Multiscale Modelling and Inverse Problems

J. Nolen, G.A. Pavliotis and A.M. Stuart

Abstract The need to blend observational data and mathematical models arises in many applications and leads naturally to inverse problems. Parameters appearing in the model, such as constitutive tensors, initial conditions, boundary conditions, and forcing can be estimated on the basis of observed data. The resulting inverse problems are often ill-posed and some form of regularization is required. These notes discuss parameter estimation in situations where the unknown parameters vary across multiple scales. We illustrate the main ideas using a simple model for groundwater flow.

We will highlight various approaches to regularization for inverse problems, including Tikhonov and Bayesian methods. We illustrate three ideas that arise when considering inverse problems in the multiscale context. The first idea is that the choice of space or set in which to seek the solution to the inverse problem is intimately related to whether a homogenized or full multiscale solution is required. This is a choice of regularization. The second idea is that, if a homogenized solution to the inverse problem is what is desired, then this can be recovered from carefully designed observations of the full multiscale system. The third idea is that the theory of homogenization can be used to improve the estimation of homogenized coefficients from multiscale data.

J. Nolen

Department of Mathematics, Duke University Durham, NC 27708, USA , e-mail: `nolen@math.duke.edu`

G.A. Pavliotis

Department of Mathematics Imperial College London London SW7 2AZ, UK e-mail: `g.pavliotis@imperial.ac.uk`

A.M. Stuart

Mathematics Institute Warwick University Coventry CV4 7AL, UK e-mail: `A.M.Stuart@warwick.ac.uk`

1

1 Introduction

The objective of this overview is to demonstrate the important role of multiscale modelling in the solution of inverse problems for differential equations. The main inverse problem we discuss is that of determining unknown parameters by matching observed data to a differential equation model involving those parameters. The unknown parameters may be functions, in general, and they may have variation over multiple (length) scales. This multiscale structure makes the forward problem more challenging: numerically computing the solution to the differential equation requires very high resolution. The multiscale structure also complicates the inverse problem. Should we try to fit the data with a high-dimensional parameter, or should we seek a low-dimensional "homogenized" approximation of the parameter? If a low-dimensional parameter model is used, how should we account for the mismatch between the true parameters and the low-dimensional representation? After obtaining a solution to the inverse problem, one typically wants to make further predictions using whatever parameter is fit to the observed data, so it is important to consider whether a low-dimensional representation of the unknown parameter is sufficient to make additional predictions.

Throughout these notes the unknown parameters will be denoted by $u \in X$; typically u is a function assumed to lie in a Banach space X. We use $y \in Y$ to denote the data (for simplicity we often take $Y = \mathbb{R}^N$) and z to denote the predicted quantity, assumed to be an element of a Banach space Z or, in some cases, a $Z-$valued random variable. The map $\mathscr{G} : X \to \mathbb{R}^N$ denotes the forward mapping from the unknown parameter to the data, and $\mathscr{F} : X \to Z$ (or $\mathscr{F} : X \times \Omega \to Z$ in the random case) denotes the forward mapping from the parameter to the prediction. We sometimes refer to $\mathscr{G}$ as the *observation operator* and $\mathscr{F}$ as the *prediction operator*. Both $\mathscr{G}$ and $\mathscr{F}$ are typically derived from a common solution operator $G : X \to P$ mapping $u \in X$ to the solution $G(u) \in P$ of a partial differential equation (PDE), where P is a Banach space. For example $\mathscr{G}$ may be derived by composing G with N linear functionals.

The ideal inverse problem is to determine $u \in X$ from knowledge of $y \in \mathbb{R}^N$ where it is assumed that $y = \mathscr{G}(u)$. In practice, however, the data y is generated from outside this clean mathematical model, so it is natural to think of the data y as being given by

$$y = \mathscr{G}(u) + \xi \tag{1}$$

for some $\xi \in \mathbb{R}^N$ quantifying model error[1] and observational noise. The value of ξ is not known, but it is common in applications to assume that some of its statistical properties are known and these can then be built into the methods used to estimate u. Once the function u is determined by solving this inverse problem, it can be used to make a prediction $z = \mathscr{F}(u)$.

We illustrate three ideas that arise when attempting to solve the inverse problem defined by (1) in the multiscale context:

---

[1] Model error can be incorporated within the set of unknown parameters u and estimated using data; however this idea is not pursued here.

- (a) The choice of the space or set in which to seek the solution to the inverse problem is intimately related to whether a low-dimensional "homogenized" solution or a high-dimensional "multiscale" solution is required for predictive capability. This is a choice of regularization.
- (b) If a homogenized solution to the inverse problem is desired, then this can be recovered from carefully designed observations of the full multiscale system.
- (c) The theory of homogenization can be used to improve the estimation of homogenized parameters from observations of multiscale data.

In Section 2 we consider in detail a worked example which exemplifies the use of multiscale methods to approximate the forward problems $\mathscr{G}$ and $\mathscr{F}$ for data and predictions; this example will be used to illustrate many of the general ideas developed in these notes, and the three ideas (a)–(c) in particular. Section 3 is devoted to a brief overview of regularization techniques for inverse problems, and to discussion of the idea (a). Section 4 is devoted to the idea (b). We study the problem of estimating a single scalar parameter in a homogenized model of groundwater flow, given data which is generated by a full multiscale model. This may be seen as a surrogate for understanding the use of real-world data (which is typically multiscale in character) to estimate parameters in simpler homogenized models. Section 5 is devoted to the idea (c). We study the use of ideas from multiscale methodology to enhance parameter estimation techniques for homogenized models. The viewpoint taken is that the statistics of the error ξ appearing in (1) can be understood using the theory of homogenization for random media; when these statistical properties depend on the unknown parameter u the noise ξ is no longer additive and its dependence on u plays an important role in the parameter estimation process.

1.1 Notation

The following notation will be used throughout. We use $|\cdot|$ to denote the Euclidean norm on $\mathbb{R}^m$ (for possibly different choices of m). We let S^d (resp. $S^{d,+}$) denote the set of symmetric (resp. positive-definite) second order tensors on $\mathbb{R}^d$. If $\Gamma \in S^{d,+}$, we define the weighted norm $|\cdot|_\Gamma = |\Gamma^{-\frac{1}{2}}\cdot|$ on $\mathbb{R}^m$. Throughout the notes, X is a Banach space, containing the functions that we wish to estimate, and E a Banach space compactly embedded into X. When studying the inverse problem from a Bayesian perspective we will use Gaussian priors on X, defined via a covariance operator $\mathscr{C}$ on a Hilbert space $H \supseteq X$, with norm $\|\cdot\|_H$. In this situation E will be the Hilbert space with norm $\|\mathscr{C}^{-\frac{1}{2}}\cdot\|_H$.

1.2 Running Example

We consider a model for groundwater flow in a medium with permeability tensor k, pressure p and Darcy velocity v (or the volume flux of water per unit area) related to the pressure via the Darcy law:

$$v = -\frac{k}{\mu}(\nabla p - \rho g \hat{e}_z) \qquad (2)$$

where μ is the fluid viscosity, ρ is the fluid density, g is the acceleration due to gravity and $\hat{e}_z$ is the unit vector in the z-direction. We choose units in which $\mu = 1$. We also assume that we have a constant density fluid and redefine the pressure by adding $\rho g z$ (z is the vertical direction) to write (2) in the form $v = -k\nabla p$. Assuming that the Darcy velocity is divergence-free, except at certain known source/sink locations, we obtain the following elliptic equation for the pressure:

$$\begin{aligned} \nabla \cdot v &= f, \quad x \in D, \\ p &= 0, \quad x \in \partial D, \\ v &= -k\nabla p \end{aligned} \qquad (3)$$

where $D \subset \mathbb{R}^d$ is an open and bounded set with regular boundary, and f is assumed to be known. The permeability tensor field k, however, is assumed to be unknown and must be determined from data. In order to make the elliptic PDE (3) for the pressure p well-posed, we assume that the permeability tensor $k(x)$ is an element of $S^{d,+}$ and so we write it as the (tensor) exponential: $k(x) = \exp(u(x))$, $u \in S^d$. It is natural to view u as an element of $X := L^\infty(D; S^d)$ and to consider weak solutions of (3) with $f \in H^{-1}(D)$. Then we have a unique solution $p \in H_0^1(D)$ satisfying

$$\|\nabla p\|_{L^2} \leq c_1 \exp(\|u\|_X)\|f\|_{H^{-1}}, \qquad (4)$$

for some $c_1 > 0$ depending only on d and D, and $\|u\|_X$ being the essential supremum of the spectral radius of the matrix $u(x)$, as x varies over D:

$$\|u\|_X = \underset{x \in D}{\text{ess-sup}} \left(\max_{\substack{\xi \in \mathbb{R}^d \\ |\xi|=1}} |u(x)\xi| \right).$$

Thus we may define $G : X \to H_0^1(D)$ by $G(u) = p$. Now consider a set of real-valued continuous linear functionals $\ell_j : H^1(D) \to \mathbb{R}$ and define $\mathscr{G} : X \to \mathbb{R}^N$ by $\mathscr{G}(u)_j = \ell_j(G(u))$. The inverse problem is to determine $u \in X$ from $y \in \mathbb{R}^N$ where it is assumed that y is given by (1). Using (4) one may show that $G : X \to H_0^1(D)$ (resp. $\mathscr{G} : X \to \mathbb{R}^N$) is Lipschitz. Indeed if p_i denotes the solution to (3) with log permeability u_i then, we have

$$\|\nabla p_1 - \nabla p_2\|_{L^2} \leq (c_1)^2 \|u_1 - u_2\|_X \exp\left(2(\|u_1\|_X + \|u_2\|_X)\right)\|f\|_{H^{-1}}. \qquad (5)$$

Study of the transport of contaminants in groundwater flow is a natural example of a useful prediction that can be made once the inverse problem is solved. To model this scenario we consider a particle $x(t) \in \mathbb{R}^d$ which is advected by the the groundwater velocity field v/ϕ, where ϕ is the porosity of the rock and v is the Darcy velocity field from (3), and subject to diffusion with coefficient 2η. Assuming that the contaminant is initially at x_{init} we obtain the stochastic differential equation (SDE):

$$dx = \frac{v(x)}{\phi}\,dt + \sqrt{2\eta}\,dW, \quad x(0) = x_{\text{init}}, \tag{6}$$

where $W(t)$ is a standard Brownian motion on $\mathbb{R}^d$. If we are interested in predicting the location of the contaminant at time T then our prediction will be the function $\mathscr{F}_\eta$ given by $\mathscr{F}_\eta(u) = x(T)$. Here for each fixed $\eta \in [0,\infty)$ the function $\mathscr{F}_\eta$ maps X into the family of $\mathbb{R}^d-$valued random variables.

2 The Forward Problem: Multiscale Properties

Some inverse problems arising in applications have the property that the forward model $\mathscr{G}$ mapping the unknown to the data will produce similar output on both highly oscillatory functions u and on appropriately chosen smoothly varying functions u. Furthermore, for some choices of prediction function $\mathscr{F}$ the predictions themselves will also be close for both highly oscillatory functions u and on appropriately chosen smoothly varying functions u. These properties can be seen from an application of multiscale analysis, and we illustrate them by considering the problem introduced in Section 1.2. There are many texts on the theory of multiscale analysis. For example, the basic homogenization theorems discussed here are developed in [6]. A recent overview of the subject, with many other references and using the same notational conventions that we adopt here, is [24].

We consider a multiscale version of the running example from Section 1.2 where the permeability tensor is $k = K^\varepsilon(x) = K(x, x/\varepsilon)$ where $K : D \times \mathbb{T}^d \to S^{d,+}$ is periodic in the second argument, $\varepsilon > 0$ a small parameter. For now we have assumed periodic dependence on the fast scale in K^ε; however we will generalize this to random dependence in later developments.

With this permeability we obtain the family of problems

$$\nabla \cdot v^\varepsilon = f, \quad x \in D, \tag{7a}$$
$$p^\varepsilon = 0, \quad x \in \partial D, \tag{7b}$$
$$v^\varepsilon = -K^\varepsilon \nabla p^\varepsilon. \tag{7c}$$

If we set $\eta = \varepsilon \eta_0$, then the transport of contaminants is given by the SDE

$$dx^\varepsilon = \frac{v^\varepsilon(x^\varepsilon)}{\phi}\,dt + \sqrt{2\eta_0 \varepsilon}\,dW, \quad x^\varepsilon(0) = x_{\text{init}}. \tag{8}$$

Standard techniques from the theory of homogenization for elliptic PDEs can be used to show that for ε small,

$$p^{\varepsilon}(x) \approx p_{a}^{\varepsilon}(x) := p_0(x) + \varepsilon p_1(x, \frac{x}{\varepsilon}) \tag{9}$$

where p_0 and p_1 are defined as follows. First we define the effective (homogenized) permeability tensor K_0 via solution of the *cell problem* for $\chi(x,y)$:

$$-\nabla_y \cdot \left(\nabla_y \chi K^T \right) = \nabla_y \cdot K^T, \quad y \in \mathbb{T}^d. \tag{10}$$

Then

$$K_0(x) = \int_{\mathbb{T}^d} Q(x,y) dy, \tag{11}$$

$$Q(x,y) = K(x,y) + K(x,y) \nabla_y \chi(x,y)^T. \tag{12}$$

In this sense we observe that the effective diffusivity $K_0(x)$ is the average of $Q(x,y)$ over the fast scale y. This is *not* equal to the average of $K(x,y)$ over y, except in trivial cases. We denote by u_0 the logarithm of K_0 so that $K_0 = \exp(u_0)$.

The function p_0 solves the (ε independent) elliptic PDE

$$\nabla \cdot v_0 = f, \quad x \in D, \tag{13a}$$

$$p_0 = g, \quad x \in \partial D, \tag{13b}$$

$$v_0 = -K_0 \nabla p_0. \tag{13c}$$

and the corrector p_1 is given by

$$p_1(x,y) = \chi(x,y) \cdot \nabla p_0(x). \tag{14}$$

Note that (10) may be written as

$$-\nabla_y \cdot \left(Q^T \right) = 0, \quad y \in \mathbb{T}^d. \tag{15}$$

This shows that Q, which is averaged to give the effective permeability tensor, is divergence-free with respect to the fast variable y.

It is possible to prove that, in the limit as $\varepsilon \to 0$, solutions to (7) converge to solutions to (13), the convergence being strong in $L^2(D)$ and weak in $H^1(D)$ [10, 1, 24]. However if we want to prove strong convergence in H^1 then we need to include information about the corrector term p_1. The following theorem and corollary summarize these ideas. For proofs see [1], or the discussion in the texts [10, 24].

Theorem 1. *Let p^{ε} and p_0 be the solutions of* (7) *and* (13). *Assume that $f \in C^{\infty}(D)$ and that $K(x,y) \in C^{\infty}(D; C_{\text{per}}^{\infty}(\mathbb{T}^d))$. Then*

$$\lim_{\varepsilon \to 0} \| p^{\varepsilon} - p_{a}^{\varepsilon} \|_{H^1} = 0. \tag{16}$$

Corollary 1. *Under the same conditions as in Theorem 1 we have*

$$\|p^\varepsilon - p_0\|_{L^2} \to 0 \quad \text{and} \quad \|\nabla p^\varepsilon - \left(I + \chi_y(\cdot, \cdot/\varepsilon)^T\right)\nabla p_0\|_{L^2} \to 0$$

as $\varepsilon \to 0$.

In fact it is frequently the case that the convergence in Theorem 1 may be obtained in a stronger topology. Reflecting this we make the following assumption.

Assumption 2 *The function p^ε converges to p_0 in $L^\infty(D)$ and its gradient converges to the gradient of $p_0 + \varepsilon p_1$ in $L^\infty(D)$ so that*

$$\lim_{\varepsilon \to 0} \|p^\varepsilon - p_a^\varepsilon\|_{W^{1,\infty}} = 0.$$

In Appendix 5.3 we prove this assumption for the one dimensional version of (7). The proof in the multidimensional case will be presented elsewhere [22]. The proof of this assumption in the multidimensional case is based on the estimates proved in [2] (in particular, Lemma 16), see also [15, Lemma 2.1].

With these limiting properties of the elliptic problem (7) at hand it is natural to ask what is the limiting behaviour of x^ε governed by (8). To answer this question we define

$$\frac{dx_0}{dt} = \frac{v_0(x_0)}{\phi}, \quad x_0(0) = x_{\text{init}}. \tag{17}$$

Notice that this ordinary differential equation (ODE) has vector field v_0 which is defined entirely through knowledge of the homogenized permeability K_0: once K_0 is known, the elliptic PDE (13) can be solved for p_0 and then v_0 is recovered from (13c). If we can show that solutions of (8) and (17) are close then this will establish that the prediction of particle transport in the model (7), (8) can be made accurately by use of only homogenized information about the permeability.

In proving such a result there are a number of technical issues which arise caused by the presence of the boundary D of the domain in which the PDE (7) is posed. In particular solutions of (8) may leave D requiring a definition of the velocity field outside D. These issues disappear if we consider the case where D is itself a box of length L and is equipped with periodic boundary conditions instead of Dirichlet conditions: we may then extend all fields to the whole of $\mathbb{R}^d$ by periodicity. In this case, the homogenization theory for (7) with (7b) replaced by periodic boundary conditions is identical to that given above, except that (13b) is also replaced by periodic boundary conditions. We write $D = (L\mathbb{T})^d$ and adopt this periodic setting for the next theorem, which is proved in Appendix 5.3:

Theorem 3. *Let $x^\varepsilon(t)$ and $x_0(t)$ be the solutions to equations (8) and (17), with velocity fields extended from $D = (L\mathbb{T})^d$ to $\mathbb{R}^d$ by periodicity, and assume that Assumption 2 holds. Assume also that $f \in C^\infty(D)$ and that $K(x,y) \in C^\infty(D; C_{\text{per}}^\infty(\mathbb{T}^d))$. Then*

$$\lim_{\varepsilon \to 0} \mathbb{E} \sup_{0 \le t \le T} \|x^\varepsilon(t) - x_0(t)\| = 0.$$

In summary, this example exhibits the property that, if the length scale ε is small, the data generated from K^ε and K_0 may appear very similar due to homogenization effects. Therefore, when trying to infer parameters from data, it is difficult to distinguish between K^ε and K_0 without some form of regularization or prior assumptions about the form of the parameter. On the other hand, Theorem 3 shows that knowing only K_0 is sufficient to make accurate predictions of the trajectories of (8).

3 Regularization of Inverse Problems

In this section we describe various approaches to regularizing inverse problems, motivating them by reference to the multiscale example in the previous section. The approach to regularizing which is described in Section 3.2 is developed in detail in [5]. The Tikhonov regularization approach from Section 3.3 is developed in detail in [16, 17]. Both of these regularization approaches are specific examples of the general set-up often called PDE constrained optimization, which we discuss in Section 3.4; this subject is overviewed in [18]. An overview of the Bayesian approach to inverse problems, a subject that we outline in Section 3.5, is given in [26] and [17].

3.1 Set-Up

Our objective here is to determine u, given y, where u and y are related by (1). We assume that, whilst the actual value of ξ is not available, it is reasonable to view it as a single draw from a statistical distribution whose properties are known to us. To be concrete we assume that ξ is drawn from a mean zero Gaussian random variable with covariance Γ: we write this as $\xi \sim N(0,\Gamma)$. We make the following continuity assumption concerning the observation operator $\mathscr{G}$. Note that this (local) Lipschitz condition also implies an (exponential in $\|u\|_X$) bound on $|\mathscr{G}(u)|$.

Assumption 4 *There are constants $c_1, c_2 > 0$ such that, for $u_i \in X$ with $\|u_i\|_X < r, i = 1, 2$,*

$$|\mathscr{G}(u_1) - \mathscr{G}(u_2)| \leq c_1 \exp(c_2 r)\|u_1 - u_2\|_X.$$

In general the inverse problems such as that given by (1) with $\xi = 0$ are hard to solve: they may have no solutions, multiple solutions and solutions may exhibit sensitive dependence on initial data. For this reason it is natural to seek a least squares approach to finding functions u which best explain the data. In view of the assumed structure on ξ a natural least squares functional is

$$\Phi(u) = \frac{1}{2}|y - \mathscr{G}(u)|_\Gamma^2. \tag{18}$$

The weighting by Γ in the Euclidean norm induces a normalization on the model-data mismatch. This normalization is given by the assumed standard deviations of the noise in a coordinate system defined by the eigenbasis for Γ.

Example 1. Consider the running example of Section 1.2. Equation (5) shows that Assumption 4 holds in this case, noting that $\mathscr{G}(u)_j = \ell_j(p)$ for some linear functional ℓ_j on $H^1(D)$, with the choice $X = L^\infty(D; S^d)$, provided $f \in H^{-1}$. We use this example to illustrate why inverse problems are, in general, hard.

Assume that the linear functionals ℓ_j satisfy the property that $\ell_j(p^\varepsilon - p_0) \to 0$ as $\varepsilon \to 0$. This occurs if they are linear functionals on $L^2(D)$, by Theorem 1 or if Assumption 2 holds, if they are linear functionals on $C(\overline{D})$. Writing this in terms of $\mathscr{G}$ we have $|\mathscr{G}(u^\varepsilon) - \mathscr{G}(u_0)| \to 0$ as $\varepsilon \to 0$. (Note that this occurs even though u^ε and u_0 are not themselves close.) Hence there is an uncountable family of functions (indexed by all ε sufficiently small) which all return approximately the same value of $\Phi(u^\varepsilon)$ and thus simply minimizing Φ may be very difficult. Furthermore, there may be minimizing sequences which do not converge. For example fix a particular realization of the data given by $y = \mathscr{G}(u_0)$ where u_0 is the homogenized log permeability. Then $\Phi(u^\varepsilon) \geq 0$ for all $\varepsilon > 0$ and $\Phi(u^\varepsilon) \to 0$ as $\varepsilon \to 0$, since

$$|\Phi(u^\varepsilon)| = \frac{1}{2}|y - \mathscr{G}(u^\varepsilon)|_\Gamma^2 = \frac{1}{2}|\mathscr{G}(u_0) - \mathscr{G}(u^\varepsilon)|_\Gamma^2 \tag{19}$$

On the other hand, u^ε does not converge in X as $\varepsilon \to 0$.

$\square$

In order to overcome the difficulties demonstrated in this example regularization is needed. In the remaining sections we discuss various regularizations, in general, illustrating ideas by returning to the running example.

3.2 Regularization by Minimization Over a Convex, Compact Set

Recall that E is a Banach space compactly embedded into X. Let $E_{\mathrm{ad}} = \{u \in E : \|u\|_E \leq \alpha\}$. Then E_{ad} is a closed convex and bounded set in E and, as such, any sequence in E_{ad} must contain a weakly convergent subsequence with limit in E_{ad} (see, for example, Theorem 1.17 in [18]). Now consider the minimization problem

$$\overline{\Phi} = \inf_{u \in E_{\mathrm{ad}}} \Phi(u). \tag{20}$$

Theorem 5. *Any minimizing sequence $\{u^n\}_{n \in \mathbb{Z}^+}$ for (20) contains a weakly convergent subsequence in E with limit $\overline{u} \in E_{\mathrm{ad}}$ which attains the infimum: $\Phi(\overline{u}) = \overline{\Phi}$.*

Proof. This is a classical theorem from the field of optimization; see [18] for details and context. Since $\{u^n\}$ is contained in E_{ad} we deduce the existence of a subsequence (which for convenience we relabel as $\{u^n\}$) with weak limit $\overline{u} \in E_{\mathrm{ad}}$. Thus $u^n \rightharpoonup \overline{u}$ in E. Hence, by compactness, $u^n \to u$ in X. By Assumption 4 we deduce that

$\Phi : E \to \mathbb{R}$ is weakly continuous. By definition, for any $\delta > 0$ there exists $N = N(\delta)$ such that

$$\overline{\Phi} \leq \Phi(u_n) \leq \overline{\Phi} + \delta, \quad \forall n \geq N.$$

By weak continuity of $\Phi : E \to \mathbb{R}$ we deduce that

$$\overline{\Phi} \leq \Phi(\overline{u}) \leq \overline{\Phi} + \delta.$$

The result follows since δ is arbitrary. $\quad\square$

Example 2. Consider the running example of Section 1.2. Let A denote a fixed symmetric positive-definite tensor A so that $\log(A)$ is defined. We define the subspace of tensor valued functions of the form $u' = uI + \log(A)$, for some constant $u \in \mathbb{R}$ noting that then $\exp(u') = \exp(u)A$. By Lipschitz continuity of $\mathscr{G}$ in $u' \in X$ we deduce (abusing notation) Lipschitz continuity of $\mathscr{G}$ viewed as a function of $u \in \mathbb{R}$. We define

$$E_{\mathrm{ad}} = \{u \in \mathbb{R} : |u| \leq \alpha\}. \tag{21}$$

We may take the norm $\| \cdot \|_E = |u|$. Thus the problem (20) attains its infimum for some $\overline{u} \in E_{\mathrm{ad}}$. The regularization of seeking to minimize Φ over E_{ad} corresponds to looking for solution over a one-parameter set of tensor fields, in which the free parameter is bounded by α. Note that such a solution set automatically rules out the oscillating minimizing sequences which were exhibited in Example 1. $\quad\square$

3.3 Tikhonov Regularization

Instead of regularizing by seeking to minimize Φ over a bounded and convex subset of a compact set E in X, we may instead adopt the Tikhonov approach to regularization. We consider the minimization problem

$$\overline{I} = \inf_{u \in E} I(u), \tag{22}$$

where

$$I(u) = \frac{\lambda}{2} \|u\|_E^2 + \Phi(u). \tag{23}$$

Theorem 6. *Any minimizing sequence $\{u^n\}_{n \in \mathbb{Z}^+}$ for (22) contains a weakly convergent subsequence in E with limit $\overline{u}$ which attains the infimum: $I(\overline{u}) = \overline{I}$.*

Proof. This is a classical theorem from the calculus of variations; see [12] for details and context. Since $\{u^n\}$ is a minimizing sequence and $\Phi \geq 0$, we deduce that for any $\delta > 0$ there exists $N = N(\delta)$ such that

$$\frac{\lambda}{2} \|u_n\|_E^2 \leq \overline{I} + \delta, \quad \forall n \geq N.$$

From this it follows that $\{u^n\}_{n \in \mathbb{Z}^+}$ is bounded in E and hence contains a weak limit $\overline{u}$, along a subsequence which, for convenience, we relabel as $\{u^n\}$. The weak continuity of $\Phi : E \to \mathbb{R}$, together with weak lower semicontinuity of the function $\|\cdot\|_E^2 \to \mathbb{R}$ implies the weak lower semicontinuity of $I : E \to \mathbb{R}$. Hence

$$I(\overline{u}) \leq \liminf_{n \to \infty} I(u_n) \leq \overline{I}.$$

Since $I(\overline{u}) \geq \overline{I}$, the result follows. $\square$

Example 3. Consider the running example of Section 1.2. Let $E = H^s(D; S^d)$ and note that E is compact in $X = L^\infty(D; S^d)$ for $s > d/2$. Thus the problem (22) attains its infimum for some $\overline{u} \in E$. As with the example from the previous section the regularization rules out highly oscillating minimizing sequences such as those seen in Example 1. The choice of the parameter λ will effect how much oscillation is allowed in any minimizing sequence. $\square$

3.4 PDE Constrained Optimization

The regularizations imposed in the two previous subsections involed the imposition of constraints on the input u to a PDE model and the resulting minimizations were expressed in terms of u alone. For at least two reasons it is sometimes of interest to formulate the minimization problem simultaneously over the input variable u, together with the solution of the PDE $p = G(u) \in P$: firstly computational algorithms which work to find (p, u) in $P \times X$ can be more effective than working entirely in terms of $u \in X$; and secondly regularization constraints may be imposed on the variable p as well as on u. If $J : P \times X \to \mathbb{R}$ then this leads to constrained minimization problems of the form

$$\min_{(p,u) \in P \times X} J(p, u) : \ p = G(u), c(p, u) \in \mathcal{K} \tag{24}$$

where $\mathcal{K}$ denotes the constraints imposed on both the input u and on the output p from the PDE model. Typically the observation operator $\mathcal{G} : X \to \mathbb{R}^N$ is found from G and then the information in Φ can be built into the definition of J.

Example 4. Consider the running example from Section 1.2 and assume that the observational noise $\xi \sim N(0, \gamma^2 I)$. Define

$$J(p, u) = \frac{1}{2\gamma^2} \sum_{j=1}^N |y - \ell_j(p)|^2 + \frac{\lambda_1}{2} \|u\|_{H^s}^2 + \frac{\lambda_2}{2} \|p\|_P^2$$

for some $s > d/2$. Choosing $\lambda_1 = \lambda$ and $\lambda_2 = 0$, together with $c(p, u) = (p, u)$ and $\mathcal{K} = P \times X$ we obtain from (24) the minimization from Example 3 in the case $\Gamma = \gamma^2 I$. Choosing $\lambda_1 = \lambda_2 = 0$, $c(p, u) = (p, u)$ and $\mathcal{K} = P \times E_{\mathrm{ad}}$ from Example 2 we recover that example. Choosing $\lambda_2 \neq 0$ and/or choosing the constraint set $\mathcal{K}$

to impose constraints on p leads to minimization in which the output p of the PDE model is constrained as well as the input u that we are trying to estimate. $\quad\square$

3.5 Bayesian Regularization

The preceding regularization approaches have a nice mathematical structure and form a natural approach to the inverse problem when a unique solution is to be expected. But in many cases it may be interesting or important to find a large class of solutions, and to give relative weights to their importance. This allows, in particular, for predictions which quantify uncertainty. The Bayesian approach to regularization does this by adopting a probabilistic framework in which the solution to the inverse problem is a probability measure on X, rather than a single element of X.

We think of $(u, y) \in X \times \mathbb{R}^N$ as a random variable. Our goal is to find the distribution of u given y, often denoted by $u|y$. We define the joint distribution of (u, y) as follows. We assume that u and ξ appearing in (1) are indepenent mean zero Gaussian random variables, supported on X and $\mathbb{R}^N$ respectively, with covariance operator $\frac{1}{\lambda}\mathscr{C}$ and covariance matrix Γ respectively. By equation (1), the distribution of y given u, denoted $y|u$, is Gaussian $N(\mathscr{G}(u), \Gamma)$. The measure $\mu_0 = N(0, \frac{1}{\lambda}\mathscr{C})$ is known as the *prior* measure. It is most natural to define the measure μ_0 on a Hilbert space $H \supseteq X$. Under suitable conditions on $\mathscr{C}$, we have $\mu_0(X) = 1$. This means that under the measure μ_0, $u \in X$ almost surely so that $\mathscr{G}(u)$ is well-defined, almost surely. If $\mu_0(X) = 1$, it follows that the Hilbert space E with norm $\|\cdot\|_E = \|\mathscr{C}^{-1/2} \cdot \|_H$ is compactly embedded into X. The space E is known as the Cameron-Martin space. In the infinite dimensional setting, functions drawn from μ_0 are almost surely not in the Cameron-Martin space. See [9, 20] for detailed discussion of Gaussian measures on infinite dimensional spaces.

When solving the inverse problem, the aim is to find the posterior measure $\mu^y(du) = \mathbb{P}(du|y)$, and to obtain information about likely candidate solutions to the inverse problem from it. Informal application of Bayes' theorem gives

$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u)\mu_0(u). \tag{25}$$

The probability density function for $\mathbb{P}(y|u)$ is, using the property of Gaussians, proportional to

$$\exp\left(-\frac{1}{2}|y - \mathscr{G}(u)|_\Gamma^2\right) = \exp\left(-\Phi(u)\right).$$

The infinite dimensional analogue of this result is to show that μ^y is absolutely continuous with respect to μ_0 with Radon-Nikodym derivative relating posterior to prior as follows:

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\exp\left(-\Phi(u)\right). \tag{26}$$

Here $\Phi(u)$ is given by (18) and $Z = \int_X \exp\left(-\Phi(u)\right)\mu_0(du)$. The meaning of the formula (26) is that expectations under the posterior measure μ^y can be rewritten

as weighted expectations with respect to the prior: for a function $\mathscr{F}$ on X we may write

$$\int_X \mathscr{F}(u)\mu^y(du) = \int_X \frac{1}{Z}\exp\bigl(-\Phi(u)\bigr)\mathscr{F}(u)\mu_0(du).$$

Theorem 7. *([11]) Assume that $\mu_0(X) = 1$. Then μ^y is absolutely continuous with respect to μ_0 with Radon-Nikodym derivative given by* (26). *Furthermore the measure μ^y is locally Lipschitz in the data y with respect to the Hellinger metric: there is a constant $C = C(r)$, such that, for all y, y' with $\max\{|y|, |y'|\} \leq r$,*

$$d_{\text{HELL}}(\mu^y, \mu^{y'}) \leq C|y - y'|. \tag{27}$$

If μ, ν are probability measures that are absolutely continuous with respect to the probability measure ρ, then the Hellinger metric is defined as

$$d_{\text{HELL}}(\mu, \nu)^2 = \frac{1}{2}\int\left(\sqrt{\frac{d\mu(u)}{d\rho}} - \sqrt{\frac{d\nu(u)}{d\rho}}\right)^2 \rho(du).$$

For any function of u which is square integrable with respect to both μ and ν it may be shown that the difference in expectations of that function, under μ and under ν, is bounded above by the Hellinger distance. In particular, this theorem shows that the posterior mean and covariance operators corresponding to data sets y and y' are $\mathscr{O}(|y - y'|)$ apart.

The choice of prior μ_0, relates directly to the regularization of the inverse problem. To see this we note that since the operator $\mathscr{C}$ is necessarily positive and self-adjoint we may write down the complete orthonormal system

$$\frac{1}{\lambda}\mathscr{C}\phi_m = \sigma_m^2\phi_m, \quad m \in \mathbb{Z}^+, \qquad \lim_{m\to\infty}\sigma_m = 0. \tag{28}$$

Then $u \sim \mu_0$ can be written via the Karhunen-Loève expansion as

$$u(x) = \sum_{m\in\mathbb{Z}^+} \sigma_m \eta_m \phi_m(x) \tag{29}$$

where the η_m form an i.i.d. sequence of unit Gaussian random variables. We may regularize the inverse problem by modifying the decay rate of σ_m. For example, choosing $\sigma_m = 0$ for $m \notin \mathscr{M}$, where $\mathscr{M} \subset \mathbb{Z}^+$ has finite cardinality restricts the solution of the inverse problem to a finite dimensional set, and is hence a regularization. More generally, the rate of decay of the σ_m (which are necessarily summable as $\mathscr{C}$ is trace class) will effect the almost sure regularity properties of functions drawn from μ_0 and, by absolute continuity of μ^y with respect to μ_0, of functions drawn from μ^y.

In the case that X is a subset of $H = L^2(D)$ with $D \subset \mathbb{R}^d$, the operator $\mathscr{C}$ may be identified with an integral operator:

$$\frac{1}{\lambda}(\mathscr{C}\phi)(x_1) = \int_D c(x_1, x_2)\phi(x_2)dx_2$$

for some kernel $c(x_1, x_2)$. The regularity of $c(x_1, x_2)$ determines the decay rate of σ_m [19]. If $\mathscr{C} = (-\Delta)^\alpha$ then the corresponding measure μ_0 has the property that samples are almost surely in the Sobolev space H^s and in the Hölder space C^s for all $s < \alpha - \frac{d}{2}$ (see [13] for more details). In particular, if $\alpha > d/2$, then $\mu(X) = 1$ when $X = L^\infty(D)$.

Priors which charge functions with a multiscale character can be built in this Gaussian context. One natural way to do this is to choose $\mathscr{M}$ as above so that it contains two distinct sets of functions varying on length scales of $\mathscr{O}(1)$ and $\mathscr{O}(\varepsilon)$ respectively. A second natural way is to choose a covariance function $c = c^\varepsilon$ which has two scales.

The formula (26) shows quite clearly how regularization works in the Bayesian context: the main contribution to the expectation will come from places where Φ is close to its minimum value and where μ_0 is concentrated; thus minimizing Φ is important, but this minimization is regularized through the properties of the measure μ_0. We now develop this intuitive concept further by linking the Bayesian approach to Tikhonov regularization and the functional I given by (23).

Given $z \in E$ and $\delta \ll 1$ define the small ball probability

$$J^\delta(z) = \mathbb{P}^{\mu^y}\big(\|u - z\|_X < \delta\big).$$

Note that this ball is in X but centred at a point $z \in E$, with E (the Cameron-Martin space) compact in X. It is natural to ask where $J^\delta(z)$ is maximized as a function of z and placing z in E allows us to answer this question. Furthermore we then see a connection between the Bayesian approach and the Tikhonov approach to regularization. The next theorem shows that small balls centred at minimizers of (23) will have maximal relative probability under the Bayesian posterior measure, in the small ball limit $\delta \to 0$.

Theorem 8. *([14]) Assume that $\mu_0(X) = 1$. Then*

$$\lim_{\delta \to 0} \frac{J^\delta(z_1)}{J^\delta(z_2)} = \exp\big(I(z_2) - I(z_1)\big).$$

In the Bayesian context the solution of the Tikhonov regularized problem is known as the Maximum A Posteriori estimator (MAP estimator) [7, 17].

4 Large Data Limits

In the previous section we showed how regularization plays a significant role in the solution of inverse problems. Choosing the correct regularization is part of the overall modelling scenario in which the inverse problem is embedded, as we demonstrated in the running example of Section 1.2. In some situations it may be suitable to look for the solution of the inverse problem over a small finite set of parameters,

whilst in others it may be desirable to look over a larger, even infinite dimensional set, in which oscillations are captured.

This section is devoted entirely to inverse problems where a single scalar parameter is sought and we study whether or not this parameter is correctly identified when a large amount of noisy data is available. The development is tied specifically to the running example, namely the PDE (3). For a fixed permeability coefficient generating the data, Fitzpatrick has also studied the consistency and asymptotic normality of maximum likelihood estimates in the large data limit [17]. Related work on parameter estimation in the context stochastic differential equations (SDEs) may be found in [25, 23].

4.1 The Statistical Model

We consider the problem of estimating a single scalar parameter $u \in \mathbb{R}$ in the elliptic PDE

$$
\begin{aligned}
\nabla \cdot v &= f, \quad x \in D, \\
p &= 0, \quad x \in \partial D, \\
v &= -\exp(u)A\nabla p
\end{aligned}
\tag{30}
$$

where $D \subset \mathbb{R}^d$ is bounded and open, and $f \in H^{-1}$ as well as the constant symmetric matrix A are assumed to be known. We let $G : \mathbb{R} \to H_0^1(D)$ be defined by $G(u) = p$. Then using the same linear functionals as in the running example from Section 1.2 we may construct the observation operator $\mathscr{G} : \mathbb{R} \to \mathbb{R}^N$ defined by $\mathscr{G}(u)_j = \ell_j(G(u))$. Our aim is to solve the inverse problem of determining u given y satisfying (1). For simplicity we assume that $\xi \sim N(0, \gamma^2 I)$ which implies that the observational noise on each linear functional is i.i.d. $N(0, \gamma^2)$. Since u is finite dimensional we will simply minimize Φ given by (18): no further regularization is needed because u is already finite dimensional.

Notice that the solution p of (30) is linear in $\exp(-u)$ and that we may write $G(u) = \exp(-u)p^\star$ where $p^\star$ solves

$$
\begin{aligned}
\nabla \cdot v &= f, \quad x \in D, \\
p^\star &= 0, \quad x \in \partial D. \\
v &= -A\nabla p^\star
\end{aligned}
\tag{31}
$$

Note that $\mathscr{G}(u)_j = \exp(-u)\ell_j(p^\star)$ so that the least squares functional (18) has the form

$$
\Phi(u) = \frac{1}{2\gamma^2} \sum_{j=1}^{N} |y_j - \mathscr{G}_j(u)|^2 = \frac{1}{2\gamma^2} \sum_{j=1}^{N} |y_j - \exp(-u)\ell_j(p^\star)|^2.
$$

It is straightforward to see that Φ has a unique minimizer $\overline{u}$ satisfying

$$\exp(-\overline{u}) = \frac{\sum_{j=1}^{N} y_j \ell_j(p^\star)}{\sum_{j=1}^{N} \ell_j(p^\star)^2}. \tag{32}$$

It is now natural to ask whether, for large N, the estimate $\overline{u}$ is close to the desired value of the parameter. We study two situations: the first where the data is generated by the model which is used to fit the data; and the second where the data is generated by a multiscale model whose homogenized limit gives the model which is used to fit the data.

4.2 Data From the Homogenized Model

We define $p_0 = \exp(-u_0)p^\star$ so that p_0 solves (30) with $u = u_0$.

Assumption 9 *We assume that the data y is generated from noisy observations generated by the statistical model:*

$$y_j = \ell_j(p_0) + \xi_j$$

where $\{\xi_j\}$ form an i.i.d. sequence of random variables distributed as $N(0, \gamma^2)$.

Theorem 10. *Let Assumptions 9 hold and assume that* $\liminf_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \ell_j(p^\star)^2 \geq L > 0$ *as $N \to \infty$. Then ξ-almost surely*

$$\lim_{N\to\infty} |\exp(-\overline{u}) - \exp(-u_0)| = 0.$$

Proof. Substituting the assumed expression for the data from Assumption 9 into the formula (32) gives

$$\exp(-\overline{u}) = \exp(-u_0) + I_1$$

where

$$I_1 = \frac{\frac{1}{N}\sum_{j=1}^{N} \xi_j \ell_j(p^\star)}{\frac{1}{N}\sum_{j=1}^{N} \ell_j(p^\star)^2}.$$

Therefore,

$$\mathbb{E}[I_1^2] = \frac{\gamma^2}{\sum_{j=1}^{N} \ell_j(p^*)^2} \leq \frac{2\gamma^2}{NL} \tag{33}$$

for N sufficiently large. Since I_1 is Gaussian we deduce that $\mathbb{E}I_1^{2p} = \mathcal{O}(N^{-p})$ as $N \to \infty$. Application of the Borel-Cantelli lemma shows that I_1 converges almost surely to zero as $N \to \infty$. $\square$

This shows that, in the large data limit, random observational error may be averaged out and the true value of the parameter recovered, in the idealized scenario where the data is taken from the statistical model used to identify the parameter. The condition that $L > 0$ prevents additional observation noise from overwhelming the

information obtained from additional measurements as $N \to \infty$. It is a simple explicit example of what is known as *posterior consistency* [8] in the theory of statistics.

4.3 Data From the Multiscale Model

In practice, of course, real data does not come from the statistical model used to estimate parameters. In order to probe the effect that this can have on posterior consistency we study the situation where the data is taken from a multiscale model whose homogenized limit falls within the class used in the statistical model to estimate parameters. Again we define $p_0 = \exp(-u_0)p^\star$ and we now define p^ε to solve (7) with K^ε chosen so that the homogenized coefficient associated with this family is $K_0 = \exp(u_0)A$.

Assumption 11 *We assume that the data y is generated from noisy observations of a multiscale model:*

$$y_j = \ell_j(p^\varepsilon) + \xi_j$$

with p^ε as above and the $\{\xi_j\}$ an i.i.d. sequence of random variables distributed as $N(0, \gamma^2)$.

Theorem 12. *Let Assumptions 11 hold and assume that that the linear functionals ℓ_j are chosen so that*

$$\lim_{\varepsilon \to 0} \limsup_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} |\ell_j(p^\varepsilon - p_0)|^2 = 0 \tag{34}$$

and $\liminf_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \ell_j(p^\star)^2 \geq L > 0$ as $N \to \infty$. Then $\xi-$ almost surely

$$\lim_{\varepsilon \to 0} \lim_{N \to \infty} |\exp(-\overline{u}) - \exp(-u_0)| = 0.$$

Proof. Notice that the solution of the homogenized equation is $p_0 = \exp(-u_0)p^\star$. We write

$$y_j = \ell_j(p_0) + \ell_j(p^\varepsilon - p_0) + \xi_j$$
$$= \exp(-u_0)\ell_j(p^\star) + \ell_j(p^\varepsilon - p_0) + \xi_j.$$

Substituting this into the formula (32) gives

$$\exp(-\overline{u}) = \exp(-u_0) + I_1 + I_2^\varepsilon$$

where I_1 is as defined in the proof of Theorem 10 and is independent of ε, and

$$I_2^\varepsilon = \frac{\sum_{j=1}^{N} \ell_j(p^\varepsilon - p_0)\ell_j(p^\star)}{\sum_{j=1}^{N} \ell_j(p^\star)^2}.$$

The Cauchy-Schwarz inequality gives

$$|I_2^\varepsilon| \leq \frac{\left(\sum_{j=1}^N |\ell_j(p^\varepsilon - p_0)|^2\right)^{1/2}}{\left(\sum_{j=1}^N \ell_j(p^\star)^2\right)^{1/2}} \leq \left(\frac{2}{NL}\sum_{j=1}^N |\ell_j(p^\varepsilon - p_0)|^2\right)^{1/2}$$

for N sufficiently large. As in the proof of Theorem 10 we have, ξ-almost surely,

$$\lim_{N\to 0} |\exp(-\overline{u}) - \exp(-u_0) - I_2^\varepsilon| = 0.$$

From this and (34) the desired result now follows.   □

The assumption (34) encodes the idea that, for small ε, the linear functionals used in the observation process return nearby values when applied to the solution p^ε of the multiscale model or to the solution p_0 of the homogenized equation. In particular, Corollary 1 implies that if $\{\ell_j(p)\}_{j=1}^\infty$ is a family of bounded linear functionals on $L^2(D)$, uniformly bounded in j, then (34) will hold. On the other hand, we may choose linear functionals that are bounded as functionals on $H^1(D)$ yet unbounded on $L^2(D)$. In this case Theorem 1 shows that (34) may not hold and the correct homogenized coefficient may not be recovered, even in the large data limit. An analogous phenomenon occurs in inference for SDEs where if the observations of a multiscale diffusion are too frequent (relative to the fast scale) then the correct homogenized coefficients are not recovered [25, 23].

5 Exploiting Multiscale Properties Within Inverse Estimation

In this section we describe how ideas from homogenization theory can be used to improve the estimation of parameters in homogenized models. We consider a regime where the unknown parameter has small-scale fluctuations that may be characterized as random. In this case, if we attempt to recover the homogenized parameter the error ξ appearing in (1) is affected by the model mismatch. This is because the simplified, low-dimensional parameter used to fit the data is different from the true unknown coefficient. So, even when there is no observational noise, the error ξ has a statistical structure. Nevertheless, homogenization theory predicts that this discrepancy between $G(u)$ and y associated with model mismatch will have a universal statistical structure which can be exploited in the inverse problem, as we now describe.

The specific ideas described here were developed by Nolen and Papanicolaou in [21] for one dimensional elliptic problems, including the groundwater flow problem that we study here. Bal and Ren [4] have employed similar ideas in the study of Sturm-Liouville problems with unknown potential. We begin by describing in Section 5.1 the homogenization and fluctuation theory for the case that the (scalar) permeability $k(x)$ is random. Then, in Section 5.2 we show how these ideas can be

used to develop an improved estimator for the homogenized permeability coefficient. We conclude with numerical results in Section 5.3.

5.1 The Model

In this section we will present the approach of [21] in the simplest possible setting. We consider the two-point boundary value problem

$$-\frac{d}{dx}\left(\exp(u(x))\frac{dp}{dx}\right) = f(x), \quad x \in [-1,1], \tag{35a}$$

$$p(-1) = p(1) = 0. \tag{35b}$$

This is, of course, (3) in the one-dimensional setting $d = 1$.

It is assumed that the coefficient $k(x) = \exp(u(x))$ is a single realization of a stationary, ergodic and mixing random field $k(x, \omega)$. Furthermore it is assumed that k^{-1} can be decomposed into a slowly varying non-random component, together with a random, rapidly oscillating component:

$$\frac{1}{k(x,\omega)} = \frac{1}{k_0(x)} + \sigma\mu\left(\frac{x}{\varepsilon}, \omega\right), \tag{36}$$

where $\mu(x, \omega)$ is a stationary, mean zero random field with covariance

$$R(x) = \mathbb{E}(\mu(x+y)\mu(y)).$$

We assume that $R(0) = 1$ and $\int_{\mathbb{R}} R(x)\,dx = 1$. Thus, σ^2 and ε are the (given) variance and correlation length of the fluctuations. We are interested in the case where $\varepsilon \ll 1$ so that the random fluctuations are rapid.

The solution $p = p_\varepsilon(x, \omega)$ of (35) depends on $\varepsilon > 0$ and on the realization of $k(x, \omega)$. However, in the limit as $\varepsilon \to 0$, p_ε coverges to $p_0(x)$ which is the solution of the homogenized Dirichlet problem

$$-\frac{d}{dx}\left(k_0(x)\frac{d}{dx}p_0\right) = f(x), \quad x \in [-1,1], \tag{37a}$$

$$p_0(-1) = p_0(1) = 0. \tag{37b}$$

Observe that the homogenized coefficient is the harmonic mean of k: $k_0(x) = \mathbb{E}[k^{-1}]^{-1}$. Moreover, in the limit as $\varepsilon \to 0$, the solution p_ε has Gaussian fluctuations about its asymptotic limit [3]. Specifically, one can prove that

$$\frac{p_\varepsilon(x,\omega) - p_0(x)}{\varepsilon^{1/2}} \to \sigma \int_D Q(x,y;k_0)v_0(y;k_0)\,dW_y(\omega) \tag{38}$$

in distribution as $\varepsilon \to 0$, where $W_y(\omega)$ is a Brownian random field, which is a Gaussian process. Here $v_0(x;k_0) = k_0(x)p_0(x)$, and the kernel $Q(x,y;k_0)$ is then related to the Green's function for the one dimensional system:

$$\begin{pmatrix} p_x \\ v_x \end{pmatrix} - \begin{pmatrix} 0 & 1/k_0(x) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} p \\ v \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}.$$

If the 2×2 Green's matrix for this system is $G(x,y;k_0) : D \times D \to \mathbb{R}^2 \otimes \mathbb{R}^2$, then $Q(x,y;k_0) = G_{1,1}(x,y;k_0)$. The important point here is that the integral

$$I(x,\omega) = \sigma \int_D Q(x,y;k_0)v_0(y;k_0)\,dW_y(\omega)$$

which appears on the right side of (38) is a centered Gaussian random variable with covariance

$$\mathbb{E}[I(x)I(z)] = \sigma^2 \int_D Q(x,y;k_0)v_0(y;k_0)^2 Q(y,z;k_0)\,dy.$$

This covariance depends on k_0. The asymptotic theory given by the limit theorem (38) gives us a good approximation of the statistics of $p_\varepsilon(x,\omega)$ even when there is no observation noise, and shows that the fluctuations depend on k_0. In this simple case presented here, Q can be computed explicitly. In other cases, it can be computed numerically; see [21] for more details.

5.2 Enhanced Estimation

We now show how this asymptotic theory can be used to enhance estimation of the homogenized parameter $k_0(x)$. The inverse problem is to identify the parameter $k_0(x)$ in the model

$$-\frac{d}{dx}\left(k_0(x)\frac{d}{dx}p_0\right) = f(x), \quad x \in [-1,1], \tag{39a}$$

$$p_0(-1) = p_0(1) = 0. \tag{39b}$$

We take the viewpoint that the data actually come from observations of $p_\varepsilon(x,\omega)$, which is the solution of the multiscale model (35) with $k(x,\omega)$ given by (36), so there is a discrepancy between the model used to fit the data and the true model which generates the data. Now the outstanding modelling issue is the choice of statistical model for the error ξ in (1).

Suppose we make noisy observations of $p_\varepsilon(x_j)$ at points $\{x_j\}_{j=1}^N$ distributed throughout the domain. Then the measurements are

$$y_j = p_\varepsilon(x_j,\omega) + \xi_j, \quad j = 1,\ldots,N$$

where $\xi_j \sim N(0, \gamma^2)$ are mutually independent, representing observation noise. The limit (38) we have just described tells us that for ε small, these measurements are approximated well by

$$y_j \approx p_0(x_j) + \xi_j',$$

where $\{\xi_j'\}_{j=1}^N$ are Gaussian random variables with mean zero and covariance

$$C_{j,\ell}(k_0, \varepsilon) = \mathbb{E}[\xi_j' \xi_\ell'] = \gamma^2 \delta_{j,\ell} + \varepsilon \sigma^2 \int_D Q(x_j, y; k_0) v_0(y; k_0)^2 Q(x_\ell, z; k_0) \, dy \quad (40)$$

Therefore, we model the observations as

$$y_j \approx \mathscr{G}(k_0) + \xi_j', \quad j = 1, \ldots, N$$

where $\mathscr{G}(k_0) = p_0(x_j; k_0)$ with p_0 being the solution of (39). The modified statistical error ξ' has two components. The first term $\gamma^2 \delta_{j,\ell}$ is due to observation error. The second term comes from the asymptotic theory and is associated with the random microstructure in the true parameter $k(x, \omega)$. Of course, if ε is very small, relative to γ^2, then the observation noise dominates (40). In this case, the observations of p_ε may be very close to observations of the homogenized solution p_0, and we might simply assume that $\xi' \sim N(0, \gamma^2 I)$, ignoring the error associated with the model mismatch. On the other hand, if γ^2 is small relative to ε then the statistical error ξ' is dominated by the model mismatch. In this case, homogenization theory gives us an asymptotic approximation of the true covariance structure of ξ', which is quite different from $N(0, \gamma^2 I)$. See [21] for a discussion of some properties of the covariance matrix $C(k_0, \varepsilon)$.

Using the covariance (40), we make the approximation

$$\mathbb{P}(y|k_0) \approx \frac{1}{\sqrt{2\pi |C(k_0; \varepsilon)|}} \exp\left(-\frac{1}{2}\big(y - \mathscr{G}(k_0)\big)^T C(k_0; \varepsilon)^{-1} \big(y - \mathscr{G}(k_0)\big)\right),$$

where $|\cdot|$ denotes the determinant. The parameter $k_0(x)$ is a function, in general, and we may place a Gaussian prior μ_0 on $u_0(x) = \log k_0(x)$. Application of Bayes' theorem (25) (with k_0 replacing u) gives that

$$\mathbb{P}(k_0|y) \propto \frac{1}{\sqrt{2\pi |C(k_0; \varepsilon)|}} \exp\left(-\frac{1}{2}\big(y - \mathscr{G}(k_0)\big)^T C(k_0; \varepsilon)^{-1} \big(y - \mathscr{G}(k_0)\big)\right) \mu_0(\log k_0)$$

where the constant of proportionality is independent of k_0. The maximum a posteriori estimator (MAP) is then found as the function $k_0(x)$ which maximizes $\mathbb{P}(k_0|y)$ which is the same as minimizing $I(k_0) = -\ln\big(\mathbb{P}(k_0|y)\big)$. The key contribution of homogenization theory is to correctly identify the noise structure which has covariance $C(k_0; \varepsilon)$ depending on $k_0(x)$, the parameter to be estimated.

5.3 Numerical Results

In this section we demonstrate the results of a numerical computation that show some advantage to using the homogenization theory as we have just described. Given noisy observations of $p_\varepsilon(x_j)$ we may compute the MAP estimator $\hat{k}_1$ using (41) with covariance $C(k_0; \varepsilon)$ given by (40):

$$\hat{k}_1 = \mathrm{argmax}_{k_0} \ \frac{1}{\sqrt{2\pi |C(k_0;\varepsilon)|}} \exp\left(-\frac{1}{2}(y-\mathscr{G}(k_0))^T C(k_0;\varepsilon)^{-1}(y-\mathscr{G}(k_0))\right)\mu_0(\log k_0),$$
(41)

On the other hand, we might ignore the effect of the random microstructure and simply use $C = \gamma^2 I$, accounting only for observation noise:

$$\hat{k}_2 = \mathrm{argmax}_{k_0} \ \frac{1}{\sqrt{2\pi |\gamma^2 I|}} \exp\left(-\frac{1}{2}\gamma^{-2}|y - \mathscr{G}(k_0)|^2\right)\mu_0(\log k_0). \qquad (42)$$

Both estimates $\hat{k}_1$ and $\hat{k}_2$ are random variables, depending on the random data observed, but we should hope that $\hat{k}_1$ gives us a better approximation of k_0, since it makes use of the true covariance (40). Indeed for simple linear statistical models, it is easy to see that an efficient estimator, which realizes the theoretically optimal variance given by the Cramér-Rao lower bound, may be obtained by using the true covariance of the data; however, using the incorrect covariance may lead to an estimate with significantly higher variance than the theoretical optimum. See [21] for more discussion of this point. The present setting is highly nonlinear and the variance of the estimates $\hat{k}_1$ and $\hat{k}_2$ cannot be computed explicitly, since $C(k_0, \varepsilon)$ depends on k_0 in a nonlinear way through solution of the PDE. Nevertheless the numerical results are consistent with the expectation that approximation of the true covariance (through homogenization theory) yields a MAP estimator that has smaller variance, relative to the estimate that makes no use of the homogenization theory (see Figure 3).

In Figure 1 we show one realization of the true coefficient $k(x, \omega)$ which was used to generate the data. The highly-oscillatory graph represents the true coefficient $k(x, \omega)$ with variation on many scales. The slowly-varying harmonic mean $k_0(x)$ also is displayed here as the thick curve; this function k_0 is what we attempt to estimate. The data was generated as follows. Using one realization of $k(x, \omega)$ and given forcing f, we solve the Dirichlet boundary value problem (35). The observation data involves point-wise evaluation of $p^\varepsilon(x_j)$ at points $\{x_j\}_{j=1}^N$ spaced uniformly across the domain, plus independent observation noise $N(0, \gamma^2)$ at each point of observation. Using this data, we compute estimates $\hat{k}_1$ and $\hat{k}_2$ by minimizing (41) and (42), respectively. For the computation shown here, the function $k_0(x)$ is parameterized by the first three coefficients in a Fourier series expansion. So, computing $\hat{k}_1$ and $\hat{k}_2$ involves an optimization in $\mathbb{R}^3$. To evaluate $\mathbb{P}(k_0|y)$ at each step in the minimization algorithm, we must solve the forward problem (39) with the current estimate of k_0, and in the case of $\hat{k}_1$ we must also compute $C(k_0, \varepsilon)$. See [21] for more details about this computation.
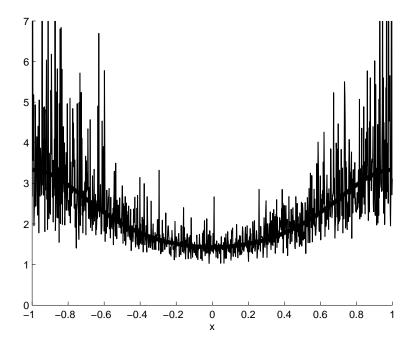
Fig. 1 The thin erratic curve is one realization of the true coefficient $k^\varepsilon(x, \omega)$. The thick curve is the slowly-varying harmonic mean $k_0(x)$. This realization was used to generate the data.

Figure 2 compares the estimate $\hat{k}_1(x)$ with the true function $k_0(x)$. Since the estimate $\hat{k}_1(x)$ is a random function, we performed the experiment many times (generating new $k(x, \omega)$ to compute each estimate $\hat{k}_1$) and display the results of 100 experiments. The data for $\hat{k}_2$ is qualitatively similar. Nevertheless, the pointwise variance $Var[\hat{k}_1(x)]$ is smaller than $Var[\hat{k}_2(x)]$, as shown in Figure 3. This is consistent with the linear estimation theory for which knowledge of the true data covariance yields an estimate with optimal variance.

Appendix 1

In this Appendix we prove Theorem 3 which, recall, applies in the case where (7b) and (13b) are replaced by periodic conditions on $D = (L\mathbb{T})^d$.
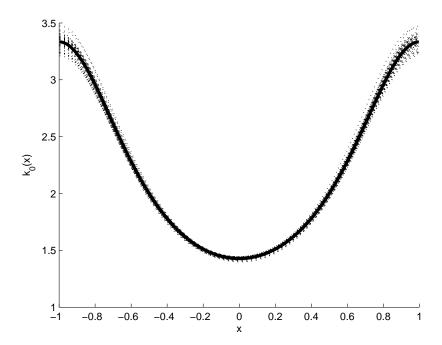
Fig. 2 The thick curve is the true k_0. The dashed series represent 100 independent realizations of the estimate $\hat{k}_1$.

Theorem 13. *Let $x^\varepsilon(t)$ and $x_0(t)$ be the solutions to equations* (8) *and* (17), *with velocity fields extended from $D = (L\mathbb{T})^d$ to $\mathbb{R}^d$ by periodicity, and assume that Assumption 2 holds. Assume also that $f \in C^\infty(D)$ and that $K(x,y) \in C^\infty(D;C^\infty_{\text{per}}(\mathbb{T}^d))$. Then*

$$\lim_{\varepsilon \to 0} \mathbb{E} \sup_{0 \le t \le T} \|x^\varepsilon(t) - x_0(t)\| = 0.$$

Proof. To simplify the notation we will set the porosity of the rock to be equal to 1, $\phi = 1$. Recall that $v^\varepsilon(x) = K^\varepsilon(x)\nabla p^\varepsilon(x)$. Our first observation is that, for $p_a^\varepsilon(x)$ given by (9),

$$K^\varepsilon(x)\nabla p^\varepsilon(x) = K^\varepsilon(x)\nabla p_a^\varepsilon(x) - \delta^\varepsilon(x) \tag{43}$$

where

$$\delta^\varepsilon(x) = -K^\varepsilon(x)\nabla\left(p^\varepsilon(x) - p_a^\varepsilon(x)\right). \tag{44}$$

From Assumption 2 we deduce that

$$\lim_{\varepsilon \to 0} \|\delta^\varepsilon(x)\|_{L^\infty} = 0.$$

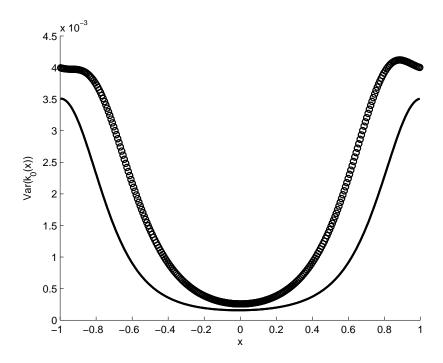From the definition of $p_a^\varepsilon(x)$ it follows that

Fig. 3 The upper series (o) is the empirical variance $Var[\hat{k}_2(x)]$. The lower series (-) is $Var[\hat{k}_1(x)]$. Both quantities were computed using 500 samples.

$$K^\varepsilon(x)\nabla p_a^\varepsilon(x) = Q^\varepsilon(x)\nabla p_0(x) - \varepsilon\delta_1^\varepsilon(x)$$

where

$$\delta_1^\varepsilon(x) = -K^\varepsilon(x)\nabla_x p_1(x, x/\varepsilon), \quad Q^\varepsilon(x) = Q(x, x/\varepsilon). \tag{45}$$

From the definition of p_1 in (14) we see that

$$\|\delta_1^\varepsilon(x)\|_{L^\infty} \leq C.$$

Putting (43) and (45) together we see that

$$v^\varepsilon(x) = -Q^\varepsilon(x)\nabla p_0(x) + \delta^\varepsilon(x) + \varepsilon\delta_1^\varepsilon(x)$$

and we see from (44) and (45) that the perturbations of $v^\varepsilon(x)$ from $Q^\varepsilon(x)\nabla p_0(x)$ are small; it is thus natural to expect a limit theorem for x^ε solving (8) which is Lagrangian transport in an appropriately averaged version of $Q^\varepsilon(x)\nabla p_0(x)$. Furthermore, since $Q(x,y)$ is divergence free in the fast y coordinate, by (15), it is natural to expect that the appropriate average is Lebesgue measure. We now demonstrate that this is indeed the case.

From (8) we deduce that

$$x^\varepsilon(t) = x(0) + \int_0^t \left(-Q^\varepsilon(x)\nabla p_0(x(s)) + \delta^\varepsilon(x(s)) + \varepsilon\delta_1^\varepsilon(x(s))\right) ds + \sqrt{2\eta_0\varepsilon}\, W(t).$$
$$\tag{46}$$

Define now $V(x,y) = -Q(x,y)\nabla p_0(x)$ and consider the system of SDEs

$$\frac{dx}{dt} = (V(x,y) + \delta^\varepsilon(x) + \varepsilon\delta_1^\varepsilon(x)) + \sqrt{2\eta_0\varepsilon}\,\frac{dW}{dt}, \tag{47a}$$

$$\frac{dy}{dt} = \frac{1}{\varepsilon}\left(V(x,y) + \delta^\varepsilon(x)\right) + \delta_1^\varepsilon(x) + \sqrt{\frac{2\eta_0}{\varepsilon}}\,\frac{dW}{dt}. \tag{47b}$$

Since $y = x/\varepsilon$ we see that $x(t)$, the solution of (47) is equal to $x^\varepsilon(t)$ appearing in (46).

The process $\{x(t), y(t)\}$ is Markov with generator

$$\begin{aligned}
\mathscr{L} &= \frac{1}{\varepsilon}\left(\left(V(x,y) + \delta^\varepsilon(x)\right)\cdot\nabla_y + \eta_0\Delta_y\right) \\
&\quad + \left(\left(V(x,y) + \delta^\varepsilon(x)\right)\cdot\nabla_x + \delta_1^\varepsilon(x)\cdot\nabla_y + \eta_0\nabla_x\cdot\nabla_y + \eta_0\nabla_y\cdot\nabla_x\right) \\
&\qquad + \varepsilon\eta_0\Delta_x + \varepsilon\delta_1^\varepsilon(x)\cdot\nabla_x \\
&=: \frac{1}{\varepsilon}\left(\mathscr{L}_0 + \delta^\varepsilon(x)\cdot\nabla_y\right) + \mathscr{L}_1 + \varepsilon\mathscr{L}_2.
\end{aligned}$$

Consider now the Poisson equation

$$-\mathscr{L}_0\Phi = V(x,y) - v_0(x) \tag{48}$$

with (see (13)(c))

$$v_0(x) = \int_{\mathbb{T}^d} V(x,y)\,dy.$$

Equation (48) is posed on $\mathbb{T}^d$ with periodic boundary conditions. Notice that x enters merely as a parameter in this equation. The operator $\mathscr{L}_0$ is uniformly elliptic on $\mathbb{T}^d$ and the right hand side averages to 0, hence, by Fredholm's alternative this equation has a solution which is unique, up to constants. We fix this constant by requiring that $\int_{\mathbb{T}^d}\Phi(x,y)\,dy = 0$. We define $\Phi^\varepsilon(x) := \Phi(x,x/\varepsilon)$ and similarly for $\mathscr{L}_i\Phi^\varepsilon(x)$. Applying Itô's formula to Φ and evaluating at $y = x/\varepsilon$ we obtain

$$\begin{aligned}
d\Phi^\varepsilon(x) &= \frac{1}{\varepsilon}\left(\mathscr{L}_0\Phi^\varepsilon + \delta^\varepsilon(x)\cdot\nabla_y\Phi(x,x/\varepsilon)\right) dt + \mathscr{L}_1\Phi^\varepsilon\, dt + \varepsilon\mathscr{L}_2\Phi^\varepsilon\, dt \\
&\quad + \sqrt{\frac{2\eta_0}{\varepsilon}}\nabla_y\Phi^\varepsilon\, dW + \sqrt{2\eta_0\varepsilon}\nabla_x\Phi^\varepsilon\, dW \\
&= -\frac{1}{\varepsilon}\left(V(x,x/\varepsilon) - v_0(x) + \delta^\varepsilon(x)\cdot\nabla_y\Phi(x,x/\varepsilon)\right) dt + \mathscr{L}_1\Phi^\varepsilon + \varepsilon\mathscr{L}_2\Phi^\varepsilon\, dt \\
&\quad + \sqrt{\frac{2\eta_0}{\varepsilon}}\nabla_y\Phi^\varepsilon\, dW + \sqrt{2\eta_0\varepsilon}\nabla_x\Phi^\varepsilon\, dW.
\end{aligned}$$

Consequently,

$$\int_0^t V\big(x(s),y(s)\big)\,ds - \int_0^t v_0(x(s))\,ds$$

$$= \int_0^t \Big(\delta^\varepsilon(x(s))\cdot\nabla_y\Phi(x(s),x(s)/\varepsilon) + \varepsilon\mathscr{L}_1\Phi^\varepsilon(x(s)) + \varepsilon^2\mathscr{L}_2\Phi^\varepsilon(x(s))\Big)\,ds$$

$$- \varepsilon\Big(\Phi^\varepsilon(x^\varepsilon(t)) - \Phi^\varepsilon(x^\varepsilon(0))\Big) + \sqrt{\varepsilon}M^\varepsilon(t),$$

where

$$M^\varepsilon(t) := \int_0^t \Big(\sqrt{2\eta_0}\nabla_y\Phi^\varepsilon + \varepsilon\sqrt{2\eta_0}\nabla_x\Phi^\varepsilon\Big)\,dW.$$

Since $\Phi(x,y)$ is periodic in both coordinates we have that

$$\|\nabla_y\Phi(x,x/\varepsilon)\|_{L^\infty} \le C, \quad \|\Phi^\varepsilon(x)\|_{L^\infty} \le C, \quad \|\mathscr{L}_1\Phi^\varepsilon\|_{L^\infty} \le C, \quad \|\mathscr{L}_1\Phi^\varepsilon\|_{L^\infty} \le C$$

and

$$\mathbb{E}\|M^\varepsilon(t)\|^p \le C, \quad p \ge 1. \tag{49}$$

We combine the above calculations to obtain

$$x^\varepsilon(t) = x(0) + \int_0^t v_0(x^\varepsilon(s))\,ds + H^\varepsilon(t) + \sqrt{\varepsilon}\tilde{M}^\varepsilon(t),$$

where

$$H^\varepsilon(t) := -\varepsilon\Big(\Phi^\varepsilon(x^\varepsilon(t)) - \Phi^\varepsilon(x^\varepsilon(0))\Big) + \int_0^t \big(\delta^\varepsilon(x^\varepsilon(s)) + \varepsilon\delta_1^\varepsilon(x^\varepsilon(s))\big)\,ds$$

$$+ \int_0^t \Big(\delta^\varepsilon(x(s))\cdot\nabla_y\Phi(x(s),x(s)/\varepsilon) + \varepsilon\mathscr{L}_1\Phi^\varepsilon(x(s)) + \varepsilon^2\mathscr{L}_2\Phi^\varepsilon(x(s))\Big)\,ds$$

and

$$\tilde{M}^\varepsilon(t) = M^\varepsilon(t) + \sqrt{2\eta_0}W(t).$$

Our estimates imply that

$$\lim_{\varepsilon\to0}\mathbb{E}\sup_{t\in[0,T]}|H^\varepsilon(t)| = 0.$$

Furthermore, estimate (49), together with the Burkhölder-Davis-Gundy inequality imply that

$$\mathbb{E}\sup_{t\in[0,T]}|\tilde{M}^\varepsilon(t)| \le C.$$

On the other hand,

$$x(t) = x(0) + \int_0^t v_0(x(s))\,ds.$$

Set $\theta(T) := \mathbb{E}\sup_{t\in[0,T]}|x^\varepsilon(t) - x(t)|$. Because v_0 is periodic it is in fact globally Lipschitz so that we obtain

$$\theta(T) \le C\int_0^T \theta(t)\,dt + h^\varepsilon(T),$$

where

$$\lim_{\varepsilon \to 0} h^{\varepsilon}(T) = 0.$$

We use Gronwall's inequality to deduce

$$\theta(T) \le h^{\varepsilon}\left(1 + CTe^{CT}\right),$$

from which the claim follows.   □

Appendix 2

In this appendix we study the homogenization problem (7) in one dimension. In this case we can calculate the homogenized coefficient explicitly and to prove Assumption 2. More details can be found in [24, Ch. 12].

The Homogenized Equations

We take $d = 1$ in (7) and set $D = [0,L]$. Then the Dirichlet problem (7) reduces to a two–point boundary value problem:

$$-\frac{d}{dx}\left(\exp\left(u\left(x,\frac{x}{\varepsilon}\right)\right)\frac{dp^{\varepsilon}}{dx}\right) = f \quad \text{for} x \in (0,L), \tag{50a}$$

$$p^{\varepsilon}(0) = p^{\varepsilon}(L) = 0. \tag{50b}$$

We assume that $u(x,y)$ is smooth in both of its arguments and periodic in y with period 1. Furthermore, we assume that this function is bounded from above and below. Consequently, there exist constants $0 < \alpha \le \beta < \infty$ such that

$$\alpha \le \exp(u(x,y)) \le \beta, \quad \forall y \in [0,1]. \tag{51}$$

We also assume that f is smooth.

The cell problem becomes a boundary value problem for an ordinary differential equation with periodic boundary conditions. Introducing the notation $k(x,y) := \exp(u(x,y))$, the cell problem can be written as

$$-\frac{\partial}{\partial y}\left(k(x,y)\frac{\partial \chi}{\partial y}\right) = \frac{\partial k(x,y)}{\partial y}, \quad \text{for } y \in (0,1), \tag{52a}$$

$$\chi \text{ is 1–periodic}, \quad \int_0^1 \chi(x,y)\,dy = 0. \tag{52b}$$

Notice that the macrovariable x enters the cell problem (52) as a parameter. Since $d = 1$ we only have one effective coefficient which is given by the one dimensional

version of (11),(12), namely

$$k_0(x) = \int_0^1 \left(k(x,y) + k(x,y)\frac{\partial \chi}{\partial y}(x,y) \right) dy$$

$$= \left\langle k(x,y)\left(1 + \frac{\partial \chi}{\partial y}(x,y) \right) \right\rangle \tag{53}$$

where we have introduced the notation $\langle \phi(x,y) \rangle := \int_0^1 \phi(x,y)\,dy$. The homogenized equation is then

$$-\frac{d}{dx}\left(k_0(x)\frac{dp_0}{dx} \right) = f, \quad x \in (0,L), \tag{54a}$$

$$p(0) = p(L) = 0. \tag{54b}$$

Explicit Solution of the Cell Problem

Equation (52a) can be solved exactly. After integrating the equation and applying the periodic boundary conditions, we obtain

$$\chi(x,y) = -y + c_1 \int_0^y \frac{1}{k(x,y)}\,dy + c_2,$$

with

$$c_1(x) = \frac{1}{\int_0^1 \frac{1}{k(x,y)}\,dy} = \langle k(x,y)^{-1} \rangle^{-1}.$$

Therefore, from (53) we obtain:

$$k_0(x) = \langle k(x,y)^{-1} \rangle^{-1}. \tag{55}$$

The constant c_2 is irrelevant. This is the formula which gives the homogenized coefficient in one dimension. It shows clearly that, even in this simple one–dimensional setting, the homogenized coefficient is not found by simply averaging the unhomogenized coefficients over a period of the microstructure. Rather, the homogenized coefficient is the *harmonic average* of the unhomogenized coefficient. It is quite easy to show that $k_0(x)$ is bounded from above by the average of $k(x,y)$. Notice that the homogenized coefficient can be written in the form

$$k_0(x) = e^{u_0(x)}, \quad \text{where} \quad u_0(x) = \log\left(\langle \exp(-u(x,y)) \rangle^{-1} \right). \tag{56}$$

Error Estimates in $W^{1,\infty}$

The fact that we can obtain an explicit formula for the solution of the boundary value problem (50) as well as for the solution of the cell problem (52) enables us to prove Assumption 2.

Proposition 14 *Let $p^\varepsilon(x)$ be the solution of the two-point boundary value problem (50) where the log permeability $u(x,y)$ is smooth in both of its arguments and satisfies (51). Let $k(x,y) = \exp(u(x,y))$ and define*

$$v^\varepsilon(x) = k\left(x, \frac{x}{\varepsilon}\right)\frac{dp^\varepsilon}{dx}(x)$$

and

$$V(x,y) = k(x,y)\left(1 + \frac{\partial \chi}{\partial y}(x,y)\right)\frac{dp_0}{dx}(x),$$

where $p_0(x)$ is the solution of the homogenized equation (54). Then

$$\lim_{\varepsilon \to 0}\|v^\varepsilon(x) - V(x, x/\varepsilon)\|_{L^\infty} = 0. \tag{57}$$

Notice that, by (14), the corrector $p_1(x,y) = \chi(x,y)\frac{dp_0}{dx}(x)$. Hence, using the bound (51) from below on a, together with the definition (9) of p_a^ε, this theorem delivers the following immediate corollary:

Corollary 2. *Under the assumptions of Proposition 14 we have*

$$\lim_{\varepsilon \to 0}\|p^\varepsilon - p_a^\varepsilon\|_{W^{1,\infty}} = 0.$$

Proof of Proposition 14. We have that

$$\frac{d\chi}{dy}(x,y) = -1 + \frac{k_0(x)}{k(x,y)}.$$

Consequently

$$V(x,y) = k_0(x)\frac{dp_0}{dx}(x).$$

Define a function F by $F'(z) = f(z)$. We solve the homogenized equation to obtain

$$k_0(x)\frac{dp_0}{dx}(x) = -F(x) + c,$$

with

$$c = \frac{\int_0^L k_0^{-1}(z)F(z)\,dz}{\int_0^L k_0^{-1}(z)\,dz}.$$

Similarly, from (50) we obtain

$$k\left(x,\frac{x}{\varepsilon}\right)\frac{dp^{\varepsilon}}{dx} = -F(x) + c^{\varepsilon},$$

with

$$c^{\varepsilon} = \frac{\int_0^L k^{-1}(z,z/\varepsilon)F(z)\,dz}{\int_0^L k^{-1}(z,z/\varepsilon)\,dz}.$$

From the above calculations we deduce that

$$\|v^{\varepsilon}(x) - V(x,x/\varepsilon)\|_{L^{\infty}} = |c - c^{\varepsilon}|.$$

It suffices to show that $|c - c^{\varepsilon}| = \mathcal{O}(\varepsilon)$. This will follow from the fact that

$$\int_0^L k^{-1}(z,z/\varepsilon)G(z) = \int_0^L k_0^{-1}(z)G(z)dz + \mathcal{O}(\varepsilon)$$

for any smooth function G, as $\varepsilon \to 0$. To see this, define integer N and $\delta \in [0,\varepsilon)$ uniquely by the identity

$$L = N\varepsilon + \delta. \tag{58}$$

Then note that, using the uniform bounds on $k(x,y)$ from below, together with uniform (in y) Lipschitz properties of $a(\cdot,y)$ and G, we have for $z_n = n\varepsilon$,

$$\begin{aligned}
\int_0^L k^{-1}(z,z/\varepsilon)G(z)dz &= \sum_{n=0}^{N-1}\int_{n\varepsilon}^{(n+1)\varepsilon} k^{-1}(z_n,z/\varepsilon)G(z_n)dz + \mathcal{O}(\varepsilon) \\
&= \sum_{n=0}^{N-1}\int_{n\varepsilon}^{(n+1)\varepsilon} k_0^{-1}(z_n)G(z_n)dz + \mathcal{O}(\varepsilon) \\
&= \sum_{n=0}^{N-1}\int_{n\varepsilon}^{(n+1)\varepsilon} k_0^{-1}(z)G(z)dz + \mathcal{O}(\varepsilon) \\
&= \int_0^L k_0^{-1}(z)G(z)dz + \mathcal{O}(\varepsilon).
\end{aligned}$$

This completes the proof. □

References

1. G. Allaire. Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23(6):1482–1518, 1992.
2. M. Avellaneda and F.-H. Lin. Compactness methods in the theory of homogenization. *Comm. Pure Appl. Math.*, 40(6):803–847, 1987.
3. G. Bal, J. Garnier, S. Motsch, and V. Perrier. Random integrals and correctors in homogenization. *Asymptotic Analysis*, 59:1–26, 2008.
4. G. Bal and K. Ren. Physics-based models for measurement correlations: application to an inverse sturm-liouville problem. *Inverse Problems*, 25:055006, 2009.
5. H.T. Banks and K. Kunisch. *Estimation Techniqiues for Distributed Parameter Systems*. Birkhäuser, 1989.

6. A. Bensoussan, J.-L. Lions, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structures*, volume 5 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1978.
7. J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 1980.
8. P.J. Bickel and K.A. Doksum. *Mathematical Statistics*. Prentice-Hall, 2001.
9. V.I. Bogachev. *Gaussian Meausures*. American Mathematical Society, 1998.
10. D. Cioranescu and P. Donato. *An Introduction to Homogenization*. Oxford University Press, New York, 1999.
11. S.L. Cotter, M. Dashti, J.C. Robinson, and A.M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25:115008, 2009.
12. B. Dacarogna. *Direct Methods in the Calculus of Variations*. Springer, New York, 1989.
13. G. DaPrato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions.* Cambridge University Press, 1992.
14. M. Dashti and A.M. Stuart. *In preparation*, 2010.
15. Y. R. Efendiev, Th. Y. Hou, and X.-H. Wu. Convergence of a nonconforming multiscale finite element method. *SIAM J. Numer. Anal.*, 37(3):888–910 (electronic), 2000.
16. H.K. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer, 1996.
17. B. Fitzpatrick. Bayesian analysis in inverse problems. *Inverse Problems*, 7:675–702, 1991.
18. M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbirch. *Optimization with PDE Constraints*, volume 23 of *Mathematics Modelling: Theory and Applications*. Springer, 2009.
19. H. König. *Eigenvalue Distribution of Compact Operators.* Birkhäuser Verlag, Basel, 1986.
20. M.A. Lifshits. *Gaussian Random Functions*, volume 322 of *Mathematics and its Applications*. Kluwer, Dordrecht, 1995.
21. J. Nolen and G. Papanicoulaou. Inverse problems. *Fine scale unertainty in parameter estimation for elliptic equations*, 25:115021, 2009.
22. J. Nolen, G.A. Pavliotis, and A.M. Stuart. *In preparation*, 2010.
23. A. Papavasiliou, G.A. Pavliotis, and A.M Stuart. Maximum likelihood estimation for multiscale diffusions. *Stoch. Proc. and Applics.*, 2009.
24. G. A. Pavliotis and A. M. Stuart. *Multiscale Methods*, volume 53 of *Texts in Applied Mathematics*. Springer, 2008. Averaging and Homogenization.
25. G.A. Pavliotis and A.M Stuart. Parameter estimation for multiscale diffusions. *J. Stat. Phys.*, 127(4):741–781, 2007.
26. A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19, 2010.