intended and a binary switch is needed, problem (6.22) can be transformed into a constrained optimization problem over integral values

$$\operatorname*{arg\,inf}_{\gamma \in \{0,1\}^{n+1},\, \sum_i \gamma_i = 1} \sum_{K \in \mathbb{T}_\delta(\Omega)} \left| \left((\bar{A}^\delta(\gamma) - A^\varepsilon) \nabla u^\delta(\bar{A}^\delta(\gamma)), \nabla z^\varepsilon \right)_K \right|^2. \tag{6.23}$$

6.3 An efficient post-processing strategy

The optimization problem (6.6) can be used as an efficient post-processing strategy that does not require—with the exception of an initial model—any additional a priori knowledge on effective models. For this, fix a macroscale and a sampling discretization $\mathbb{T}_H(\Omega)$ and $\mathbb{T}_\delta(\Omega)$, as well as initial effective coefficients $\bar{A}^{\delta,0}$. The dual solution is again either approximated with a coarse dual solution according to (4.57) or with the help of the effective dual solution and a local enhancement $z^\delta + \sum z_K^\delta$, given in Definition 4.27.

Remark 6.10. Within the optimization framework (6.6), as no further sampling of $\bar{A}^\delta$ by *primal reconstruction* process is necessary, the computation of the primal problem is very cheap. This justifies to also use more sophisticated, but also more expensive, reconstruction approaches for the dual solution. A natural candidate for an improved reconstruction approach is the VMM. In particular, with the fact in mind, that the *correct local behavior* of ∇z^ε is required for a

quantitative η_K^δ, VMM approaches with higher orthogonality requirements are a natural candidate (see also Section 2.1.4 and Remark 4.22).

In summary, this results in the following general optimization strategy formulated for the case of a reduced, locally enhanced approximation of the dual solution:

Definition 6.11 (Reduced, locally enhanced model-optimization problem)**.** Let $\mathbb{T}_\delta(\Omega)$ be a fixed sampling mesh and $\mathbb{T}_H(\Omega)$ a fixed macroscale discretization. Fix a microscale discretization $\{\mathbb{T}_h(K) : K \in \mathbb{T}_\delta(\Omega)\}$ as well and let $\bar{A}^{\delta,0} : \mathbb{T}_\delta(\Omega) \to \mathbb{R}^{d\times d}$ be an initial effective model. The *reduced, locally enhanced model-optimization problem* reads: Find a solution $\bar{A}^{\delta,\mathrm{opt}} \in \mathscr{A}^\delta$ of

$$\begin{aligned}\operatorname*{arg\,inf}_{\bar{A}^\delta\in\mathscr{A}^\delta} \sum_{K\in\mathbb{T}_\delta(\Omega)} \left|\left((\bar{A}^\delta - A^\varepsilon)\nabla U(\bar{A}^\delta), \nabla(\tilde{Z}+\tilde{Z}_K)(\bar{A}^\delta)\right)_K\right|^2 \\ + \alpha_K \left\|\bar{A}_K^\delta - \bar{A}_K^{\delta,0}\right\|^2_{\mathbb{R}^{d\times d}} \end{aligned} \tag{6.24}$$

with $U, \tilde{Z} \in V^H(\Omega)$, and $\tilde{Z}_K \in V^h(K)$ subject to the side conditions:

$$\left(\bar{A}^\delta \nabla U(\bar{A}^\delta), \nabla\varphi\right) = (f,\varphi) \quad \forall \varphi \in V^H(\Omega), \tag{6.25}$$

$$\left(\bar{A}^\delta \nabla\varphi, \nabla\tilde{Z}(\bar{A}^\delta)\right) = \langle j,\varphi\rangle \quad \forall\, \varphi \in V^H(\Omega), \tag{6.26}$$

$$\left(A^\varepsilon \nabla\varphi, \nabla\tilde{Z} + \nabla\tilde{Z}_K\right)_K = \langle j,\varphi\rangle \quad \forall\, \varphi \in V^h(K). \tag{6.27}$$

Remark 6.12. Here, in contrast to the first two model-adaptation strategies, a local enhancement is actually necessary, as will be evidenced by the numerical results given below.

Remark 6.13. A *microscale discretization* (with respect to h) only enters the modified model-optimization problem by means of an *approximation quality* of the dual solution. Either due to the coarse approximation, say $\tilde{Z} \in V^h(\Omega)$, or in terms of resolution of the reconstruction $\tilde{Z}_K \in V^h(K)$. In case of the local enhancement strategy, the artificial boundary conditions in (6.27) introduce an additional *systematic error* such that a solution of (6.24) is usually slightly different from a solution of (6.6).

Remark 6.14. A posteriori error estimation with a *reduced, locally enhanced* strategy given in Definitions 4.27 and 4.28 had the disadvantage that for *both*, the primal problem (in form of sampling $A^{\delta,h}$) and the dual problem (by means of a local reconstruction according to Definition 4.28) an expensive reconstruction

has to be computed. In contrast, the model-optimization framework of Definition 6.11 only requires the local reconstruction of the dual solution. The only exception is the computation of the initial model $\bar{A}^{\delta,0}$. But given the fact that the effective model is subject to an optimization procedure, this computation is usually relatively cheap.

As was already discussed in Chapter 4, the error identity lifts the question of suitable approximation in terms of a quantity of interest (for the primal problem) to the question of suitable approximation properties of the localization technique for the dual problem (see Remark 4.22). The latter is typically measured in the L^2-norm of the gradient of the error of the dual approximation, for which—depending on the localization approach—strong approximation properties are available, and not in the quantity of interest itself.

Thus, the proposed optimization framework can be interpreted as a multiscale method in its own right, where a reconstruction process is used for the dual solution. The *modeling aspect* of the optimization problem lies in the *choice of functional j* as quantity of interest, as well as (in case of a local enhancement) the choice of local reconstruction of the dual solution.

Analogously to Proposition 6.6 we formulate the following result:

Proposition 6.15. Let $\tilde{\mathscr{F}}$ be the modified cost functional of (6.24). Then, in full analogy of the result for $\mathscr{F}$ in Proposition 6.6, the functional dependency of $\tilde{\mathscr{F}}(\bar{A}^\delta)$ is also Gâteaux-differentiable and it holds true that

$$\mathrm{D}\tilde{\mathscr{F}}(\bar{A}^\delta)[\delta\bar{A}^\delta] = \mathrm{D}\mathscr{F}(\bar{A}^\delta, U, \tilde{Z})[\delta\bar{A}^\delta] + \sum_{K\in\mathbb{T}_\delta(\Omega)} 2\eta_K^\delta\big((\bar{A}^\delta - A^\varepsilon)\nabla U, \nabla(\mathrm{D}\tilde{Z} + \mathrm{D}\tilde{Z}_K)(\bar{A}^\delta)[\delta\bar{A}^\delta]\big), \quad (6.28)$$

with $\mathrm{D}\tilde{Z} \in V^H(K)$ being defined as the solution of

$$\big(\bar{A}^\delta\nabla\varphi, \nabla\mathrm{D}\tilde{Z}(\bar{A}^\delta)[\delta\bar{A}^\delta]\big) + \big(\delta\bar{A}^\delta\nabla\varphi, \nabla\tilde{Z}(\bar{A}^\delta)\big) = 0 \quad \forall\varphi\in V^H(\Omega), \quad (6.29)$$

and $\mathrm{D}\tilde{Z}_K \in V^h(K)$ solving

$$\big(A^\varepsilon\nabla\varphi, \nabla\mathrm{D}\tilde{Z}_K(\bar{A}^\delta)[\delta\bar{A}^\delta]\big)_K + \big(A^\varepsilon\nabla\varphi, \nabla\delta\tilde{Z}(\bar{A}^\delta)[\delta\bar{A}^\delta]\big)_K = 0 \quad \forall\varphi\in V^h(K). \quad (6.30)$$

Proof. The first part of the statement is already proved in Proposition 6.6. The additional terms arise from the derivatives of Equations (6.26) and (6.27). □

6.4 Implementational aspects

The optimization problem (6.6) and its modified variant (6.24) contain strongly nonlinear side conditions, where computing the Gâteaux-derivative for a given direction $\delta\bar{A}^{\delta}$ alone already involves solving the variational equation (6.13), and, depending on the reconstruction approach, also (6.29) and (6.30). Consequently, a straightforward application of the *Newton method* to solve the optimization problem,

$$\begin{cases} \bar{A}^{\delta,i+1} \leftarrow \bar{A}^{\delta,i} + \delta\bar{A}^{\delta,i}, \\ \mathrm{D}^2\mathscr{F}(\bar{A}^{\delta,i})[\delta B][\delta\bar{A}^{\delta,i}] = -\mathrm{D}\mathscr{F}(\bar{A}^{\delta,i})[\delta B] \quad \forall \delta B, \end{cases} \tag{6.31}$$

has to be avoided, because ensuring above variational equation to hold true for every direction δB implies the construction of the full gradient $\nabla\mathscr{F}$ (by interpreting $\mathscr{F} : \mathscr{A}^{\delta} \cong \mathbb{R}^{d\times d\times n} \to \mathbb{R}$), which means solving $d \times d \times n$ times Equation (6.13), etc. Further, actually determining the direction $\delta\bar{A}^{\delta,i}$ effectively squares this complexity.

In order to avoid computing the second order derivatives $\mathrm{d}^2\mathscr{F}(\bar{A}^{\delta,i})$ a *modified Gauß-Newton method*[70,73] is used. For this, we reformulate the optimization problem slightly:

Lemma 6.16. Introduce an index $(K,i,j) \in \mathbb{T}_{\delta} \times \mathbb{R}^{d\times d}$ and define the vector-valued function

$$\mathscr{G} \coloneqq \left\{ \left(\tilde{\eta}_K\right)_K, \left(g_{Kij}\right)_{Kij} \right\}, \tag{6.32}$$

with

$$\tilde{\eta}_K \coloneqq \left((\bar{A}^{\delta} - A^{\varepsilon}) \nabla U(\bar{A}^{\delta}), \nabla(\tilde{Z} + \tilde{Z}_K) \right)_K, \tag{6.33}$$

$$g_{Kij} \coloneqq \sqrt{\alpha_K} \left(\bar{A}^{\delta}_{K,ij} - \bar{A}^{\delta,0}_{K,ij} \right). \tag{6.34}$$

Then, the modified optimization problem (6.24) can equivalently be expressed as the minimization of the squared *Euclidian norm* $|.|$ of $\mathscr{G}$:

$$\operatorname*{arg\,inf}_{\bar{A}^{\delta}\in\mathscr{A}^{\delta}} \left|\mathscr{G}\right|^2 = \operatorname*{arg\,inf}_{\bar{A}^{\delta}\in\mathscr{A}^{\delta}} \sum_{K\in\mathbb{T}_{\delta}(\Omega)} \left\{ \tilde{\eta}_K^2 + \sum_{ij} g_{Kij}^2 \right\}. \tag{6.35}$$

Further, for a given index (K,i,j) let $\delta\bar{A}^{\delta}(Kij) : \mathbb{T}_{\delta} \to \mathbb{R}^{d\times d}$ be defined as the value

$$\left(\delta\bar{A}^{\delta}_Q\right)_{mn} \coloneqq \delta_{QK}\,\delta_{mi}\,\delta_{ni}, \tag{6.36}$$

for a cell $Q \in \mathbb{T}_\delta(\Omega)$, where δ_{QK} denotes *Kronecker's delta*. Define the short notation

$$\mathrm{D}_{Kij}\tilde{\eta}_Q \coloneqq \mathrm{D}\tilde{\eta}_Q[\delta\bar{A}^\delta(Kij)], \tag{6.37}$$

$$\mathrm{D}_{Kij}g_{Qmn} \coloneqq \mathrm{D}g_{Qmn}(\bar{A}^\delta)[\delta\bar{A}^\delta(Kij)]. \tag{6.38}$$

Lemma 6.17. By virtue of Propositions 6.6 and 6.15 it holds true that

$$\begin{aligned}\mathrm{D}_{Kij}\tilde{\eta}_Q \;=\; \delta_{QK}\int_Q \nabla_j U\,\nabla_i(\tilde{Z}+\tilde{Z}_Q)\,\mathrm{d}x \\ + \int_Q (\bar{A}^\delta - A^\varepsilon)\nabla \mathrm{D}_{Kij}U\cdot\nabla(\tilde{Z}+\tilde{Z}_Q)\,\mathrm{d}x \\ + \int_Q (\bar{A}^\delta - A^\varepsilon)\nabla U\cdot\nabla \mathrm{D}_{Kij}(\tilde{Z}+\tilde{Z}_Q)\,\mathrm{d}x,\end{aligned} \tag{6.39}$$

as well as

$$\mathrm{D}_{Kij}g_{Qmn} = \delta_{QK}\,\delta_{mi}\,\delta_{ni}\,\sqrt{\alpha_Q}. \tag{6.40}$$

With these prerequisites at hand, a modified Gauß-Newton iteration following a discussion by Levenberg[70] and Marquardt[73] is defined:

Definition 6.18 (Gauß-Newton iteration)**.** Let $\mathscr{J}$ denote the Jacobian matrix of $\mathscr{G}$,

$$\mathscr{J} = \Big\{\big(\mathrm{D}_{Kij}\tilde{\eta}_Q\big)_Q^{Kij}, \big(\mathrm{D}_{Kij}\tilde{g}_{Qmn}\big)_{Qmn}^{Kij}\Big\}. \tag{6.41}$$

Given a penalty $\lambda \geq 0$ and starting from an initial effective model $\bar{A}^{\delta,0}$ the modified Gauß-Newton iteration reads

$$\begin{cases}\bar{A}^{\delta,n+1} \leftarrow \bar{A}^{\delta,n} + \delta\bar{A}^{\delta,n}, \\ \big(\mathscr{J}\mathscr{J}^T(\bar{A}^{\delta,n}) + \lambda\,\mathrm{Id}\big)\delta\bar{A}^{\delta,n} = -\mathscr{J}\mathscr{G}^T(\bar{A}^{\delta,n}).\end{cases} \tag{6.42}$$

The penalization term $\lambda\,\mathrm{Id}$ acts as a dampening term in the Gauß-Newton method to stabilize the iteration and to reduce the influence of approximation errors of the Jacobian $\mathscr{J}$. Depending on the situation, it will be chosen between $0-1$ times the mean value of the diagonal elements of $\mathscr{J}\mathscr{J}^T$.

Reduction of computational complexity The computationally expensive part of computing the Jacobi matrix $\mathscr{J}$ are the non-local responses $\mathrm{D}_{Kij}U$, $\mathrm{D}_{Kij}\tilde{Z}$, and $\mathrm{D}_{Kij}\tilde{Z}_Q$ that have to be computed for each choice (K,i,j) individually according to (6.13), (6.29), and (6.30). Another aspect that has to be kept in mind is the fact that $\mathscr{J}\mathscr{J}^T$ is actually a dense matrix of size $N \times N$ with $N = \left|\mathbb{T}_\delta(\Omega)\right|(1+d^2)$. Storing such a matrix, even for moderate sizes of the sampling mesh $\mathbb{T}_\delta(\Omega)$, is computationally infeasible. Thus, a reduction strategy to efficiently approximate $\mathscr{J}$ is necessary.

The microscale response $\mathrm{D}_{Kij}U$ is given by (cf. Equation 6.13):

$$\left(\bar{A}^\delta \nabla \mathrm{D}_{Kij}U, \nabla\varphi\right) = -\int_K \nabla_j U \nabla_i \varphi \,\mathrm{d}x \quad \forall \varphi \in V^H(\Omega). \tag{6.43}$$

The right hand side of this equation is highly localized, consequently, the contribution of

$$\int_Q (\bar{A}^\delta - A^\varepsilon)\nabla \mathrm{D}_{Kij}U \cdot \nabla(\tilde{Z} + \tilde{Z}_Q)\,\mathrm{d}x \tag{6.44}$$

rapidly decreases the farther Q is away from K—and can be neglected at some point. A sensible compromise is, for example, to compute above contribution only for the case $K = Q$, or alternatively, as a more precise strategy, only if Q belongs to a small patch around K, e. g., if $\overline{K} \cap \overline{Q} \neq \emptyset$. All of this choices result in a block diagonal matrix $\tilde{\mathscr{J}}$ whose *band size* is independent of $|\mathbb{T}_\delta(\Omega)|$.

The microscale response $\mathrm{D}\tilde{Z}$, $\mathrm{D}\tilde{Z}_Q$ in contrast will just be neglected entirely:

$$\int_Q (\bar{A}^\delta - A^\varepsilon)\nabla U \cdot \nabla \mathrm{D}_{Kij}(\tilde{Z} + \tilde{Z}_Q)\,\mathrm{d}x \;\approx\; 0. \tag{6.45}$$

The reasoning behind this choice is the fact that in case of a fully resolved dual solution z^ε, or the coarse approximation strategy, such finescale response does not exist at all, and further, that the optimization problem (6.24) is designed to be an approximation of (6.6); so a derivative approximation (probably) closer to the original problem is actually desirable.

In summary, the following approximation strategies of the derivative $\mathrm{D}_{Kij}\tilde{\eta}_Q$ will be considered:

Definition 6.19 (Approximative Jacobian)**.** The derivative $\mathrm{D}_{Kij}\tilde{\eta}_Q$ is either approximated by ignoring all microscale responses $\mathrm{D}_{Kij}U$ entirely, i. e.,

$$\mathrm{D}_{Kij}\tilde{\eta}_Q \;\approx\; \delta_{QK}\int_Q \nabla_j U\, \nabla_i(\tilde{Z} + \tilde{Z}_Q)\,\mathrm{d}x, \tag{6.46}$$

or by taking the local response of $\mathrm{D}_{Kij}U$ on K itself into account:

$$\mathrm{D}_{Kij}\tilde{\eta}_Q \approx \delta_{QK}\int_Q \nabla_j U\,\nabla_i(\tilde{Z}+\tilde{Z}_Q)\,\mathrm{d}x + \delta_{QK}\int_Q (\bar{A}^\delta - A^\varepsilon)\nabla \mathrm{D}_{Kij}U\cdot\nabla(\tilde{Z}+\tilde{Z}_Q)\,\mathrm{d}x. \quad (6.47)$$

Finally, as a more precise reconstruction approach, define a patch $\omega(K) := \{Q \in \mathbb{T}_\delta(\Omega) : \overline{K}\cap\overline{Q}\neq\emptyset\}$ and let $I_{Q\omega(K)}$ be the indicator function that is equal to 1 for $Q\in\omega(K)$ and 0 otherwise. Now set

$$\mathrm{D}_{Kij}\tilde{\eta}_Q \approx \delta_{QK}\int_Q \nabla_j U\,\nabla_i(\tilde{Z}+\tilde{Z}_Q)\,\mathrm{d}x + I_{Q\omega(K)}\int_Q (\bar{A}^\delta - A^\varepsilon)\nabla \mathrm{D}_{Kij}U\cdot\nabla(\tilde{Z}+\tilde{Z}_Q)\,\mathrm{d}x. \quad (6.48)$$

One last obstacle for the patch-centered reconstruction (6.47) remains. Namely, that the response $D_{Kij}U$ is needed in combination with the microscale reconstruction $\tilde{Z}_Q$. In an efficient algorithm, fine-scale reconstructions of such kinds cannot be stored for further use but are local to the computation on the current sampling region Q. One way to mitigate this problem is to not use a finescale reconstruction Z_Q defined on Q, but to use a slightly more expensive $\tilde{Z}_{\omega(K)}$ defined on the patch $\omega(K)$ around K with patch-depth 1. This allows for an efficient assembly as described in Algorithm 6.

Finally, a model-optimization algorithm can be defined, see Algorithm 7.

Remark 6.20. The implementational details for the optimization framework discussed in this subsection are readily transferable to the case of improved reconstruction approaches with the VMM that were briefly mentioned in Remark 6.10. On an abstract level those also take the form of a global approximation $z^H \in V^H$ (corresponding to $\tilde{Z}$) and localized reconstructions $z_i^f \in V_i^f$ (corresponding to $\tilde{Z}_K$). Thus, an application to this type of reconstruction with a slightly different partition (nodal based versus cell based) is straightforward. Further, z_i^f and z^H do not depend on $\bar{A}^\delta$ which removes the necessity for the approximation (6.45).

Remark 6.21. Due to the fact that $\mathscr{J}$ is always approximated with a substantially reduced variant, the value $||\mathscr{J}||$ does not provide a good stopping criterion with $||\mathscr{J}|| \ll 1$. Instead, it is better to use the approximative estimator value $|\tilde{\theta}^\delta|$ directly. For example, stop if $|\tilde{\theta}^\delta|$ is reduced to 1 % of its initial value.

Algorithm 6: Assembly of $\{\tilde{\eta}_K\}$ and $\mathscr{J}$

– Set up $\mathbb{T}_H(\Omega)$ and assemble matrix A: $A_{\nu\mu} = (\bar{A}^{\delta,i}\nabla\varphi_\mu, \nabla\varphi_\nu)$.
– Compute matrix decomposition of A: $LU = A$.
for $K \in \mathbb{T}_\delta(\Omega)$ **do**
  – Assemble $\omega(K)$ and compute $\tilde{Z}_K \in V^h(\omega(K))$.
  – Compute $\tilde{\eta}_K^\delta$ with (4.64) and (4.65). **for** $i = 1, \dots, d$ **do**
    for $j = 1, \dots, d$ **do**
      – Compute response $\mathrm{D}_{Kij}U$ with above decomposition.
      for $Q \in \omega(K)$ **do**
        – Compute contribution $\mathrm{D}_{Kij}\tilde{\eta}_Q$ according to (6.47) for $\mathscr{J}$.

Algorithm 7: Model-optimization algorithm

– Compute initial model $\bar{A}^{\delta,0}$.

– Solve primal and (reduced) dual problem $U(\bar{A}^{\delta,0})$, $\tilde{Z}(\bar{A}^{\delta,0})$ with the help of (3.25) and (4.58), respectively.

while *stopping criterion not reached* **do**

– Compute the error estimator and local indicators $\{\tilde{\eta}_K\}$

$$\tilde{\theta}^\delta = \sum_{K\in\mathbb{T}_\delta(\Omega)} \tilde{\eta}_K^\delta,$$

as well as, the Jacobian $\mathscr{J}$ with Algorithm 6.

– Solve

$$\big(\mathscr{J}\mathscr{J}^T(\bar{A}^{\delta,n}) + \lambda\mathrm{Id}\big)\delta\bar{A}^{\delta,n} = -\mathscr{J}\mathscr{G}^T(\bar{A}^{\delta,n}).$$

– Update model: $\bar{A}^{\delta,n+1} \leftarrow \bar{A}^{\delta,n} + \delta\bar{A}^{\delta,n}$.

– Solve primal and (reduced) dual problem $U(\bar{A}^{\delta,n+1})$, $\tilde{Z}(\bar{A}^{\delta,n+1})$ with the help of (3.25) and (4.58), respectively.

6.5 Numerical results for the model-optimization strategy

With the help of the elliptic model problem, a series of short numerical tests shall be conducted in order to examine specific behavior and aspects of the model-optimization approach that was proposed in the previous section. In particular, the dependence of the optimization result on the initial value $\bar{A}^{\delta,0}$, on the size of the sampling discretization $\mathbb{T}_\delta(\Omega)$, and on the strength of the regularization parameters α_K is examined for the global functional j_1, as well as the local variant j_2. The test is conducted for for the random coefficients (Section 6.5.1) for both, the fully resolved dual solution z^ε, as well as the reduced, locally enhanced variant $z^\delta + \sum z_K^\delta$. Finally, Section 6.5.2 concludes with a counterexample for large model-deviation when starting from an insufficient initial model $\bar{A}^{\delta,0}$.

6.5.1 Parameter study for random coefficients

The purpose of the first numerical test is to examine the stability of the optimization approach for a variety of differently chosen discretization and optimization parameters. In particular, the feasibility of using the reduced, locally enhanced approximation approach within the optimization framework shall be assessed as this property is essential for the optimization approach to be computationally feasible and thus comparable to VMM or HMM approaches.

Consider the computational domain Ω with the random microstructure introduced in Section 3.5 together with the global and local functionals

$$\langle j_1, \varphi \rangle = \int_\Omega \varphi \, \mathrm{d}x, \quad \langle j_2, \varphi \rangle = \varphi(x_0), \tag{6.49}$$

that were examined in Section 4.5. For a fixed choice of $65.5K$ macrocells and a microscale resolution of $h = 2^{-12}$, a parameter study is conducted with sampling discretizations $\delta = 2^{-3}$ and $\delta = 2^{-4}$, a choice of mild penalty with $\lambda = 0.1\,m$ and regularization $\alpha_K = 0.001$ and strong penalty $\lambda = 1.0\,m$ and regularization $\alpha_K = 0.01$, where m is the absolute mean value of the diagonal entries of the matrix $\mathscr{J}^T\mathscr{J}$, see (6.42). The optimization algorithm is run for the full optimization strategy (6.48), as well as the reduced variant (6.46) for both types of reconstruction approaches for the dual solution: fully resolved z^ε and the reduced, local enhanced variant $z^\delta + \sum z_K^\delta$. With reference values of $\langle j_1, u_{\mathrm{ref}}^\varepsilon \rangle \approx 0.14641$ and $\langle j_2, u_{\mathrm{ref}}^\varepsilon \rangle \approx 0.189403$ the initial model errors are in the range of 40 % for the arithmetic average and around 1 % for the geometric average.

Remark 6.22. An application of the model-optimization framework to an advection-diffusion example with dominant finescale transport and actually *large* initial model error is given in Section 6.6.

For each choice of parameters, Tables 6.1 and 6.2 show the final error after a fixed number of 15 optimization cycles for periodic and random coefficients.

The first observation that can be made is that in almost all cases the model-optimization approach is able to consistently reduce the initial error to around 1 % for the arithmetic average as initial model, and well below 1 % for the geometric average (that already provides a very good initial model), respectively. The slight difference in the final value can be explained by the different impact of the chosen regularization α_K.

More importantly, the reduced, locally enhanced variant $z^\delta + \sum z_K^\delta$ with increased patch size (and thus reduced impact of the artificial Dirichlet boundary conditions of the reconstruction problems) leads to comparable results very similar to the results for the full variant z^ε. As discussed, this observation is crucial for the proposed optimization approach to be computationally competitive.

There are two notable exceptions to this general observation. The first one is for the case of the arithmetic average as initial model and strong regularization ($\alpha_K = 10^{-2}, 10^{-3}$). Here, the regularization term prevents further convergence to an acceptable optimal model (cf., e. g., Table 6.1a, $\delta = 2^{-4}$, for z^ε and $z^\delta + \sum z_K^\delta$). The effect is more pronounced for the case of small sampling cells ($\delta = 2^{-4}$).

The second exception occurs for the reduced model-optimization in combination with the reduced, locally enhanced dual approximation (cf. Table 6.1c–d and 6.2c–d). Here, the higher influence of the artificial boundary conditions of the dual approximation leads to a consistent overall error of around 3 % ($\delta = 2^{-3}$) and 6 % ($\delta = 2^{-4}$) independent of starting model and quantity of interest.

Remark 6.23. The decrease of precision in case of the reduced optimization is in fact due to the increased influence of the artificial boundaries and not the patch size. This is evidenced by the fact that the result for the reduced, locally enhanced dual reconstruction with increased patch size is practically identical for $\delta = 2^{-3}$, $\delta = 2^{-4}$ (and $\delta = 2^{-5}$, which is not shown in the tables).

The results for the full dual solution z^ε (and the case of reduced optimization strategy), on the other hand, show that the reduced reconstruction (6.46) gives practically the same result for both examined quantities of interest compared to the full reconstruction (6.48).

Table 6.1. Parameter study for the random permeability and the global functional j_1. For each choice the absolute and relative error after cycle 15 of the optimization algorithm are shown.

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	5.9e-2 (40 %)	5.9e-2 (40 %)	5.9e-2 (40 %)	5.9e-2 (40 %)
$\alpha_K = 10^{-2}$	15	6.0e-3 (4.1 %)	2.7e-2 (18 %)	7.1e-3 (4.8 %)	3.6e-2 (24 %)
$\alpha_K = 10^{-3}$	15	9.6e-4 (0.7 %)	1.4e-3 (0.9 %)	9.1e-4 (0.6 %)	1.4e-3 (1.0 %)
$\alpha_K = 10^{-4}$	15	9.3e-4 (0.6 %)	9.9e-4 (0.7 %)	8.7e-4 (0.6 %)	9.6e-4 (0.7 %)

(a) Full model-optimization (6.48), arithmetic average $\bar{A}^{\delta,0}$

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	1.8e-3 (1.3 %)	1.3e-3 (0.9 %)	1.8e-3 (1.3 %)	1.3e-3 (0.9 %)
$\alpha_K = 10^{-2}$	15	7.5e-4 (0.5 %)	4.2e-4 (0.3 %)	6.7e-4 (0.5 %)	3.7e-4 (0.3 %)
$\alpha_K = 10^{-3}$	15	8.0e-4 (0.6 %)	7.9e-4 (0.5 %)	7.1e-4 (0.5 %)	7.2e-4 (0.5 %)

(b) Full model-optimization (6.48), geometric average $\bar{A}^{\delta,0}$

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	5.9e-2 (40 %)	5.9e-2 (40 %)	5.9e-2 (40 %)	5.9e-2 (40 %)
$\alpha_K = 10^{-2}$	15	5.6e-2 (3.8 %)	2.6e-2 (17 %)	1.1e-2 (7.7 %)	3.8e-2 (26 %)
$\alpha_K = 10^{-3}$	15	5.8e-4 (0.4 %)	1.3e-3 (0.9 %)	4.8e-3 (3.2 %)	8.7e-3 (5.9 %)
$\alpha_K = 10^{-4}$	15	9.0e-4 (0.6 %)	9.7e-4 (0.7 %)	4.7e-3 (3.2 %)	8.2e-3 (5.6 %)

(c) Reduced model-optimization (6.46), arithmetic average $\bar{A}^{\delta,0}$

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	1.8e-3 (1.3 %)	1.3e-3 (0.9 %)	1.8e-3 (1.3 %)	1.3e-3 (0.9 %)
$\alpha_K = 10^{-2}$	15	7.6e-4 (0.5 %)	4.2e-4 (0.3 %)	4.4e-3 (3.0 %)	5.9e-3 (4.0 %)
$\alpha_K = 10^{-3}$	15	8.0e-4 (0.5 %)	7.9e-4 (0.5 %)	4.6e-3 (3.1 %)	8.0e-3 (5.4 %)

(d) Reduced model-optimization (6.46), geometric average $\bar{A}^{\delta,0}$

Table 6.2. Parameter study for the random permeability and the local functional j_2. For each choice the absolute and relative error after cycle 15 of the optimization algorithm are shown.

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	7.5e-2 (40 %)	7.4e-2 (39 %)	7.5e-2 (40 %)	7.4e-2 (39 %)
$\alpha_K = 10^{-2}$	15	3.5e-3 (1.9 %)	6.9e-4 (0.4 %)	3.0e-3 (1.6 %)	4.2e-3 (2.2 %)
$\alpha_K = 10^{-3}$	15	1.8e-3 (1.0 %)	1.0e-3 (0.6 %)	1.6e-3 (0.8 %)	7.1e-3 (3.7 %)
$\alpha_K = 10^{-4}$	15	1.8e-3 (1.0 %)	3.0e-3 (1.6 %)	1.5e-3 (0.8 %)	6.8e03 (3.6 %)

(a) Full model-optimization (6.48), arithmetic average $\bar{A}^{\delta,0}$

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	2.1e-3 (1.1 %)	2.2e-3 (1.2 %)	2.1e-3 (1.1 %)	2.2e-3 (1.2 %)
$\alpha_K = 10^{-2}$	15	6.0e-4 (0.3 %)	6.2e-4 (0.3 %)	8.7e-4 (0.5 %)	9.5e-4 (0.5 %)
$\alpha_K = 10^{-3}$	15	6.9e-4 (0.4 %)	7.2e-4 (0.4 %)	9.5e-4 (0.5 %)	1.0e-3 (0.5 %)

(b) Full model-optimization (6.48), geometric average $\bar{A}^{\delta,0}$

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	7.5e-2 (40 %)	7.4e-2 (39 %)	7.5e-2 (40 %)	7.4e-2 (39 %)
$\alpha_K = 10^{-2}$	15	3.6e-3 (1.9 %)	7.3e-3 (3.9 %)	8.5e-4 (4.5 %)	1.7e-2 (8.9 %)
$\alpha_K = 10^{-3}$	15	2.6e-3 (1.4 %)	2.1e-3 (1.1 %)	7.6e-3 (4.0 %)	1.2e-2 (6.3 %)
$\alpha_K = 10^{-4}$	15	2.6e-3 (1.4 %)	1.9e-3 (1.0 %)	7.6e-3 (4.0 %)	1.3e-2 (6.7 %)

(c) Reduced model-optimization (6.46), arithmetic average $\bar{A}^{\delta,0}$

		z^ε		$z^\delta + \sum z_K^\delta$	
	cl.	$\delta = 2^{-3}$	$\delta = 2^{-4}$	$\delta = 2^{-3}$	$\delta = 2^{-4}$
	1	2.1e-3 (1.1 %)	2.2e-3 (1.2 %)	2.1e-3 (1.1 %)	2.2e-3 (1.2 %)
$\alpha_K = 10^{-2}$	15	6.2e-4 (0.3 %)	4.1e-4 (0.2 %)	6.1e-3 (3.2 %)	1.0e-2 (5.3 %)
$\alpha_K = 10^{-3}$	15	7.1e-4 (0.4 %)	5.0e-4 (0.3 %)	6.2e-3 (3.2 %)	1.0e-2 (5.5 %)

(d) Reduced model-optimization (6.46), geometric average $\bar{A}^{\delta,0}$

6.5.2 Counterexample for large model-deviation

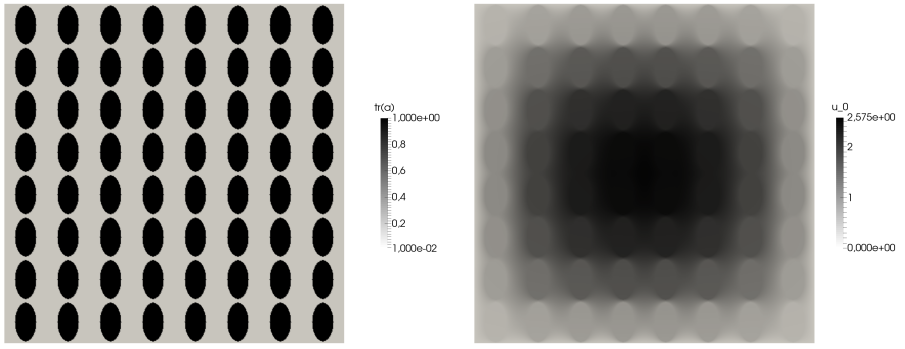
The optimization problem (Definition 6.1) itself is not uniquely solvable. In fact, with the regularization term introduced in (6.6), as well as with the Gauß-Newton iteration with approximate Jacobian that was introduced, the outcome of the minimization algorithm is dependent on specific properties being present in the initial model $\bar{A}^{\delta,0}$. As a numerical test case, consider periodic coefficients given by regularly located, small ellipses with periodicity $\varepsilon = 2^{-3}$ and high permeability ($K^\varepsilon = 1$ compared to $K^\varepsilon = 0.01$ outside) on the unit square Ω (see Figure 6.2a–b). Two choices of initial models are used: the geometric average, as well as a sampling with the cell problem (3.10) of the homogenization sampling strategy. The algorithm is run for a sampling region consisting of just a single sampling cell ($\delta = 1$) and for a fine sampling discretization with $\delta = \varepsilon = 2^{-3}$. The global functional j_1 given by (4.67) is chosen as quantity of interest. As can be seen from the numerical results given in Table 6.3, for all choices of initial model (that exhibit 37 % and 2.6 % relative error, respectively) the optimization algorithm converges to the almost same result close to the reference value with a relative error of 0.03 % (which is basically entirely dominated by macro- and microscale discretization errors present in the primal and dual approximation).

However, when examined in the *picture norm* (see Figure 6.2c–f) an important difference appears in the sense that the "optimal" solution for $\delta = 1$ and simple averaging as initial model is almost isotropic, where, in fact, for the reference solution and for the result obtained by starting with $\bar{A}^{\delta,0}$ computed by (3.10) this isn't the case. This behavior gets less pronounced with a refined sampling discretization ($\delta = 2^{-3}$).

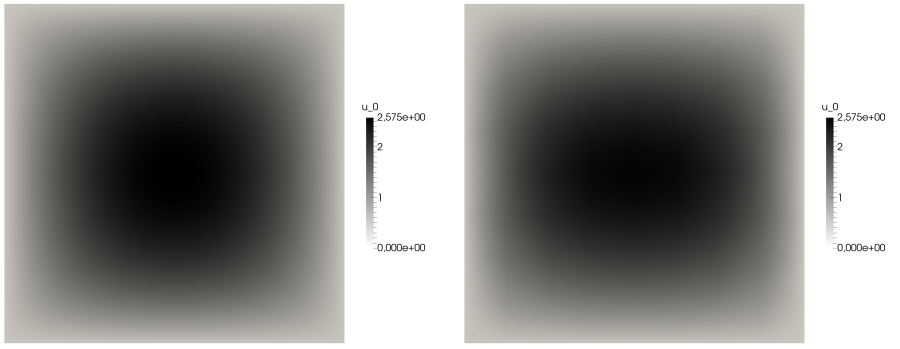
In order to avoid such a pathologic behavior, it is necessary to either use an initial model that already captures such important properties and further optimize it (as was done in this example), or, to use a (possibly nonlinear) functional capable of capturing such anisotropy in the optimization problem.

Table 6.3. Model-optimization results for periodic coefficients with ellipses (see Figure 6.2) for different initial models (geometric average and homogenization) and different sampling grid sizes

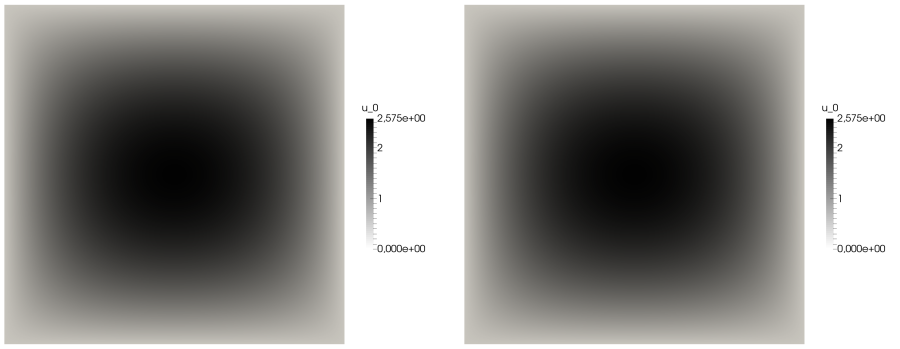
	Homogenization		Geometric average	
cycle.	$\delta = 2^{-0}$	$\delta = 2^{-3}$	$\delta = 2^{-0}$	$\delta = 2^{-3}$
1	3.2e-2 (2.6 %)	3.2e-2 (2.6 %)	5.6e-1 (45 %)	5.6e-1 (45 %)
15	4.1e-4 (0.03 %)	4.1e-4 (0.03 %)	4.2e-4 (0.03 %)	3.9e-4 (0.03 %)



(a) Periodic coefficients (left) and corresponding reference solution (right)



(b) Optimal solution for initial model with geometric average, $\delta = 1$ (left) and $\delta = 2^{-3}$ (right)



(c) Optimal solution for initial model computed with HMM, $\delta = 1$ (left) and $\delta = 2^{-3}$ (right)

Figure 6.2. A periodic microstructure consisting of ellipses on the unit square and its reference solution (a). Panel (b) and (c) show the final results for the optimization algorithm for different initial models and sampling grid sizes. The optimal solution in case of an initial model with geometric average only exhibit isotropic effective coefficients $\bar{A}^{\delta}$. This is clearly visible in case of only one sampling cell $\delta = 1$.

6.6 An advection-diffusion example with dominant transport

As a further test case consider an *advection-diffusion* problem

$$\gamma(\nabla u^\varepsilon, \nabla\varphi) + (\boldsymbol{b}^\varepsilon \cdot \nabla u^\varepsilon, \varphi) = (f, \varphi) \quad \forall \varphi \in V \tag{6.50}$$

driven by a divergence-free vector field $\boldsymbol{b}^\varepsilon \in H^{1,\infty}(\Omega)^d$, i. e. $\nabla \cdot \boldsymbol{b}^\varepsilon = 0$ a. e. on Ω and $\boldsymbol{b}^\varepsilon \equiv 0$ on $\partial\Omega$, together with a positive scaling factor $\gamma \in \mathbb{R}^+$. This time, the multiscale character is given by $\boldsymbol{b}^\varepsilon$ that shall consist of small (but strong) eddies.

6.6.1 Periodic coefficients

For a first numerical test, the influence of $\boldsymbol{b}^\varepsilon$ is assumed to be entirely *local*, i. e., no *effective macroscopic transport* shall occur. More precisely, assume

$$\fint_K \boldsymbol{b}^\varepsilon = 0 \quad \forall K \in \mathbb{T}_\delta(\Omega). \tag{6.51}$$

Due to the absence of an averaged transport, the small eddies only influence the macroscopic behavior by means of an additional *effective diffusivity*. Consequently, let the task be to find effective (diffusion) constants $\bar{A}^\delta : \mathbb{T}_\delta(\Omega) \to \mathbb{R}^{d\times d}$ such that the solution u^δ of the usual effective problem

$$(\bar{A}^\delta \nabla u^\delta, \nabla\varphi) = (f, \varphi) \quad \forall \varphi \in V \tag{6.52}$$

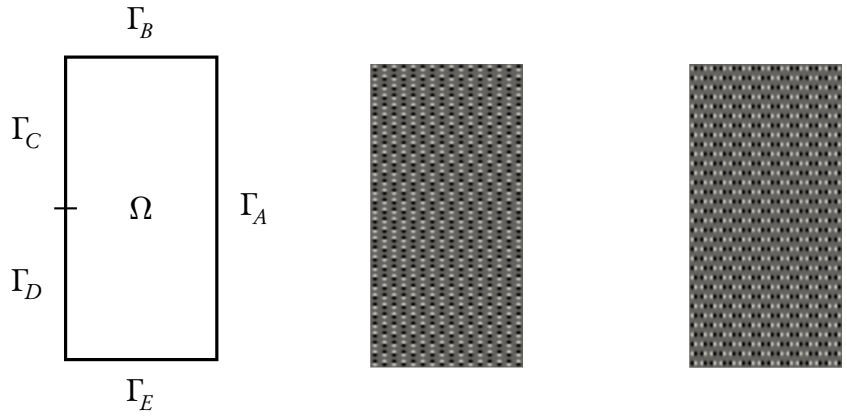
is a good approximation of above u^ε in some quantity of interest.

The only significant change in the model-adaptation framework for above advection-diffusion problem is the occurrence of an additional term $(\boldsymbol{b}^\varepsilon \cdot \nabla u^\varepsilon, z^\varepsilon)$ in the error identity (4.6) that now splits into

$$\begin{aligned}\langle j, u^\varepsilon\rangle - \langle j, U\rangle &= \underbrace{\big(f, z^\delta\big) - \big(A^{\delta,h}\nabla U, \nabla z^\delta\big)}_{=:\theta^H} + \underbrace{\big(A^{\delta,h}\nabla U, \nabla z^\delta\big) - \big(A^\delta \nabla U, \nabla z^\delta\big)}_{=:\theta^h} \\ &\quad + \underbrace{\big(A^\delta \nabla u^\delta, \nabla z^\varepsilon\big) - \gamma\big(\nabla u^\delta, \nabla z^\varepsilon\big) - \big(\boldsymbol{b}^\varepsilon \cdot \nabla u^\delta, z^\varepsilon\big)}_{=:\theta^\delta}.\end{aligned} \tag{6.53}$$

This leads to a local model-error indicator

$$\eta_K^\delta := \big(\{\gamma\,\mathrm{Id} - A^\delta\}\nabla u^\delta, \nabla z^\varepsilon\big)_K - \big(\boldsymbol{b}^\varepsilon \cdot \nabla u^\delta, z^\varepsilon\big)_K. \tag{6.54}$$



(a) Domain Ω **(b)** $\boldsymbol{b}^\varepsilon$, x-comp. **(c)** $\boldsymbol{b}^\varepsilon$, y-comp.

Figure 6.3. The computational domain Ω for the advection-diffusion test case (a) and the periodic vector field $\boldsymbol{b}^\varepsilon$ for the first numerical test (b-c).

With above assumptions on $\boldsymbol{b}^\varepsilon$ the corresponding dual problem reads

$$\gamma\,(\nabla\varphi,\nabla z^\varepsilon)-(\boldsymbol{b}^\varepsilon\cdot\nabla z^\varepsilon,\varphi)=\langle j,\varphi\rangle\quad\forall\varphi\in V. \tag{6.55}$$

A rectangular domain Ω is chosen (see Figure 6.3) with homogeneous Dirichlet boundary conditions on Γ_D, homogeneous Neumann conditions on Γ_A, Γ_B and Γ_C, and $\gamma\,\partial_{\boldsymbol{n}}u^\varepsilon\equiv 1$ on Γ_E. The source term is set to $f\equiv 0$ and the quantity of interest is chosen to be

$$\langle j,\varphi\rangle=\int_{\Gamma_B}\varphi\,\mathrm{d}o_x. \tag{6.56}$$

In spirit of Definition 4.19, a reduced dual problem with a local enhancement can be defined

$$\left(\bar{A}^\delta\nabla\varphi,\nabla z^\delta\right)=\langle j,\varphi\rangle\quad\forall\varphi\in V, \tag{6.57}$$

$$\gamma\left(\nabla\varphi,\nabla(z^\delta+z_K^\delta)\right)_K-\left(\boldsymbol{b}^\varepsilon\cdot\nabla(z^\delta+z_K^\delta),\varphi\right)_K=\langle j,\varphi\rangle\quad\forall\varphi\in V(K). \tag{6.58}$$

Here, the local reconstruction $z_K^\delta\in V(K)$ has homogeneous Dirichlet conditions on *interior* boundary parts ∂K but shall have homogeneous Neumann conditions on all Neumann boundaries of the primal problem, i. e., on boundaries $\partial K\cap\Gamma_i$ with $i=A,B,C,E$.

For the first numerical test, an artificial, periodic vector field is constructed that fulfills (6.51): Consider the function $\boldsymbol{b} \in C^2(\Omega, \mathbb{R}^2)$ defined by

$$\boldsymbol{b}(x,y) \coloneqq \frac{1}{2} \begin{cases} (1-2r+r^2)\begin{pmatrix} -y \\ x \end{pmatrix}, & \text{if } r \leq 1, \\ 0, & \text{otherwise,} \end{cases} \tag{6.59}$$

with $r = \sqrt{x^2+y^2}$. With the help of the rescaling $\hat{\boldsymbol{x}} = \boldsymbol{x}/\varepsilon \bmod Y$ into the unit cell Y define $\boldsymbol{b}^\varepsilon : \Omega \to \mathbb{R}^2$ (see Figure 6.3) by

$$\boldsymbol{b}^\varepsilon(\boldsymbol{x}) = s\,(-1)^{\lfloor x/\varepsilon \rfloor}(-1)^{\lfloor y/\varepsilon \rfloor}\,\boldsymbol{b}\Big(2 \times \hat{\boldsymbol{x}} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}\Big), \tag{6.60}$$

with a scaling factor $s \in \mathbb{R}^+$. For the choice $\varepsilon = 2^{-5}$, $\gamma = 0.01$, and $s = 1000$, a reference computation with 8.39×10^6 degrees of freedom yields the result $\langle j, u^\varepsilon_{\text{ref}} \rangle \approx 1.781$. A uniform sampling mesh with 32 sampling regions is chosen, as well as a macroscale discretization of 1.3×10^5 cells and a (fully resolved) microscale discretization with $h = 2^{-11}$. The optimization framework is run for a fully resolved dual solution ("full") with 2.1×10^6 cells as well as the reduced, locally enhanced variant given in Definition 4.28 ("enhanced"). As stopping criterion a reduction of $|\tilde{\theta}^\delta|$ to less than 1 % of the initial value is chosen with a penalty $\lambda = 1.0 \times m$, where m is the absolute mean value of the diagonal entries of the matrix $\mathscr{J}^T\mathscr{J}$, see (6.42), and a very small regularization $\alpha_K = 0.1$ (compared to $|\tilde{\theta}^\delta|^2/|\bar{A}^\delta_K|^2 \sim 1000$). The numerical results are given in Table 6.4.

To examine the numerical stability of the optimization algorithm the computation is actually run for 15 adaptation cycles well beyond the stopping criterion that is reached with step 7 for the full dual solution and with step 6 for the local enhancement strategy. The initial error of 142 % in the target functional with a starting model $\bar{A}^\delta_K = \gamma\,\mathrm{Id}$ can be reduced to under 1 % for the full dual reconstruction and to under 3 % for the localized variant. Further, the adaptation cycle remains stable well beyond the stopping criterion was reached.

Reference, initial and final (for step 7 and 6, respectively) solutions are depicted in Figure 6.4. As can be seen from the numerical results, the microscale advection due to $\boldsymbol{b}^\varepsilon$ leads to a locally increased value for $\bar{A}^\delta_K$ in the range $0.01 - 0.02$ compared to the initial choice $\bar{A}^\delta_K = \gamma\,\mathrm{Id} \sim 0.01$. The effective models found with the optimization approach match the reference solution quite well near the boundary Γ_B. In contrast, on the far end of Γ_B near the inhomogeneous Neumann condition on Γ_E, the effective solutions deviate from u^ε. This has to be expected as the optimization problem only minimizes the error given by an integral over Γ_B.

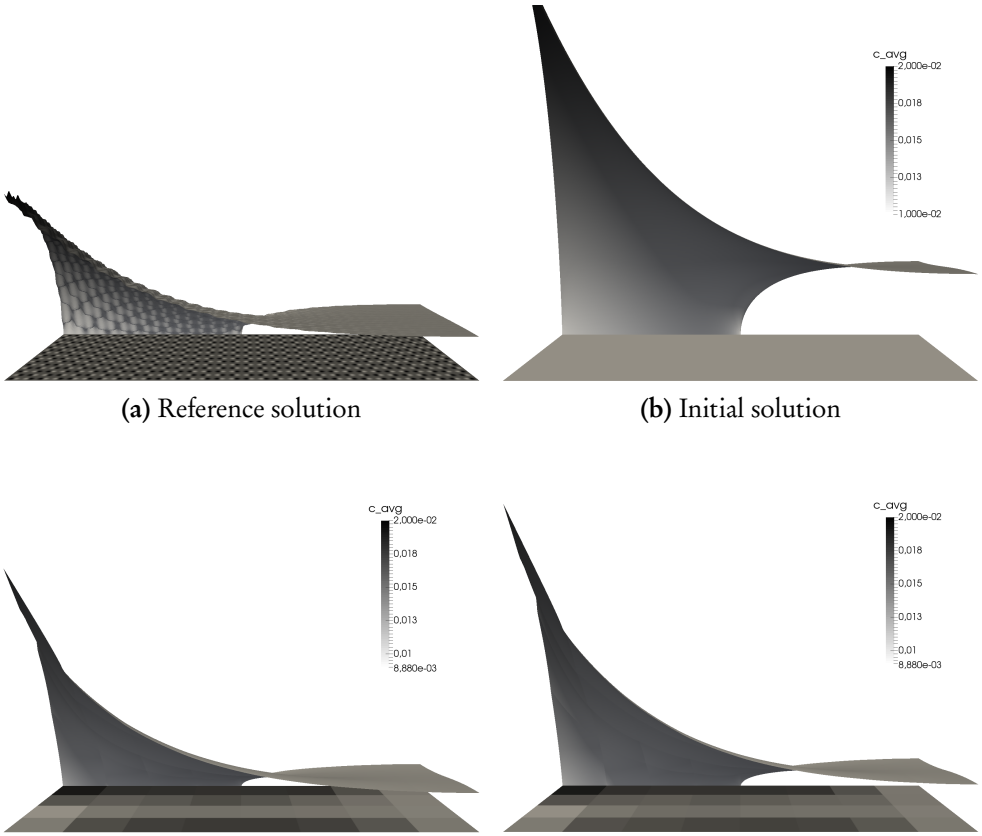
Table 6.4. Results for the model-optimization algorithm (Algorithm 7) applied to the advection-diffusion problem (6.50) with periodic advection (6.60) with fully resolved dual solution (a) and for the reduced, locally enhanced variant (b). After steps 7 and 6, respectively, the estimator $|\tilde{\theta}^{\delta}|$ is reduced to less than 1 % of its initial value.

	$L^2(\Omega)$	$\lvert\langle j, U\rangle\rvert$	$\lvert\langle j, u^{\varepsilon}-U\rangle\rvert$	$\lvert\tilde{\theta}^{\delta}\rvert$	I_{eff}	I_{loc}
1	5.77e+0	4.31e+0	-2.53e+0 (142 %)	-2.55e+0	1.01	1.67
2	2.82e+0	2.58e+0	-7.96e-1 (44,7 %)	-8.07e-1	1.01	2.13
3	1.77e+0	2.03e+0	-2.52e-1 (14.1 %)	-2.60e-1	1.03	2.86
4	1.46e+0	1.87e+0	-8.74e-2 (4.91 %)	-9.52e-2	1.07	4.16
5	1.38e+0	1.82e+0	-3.63e-2 (2.04 %)	-4.39e-2	1.17	5.69
6	1.35e+0	1.80e+0	-1.83e-2 (1.03 %)	-2.59e-2	1.34	6.91
7	1.34e+0	1.79e+0	-1.02e-2 (0.57 %)	-1.77e-2	1.61	7.86
14	1.33e+0	1.78e+0	4.92e-3 (0.28 %)	-2.49e-3	-0.25	19.7
15	1.33e+0	1.78e+0	5.40e-3 (0.31 %)	-2.00e-3	-0.14	21.81

(a) Model-optimization algorithm with fully resolved dual solution

	$L^2(\Omega)$	$\lvert\langle j, U\rangle\rvert$	$\lvert\langle j, u^{\varepsilon}-U\rangle\rvert$	$\lvert\tilde{\theta}^{\delta}\rvert$	I_{eff}	I_{loc}
1	5.77e+0	4.31e+0	-2.53e+0 (142 %)	-6.18e+0	2.44	1.66
2	2.82e+0	2.57e+0	-7.93e-1 (44,5 %)	-1.25e+0	1.58	2.29
3	1.75e+0	2.03e+0	-2.47e-1 (13.8 %)	-3.95e-1	1.60	2.65
4	1.44e+0	1.84e+0	-6.17e-2 (3.47 %)	-1.50e-1	2.43	3.22
5	1.36e+0	1.78e+0	5.85e-4 (0.33 %)	-6.57e-2	-112	4.02
6	1.33e+0	1.76e+0	2.31e-2 (1.30 %)	-3.37e-2	-1.46	5.04
14	1.32e+0	1.74e+0	4.30e-2 (2.42 %)	-3.78e-3	-0.09	9.90
15	1.32e+0	1.74e+0	4.34e-2 (2.44 %)	-3.31e-3	-0.08	10.1

(b) Model-optimization algorithm with reduced, locally enhanced dual solution



(a) Reference solution **(b)** Initial solution

(c) Fully resolved dual solution **(d)** Locally enhanced dual solution

Figure 6.4. 3 dimensional plot of the reference solution (a), the initial solution $u^{\delta,0}$ of the optimization problem (b), and the solutions of the final models for fully resolved (c) and reduced, locally enhanced (d) dual solution. The figures are a 3 dimensional view from the right side Γ_A onto Ω. The height is given by the value u^{ε}, $u^{\delta,i}$, the scale is kept the same. Onto Ω itself the values $|\bar{A}^{\delta,i}|$ are plotted, except for the reference solution, where b_x is depicted.

6.6.2 Random coefficients

As a last numerical test, the model-optimization approach shall be tested on the advection-diffusion problem (6.50) with a random advection field $\boldsymbol{b}^\varepsilon$. In order to construct a divergence-free vector field that also fulfills $\boldsymbol{b}^\varepsilon \equiv 0$ on $\partial\Omega$, a potential $\Psi^\varepsilon \in C^2(\Omega)$ is constructed first with the help of a *Hermite interpolation*: Fix a uniform, fine-scale mesh $\mathbb{T}_\varepsilon(\Omega)$ and assign random, Gaussian-distributed values to its vertices. In order to ensure the boundary condition $\boldsymbol{b}^\varepsilon \equiv 0$ on $\partial\Omega$, the individual values of the vertices at the boundary are replaced by the mean value of the random distribution. Now, for all $K \in \mathbb{T}_\varepsilon(\Omega)$ define Ψ^ε_K to be given by

$$\Psi^\varepsilon_K(\hat{\boldsymbol{x}}) = \sum_{i,j=0}^{1} v_{ij}\, p_i(\hat{x}) p_j(\hat{y}), \tag{6.61}$$

where v_{ij} denote the individual, assigned nodal values, and p_i are the *Hermite polynomials*

$$p_0(x) = 1 + 2x^3 - 3x^2, \tag{6.62}$$
$$p_1(x) = -2x^3 + 3x^2. \tag{6.63}$$

$\hat{\boldsymbol{x}}$ is the already introduced rescaling to the unit cell, $\hat{\boldsymbol{x}} = \boldsymbol{x}/\varepsilon \mod Y$. Now define

$$\boldsymbol{b}^\varepsilon(\boldsymbol{x}) := \operatorname{curl} \Psi^\varepsilon = -\partial_y \Psi^\varepsilon\, \boldsymbol{e}_x + \partial_x \Psi^\varepsilon\, \boldsymbol{e}_y. \tag{6.64}$$

Due to the random nature of the advection field, the strong condition (6.51) of a vanishing mean value of $\boldsymbol{b}^\varepsilon$ over sampling regions cannot hold true. Consequently, the full effective advection-diffusion problem of Section 2.2.5 has to be used. Therefore, for a given sampling discretization $\mathbb{T}_\delta(\Omega)$ define

$$\boldsymbol{b}^\delta : \mathbb{T}_\delta(\Omega) \to \mathbb{R}^d, \tag{6.65}$$
$$\boldsymbol{b}^\delta_K := \fint_K \boldsymbol{b}^\varepsilon \,\mathrm{d}x \quad \text{for } K \in \mathbb{T}_\delta(\Omega). \tag{6.66}$$

This leads to the *effective advection-diffusion problem*

$$(\bar{A}^\delta \nabla u^\delta, \nabla\varphi) + (\boldsymbol{b}^\delta \cdot \nabla u^\delta, \varphi) = (f, \varphi) \quad \forall \varphi \in V. \tag{6.67}$$

The error identity (4.6) now features additional contributions given by b^δ:

$$\begin{aligned}\langle j, u^\varepsilon\rangle - \langle j, U\rangle = &\underbrace{\big(f, z^\delta\big) - \big(A^{\delta,h}\nabla U, \nabla z^\delta\big) - \big(\boldsymbol{b}^{\delta,h}\cdot\nabla U, z^\delta\big)}_{=:\theta^H}\\ &+\underbrace{\big(A^{\delta,h}\nabla U, \nabla z^\delta\big) - \big(A^\delta\nabla U, \nabla z^\delta\big) - \big((\boldsymbol{b}^\delta - \boldsymbol{b}^{\delta,h})\cdot\nabla U, z^\delta\big)}_{=:\theta^h}\\ &+\underbrace{\big(A^\delta\nabla u^\delta, \nabla z^\varepsilon\big) - \gamma\big(\nabla u^\delta, \nabla z^\varepsilon\big) - \big((\boldsymbol{b}^\varepsilon - \boldsymbol{b}^\delta)\cdot\nabla u^\delta, z^\varepsilon\big)}_{=:\theta^\delta}\end{aligned} \tag{6.68}$$

along with model-error indicators

$$\eta_K^\delta := \big(\{\gamma\,\mathrm{Id} - A^\delta\}\nabla u^\delta, \nabla z^\varepsilon\big)_K - \big((\boldsymbol{b}^\varepsilon - \boldsymbol{b}^\delta)\cdot\nabla u^\delta, z^\varepsilon\big)_K. \tag{6.69}$$

For the choice $\varepsilon = 2^{-8}$, $\gamma = 0.1$, as well as values of the advection field with magnitude in the range $0-300$, a reference computation 8.39×10^6 degrees of freedom yields the result $\langle j, u^\varepsilon_{\mathrm{ref}}\rangle \approx 0.2170$. Otherwise, identical discretization and stabilization parameters as for the periodic case are chosen.

The optimization framework is once again run for a fully resolved dual solution ("full") with 2.1×10^6 cells as well as the reduced, locally enhanced variant. A slightly decreased stopping criterion to reach a reduction of $|\tilde\theta^\delta|$ to less than 5 % of the initial value is chosen. The numerical results are given in Table 6.5 and Figure 6.5. This time, after 12 and 10 steps, respectively, the stopping criterion is reached. The final error stays well below the targeted error bound of 5 %. In fact, a final error of $2-3$ % can be assumed.

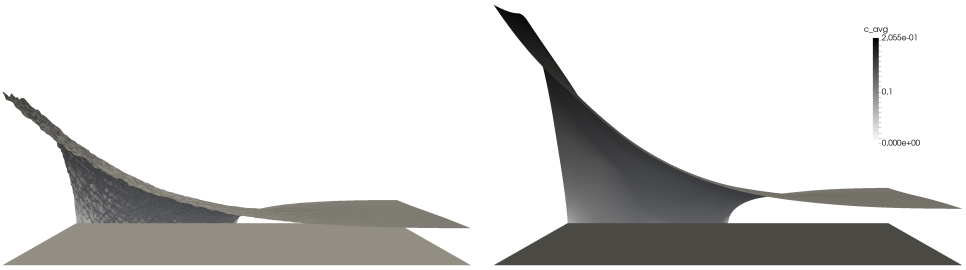
Table 6.5. Results for the model-optimization algorithm (Algorithm 7) applied to the advection-diffusion problem (6.50) with random advection (6.64) with fully resolved dual solution (a) and for the reduced, locally enhanced variant (b). After steps 12 and 10, respectively, the estimator $|\tilde{\theta}^{\delta}|$ is reduced to less than 5 % of its initial value.

	$L^2(\Omega)$	$\lvert\langle j, U\rangle\rvert$	$\lvert\langle j, u^{\varepsilon}-U\rangle\rvert$	$\lvert\tilde{\theta}^{\delta}\rvert$	I_{eff}	I_{loc}
1	4.43e-1	3.86e-1	-1.69e-1 (77.9 %)	-1.69e-1	1.00	2.21
2	3.25e-1	2.99e-1	-8.17e-2 (37.6 %)	-8.19e-2	1.00	3.28
3	2.80e-1	2.71e-1	-5.36e-2 (24.7 %)	-5.38e-2	1.00	4.24
4	2.52e-1	2.57e-1	-4.02e-2 (18.5 %)	-4.04e-2	1.00	5.07
5	2.29e-1	2.49e-1	-3.16e-2 (14.6 %)	-3.18e-2	1.00	5.86
6	2.09e-1	2.42e-1	-2.52e-2 (11.6 %)	-2.54e-2	1.00	6.71
7	1.90e-1	2.37e-1	-2.02e-2 (9.30 %)	-2.04e-2	1.00	7.68
11	1.39e-1	2.26e-1	-9.14e-3 (4.21 %)	-9.35e-3	1.00	13.2
12	1.30e-1	2.25e-1	-7.91e-3 (3.64 %)	-8.12e-3	1.00	14.7
13	1.23e-1	2.24e-1	-7.07e-3 (3.25 %)	-7.29e-3	0.99	16.0
14	1.17e-1	2.24e-1	-6.55e-3 (3.01 %)	-6.77e-3	0.99	16.9
15	1.13e-1	2.23e-1	-6.28e-3 (2.89 %)	-6.51e-3	0.99	17.3

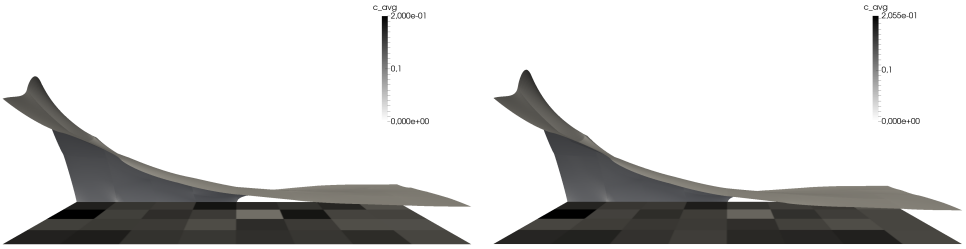
(a) model-optimization algorithm with fully resolved dual solution

	$L^2(\Omega)$	$\lvert\langle j, U\rangle\rvert$	$\lvert\langle j, u^{\varepsilon}-U\rangle\rvert$	$\lvert\tilde{\theta}^{\delta}\rvert$	I_{eff}	I_{loc}
1	4.43e-1	3.86e-1	-1.69e-1 (77.9 %)	-2.96e-1	1.76	2.15
2	3.03e-1	2.95e-1	-7.78e-2 (35.9 %)	-9.55e-2	1.23	3.21
3	2.33e-1	2.62e-1	-4.48e-2 (20.6 %)	-4.92e-2	1.10	4.01
4	1.85e-1	2.48e-1	-3.12e-2 (14.4 %)	-3.61e-2	1.16	4.07
5	1.46e-1	2.39e-1	-2.20e-2 (10.1 %)	-2.85e-2	1.30	3.85
6	1.17e-1	2.32e-1	-1.53e-2 (7.05 %)	-2.35e-2	1.53	3.70
7	9.86e-2	2.28e-1	-1.08e-2 (4.97 %)	-2.06e-2	1.90	3.80
8	8.77e-2	2.25e-1	-7.68e-3 (3.53 %)	-1.83e-2	2.39	3.84
9	8.21e-2	2.22e-1	-5.22e-3 (2.40 %)	-1.58e-2	3.04	4.20
10	7.99e-2	2.20e-1	-3.02e-3 (1.39 %)	-1.27e-2	4.18	4.91
14	9.03e-2	2.14e-1	3.14e-3 (1.44 %)	3.85e-3	1.23	17.2
15	9.70e-2	2.14e-1	3.07e-3 (1.41 %)	7.00e-3	2.28	10.4

(b) model-optimization algorithm with reduced, locally enhanced dual solution



(a) Reference solution **(b)** Initial solution



(c) Fully resolved dual solution **(d)** Locally enhanced dual solution

Figure 6.5. 3 dimensional plot of the reference solution (a), the initial solution $u^{\delta,0}$ of the optimization problem (b), and the solutions of the final models for fully resolved (c) and reduced, locally enhanced (d) dual solution. The figures are a 3 dimensional view from the right side Γ_A onto Ω. The height is given by the value u^{ε}, $u^{\delta,i}$, the scale is kept the same. Onto Ω itself the values $|\bar{A}^{\delta,i}|$ are plotted.

7 Conclusion and Outlook

In this thesis, an abstract model-adaptation framework was introduced that explicitly decouples all discretization and modeling parameters. It can be regarded as a generalization of the HMM and shares some ideas with the model adaptation approaches by Oden and Vemaganti[84–87,91] and Braack and Ern[26]. The novelty lies in the explicit decoupling of the sampling processes from the macroscopic discretization and the simultaneous treatment of discretization and model errors.

With the concept of an effective model,

$$\bar{A}^{\delta} : \mathbb{T}_{\delta}(\Omega) \to \mathbb{R}^{d\times d},$$

that was introduced in the abstract framework, different model-adaptation strategies were discussed. In addition to the category of model-switching approaches (which most of the model-adaptation strategies discussed in the literature so far fall into) a sampling-mesh adaptation strategy was proposed that uses local mesh refinement of $\mathbb{T}_{\delta}(\Omega)$ for model adaptation. Numerical tests showed that this approach can result in an efficient post-processing strategy capable of balancing model and discretization errors and, in particular, can lead to significant savings in computational complexity.

Further, as a novel approach, a model-adaptation strategy based on solving an optimization problem that minimizes the local model-error indicators derived from a DWR formulation was proposed. The optimization approach allowed to derive an efficient post-processing strategy that can be regarded as a multiscale approach in its own right. Its strength lies in the fact that it is in principle independent of strong a priori knowledge about applicability of efficient models—its efficiency is rooted in the almost quantitative behavior of the DWR method when combined with a suitable localization technique for the dual problem. The modeling aspect of the optimization problem lies in the choice of functional j as quantity of interest and the choice of localization approach for the dual problem. In this sense it lifts the question of suitable approximation in terms of a quantity of interest (for the primal problem) to the question of suitable approximation properties of the localization technique for the dual problem. The catch here is that the latter is typically measured in the L^2-norm of the

gradient of the error of the dual approximation, for which—depending on the localization approach—strong approximation properties are available, and not in the quantity of interest itself. Prototypical numerical results were given for periodic, as well as random coefficients (with specific stochastic properties), that indicate that the optimization approach combined with a localization technique that *globally* uses the same effective model as the primal problem and *locally reconstructs* finescale features of the full dual solution with the help of local cell problems does result in an efficient model-adaptation strategy.

Outlook Interpreting model adaptation as an optimization problem in the context of the DWR method has turned out to be a feasible a posteriori technique and a promising field for future research. Extensions of the model-adaptation framework are possible in various directions.

First of all, there remain principle open questions with regard to efficient approximations of the dual problem. The approaches studied in this thesis are by no means exhaustive. An extension of the methodology can be done by considering more sophisticated reconstruction approaches of the dual solution, e. g., by using a full VMM ansatz that (depending on the chosen orthogonality for the ansatz-space splitting) exhibits better approximation properties in H^1 than the heuristic approach of using globally the same effective model as for the primal problem and reconstruct finescale behavior locally, that was presented.

Further, an interesting extension of the optimization framework is to incorporate the microscale-discretization indicator for controlling the discretization parameters of the dual reconstruction. This again requires an efficient balancing strategy of model and discretization adaptation with the additional obstacle that discretization errors now also enter the adaptation process of the model—which is a far less problem in case of the model-switching or sampling-adaptation strategies because of a priori control of the sampling process.

Another promising direction for extending the current work is given by an application of the abstract model- and discretization-adaptation framework to nonlinear and time-dependent problems—especially problems exhibiting multiscale behavior in time.

Furthermore, in current research on multiscale methods there is a gap between the Physics community, that emphasizes on model derivation and simulation of large multiscale and multiphysics problems, and the mathematical community (where my current work is located), that mainly works on numerical schemes and a priori/a posteriori error control in context of simplified model problems. This is a gap that should be closed.

Acknowledgments

"If it can't be expressed in figures, it is not science; it is opinion."

— Robert A. Heinlein

I would like to take this opportunity to express my deepest gratitude to my supervisor, Prof. Dr. Dr. h. c. Rolf Rannacher, for his continued support and guidance throughout my doctoral research project. I would also like to thank him for all the opportunities and liberties he gave me with respect to research, collaboration, as well as teaching. Furthermore, I am grateful for all the fruitful discussions I had with my thesis co-advisor Prof. Dr. Guido Kanschat.

In addition, I would like to thank Prof. Dr. Wolfgang Bangerth, Prof. Dr. Malte Braack, Prof. Dr. Mitchell Luskin, Prof. Dr. Dr. h. c. Mary F. Wheeler, and Dr. Thomas Wick for their kind invitations to their respective working groups and the possibilities to present my work.

As a member of the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences, I gratefully acknowledge the financial support for travel and organization of workshops I received during my doctoral research at the Faculty of Mathematics and Computer Sciences.

I would like to express my special thanks to my colleague Stefan Frei, who happened to share an office with me for almost four years, as well as to Prof. Dr. Thomas Richter—who both were a constant source of advice in all matters imaginable. I enjoyed the numerous work-related and private activities we had over the years. Further, I would like to thank my colleagues and fellow Ph. D. students from the Numerical Analysis Group for the friendly and inspirational atmosphere and the "open doors".

Ich danke meiner Familie und meinen Freunden für ihre Unterstützung, ihren Rat und Zuspruch in allen Lebenslagen.

List of Abbreviations and Symbols

$\langle .,. \rangle$	Duality pairing	9
$(.,.)$	Scalar product of the space $L^2(\Omega)$, or $L^2(\Omega)^d$	10
$(.,.)_X$	Scalar product of an arbitrary Hilbert space X	10
$\lfloor . \rfloor$	Floor function. $\lfloor x \rfloor$ is the largest integral value $\leq x$	74
$⨍$	Mean value of integrand, $⨍_X = \frac{1}{\lvert X \rvert} \int_X$	25
$[.]_{\partial K}$	Jump over boundary ∂K	81
∇	Nabla operator, $\nabla = \sum_{i=1}^{d} \boldsymbol{e}_i \frac{\partial}{\partial x_i}$; in particular for a scalar funtion: $\nabla f = \operatorname{grad} f$, and for a vector valued function: $\nabla \cdot \boldsymbol{u} = \operatorname{div} \boldsymbol{u}$	10
a	Bilinear form $a : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$	9
(A1)	Assumption on A^ε to be Lipschitz continuous	53
(A2)	Regularity assumptions on the meshes	53
(A3)	Additional structural requirements	54
(A4)	Assumption on A^ε to obey a local periodic structure	65
(A5)	Assumption on δ to be an integral multiple of ε	65
$\bar{A}^\delta$	Region-wise constant, effective coefficients for a given sampling discretization $\mathbb{T}_\delta(\Omega)$	46
$\bar{A}^{\delta,h}$	Region-wise constant, numerically computed, effective coefficients for a given sampling discretization $\mathbb{T}_\delta(\Omega)$	50
$\bar{A}^{\delta,0}$	Initial effective coefficient $\bar{A}^\delta$ for the model-optimization problem	123
$A^{\delta,h}$	Post-processed, continuous variant of $\bar{A}^{\delta,h}$ defined with the help of a Clément-type interpolation	50
A^δ	Post-processed, continuous variant of $\bar{A}^\delta$ defined with the help of a Clément-type interpolation	48

A^ε	Coefficient matrix $A^\varepsilon \in L^\infty(\Omega)^{d\times d}$ exhibiting multiscale features	9, 24
a. e.	almost everywhere	9
α_K	Regularization parameter for $K \in \mathbb{T}_\delta(\Omega)$	123
α_ν	Scaling parameter	114
A	Two-scale coefficients $A(x,y) \in C^{0,1}\big(\Omega, L^\infty_{\mathrm{per}}(Y)^{d\times d}\big)$	24
A^0	Homogenized coefficient matrix	26
$\mathscr{A}^\delta$	Set of admissible coefficients	122
$C^k_{\mathrm{per}}(Y^\delta)$	Space of k-times differentiable, δ-periodic functions	25
$C^k(\mathbb{R}^d)$	Space of k-times differentiable functions on $\mathbb{R}^d$	25
$C^{0,1}(\Omega, X)$	Space of Lipschitz continuous functions defined on Ω with values in a Banach space X	24
D_{Kij}	Short-hand notation for $\mathrm{D}.[\delta\bar{A}^\delta(Kij)]$	131
$\mathrm{D}.[\delta X]$	Gâteaux derivative in direction δX	124
δ	Discretization parameter, $\delta > 0$, denoting the typical length scale of a sampling discretization $\mathbb{T}_\delta(\Omega)$	46
DWR	Dual weighted residual	2, 79
ε	Scaling parameter $\varepsilon > 0$ denoting fine-scale oscillations	10, 23
η^δ_K	Local model-error indicator for a sampling cell $K \in \mathbb{T}_\delta(\Omega)$	81
$\tilde{\eta}^\delta_K$	Approximate local model-error indicator	94
$\tilde{\eta}^\delta_{K,\mathrm{rec}}$	Local correction of the approximate indicator $\tilde{\eta}^\delta_K$ given by a solution of a local reconstruction problem	94
η^H_K	Local macroscale discretization error indicator for a cell $K \in \mathbb{T}_H(\Omega)$	81
η^h_K	Local microscale discretization error indicator for a sampling cell $K \in \mathbb{T}_\delta(\Omega)$	81
$\tilde{\eta}^H_K$	Approximate macroscale discretization error indicator	93
$\tilde{\eta}^h_K$	Approximate microscale discretization error indicator	93
$\mathscr{F}(\bar{A}^\delta)$	Cost function of the optimization problem	124
f^ε	Right hand side with multiscale features	7, 23

f^0	Homogenized right hand side	24
$\mathscr{G}$	Vector valued variant of the cost function $\mathscr{F}$, $\mathscr{F} = \lvert\mathscr{G}\rvert^2$	130
g_K	Greens's function associated with the solution of a local cell problem on K	14
g_{Kij}	Regularization part of the optimization problem	130
h	Discretization parameter, $h > 0$, for the fine-scale discretizations $\mathbb{T}_h(Y_K^\delta)$, $\mathbb{T}_h(K)$	49
HMM	Heterogeneous Multiscale Method	1, 26
$H^{-1}(\Omega)$	Dual space of $H_0^1(\Omega)$	18
$H^m_{\text{per}}(Y^\delta)$	Hilbert space $H^m_{\text{per}}(Y^\delta) := W^{m,2}_{\text{per}}(Y^\delta)$	25
$\tilde{H}^m_{\text{per}}(Y^\delta)$	Subspace of $H^m_{\text{per}}(Y^\delta)$ given by all functions with vanishing mean value	25
$H^1(\Omega)$	Sobolev space of functions $u \in L^2(\Omega)$ with generalized derivatives ∇u of class $L^2(\Omega)^d$	10
$H_0^1(\Omega)$	Closed subspace of $H^1(\Omega)$ fulfilling homogeneous Dirichlet boundary conditions	10
Id	Identity matrix	40
I_{eff}	Effectivity index describing the quantitative behavior of the error estimators	99
I^i	Region of influence with patch-depth i	110
I_{loc}	Indicator for the oscillatory behavior of error indicators	99
$\mathscr{J}$	Jacobian of $\mathscr{G}$	131
j	Quantity of interest (a linear and continuous functional)	44, 79
$\mathscr{L}$	Operator $\mathscr{L} : \mathscr{V} \to \mathscr{V}^*$ of the model problem	12
$\mathscr{L}^\varepsilon$	Operator with multiscale features	7, 23
$\mathscr{L}_K$	Local operator of a cell problem defined on K	14
λ	Penalty for Levenberg-Marquardt system	131
$\mathscr{L}^0$	Operator $\mathscr{L} : \mathscr{V} \to \mathscr{V}^*$ of the homogenized problem	24

$L^\infty(\Omega)$	Lebesgue space of measurable functions on Ω bounded by the essential supremum norm	10
$L^p_{\text{per}}(Y^\delta)$	Closure of $C^\infty_{\text{per}}(Y^\delta)$ with respect to $\|\|.\|\|_{L^p_{\text{per}}(Y^\delta)}$	25
$L^2(\Omega)$	Lebesgue space of measurable, square integrable functions	10
MsFEM	Multiscale Finite-Element Method	22
$\boldsymbol{n}$	Outward unit normal	81
Ω	Bounded computational domain	9
ω_i	Solution of a cell problem	26
PDE	Partial differential equation	2
φ^f	Oscillatory testfunction $\varphi^f \in \mathcal{V}^f$	12
φ^H	Macroscopic testfunction $\varphi^H \in V^H$	12
$Q_{h,K}$	Summed quadrature rule, associated with a discretization $\mathbb{T}_h$ of a sampling region Y^δ_K, or a sampling cell K	49
RVE	Representative volume element, denoted by RVE(x) if associated with a specific spatial location $x \in \Omega$	35
$\mathcal{R}^h_{K,i}$	Discretized fine-scale reconstruction operator	29
s. t.	such that	7
$\mathcal{T}$	Fine-scale solution operator $\mathcal{T} : \mathcal{V}^{f*} \to \mathcal{V}^f$	12
$\mathbb{T}_\delta$	Sampling discretization of Ω	46
$\mathbb{T}_H$	Macroscale discretization	13, 50
$\mathbb{T}_h$	Microscale discretization, typically associated with a sampling cell K, or a sampling region Y^δ_K	49
θ^δ	Model-error estimator	81
θ^H	Macroscale discretization error estimator	81
θ^h	Microscale discretization error estimator	81

U	Macroscale finite-element solution $U \in V^H$ of the fully discretized, effective problem	50
$u^{\delta,h}$	Solution of the effective, semi-discretized problem defined with the effective, numerically computed coefficients $\bar{A}^{\delta,h}$, or $A^{\delta,h}$	58
u^{δ}	Solution of the effective problem defined with the effective coefficients $\bar{A}^{\delta}$, or A^{δ}	49
u^{ε}	Solution of the model problem with multiscale character denoted by a scaling parameter ε	7, 23
u^f	Osciallatory part $u^f \in \mathcal{V}^f$ of u coming from a decomposition $u = u^H + u^f$	12
u^H	Macroscale part $u^H \in V^H$ of u coming from a decomposition $u = u^H + u^f$	12
u^0	Solution of the homogenized problem	24, 26
$\mathcal{V}$	Infinite-dimensional, separable Hilbert space	9
VMM	Variational Multiscale Method	1, 11
$\mathcal{V}^*$	Dual space of $\mathcal{V}$	9
$\mathcal{V}^{\varepsilon}$	Infinite-dimensional, separable Hilbert space, exhibiting multiscale features	7
$\mathcal{V}^f$	Infinite-dimensional space of microscale fluctuations	11
$\mathcal{V}_B^f$	Space of cell-wise bubbles	13
V^f	Finite-dimensional subspace of $\mathcal{V}^f$	12
V^H	Finite-dimensional subspace given by a macroscale finite-element discretization	11
V^h	Finite-dimensional subspace given by a microscale finite-element discretization	18
$W^{m,p}_{\text{per}}(Y^{\delta})$	Closure of $C^{\infty}_{\text{per}}(Y^{\delta})$ with respect to $\|.\|_{W^{m,p}(Y^{\delta})}$	25
Y	Unit cell $Y = [0,1]^d$	25
Y^{δ}	Rescaled unit cell $Y^{\delta} := \delta Y$	25
Y^{δ}_K	Sampling region, i. e. a rescaled unit cell Y^{δ} centered at a sampling cell K	47

z^δ	Solution of the effective dual problem defined with $\bar{A}^\delta$, or A^δ	88
z_K^δ	Local reconstruction of the dual solution for a sampling cell $K \in \mathbb{T}_\delta(\Omega)$	90
z^ε	Solution of the full dual problem (defined with A^ε)	79
$\tilde{Z}$	Finite-element approximation of specific dual solutions	90

Bibliography

[1] Standard for Programming Language C++, 2011. ISO/IEC 14882:2011.

[2] Assyr Abdulle. On A Priori Error Analysis of Fully Discrete Heterogeneous Multiscale FEM. *SIAM Journal of Multiscale Modeling and Simulation*, 4(2):447–459, 2005.

[3] Assyr Abdulle. A Priori and a Posteriori Error Analysis for Numerical Homogenization: A Unified Framework. In *Multiscale Problems*, volume 16 of *Series in Contemporary Applied Mathematics*, pages 280–305, 2011.

[4] Assyr Abdulle and Achim Nonnenmacher. A Posteriori Error Estimate In Quantities of Interest for the Finite Element Heterogeneous Multiscale Method. *Numerical Methods for Partial Differential Equations*, 29(5):1629–1656, 2013.

[5] Assyr Abdulle and Gilles Vilmart. Analysis of The Finite Element Heterogeneous Multiscale Method for Quasilinear Elliptic Homogenization Problems. *Mathematics of Computation*, 83:513–536, 2014.

[6] Robert A. Adams. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. 2nd edition, 2003.

[7] Grégoire Allaire. Homogenization and Two-Scale Convergence. *SIAM Journal on Mathematical Analysis*, 23(6):1482–1518, 1992.

[8] Grégoire Allaire and Rafael Orive. Homogenization of periodic non self-adjoint problems with large drift and potential. *ESAIM: Control, Optimisation and Calculus of Variations*, 13(4):735–749, 2007.

[9] Grégoire Allaire and Andrey L. Piatnitski. Homogenization of nonlinear reaction-diffusion equations with a large reaction term. *Annali dell'Università di Ferrara. Sezione VII*, 56(1):141–161, 2010.

[10] Grégoire Allaire and Anne-Lise Raphael. Homogenization of a Convection-Diffusion Model with Reaction in a Porous Medium. *Comptes Rendus Mathématique*, 344(8):523–528, 2007.

[11] Todd Arbogast and Kirsten J. Boyd. Subgrid Upscaling and Mixed Multiscale Finite Elements. *Journal of Numerical Analysis*, 44(3):1150–1171, 2006.

[12] Todd Arbogast and Steven L. Bryant. A Two-Scale Numerical Subgrid Technique for Waterflood Simulations. *SPE Journal*, 7(4):446–457, 2002.

[13] Ivo M. Babuška. Homogenization and its Application. Mathematical and Computational Problems. In *Numerical Solution of Partial Differential Equations*, volume 3 of *Proceedings of the Third Symposium on the Numerical Solution of Partial Differential Equations, SYNSPADE*, pages 363–379, 1976.

[14] Ivo M. Babuška and John E. Osborn. Generalized Finite Element Methods: Their Performance and Their Relation to Mixed Methods. *SIAM Journal of Numerical Analysis*, 20(3):510–536, 1983.

[15] Wolfgang Bangerth, Timo Heister, Luca Heltai, Guido Kanschat, Martin Kronbichler, M., Bruno Turcksin, and Toby D. Young. The `deal.II` Library, Version 8.1. *arXiv preprint*, 2013.

[16] Wolfgang Bangerth, Timo Heister, Luca Heltai, Guido Kanschat, Martin Kronbichler, M., Bruno Turcksin, and Toby D. Young. The `deal.II` Library, Version 8.2. *Archive of Numerical Software*, 3, 2015.

[17] Claus G. Bayreuther. *Mehrskalenmodelle in der Festkörpermechanik und Kopplung von Mehrgittermethoden mit Homogenisierungsverfahren*. PhD thesis, Universität Stuttgart, Germany, 2004.

[18] Roland Becker and Malte Braack. Multigrid Techniques for Finite Elements on Locally Refined Meshes. *Numerical Linear Algebra with Applications*, 7:363–379, 2000.

[19] Roland Becker and Rolf Rannacher. A Feed-Back Approach to Error Control in Finite Element Methods: Basic Analysis and Examples. *East-West Journal of Numerical Mathematics*, 4:237–264, 1996.

[20] Roland Becker and Rolf Rannacher. Weighted A Posteriori Error Control in FE Methods. In *Proceedings of ENUMATH-97*, pages 621–637, 1998. Lecture at ENUMATH-95, Paris, September 18-22, 1995.

[21] Roland Becker and Rolf Rannacher. An Optimal Control Approach to A Posteriori Error Estimation in Finite Element Methods. *Acta Numerica*, 10:1–102, 2001.

[22] Alain Bensoussan, Jacques-Louis Lions, and George Papanicolaou. *Asymptotic Analysis for Periodic Structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland, 1st edition, 1978.

[23] Heribert Blum and Rolf Rannacher. Extrapolation Techniques for Reducing the Pollution Effect of Reentrant Corners in the Finite Element Method. *Numerische Mathematik*, 52:539–564, 1988.

[24] Matthias Blumenfeld. *Eigenlösungen von gemischten Interface-Problemen auf Gebieten mit Ecken – ihre Regularität und Approximation mittels Finiter Elemente*. PhD thesis, Freie Universität Berlin, Germany, 1983.

[25] Matthias Blumenfeld. The Regularity of Interface-Problems on Corner-Regions. In *Singularities and Constructive Methods for Their Treatment*, volume 1121 of *Lecture Notes in Mathematics*, pages 38–54, 1985.

[26] Malte Braack and Alexandre Ern. A Posteriori Control of Modeling Errors and Discretization Errors. *Multiscale Modeling and Simulation*, 1 (2):221–238, 2003.

[27] Malte Braack and Thomas Richter. Solutions of 3D Navier-Stokes Benchmark Problems With Adaptive Finite Elements. *Computers and Fluids*, 35:27–392, 2006.

[28] Susanne C. Brenner and L. Ridgway Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, 3rd edition, 2008.

[29] Franco Brezzi. Interacting with the Subgrid World. *Numerical Analysis*, pages 69–82, 1999.

[30] Franco Brezzi, Leopoldo P. Franca, Thomas J.R. Hughes, and Alessandro Russo. $b = \int g$. *Computer Methods in Applied Mechanics and Engineering*, 145:329–339, 1997.

[31] W.T. Jr. Cardwell and R.L. Parsons. Average Permeability of Heterogeneous Oil Sands. *Transactions of the AIME*, 160(1):34–42, 1945.

[32] Zhiming Chen and Thomas Y. Hou. A Mixed Multiscale Finite Element Method for Elliptic Problems with Oscillating Coefficients. *Mathematics of Computation*, 72(242):541–576, 2003.

[33] Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*, volume 40 of *Classics im Applied Mathematics*. SIAM, 2nd edition, 2002.

[34] Doina Cioranescu and Patrizia Donato. *An Introduction to Homogenization*, volume 17 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, 1st edition, 1999.

[35] Philippe Clément. Approximation by Finite Element Functions using Local Regularization. *RAIRO Analyse Numérique*, 9:77–84, 1975.

[36] Jelmer M. Cnossen, Hester Bijl, Marc I. Gerritsma, and Barry Koren. Aspects of Goal-Oriented Model-Error Estimation in Convection-Diffusion Problems. In *European Conference on Computational Fluid Dynamics*, 2006.

[37] Timothy A. Davis. Algorithm 915, SuiteSparseQR: Multifrontal Multithreaded Rank-Revealing Sparse QR Factorization, 2011.

[38] Timothy A. Davis, Patrick R. Amestoy, Iain S. Duff, et al. SuiteSparse 4.2.1, A Suite of Sparse Matrix Software, 2013.

[39] Ennio De Giorgi. G-operators and Γ-convergence. In *Plenary Lectures and Ceremonies*, volume 1 of *Proceedings of the International Congress of Mathematicians*, pages 1175–1191, 1984.

[40] Peter H. Dederichs and Rudolf Zeller. Variational Treatment of the Elastic Constants of Disordered Materials. *Zeitschrift für Physik*, 259:103–116, 1972.

[41] Manfred Dobrowolski. *Numerical Treatment of Elliptic Interface and Corner Problems*. Habilitationsschrift, Rheinische Friedrich-Wilhelms-Universität Bonn, Germany, 1981.

[42] Weinan E and Björn Engquist. The Heterogeneous Multiscale Methods. *Communications in Mathematical Sciences*, 1(1):87–132, 2003.

[43] Weinan E and Björn Engquist. Multiscale Modeling and Computation. *Notices of the American Mathematical Society*, 50(9):1062–1070, 2003.

[44] Weinan E, Pingbing Ming, and Pingwen Zhang. Analysis of the Heterogeneous Multiscale Method for Elliptic Homogenization Problems. *Journal of the American Mathematical Society*, 18:121–156, 2005.

[45] Weinan E, Björn Engquist, Xiantao Li, Weiqing Ren, and Eric Vanden-Eijnden. The Heterogeneous Multiscale Methods: A review. *Communications in Computational Physics*, 2(3):367–450, 2007.

[46] Wiktor Eckhaus. *Matched Asymptotic Expansions and Singular Perturbations*, volume 6 of *Mathematics Studies*. North-Holland, 1973.

[47] Yalchin Efendiev and Thomas Y. Hou. *Multiscale Finite Element Methods, Theory and Applications*, volume 4 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, 2009.

[48] Yalchin Efendiev, Victor Ginting, and Thomas Y. Hou. Multiscale Finite Element Methods for Nonlinear Problems and their Applications. *Communications in Mathematical Sciences*, 2(4):553–589, 2004.

[49] Marc G.D. Geers, Varvara G. Kouznetsova, and W. A.M. Brekelmans. Multi-scale computational homogenization: Trends and challenges. *Journal of Computational and Applied Mathematics*, 234(7):2175–2182, 2010.

[50] Paul Germain. The Method of Virtual Power in Continuum Mechanics. Part 2: Microstructure. *SIAM Journal on Applied Mathematics*, 25(3): 556–575, 1973.

[51] Pièrre Grisvard. *Elliptic Problems in Nonsmooth Domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman Advanced Publishing Program, 1985.

[52] Patrick Henning. *Heterogeneous Multiscale Finite Element Methods for advection-diffusion and nonlinear elliptic multiscale problems*. PhD thesis, Westfälische Wilhelms-Universität Münster, Germany, 2011.

[53] Patrick Henning and Mario Ohlberger. A Note on Homogenization of Advection-Diffusion Problems with Large Expected Drift. *Zeitschrift für Analysis und ihre Anwendungen*, 30(3):319–339, 2011.

[54] Patrick Henning and Mario Ohlberger. Error Control and Adaptivity for Heterogeneous Multiscale Approximations of Nonlinear Monotone Problems. *Discrete and Continuous Dynamical Systems*, 8(1), 2015.

[55] Patrick Henning, Mario Ohlberger, and Ben Schweizer. An Adaptive Multiscale Finite Element Method. Technical Report 05/12 - N, FB 10 , Universität Münster, 2012.

[56] Rodney Hill. Elastic Properties of Reinforced Solids: Some Theoretical Principles. *Journal of the Mechanics and Physics of Solids*, 11:357–372, 1963.

[57] Rodney Hill. On Constitutive Macro-Variables for Heterogeneous Solids at Finite Strain. *Proceedings of the Royal Society*, A326:131–147, 1972.

[58] Viet Ha Hoang and Christoph Schwab. High-Dimensional Finite Elements for Elliptic Problems with Multiple Scales. 3(1):168–194, 2005.

[59] Thomas Y. Hou and Xiao-Hui Wu. A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media. *Journal of Computational Physics*, 134:169–189, 1997.

[60] Thomas J.R. Hughes. Multiscale Phenomena: Green's functions, the Dirichlet-to-Neumann Formulation, Subgrid Scale Models, Bubbles and the Origins of Stabilized Methods. *Computer Methods in Applied Mechanics and Engineering*, 127(1):387–401, 1995.

[61] Thomas J.R. Hughes and Giancarlo Sangalli. Variational Multiscale Analysis: The Fine-Scale Green's Function, Projection, Optimization, Localization, and Stabilized Methods. *SIAM Journal of Numerical Analysis*, 45(2):539–557, 2007.

[62] Thomas J.R. Hughes and James R. Stewart. A space-time formulation for multiscale phenomena. *Journal of Computational and Applied Mathematics*, 74(1-2):217–229, 1996.

[63] Thomas J.R. Hughes, Gonzalo R. Feijóo, Luca Mazzei, and Jean-Baptiste Quincy. The Variational Multiscale Method—a Paradigm for Computational Mechanics. *Computer Methods in Applied Mechanics and Engineering*, 166(1-2):3–24, 1998.

[64] Vladimir Aleksandrovich Kondrat'ev. Boundary Value Problems for Elliptic Equations in Domains with Conical or Angular Points. *Transactions of the Moscow Mathematical Society*, 16:227–313, 1967.

[65] Ekkehart Kröner. Bounds for Effective Elastic Moduli of Disordered Materials. *Journal of the Mechanics and Physics of Solids*, 25(2):137–155, 1977.

[66] Paco A. Lagerstrom. *Matched Asymptotic Expansions: Ideas and Techniques*, volume 78 of *Applied Mathematical Sciences*. Springer, 1988.

[67] Mats G. Larson and Axel Målqvist. Adaptive Variational Multiscale Methods Based on A Posteriori Error Estimation. In *Proceedings of the 4th European Congress on Computational Methods in Applied Sciences and Engineering*, 2004.

[68] Mats G. Larson and Axel Målqvist. Adaptive Variational Multiscale Methods Based on A Posteriori Error Estimation: Duality Techniques for Elliptic Problems. In *Multiscale Methods in Science and Engineering*, volume 44 of *Lecture notes in Computational Science and Engineering*, pages 181–193, 2005.

[69] Mats G. Larson and Axel Målqvist. An Adaptive Variational Multiscale Method for Convection–Diffusion Problems. *Communications in Numerical Methods in Engineering*, 25(1):65–79, 2009.

[70] Kenneth Levenberg. A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.

[71] D. Li, B. Beckner, and A. Kumar. SPE 56554 - A New Efficient Averaging Technique for Scaleup of Multimillion-Cell Geologic Models. *SPE Papers*, pages 495–510, 1999.

[72] M. and Rolf Rannacher. Duality-Based Adaptivity in Heterogeneous Multiscale Finite Element Discretization. *Journal of Numerical Mathematics*, submitted, 2014.

[73] Donald W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society of Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[74] Eduard Marušić-Paloka and Andrey L. Piatnitski. Homogenization of a Nonlinear Convection-Diffusion Equation with Rapidly Oscillating Coefficients and Strong Convection. *Journal of the London Mathematical Society*, 72(2):391–409, 2005.

[75] Ana-Maria Matache and Christoph Schwab. Generalized FEM for Homogenization Problems. *Multiscale and Multiresolution Methods*, 20:197–237, 2002.

[76] Ana-Maria Matache and Christoph Schwab. Two-Scale FEM for Homogenization Problems. *Mathematical Modelling and Numerical Analysis*, 36 (4):536–572, 2002.

[77] Donald McAlister. The Law of the Geometric Mean. *Proceedings of the Royal Society of London*, 29:367–376, 1879.

[78] Christian Miehe and Claus G. Bayreuther. On Multiscale FE Analyses of Heterogeneous Structures: From Homogenization to Multigrid Solvers. *International Journal for Numerical Methods in Engineering*, 71:1135–1180, 2007.

[79] Jürgen Moser. On Harnack's Theorem for Elliptic Differential Equations. *Communications on Pure and Applied Mathematics*, 14:577–591, 1961.

[80] François Murat and Luc Tartar. H-Convergence. In *Topics in the Mathematical Modelling of Composite Materials*, volume 31 of *Progress in Nonlinear Differential Equations and Their Applications*, pages 21–43. Birkhäuser, 1997.

[81] Axel Målqvist. *Adaptive Variational Multiscale Methods*. PhD thesis, Göteborgs Universitet, Sweden, 2005.

[82] Axel Målqvist and Daniel Peterseim. Localization of Elliptic Multiscale Problems. *Mathematics of Computation*, 83(290):2583–2603, 2014.

[83] Achim Nonnenmacher. *Adaptive Finite Element Methods for Multiscale Partial Differential Equations*. PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 2011.

[84] J. Tinsley Oden and Kumar S. Vemaganti. Estimation of Local Modeling Error and Goal-Oriented Adaptive Modeling of Heterogeneous Materials. Part I: Error Estimates and Adaptive Algorithms. *Journal of Computational Physics*, 164:22–47, 2000.

[85] J. Tinsley Oden and Kumar S. Vemaganti. Adaptive Modeling of Composite Structures: Modeling error estimation. *International Journal for Civil and Structural Engineering*, 1:1–16, 2000.

[86] J. Tinsley Oden and Kumar S. Vemaganti. Estimation of Local Modeling Error and Goal-Oriented Adaptive Modeling of Heterogeneous Materials.

Part II: A Computational Environment for Adaptive Modeling of Heterogeneous Elastic Solids. *Computer Methods in Applied Mechanics and Engineering*, 190(46–47):6089–6124, 2001.

[87] J. Tinsley Oden, Serge Prudhomme, Albert Romkes, and Paul T. Bauman. Multiscale Modeling of Physical Phenomena: Adaptive Control of Models. *Siam Journal on Scientific Computing*, 28(6):2359–2389, 2006.

[88] Mario Ohlberger. A Posteriori Error Estimates for the Heterogeneous Multiscale Finite Element Method for Elliptic Homogenization Problems. *Multiscale Modeling and Simulation*, 4(1):88–114, 2005.

[89] Martin Parr. *Langweilige Postkarten: Boring Postcards Germany*. Phaidon Press, 2001.

[90] Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale Methods: Averaging and Homogenization*, volume 53 of *Texts in Applied Mathematics*. Springer, 2007.

[91] Albert Romkes and Tristan C. Moody. Local Goal-Oriented Estimation of Modeling Error for Multi-Scale Modeling of Heterogeneous Elastic Materials. *International Journal for Computational Methods in Engineering Science and Mechanics*, 8(4):201–209, 2007.

[92] Enrique Sanchez-Palencia. *Non-Homogeneous Media and Vibration Theory*, volume 127 of *Lecture Notes in Physics*. Springer, 1st edition, 1980.

[93] Juliusz P. Schauder. Über lineare elliptische Differentialgleichungen zweiter Ordnung. *Mathematische Zeitschrift*, 38(1):257–282, 1934.

[94] L. Ridgway Scott and Shangyou Zhang. Finite Element Interpolation of Nonsmooth Functions Satisfying Boundary Conditions. *Mathematics of Computation*, 54:483–493, 1990.

[95] Sergio Spagnolo. Sul limite delle soluzioni di Cauchy relativi all'equazione del calore. *Annali Della Scuola Normale Superiore di Pisa*, 21(4):657–699, 1967.

[96] Sergio Spagnolo. Sulla convergenza di soluzioni di equizioni paraboliche ed ellitiche. *Annali Della Scuola Normale Superiore di Pisa*, 22(4):571–597, 1968.

[97] Luc Tartar. *The General Theory of Homogenization*, volume 7 of *Lecture Notes of the Unione Matematica Italiana*. Springer, 2010.

[98] Neil S. Trudinger. Linear Elliptic Operators with Measurable Coefficients. *Annali della Scuola Normale Superiore di Pisa*, 27(2):265–308, 1973.

[99] Hans-Jörg Vogel. QuantIm 4.01b, C/C++ Library for Scientific Image Processing, 2008.

[100] J.E. Warren and H.S. Price. Flow in Heterogeneous Porous Media. *Society of Petroleum Engineers Journal*, 1(3):153–169, 1961.

[101] Tarek I. Zohdi and Peter Wriggers. *Introduction to Computational Micromechanics*, volume 20 of *Lecture Notes in Applied and Computational Mechanics*. Springer, 2005.

[102] Bedii I. Özdemir, W. A.M. Brekelmans, and Marc G.D. Geers. FE^2-Computational Homogenization for the Thermo-Mechanical Analysis of Heterogeneous Solids. *Computer Methods in Applied Mechanics and Engineering*, 198(3-4):602–613, 2008.

Colophon

This document is typeset in URW Garamond
with LaTeX 2_ε using the TeX *Live* distribution.