Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

JON COCKAYNE, CHRIS J. OATES, TIM SULLIVAN, MARK GIROLAMI

Probabilistic Meshless Methods for Partial Differential Equations and Bayesian Inverse Problems

Jon Cockayne, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, United Kingdom. j.cockayne@warwick.ac.uk
Chris J. Oates, University of Technology Sydney, 638 Jones Street, Ultimo, NSW 2007, Australia. christopher.oates@uts.edu.au
Tim Sullivan, Free University of Berlin and Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany. sullivan@zib.de
Mark Girolami, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, United Kingdom. m.girolami@warwick.ac.uk

ZIB Report 16-31 (Mai 2016)

Zuse Institute Berlin
Takustr. 7
D-14195 Berlin

Telefon: +49 30-84185-0
Telefax: +49 30-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Probabilistic Meshless Methods for Partial Differential Equations and Bayesian Inverse Problems

Jon Cockayne[1], Chris J. Oates[2,3], Tim Sullivan[4] and Mark Girolami[1,5]
[1]University of Warwick, UK
[2]University of Technology Sydney, Australia
[3]ARC Centre of Excellence for Mathematical and Statistical Frontiers
[4]Free University of Berlin and Zuse Institute Berlin, Germany
and [5]The Alan Turing Institute for Data Science

25. Mai 2016

Zusammenfassung

This paper develops a class of meshless methods that are well-suited to statistical inverse problems involving partial differential equations (PDEs). The methods discussed in this paper view the forcing term in the PDE as a random field that induces a probability distribution over the residual error of a symmetric collocation method. This construction enables the solution of challenging inverse problems while accounting, in a rigorous way, for the impact of the discretisation of the forward problem. In particular, this confers robustness to failure of meshless methods, with statistical inferences driven to be more conservative in the presence of significant solver error. In addition, (i) a principled learning-theoretic approach to minimise the impact of solver error is developed, and (ii) the challenging setting of inverse problems with a non-linear forward model is considered. The method is applied to parameter inference problems in which non-negligible solver error must be accounted for in order to draw valid statistical conclusions.

1 Introduction

1.1 Inverse Problems and Numerical Computation

Scientific inquiry is driven by mathematical models of physical phenomena. Typically these models involve finite- or infinite-dimensional unknown parameters. In order to make use of these models for prediction one must first estimate unknown parameters on the basis of data. The mathematical literature refers to such problems as "inverse problems". Several important challenges are raised by the increasing sophistication of mathematical models built to describe physical phenomena. In particular, numerical computation associated with inverse problems typically requires repeated simulation from the "forward" problem, for different values of the unknown parameter.

"Meshless" methods [Fasshauer, 1997, Franke and Schaback, 1998, Hon and Schaback, 2008] are characterised by their not relying upon construction of a mesh over the solution domain. There is extensive empirical evidence in support of meshless methods in the solution of PDEs in situations with strong boundary effects, or when the domain itself is time-dependent, and especially when the evolutionary rates are large [Li et al., 2010, 2012]. Indeed, meshless methods have proven successful in settings where application of better-understood approaches based on finite element analysis (FEA) is challenging; for example, in ballistics problems the domain is time-dependent, so that re-meshing is required at regular time intervals, owing to the rapid deformation of the material at point of impact. The cost of such remeshing is noted in Belytschko et al. [2001]; this is not only a compuational burden at run-time, but also in post-processing and algorithmic complexity.

Compared to FEA, theoretical analysis of meshless methods has been relatively limited, with main contributions including Behrens and Iske [2002], Lorentz et al. [2003], Wendland [2005], Cialenco et al. [2012], and Chi et al. [2013]. Behrens and Iske [2002], Chi et al. [2013], and Lorentz et al. [2003] each focus on application of meshless methods in specific problems; elastodynamics in Chi et al. [2013] and

time-evolving problems in Behrens and Iske [2002] and Lorentz et al. [2003]. In particular there has been little investigation of the suitability of meshless methods in the context of inverse problems.

1.2 Contribution

The main contribution of this paper is to propose a rigorous mathematical and statistical framework for the solution of inverse problems based on meshless collocation with radial basis functions [Fasshauer, 1997]. To our knowledge, the only investigation of these methods in the context of inverse problems was Fasshauer and Ye [2013], though recently an extensive analysis of the forward problem was provided by Owhadi [2016].

Central to this development is a deep connection between collocation and stochastic PDEs that casts the latter as a "stochastic relaxation" of the former [Cialenco et al., 2012]. This permits the residual error of the meshless solution to be quantified by a distribution over an appropriate function space. The use of a distribution contrasts with conventional asymptotic estimates for numerical error [see the survey in Verfürth, 2013]. The approach pursued here is rooted in the Bayesian setting and provides much needed reassurances for solutions of inverse problems that are based on meshless methods. In the Bayesian approach to inverse problems, a "prior" distribution is specified over all unknowns and propagated, via Bayes' rule, to a "posterior" distribution over variables that accounts for information contained in data. Further details on the Bayesian perspective to inverse problems can be found in Kaipio and Somersalo [2006], Stuart [2010], Sullivan [2015]. For the use of such methods in Bayesian inverse problems, a distributional quantification of error in the forward problem can be propagated into the posterior distribution of the inverse problem. A complementary, game-theoretic motivation for the probabilistic representation of solver error comes from the recent work of Owhadi [2016].

To realise this approach, in Sec. 3 we construct probabilistic analogues of PDE solvers that operate in infinite-dimensional Hilbert spaces. This enables uncertainty due to discretisation error to be reflected directly in subsequent statistical inferences. The "data-likelihood" here requires evaluation of the solution of the PDE only at locations where observations of the physical system of interest are obtained. By aiming to reduce the variance in the solution at these points, one can reduce the computational effort required to achieve a desired level of accuracy. This leads us to formulate the choice of basis functions as a problem in optimal experimental design [Pukelsheim, 2006]. A computational approach is developed to obtain optimal configurations, based on local Gaussian process approximation, the approximate co-ordinate exchange (ACE) algorithm of Overstall et al. [2015] and Bayesian Optimisation [Mockus et al., 1978].

In Sec. 7 two detailed empirical studies are provided. Firstly, the method is applied to solve an inverse problem arising in electrical impedance tomography. Secondly, an application is presented to an inverse problem with a non-linear forward model based on the steady-state Allen-Cahn equation. The latter is particularly challenging, since for certain values of the parameter the forward problem has multiple solutions. A fully Bayesian solution is presented for this problem and shown to perform well empirically. This work therefore provides a route to tackle inverse problems with a non-linear forward model, a setting rarely examined in the existing literature, and offers the possibility to consider physical phenomena of contemporary interest that are modelled by non-linear differential equations.

1.3 On the Role of Probability in Deterministic Computation

This work falls within the wider context of a movement toward Probabilistic Numerics (PN), in which numerical tasks are considered as inference tasks that are amenable to statistical analysis. Diaconis [1988] traces the foundations of PN back to Poincaré, with other landmark and foundational papers including Hull and Swenson [1966], Kadane [1985]. Recent developments include probabilistic solvers for linear systems [Hennig, 2015], differential equations [Calderhead et al., 2009, Dondelinger et al., 2013, Schober et al., 2014, Barber and Wang, 2014, Conrad et al., 2015, Macdonald et al., 2015, Owhadi, 2016, Kersting and Hennig, 2016] and probabilistic integrators [O'Hagan, 1991, Briol et al., 2016]. A recent perspective on PN is provided by Hennig et al. [2015].

A PN method returns not just an approximate solution, but a measure or samples from a measure over the space of solutions that quantifies epistemic uncertainty resulting from inability to perform infinitely many numerical calculations. The mathematical foundations of this approach are established in the Information Complexity literature [Woźniakowski, 2009]. This field studies the approximation of continuous mathematical objects under the constraints of finite digital hardware. In such situations

information is by its nature partial and potentially contaminated. This is a setting which we find ourselves in for PDE inverse problems in all but the most trivial cases; we wish to obtain an approximation to the infinite-dimensional solution of the PDE but can only use a finite-dimensional representation to do so. Furthermore our observations of the system of interest will typically be contaminated with noise and suffer from model-mismatch error.

The PN approach to numerical solution of PDEs initially concerned models for rounding error [Hull and Swenson, 1966] and indeed this is still of interest [Mosbach and Turner, 2009, Hairer et al., 2008]. However, the aim here is to model discretisation error [e.g. following Skilling, 1992]. Of particular relevance is recent work by Conrad et al. [2015], which proposed PN methods for numerical solution of PDEs by FEA, in addition to differential equations. There a principled construction was proposed for probability measures over numerical error, with this additional source of uncertainty propagated into inferences made on the inverse problem. This provides several advantages: Firstly, the inferential procedure adapts to reflect the impact of numerical errors, automatically inflating posterior variance in settings where numerical error is likely to be non-negligible. As such, statistically valid inferences can be made in the presence of non-negligible numerical error; the posterior mode might still differ from the truth, but our confidence in the result is reduced as a result of accounting for this error. This could further allow crude, inexpensive solvers can be employed in situations where a loss in posterior precision can be tolerated, reducing the overall computational burden. Secondly, often the inferences drawn on PDE parameters form the basis of further computation, for example to make a prediction on the future behaviour of the physical system. In such settings PN methods enable numerical uncertainty to be propagated through subsequent computations, avoiding over-confidence due to successive discretisations. Going further, the impact of numerical error incurred at each stage in a computational pipeline can be explored using statistical techniques, such as analysis of variance, in order to better target computational resources. The focus of the present paper on meshless methods is distinct from Conrad et al. [2015] but, as explained below, the PN approach delivers these same advantages.

1.4 Outline

The paper proceeds as follows: Sec. 2 establishes the set-up and notation, while Sec. 3 outlines the proposed "probabilistic" meshless method. Secs. 4 and 5 provide error analysis for, respectively, the forward and inverse problems. Computational considerations are discussed in Sec. 6. Then, in Sec. 6.3 the approach is extended to a class of non-linear PDEs. Finally Sec. 7 provides empirical results on the proposed approach, with discussion reserved for Sec. 8.

2 Context and Motivation

Sec. 2.1 formalises the notion of a Bayesian inverse problem. Sec. 2.2 then expands on the statistical motivation for valid inferences in the presence of numerical error.

2.1 Inverse Problems

The physical models considered here are defined via operator equations. An "inverse problem" arises when some of these operators depend upon unknown parameters that must be inferred. This can enable future predictions to be obtained under the model, or provide insight into the physical systems of interest. The Bayesian approach to inverse problems treats unknown parameters as random variables. The first step below is to set up a mathematical framework that makes both the parameter dependence and the randomness explicit.

2.1.1 Set-up and Notation

Consider a regular open, bounded subset $D \subset \mathbb{R}^d$ with boundary ∂D. Here *regular* means that D has a Lipschitz boundary, for Sobolev embedding. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and consider (Borel) measurable operators $\mathcal{A} : H(D) \times \Omega \to H_{\mathcal{A}}(D)$ and $\mathcal{B} : H(D) \times \Omega \to H_{\mathcal{B}}(\partial D)$ such that $H(D)$, $H_{\mathcal{A}}(D)$ and $H_{\mathcal{B}}(\partial D)$ are Hilbert spaces of functions respectively on D and ∂D. Consider the stochastic solution

$u(\cdot, \omega) \in H(D)$, $\omega \in \Omega$, of operator equations of the form

$$\begin{aligned} \mathcal{A}[\omega]u(\boldsymbol{x}, \omega) &= g(\boldsymbol{x}) \quad \boldsymbol{x} \in D \\ \mathcal{B}[\omega]u(\boldsymbol{x}, \omega) &= b(\boldsymbol{x}) \quad \boldsymbol{x} \in \partial D \end{aligned} \tag{1}$$

where $g \in H_{\mathcal{A}}(D)$ and $b \in H_{\mathcal{B}}(\partial D)$. To emphasise the random nature of the operators $\mathcal{A}$ and $\mathcal{B}$, the notation $\mathcal{A}[\omega]$ and $\mathcal{B}[\omega]$, $\omega \in \Omega$ was used. For concreteness, one can associate $\mathcal{A}$ with a PDE to be solved and $\mathcal{B}$ with any initial or boundary conditions. Similarly, $g \in H_{\mathcal{A}}(D)$ and $b \in H_{\mathcal{B}}(\partial D)$ can be considered as forcing and boundary terms for the PDE. For notational simplicity we will generally restrict attention to systems with two operators such as shown above, however it is trivial to extend the algorithm. of this paper to systems of more than two operators, each potentially restricted to other areas of the domain than the interior and boundary.

An inverse problem is one in which inferences are to be made for ω, on the basis of possibly noisy observations of the underlying solution $u(\cdot, \omega_0)$, where $\omega_0 \in \Omega$ denotes the "true" value of ω. Typically, in the infinite dimensional setting, both $\mathcal{A}$ and $\mathcal{B}$ depend on ω via a measurable projection (i.e. a random variable) $\theta(\omega)$ that is of direct physical interest, where $\theta : \Omega \to \Theta$ is a measurable function mapping into a separable Banach space Θ with the Borel σ-algebra $\mathfrak{B}(\Theta)$. The true value of θ, written $\theta_0 = \theta(\omega_0)$, is the object of statistical interest.

2.1.2 The Bayesian Approach

The measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ implies a push-forward measure Π_θ on $(\Theta, \mathfrak{B}(\Theta))$ that is known as the "prior" in Bayesian statistics. Data, in this paper, is used in the statistical sense and refers to a random variable $\boldsymbol{y}$ defined on $(\mathcal{Y}, \mathfrak{B}(\mathcal{Y}))$, where $\mathcal{Y}$ is a separable Banach space equipped with the Borel σ-algebra $\mathfrak{B}(\mathcal{Y})$. In this paper $\mathcal{Y} \subseteq \mathbb{R}^n$ and data arise from a distribution $\Pi_{\boldsymbol{y}}$. The Lebesgue-measurable conditional density

$$\pi(\boldsymbol{y}|\theta) \propto \exp(-\Phi(\boldsymbol{y}, \theta)), \tag{2}$$

is called the "data-likelihood", where $\Phi : \mathcal{Y} \times \Theta \to \mathbb{R}$ is a (measurable) "potential" function. Then an infinite-dimensional analogue of Bayes' theorem [Theorem 1.1 of Dashti and Stuart, 2015] implies the existence of a "posterior" distribution $\Pi_\theta^{\boldsymbol{y}}$ on $(\Theta, \mathfrak{B}(\Theta))$ that is absolutely continuous with respect to Π_θ, with Radon–Nikodým derivative

$$\frac{\mathrm{d}\Pi_\theta^{\boldsymbol{y}}}{\mathrm{d}\Pi_\theta}(\theta) = \frac{1}{Z}\exp(-\Phi(\boldsymbol{y}, \theta)), \quad Z = \int_\Theta \exp(-\Phi(\boldsymbol{y}, \theta))\Pi_\theta(\mathrm{d}\theta) \tag{3}$$

whenever $Z > 0$. In colloquial use, the "Bayesian inverse problem" entails numerical computation of (derived quantities of) the posterior distribution $\Pi_\theta^{\boldsymbol{y}}$; these are called "inferences".

2.2 Statistical Motivation: Valid Inference for PDE Models

Below we expand on the motivation for developing a more expressive quantification of numerical error in the forward solution, within the context of statistical inverse problems. For illustration, consider the Gaussian measurement error model, described by the potential function

$$\Phi(\boldsymbol{y}, \theta) = \frac{1}{2}\|\boldsymbol{y} - \mathcal{G}(\theta)\|_\Gamma^2,$$

where $\mathcal{G} : \Theta \to \mathcal{Y}$ is the parameter-to-observable map while data $\boldsymbol{y} \in \mathcal{Y}$ are collected observations. Here Γ defines an appropriate Cameron-Martin space [Dashti and Stuart, 2015, section 7.3].

A distinction is made between "linear" inverse problems, in which $\mathcal{G}(\theta)$ is linear in the parameters θ, and "non-linear" inverse problems, where linearity fails to hold. Note that linearity of $\mathcal{G}(\theta)$ is distinct from linearity in the governing equations given by $\mathcal{A}$, $\mathcal{B}$. This case has received relatively little attention; an exception is Franke and Schaback [1998], which explored a variational Bayesian approach.

Typically an analytic representation for $\mathcal{G}(\theta)$ is unavailable, so that a numerical solver is used to obtain an approximation $\hat{\mathcal{G}}(\theta)$. Inference then proceeds based on the approximate potential

$$\hat{\Phi}(\boldsymbol{y}, \theta) = \frac{1}{2}\|\boldsymbol{y} - \hat{\mathcal{G}}(\theta)\|_\Gamma^2$$

in place of the true potential Φ. For PDEs, for example, the difference between $\hat{\Phi}$ and Φ can typically be driven to negligible values by using FEA methods based on refined meshes and appropriate error diagnostics, such as Richardson extrapolation or the adaptive methods of Moon et al. [2006] and Moon et al. [2005]. These techniques combine to produce guarantees on the extent of numerical error and, hence, lead to statistically valid inference on the parameter θ [Stuart, 2010]. However, the requirement to drive error to negligible values can lead to prohibitive computational requirements.

Instead a more flexible approach is pursued, based on collocation methods, in which error in the numerical solution of the forward problem is captured probabilistically and propagated into the inverse problem. Capturing this error can permit an overall reduction in computational complexity by avoiding construction of a mesh and, in some situations, by allowing use of a coarser discretisation while still permitting valid statistical inferences.

The approach is particularly well-suited to inference for non-linear PDEs, where solution uniqueness is no longer guaranteed. Then both discretisation error and solution identification are naturally formulated as sources of epistemic uncertainty. The methods that are developed below enable these uncertainties to be propagated, in a statistically principled way, into the inverse problem and enable valid inferences to be obtained.

2.3 Related Literature

The challenge of obtaining inferences based on approximations to the likelihood function has received considerable attention in the statistics literature. Focusing on PDEs, the use of approximate likelihoods for inference has been widely explored. Several approaches start by building a smooth approximation to gradients of the solution and penalising deviation from these gradients; see Campbell [2007], Ramsay et al. [2007], Dattner and Klaassen [2015], Heinonen and d'Alché Buc [2014] and the references therein. Closer in spirit to the approach pursued below, Marzouk and Xiu [2009] used polynomial chaos expansions to approximate the likelihood function, treating the mapping from θ to $\boldsymbol{y}$ as a black box. A similar approach was also proposed in Webster et al. [1996] and Ma and Zabaras [2009]. Both Marzouk and Xiu [2009] and Ma and Zabaras [2009] apply their likelihood approximations to solve inverse problems and Marzouk and Xiu [2009] establish that the approximate and exact posterior coincide in an appropriate limiting sense. However, these papers do not take into account the approximation error when making inferences on the parameter θ, meaning that careful control of error in the forward problem is still required to avoid the problems of bias and over-confidence that were exposed in Conrad et al. [2015].

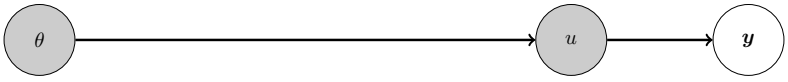
The forward problem has been the recent focus of PN methods. When this problem is described by ODEs, Conrad et al. [2015] constructed a probabilistic method for modelling discretisation error in numerical integrators, while Schober et al. [2014] revealed the underlying uncertainty model that is implied by Runge–Kutta methods. [See also the recent contribution by Kersting and Hennig, 2016]. For PDEs, Kaipio and Somersalo [2007] fitted Gaussian models for PDE errors, and Conrad et al. [2015] proposed PN methods for FEA [see also Arnold et al., 2013]. Related work by Capistrán et al. [2013] analysed the impact of discretisation error using Bayes factors.

Focusing on meshless methods, several authors have considered estimating the solution of operator equations as a fundamentally stochastic problem [e.g. Skilling, 1992, Graepel, 2003, Särkkä, 2011, Cialenco et al., 2012, Fasshauer and Ye, 2013]. The present approach is close to these papers in spirit, but the motivation was to make valid inferences for the inverse problem, which these papers do not consider. Recent work by Barber and Wang [2014] considered the inverse problem and proposed a method to obtain valid statistical inferences based on a probabilistic model for solver error. A shortcoming of Barber and Wang [2014] is an absence of theoretical analysis, as well as the restriction of attention to finite-dimensional parameters. The methods presented below address these shortcomings.

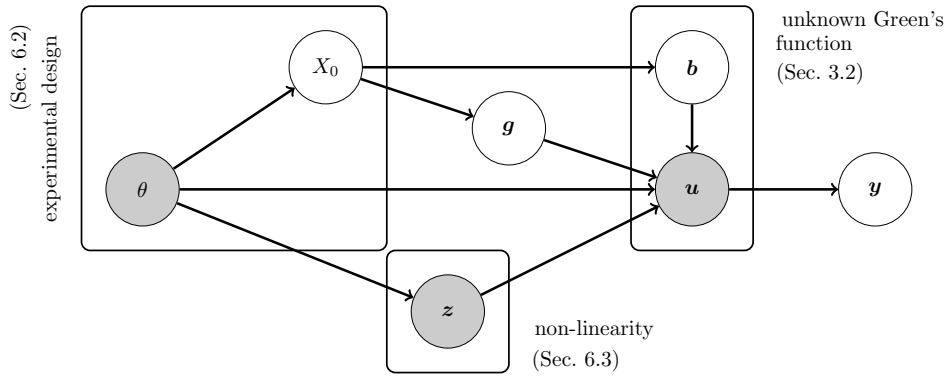
3 Probabilistic Models for the Forward Solution

In this section a "probabilistic" meshless method is formally defined. The starting point is radial basis function collocation, as studied by e.g. Fasshauer [1997] and more recently by Owhadi [2016]. Initially it is assumed that the operators $\mathcal{A}$ and $\mathcal{B}$ are linear; this will be relaxed in Section 6.3.

The method described in this and the following section is over-viewed, for reference, as a graphical model in Fig. 1.



(a) Abstract inverse problem (no discretisation error)



(b) Probabilistic meshless method

Abbildung 1: Graphical model representation. (Shaded nodes are unobserved.) (a) The abstract inverse problem, where the exact solution u can be obtained from the parameter θ and compared to observational data $\boldsymbol{y}$. (b) The probabilistic meshless method applied to the inverse problem in (a). In this framework the solution vector $\boldsymbol{u}$ is no longer a deterministic function of the parameter θ. Instead, a probabilistic model for discretisation error is integrated into inference. This aims to neutralise the inferential problems of bias and over-confidence that can result from neglecting discretisation error. The components X_0, $\boldsymbol{g}$ are defined in Sec. 3.1.2, $\boldsymbol{b}$ in Sec. 3.2 and $\boldsymbol{z}$ in Sec. 6.3 of the main text.

3.1 Probability Measures for Solutions of PDEs

Let $(\Omega', \mathcal{F}', \mathbb{P}')$ be a second probability space and consider a measurable function $g : D \times \Omega' \to \mathbb{R}$ such that for each $\omega' \in \Omega'$, $g(\cdot, \omega') \in H_{\mathcal{A}}(D)$. Similarly consider a measurable function $b : \partial D \times \Omega' \to \mathbb{R}$ such that for each $\omega' \in \Omega'$, $b(\cdot, \omega') \in H_{\mathcal{B}}(\partial D)$. The mathematical object that is studied in this section is the "doubly stochastic" solution $u(\cdot, \omega, \omega') \in H(D)$, $\omega \in \Omega$, $\omega' \in \Omega'$, of operator equations of the form

$$\begin{aligned} \mathcal{A}[\omega]u(\boldsymbol{x}, \omega, \omega') &= g(\boldsymbol{x}, \omega') \quad \boldsymbol{x} \in D \\ \mathcal{B}[\omega]u(\boldsymbol{x}, \omega, \omega') &= b(\boldsymbol{x}, \omega') \quad \boldsymbol{x} \in \partial D. \end{aligned} \tag{4}$$

The system in Eqn. 4 is a stochastic relaxation of the original inverse problem in Eqn. 1, in the sense that the deterministic forcing terms g and b are, for the purposes of this method, formally considered as random fields. This construction is justified in Owhadi [2015] as a reflection of the uncertainty, from the perspective of the numerical solver, about the value of the forcing at locations where it has not been evaluated. This uncertainty is formally modelled by the injected stochasticity.

The doubly stochastic solution u exists as a random variable that takes values $(\mathbb{P}, \mathbb{P}')$- almost surely in an appropriate function space; this will be made precise in Sec. 5.1. The immediate sections below focus on the effect of this randomisation and show how it connects at a deep level to collocation methods.

3.1.1 A Natural Reference Measure

To begin, a "reference" probability measure is devised for the unknown, infinite-dimensional solution u, based on the formal injection of stochasticity into the forcing function. The approach follows Owhadi [2015], but operates on functional Hilbert spaces rather than spaces of generalised functions.

To begin, the forcing g is formally modelled as a Gaussian stochastic process defined on $H_{\mathcal{A}}(D)$. That is, g is a stochastic process whose finite dimensional components $[g(\boldsymbol{x}_1, \cdot), \ldots, g(\boldsymbol{x}_n, \cdot)]$ are Gaussian in distribution. Recall that a Gaussian process $g : D \times \Omega' \to \mathbb{R}$ is characterised by its mean function $m : D \to \mathbb{R}$ and its covariance function $\Lambda : D \times D \to \mathbb{R}$. The notation $\Pi_g = N(m, \Lambda)$ will be used to

denote the distribution of g over $H_{\mathcal{A}}(D)$. It will be assumed that there exist fixed linear integro-differential operators $\mathcal{A}_\Lambda$ and $\mathcal{B}_\Lambda$ such that

$$\begin{aligned} \mathcal{A}_\Lambda g(\boldsymbol{x}) &= \xi(\boldsymbol{x}) \quad && \boldsymbol{x} \in D \\ \mathcal{B}_\Lambda g(\boldsymbol{x}) &= 0 && \boldsymbol{x} \in \partial D. \end{aligned} \tag{5}$$

where ξ is the standard white-noise process. Here and below we tend to leave the dependence on ω and ω' implicit to simplify the presentation. A common choice for $\mathcal{A}_\Lambda$ is the fractional Laplacian. A comprehensive background reference is Berlinet and Thomas-Agnan [2004].

Fundamental to the method is the fact that a covariance function Λ corresponds to a reproducing kernel Hilbert space (RKHS), written $H_\Lambda(D)$. It will be assumed that $H_\Lambda(D) \subseteq H_{\mathcal{A}}(D)$. By construction $H_\Lambda(D)$ contains all functions $g : D \to \mathbb{R}$ for which the norm $\|g\|_\Lambda := \|\mathcal{A}_\Lambda g\|_{L^2(D)}$ is finite. The fractional Laplacian choice for $\mathcal{A}_\Lambda$ implies that $H_\Lambda(D)$ is a standard Sobolev space. An important property of this characterisation is that the Gaussian measure assigns zero mass to the RKHS, i.e. $\Pi_g[H_\Lambda(D)] = 0$ [Berlinet and Thomas-Agnan, 2004]. This leads to an amount of additional technical detail in Sec. 3.2.

Next, randomness is propagated from the forcing term through into the solution of the PDE. Define the inner product space

$$H_k(D) := \{v \in H(D) \mid \mathcal{A}_\Lambda \mathcal{A} v \in L^2(D),\ \mathcal{B} v = 0 \text{ on } \partial D \text{ and } \mathcal{B}_\Lambda \mathcal{A} v = 0 \text{ on } \partial D\}$$

with

$$\langle u, v \rangle_k := \int_D [\mathcal{A}_\Lambda \mathcal{A} u(\boldsymbol{x})][\mathcal{A}_\Lambda \mathcal{A} v(\boldsymbol{x})] \mathrm{d}\boldsymbol{x}.$$

Under this definition $\|u\|_k^2 := \langle u, u \rangle_k = \|g\|_\Lambda^2$. Prop. 1 below establishes that $H_k(D)$ is in fact an RKHS. To elicit its reproducing kernel, focus initially on the case of a linear operator $\mathcal{A}$ and assume non-degeneracy, so that $\langle v, v \rangle_k = 0$ if and only if $v = 0$. Here it will be assumed that the problem is well-posed, meaning that, for any $g \in H_\Lambda(D)$, there exists a unique solution $u \in H_k(D)$ to the system $\mathcal{A} u = g$. Supposing that we have a Green's function G satisfying

$$\begin{aligned} \mathcal{A} G(\boldsymbol{x}, \boldsymbol{x}') &= \delta(\boldsymbol{x} - \boldsymbol{x}') \quad && \boldsymbol{x} \in D \\ \mathcal{B} G(\boldsymbol{x}, \boldsymbol{x}') &= 0 && \boldsymbol{x} \in \partial D, \end{aligned} \tag{6}$$

we define the "natural kernel" $k : D \times D \to \mathbb{R}$ via

$$k(\boldsymbol{x}, \boldsymbol{x}') := \int_D \int_D G(\boldsymbol{x}, \boldsymbol{z}) G(\boldsymbol{x}', \boldsymbol{z}') \Lambda(\boldsymbol{z}, \boldsymbol{z}') \mathrm{d}\boldsymbol{z} \mathrm{d}\boldsymbol{z}'. \tag{7}$$

Throughout, the notation $\mathcal{A} G(\boldsymbol{x}, \boldsymbol{x}')$ is used when the operator acts on the first argument $\boldsymbol{x}$, while the notation $\bar{\mathcal{A}} G(\boldsymbol{x}, \boldsymbol{x}')$ is used when the operator acts on the second argument $\boldsymbol{x}'$. The relationship between Green's functions and kernels is explored in detail by Fasshauer and Ye [2011].

Proposition 1. *Assume that $\int_D k(\boldsymbol{x}, \boldsymbol{x})^{1/2} \mathrm{d}\boldsymbol{x} < \infty$. Then $H_k(D)$ is a reproducing kernel Hilbert space and k is its reproducing kernel.*

Prop. 1 explains and justifies the $H_k(D)$ notation. The kernel k is indeed natural for this particular problem, in the sense that $H_\Lambda(D)$ is the image under $\mathcal{A}$ of $H_k(D)$. In the linear case, a realisation of g corresponds to a unique realisation of u and the randomness $\omega \in \Omega$ implies a "reference" measure Π_u over $H(D)$. Indeed, assuming zero means henceforth, we have the following:

Proposition 2. *Π_g is a centred Gaussian process with kernel Λ if and only if Π_u is a centered Gaussian process with kernel k.*

In practice one can specify either the form of Λ or the form of k, since in the linear case each fully determines the other. All proofs are provided in the Appendix.

3.1.2 A Natural Conditional Measure

Now a distribution is constructed, with respect to the reference measure, that represents epistemic uncertainty over the solution u after expending a finite amount of computational resource. The route taken is to condition the reference measure Π_u on $m_{\mathcal{A}}$ observations of the forcing function at distinct locations $X_0^{\mathcal{A}} = \{\boldsymbol{x}_{0,j}^{\mathcal{A}}\}_{j=1}^{m_{\mathcal{A}}} \subset D$. Our information on the solution u comes via the interpolation equations

$$\mathcal{A}u(\boldsymbol{x}_{0,j}^{\mathcal{A}}) = g(\boldsymbol{x}_{0,j}^{\mathcal{A}}), \quad j = 1, \ldots, m_{\mathcal{A}}.$$

Write $\boldsymbol{g}$ for the $m_{\mathcal{A}} \times 1$ vector with jth element $g(\boldsymbol{x}_{0,j}^{\mathcal{A}})$. Then the conditional process $u|\boldsymbol{g}$, denoted by $\Pi_u^{\boldsymbol{g}}$, is also Gaussian and characterised by its finite-dimensional marginals, given in Prop. 3.

For sets $X = \{\boldsymbol{x}_j\}_{j=1}^n$ and $X' = \{\boldsymbol{x}'_j\}_{j=1}^{n'}$ of states, denote by $\boldsymbol{K}(X, X')$ the $n \times n'$ matrix whose (i,j)th element is $k(\boldsymbol{x}_i, \boldsymbol{x}'_j)$. When $X = X'$ we use the more succinct notation $\boldsymbol{K}(X) = \boldsymbol{K}(X, X)$. The $n \times n'$ matrices $\mathcal{A}\boldsymbol{K}(X, X')$, $\bar{\mathcal{A}}\boldsymbol{K}(X, X')$ and $\mathcal{A}\bar{\mathcal{A}}\boldsymbol{K}(X, X')$ have respective (i,j)th entries $\mathcal{A}k(\boldsymbol{x}_i, \boldsymbol{x}'_j)$, $\bar{\mathcal{A}}k(\boldsymbol{x}_i, \boldsymbol{x}'_j)$ and $\mathcal{A}\bar{\mathcal{A}}k(\boldsymbol{x}_i, \boldsymbol{x}'_j)$.

Proposition 3 (Probabilistic meshless method I)**.** *Given $X = \{\boldsymbol{x}_j\}_{j=1}^n \subset D$, denote by $\boldsymbol{u}$ the $n \times 1$ vector with jth element $u(\boldsymbol{x}_j)$. The posterior distribution $\Pi_u^{\boldsymbol{g}}$ is Gaussian with finite dimensional marginals $\boldsymbol{u}$ given by*

$$\boldsymbol{u}|\boldsymbol{g} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where the mean and variance are

$$\boldsymbol{\mu} = \bar{\mathcal{A}}\boldsymbol{K}(X, X_0^{\mathcal{A}})[\mathcal{A}\bar{\mathcal{A}}\boldsymbol{K}(X_0^{\mathcal{A}})]^{-1}\boldsymbol{g} \tag{8}$$

$$\boldsymbol{\Sigma} = \boldsymbol{K}(X) - \bar{\mathcal{A}}\boldsymbol{K}(X, X_0^{\mathcal{A}})[\mathcal{A}\bar{\mathcal{A}}\boldsymbol{K}(X_0^{\mathcal{A}})]^{-1}\mathcal{A}\boldsymbol{K}(X_0^{\mathcal{A}}, X). \tag{9}$$

This result is well-known as a general algebraic identity; see e.g. Särkkä [2011].

This clarifies what constitutes a "probabilistic" solver; rather than returning only an approximation $\hat{u}$ to u, a probabilistic solver returns a full distribution $\Pi_u^{\boldsymbol{g}}$ where randomness represents uncertainty over the true values of u due to finite computational resources. The content of Prop. 3 will be referred to as a probabilistic meshless method (PMM). In Sec. 4 it is proven that this quantification of uncertainty is appropriate, giving rise to minimax error bounds.

An explicit "gamblet" basis for the conditional measure $\Pi_u^{\boldsymbol{g}}$ was considered in Owhadi [2016] but will not be discussed in this paper.

3.2 Connection to Meshless Methods

The presentation above assumes access to a Green's function for the PDE. In practice we cannot make this assumption as Green's functions are not generally available for nontrivial PDE systems. In this situation we must resort to an alternative choice of covariance function.

Collocation methods [Fasshauer, 1997, Fasshauer and Ye, 2011] begin by positing an RKHS for the solution u, with a kernel k which has favourable computational properties. While the natural kernel k in Eqn. 7 is unavailable, it is often straightforward to exhibit a kernel $\tilde{k}$ such that $H_k(D)$ is embedded in $H_{\tilde{k}}(D)$. A Hilbert space H is said to be "embedded" in another Hilbert space H' if $H \subseteq H'$; this implies the existence of a constant $0 < c < \infty$ such that $\|u\|_{H'} \leq c\|u\|_H$ for all $u \in H$ [Pillai et al., 2007]. In the notation of Sec. 3, set $H(D) = H_{\tilde{k}}(D)$ where the latter is an RKHS with reproducing kernel $\tilde{k}$. In this paper $\tilde{k}$ will typically be a kernel whose native space is a Sobolev space (see Assumption 1 below). The order of this space can therefore be chosen by "derivative counting", to reflect the number of (weak) derivatives that u is believed to have, based on the maximum differential order of operators in the system.

Due to the aforementioned technicality that $\mathbb{P}'$ does not place mass on the RKHS, it is necessary to introduce a third kernel $\hat{k}$. It will be required that $H_{\hat{k}}(D)$ is embedded in $H(D)$ and that $\hat{k}$ satisfy the following properties:

(i) The measure $\Pi_u = N(0, \hat{k})$ satisfies $\Pi_u[H(D)] = 1$.

(ii) The set $H_{\hat{k}}(D)$ is dense in the space $(H(D), \|\cdot\|_{\tilde{k}})$.

These conditions enable any function $u \in H_k(D) \subseteq H(D)$ to be "inferred" from data, under a prior Π_u.

One way to achieve both (i) and (ii), is to follow Cialenco et al. [2012] (Lemma 2.2), and consider the integral-type kernel below:

Proposition 4. *The integral-type kernel*

$$\hat{k}(\boldsymbol{x}, \boldsymbol{x}') = \int_D \tilde{k}(\boldsymbol{x}, \boldsymbol{z}) \tilde{k}(\boldsymbol{z}, \boldsymbol{x}') \mathrm{d}\boldsymbol{z}$$

satisfies requirements (i) and (ii).

The strategy of taking $\tilde{k}$ to be a generic kernel (e.g. a Sobolev kernel) means that functions $u \in H_{\hat{k}}$ do not automatically satisfy boundary conditions $\mathcal{B}u = b$ of the system. Instead, meshless methods incorporate boundary conditions by the introduction of additional evaluations $\mathcal{B}u(\boldsymbol{x}^{\mathcal{B}}_{0,j}) = b_j$ for a set $X^{\mathcal{B}}_0 = \{\boldsymbol{x}^{\mathcal{B}}_{0,j}\}_{j=1}^{m_{\mathcal{B}}}$ of points on the boundary ∂D. Write $\boldsymbol{b}$ for the $m_{\mathcal{B}} \times 1$ vector with jth entry $b_j = b(\boldsymbol{x}^{\mathcal{B}}_{0,j})$. Denote the full meshless design $X_0 = (X^{\mathcal{A}}_0, X^{\mathcal{B}}_0)$. Define

$$\mathcal{L} = \begin{bmatrix} \mathcal{A} \\ \mathcal{B} \end{bmatrix}, \quad \bar{\mathcal{L}} = \begin{bmatrix} \bar{\mathcal{A}} & \bar{\mathcal{B}} \end{bmatrix}.$$

Introduce the $(m_{\mathcal{A}} + m_{\mathcal{B}}) \times (m_{\mathcal{A}} + m_{\mathcal{B}})$ matrix

$$\mathcal{L}\bar{\mathcal{L}}\hat{\boldsymbol{K}}(X_0) = \begin{bmatrix} \mathcal{A}\bar{\mathcal{A}}\hat{\boldsymbol{K}}(X^{\mathcal{A}}_0, X^{\mathcal{A}}_0) & \mathcal{A}\bar{\mathcal{B}}\hat{\boldsymbol{K}}(X^{\mathcal{A}}_0, X^{\mathcal{B}}_0) \\ \bar{\mathcal{A}}\mathcal{B}\hat{\boldsymbol{K}}(X^{\mathcal{B}}_0, X^{\mathcal{A}}_0) & \mathcal{B}\bar{\mathcal{B}}\hat{\boldsymbol{K}}(X^{\mathcal{B}}_0, X^{\mathcal{B}}_0) \end{bmatrix}$$

and also the $1 \times (m_{\mathcal{A}} + m_{\mathcal{B}})$ vectors

$$\bar{\mathcal{L}}\hat{\boldsymbol{K}}(\boldsymbol{x}, X_0)^T = \begin{bmatrix} \bar{\mathcal{A}}\hat{\boldsymbol{K}}(\boldsymbol{x}, X^{\mathcal{A}}_0) & \bar{\mathcal{B}}\hat{\boldsymbol{K}}(\boldsymbol{x}, X^{\mathcal{B}}_0) \end{bmatrix}$$
$$\mathcal{L}\hat{\boldsymbol{K}}(\boldsymbol{x}, X_0)^T = \begin{bmatrix} \mathcal{A}\hat{\boldsymbol{K}}(\boldsymbol{x}, X^{\mathcal{A}}_0) & \mathcal{B}\hat{\boldsymbol{K}}(\boldsymbol{x}, X^{\mathcal{B}}_0). \end{bmatrix}$$

In each case the hat on $\hat{\boldsymbol{K}}$ indicates that entries are based on the posited kernel $\hat{k}$, as opposed to the matrix $\boldsymbol{K}$ which was based on the natural kernel k.

The posterior or "conditional" process $u|\boldsymbol{g}, \boldsymbol{b}$ is also Gaussian; denote its distribution $\Pi^{\boldsymbol{g},\boldsymbol{b}}_u$. This result is the basis for practical PMMs. Indeed, proceeding as in Prop. 3:

Proposition 5 (Probabilistic meshless method II)**.** *Write $X = \{\boldsymbol{x}_j\}_{j=1}^n \subset D$. Denote by $\boldsymbol{u}$ the $n \times 1$ vector with jth element $u(\boldsymbol{x}_j)$. Then under $\Pi^{\boldsymbol{g},\boldsymbol{b}}_u$ we have*

$$\boldsymbol{u}|\boldsymbol{g}, \boldsymbol{b} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

where the mean and variance are

$$\boldsymbol{\mu} = \bar{\mathcal{L}}\hat{\boldsymbol{K}}(X, X_0)[\mathcal{L}\bar{\mathcal{L}}\hat{\boldsymbol{K}}(X_0)]^{-1}[\boldsymbol{g}^T\ \boldsymbol{b}^T]^T \tag{10}$$
$$\boldsymbol{\Sigma} = \hat{\boldsymbol{K}}(X) - \bar{\mathcal{L}}\hat{\boldsymbol{K}}(X, X_0)[\mathcal{L}\bar{\mathcal{L}}\hat{\boldsymbol{K}}(X_0)]^{-1}\mathcal{L}\hat{\boldsymbol{K}}(X_0, X). \tag{11}$$

Here it is convenient to express the pointwise conditional mean and variance as $\mu(\boldsymbol{x})$ and $\sigma(\boldsymbol{x})^2$ respectively; i.e. as defined by Eqns. 10 and 11 with $X = \{\boldsymbol{x}\}$. Then the expression presented here for $\mu(\boldsymbol{x})$ is identical to the meshless method known as "symmetric collocation", developed by [Fasshauer, 1999]. The probabilistic interpretation of symmetric collocation was previously noted in Cialenco et al. [2012]. Compared to previous literature, the variance term $\sigma^2(\boldsymbol{x})$ will play a more central role and will enable formal quantification of numerical error.

The matrices involved here will be sparse when the kernel $\hat{k}$ has compact support. Thus, by appropriate choice of kernel, the required matrix inversion can be made competitive with efficient FEA methods. While not investigated here, meshless methods can be extended in several directions, including to multi-level methods [Fasshauer, 1999].

3.2.1 Illustrative Example: Forward Problem

To illustrate these ideas, we examine the above procedure for a simple linear PDE in one dimension, namely Poisson's Equation. Consider the system:

$$\begin{aligned} -\nabla^2 u(x) &= g(x) \quad \text{for } x \in (0,1) \\ u(x) &= 0 \qquad \text{for } x \in \{0,1\} \end{aligned}$$

This system satisfies the conditions in Sec. 3.1.1, in that the solution is known to be zero on the boundary of the domain. Futhermore the Green's function for this system is available analytically:

$$G(x,x') = \begin{cases} x(x'-1) & x > x' \\ x'(x-1) & x < x' \end{cases}$$

To proceed, place a Gaussian measure on the forcing term g, using the compactly supported polynomial of Wendland [1995]:

$$\Lambda(x,x') = \left(\max(1-\epsilon|x-x'|,0)\right)^2$$

Samples from the prior Π_g will be continuous, with no continuous derivatives, and has support wherever $|x-x'| < \epsilon^{-1}$. Associating operators in the above system with the abstract definitions, we have $\mathcal{A} := \nabla^2 = \frac{d^2}{dx^2}$, while $\bar{\mathcal{A}} = \frac{d^2}{dx'^2}$. The natural kernel

$$k(x,x') = \int_0^1 \int_0^1 G(x,z)G(x',z')\Lambda(z,z')\mathrm{d}z\mathrm{d}z'$$

has a closed-form solution since all three terms are piecewise-polynomial. The full expression is given in the Supplement. Exploiting properties of Green's functions we can also find:

$$\begin{aligned} \mathcal{A}k(x,x') &= \int_0^1 G(x',z')\Lambda(x,z')\mathrm{d}z\mathrm{d}z' \\ \bar{\mathcal{A}}k(x,x') &= \int_0^1 G(x,z)\Lambda(z,x')\mathrm{d}z\mathrm{d}z' \\ \mathcal{A}\bar{\mathcal{A}}k(x,x') &= \Lambda(x,x') \end{aligned}$$

which are similarly available in closed-form.

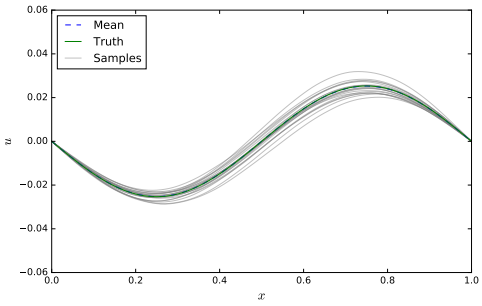
Having applied each operator in the system to the natural kernel, we must then select a set of design points x_i and evaluate the vector $\boldsymbol{g} = [g(x_i)]$ at each point. For illustration we took $m_{\mathcal{A}} = 39$ function evaluations at evenly spaced points in $(0,1)$. The conditional mean and covariance are given in Eqns. 8 and 9 respectively. In Fig. 2a the conditional mean is plotted for the above PDE with $g(x) = -\sin(2\pi x)$, along with sample paths from the full conditional measure. The covariance $\Lambda(x,x')$ is assigned a support of $\epsilon = 2.5$. This is contrasted with the closed-form solution $u(x) = (2\pi)^{-2}\sin(2\pi x)$.

Even for this most simple of examples, computation of the natural kernel is challenging. In practice collocation methods operate using a kernel such as $\hat{k}$ as given in Sec. 3.2, or even directly positing a kernel. In Fig. 2b the performance of the natural kernel k is contrasted with that of $\hat{k}$, computed from a higher-order Wendland covariance function:

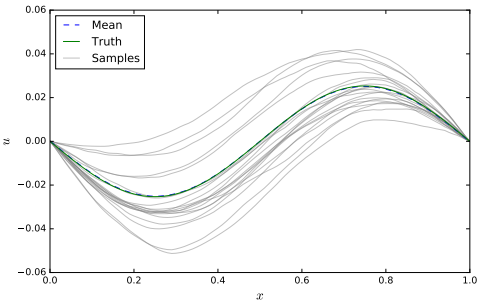
$$\begin{aligned} \tilde{k}(x,x') &= \left(\max(1-\epsilon|x-x'|,0)\right)^4 \cdot (4\epsilon|x-x'|+1) \\ \hat{k}(x,x') &= \int_0^1 \tilde{k}(x,z)\tilde{k}(z,x')dz \end{aligned}$$

This kernel is chosen to correspond to the required level of smoothness in the solution, as $\tilde{k}$ is twice-differentiable at the origin. In addition the design is augmented with $X_0^{\mathcal{B}} = \{0,1\}$ so that samples from the conditional measure satisfy the boundary conditions with probability one.

Convergence of the conditional measures $\Pi_u^{\boldsymbol{g}}$ and $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ based on these two kernels as the number of design points is increased is shown in Fig. 3. The advantage of using the (typically unavailable) natural kernel is seen to be a reduction in approximation error.

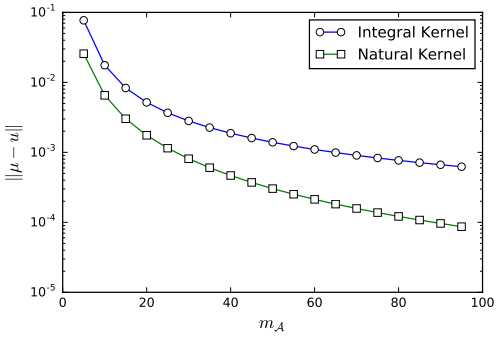


(a) $\Pi_u^{\boldsymbol{g}}$, based on k

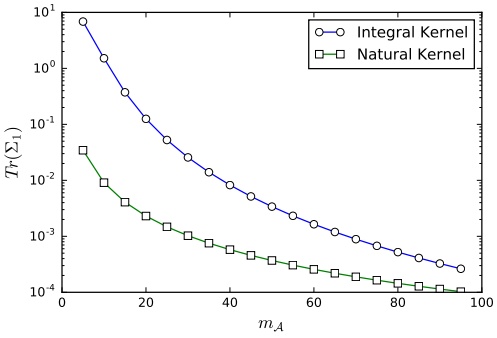


(b) $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$, based on $\hat{k}$

Abbildung 2: Probabilistic meshless methods: Comparison of conditional distributions (a) $\Pi_u^{\boldsymbol{g}}$ based on the natural kernel k and (b) $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ based on the integrated Wendland kernel $\hat{k}$. In (b) two additional observations are added at $x = 0$ and $x = 1$.



(a) Error in conditional mean, $\|\mu - u\|_2$



(b) Residual uncertainty, $\|\sigma^2\|_1$

Abbildung 3: Probabilistic meshless methods: Convergence of mean and covariance as the number m of design points is increased. Values of μ and σ were computed on a fine grid of 100 points in the domain.

3.3 The Inverse Problem

For linear PDEs with Gaussian additive noise, the data-likelihood has the closed-form expression

$$\pi(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{b},\theta) = \int \pi(\boldsymbol{y}|\boldsymbol{u})\pi(\boldsymbol{u}|\boldsymbol{g},\boldsymbol{b},\theta)\mathrm{d}\boldsymbol{u} \tag{12}$$

$$= \frac{1}{\sqrt{\det[2\pi(\boldsymbol{\Sigma}(\theta)+\boldsymbol{\Gamma})]}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}(\theta))^T(\boldsymbol{\Sigma}(\theta)+\boldsymbol{\Gamma})^{-1}(\boldsymbol{y}-\boldsymbol{\mu}(\theta))\right\}, \tag{13}$$

where dependence of the conditional distribution over $\boldsymbol{u}$ on the parameter θ is emphasised. For simplicity we have assumed the observation noise distribution $\Pi_{\boldsymbol{y}}$ is jointly Gaussian; $\boldsymbol{y}|\boldsymbol{u} \sim N(\boldsymbol{u},\boldsymbol{\Gamma})$. In Sec. 5.2 we prove that as the design set is refined, the mean $\boldsymbol{\mu}(\theta)$ converges to the true solution $\boldsymbol{u}(\theta)$, the covariance $\boldsymbol{\Sigma}(\theta)$ vanishes and $\pi(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{b},\theta)$ converges to the (unavailable) abstract data-likelihood $\pi(\boldsymbol{y}|\theta)$.

3.3.1 Illustrative Example: Inverse Problem

Returning to the example of Sec. 3.2.1, we illustrate probabilistic solution of an inverse problem. Consider the problem of estimating θ in:

$$\begin{aligned} -\nabla\cdot(\theta\nabla u(x)) &= g(x) \quad \text{for } x \in (0,1) \\ u &= 0 \qquad \text{for } x \in \{0,1\} \end{aligned}$$

An interesting observation here is that, while the kernel $\hat{k}$ is independent of the value of θ in this problem, the natural kernel does have dependence on θ through the Green's function; indeed $k(x,x';\theta) = \theta^{-2}k(x,x';1)$. This dependence would be easy to remove by simply dividing by θ, however we emphasise it here in light of theoretical considerations which will become apparent in Sec. 5. Essentially Assumption (A2) later requires that k does not depend "too strongly" on θ.

We solve this problem with $g(x) = -\sin(2\pi x)$ and data-generating parameter $\theta = 1$. Data were generated at the locations $x = 0.25$ and $x = 0.75$ by evaluating the explicit solution $u(x) = (2\pi)^{-2}\sin(2\pi x)$. This is then corrupted with Gaussian noise with covariance $\boldsymbol{\Gamma} = 0.001^2\boldsymbol{I}$.

To illustrate the advantage of a probabilistic solution to the PDE, posteriors were computed based on (a) the standard approach that ignores discretisation error and plugs an estimate $\boldsymbol{\mu}$ into the measurement error model $\pi(\boldsymbol{y}|\boldsymbol{u})$ in place of $\boldsymbol{u}$, and (b) the PMM where data-likelihood is given in Eqn. 13. The parameter θ was endowed with a standard log-Gaussian prior to ensure positivity. Posteriors are plotted in Fig. 4a, while convergence with the number of collocation points is shown in Fig. 4b.

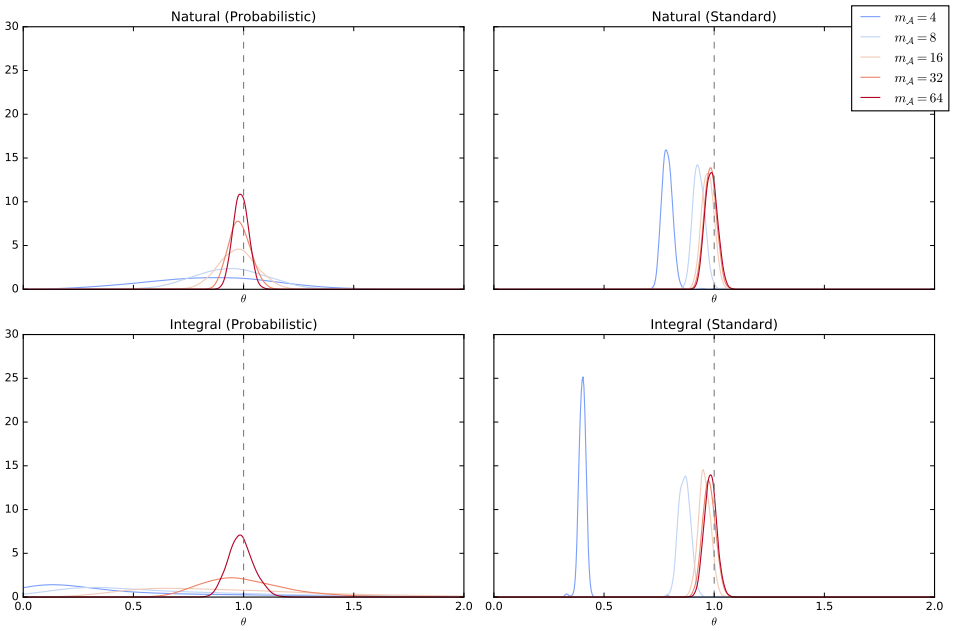
This example highlights failure of the standard approach; the posterior variance is constant, independent of the number $m_{\mathcal{A}}$ of collocation basis functions, and posterior credible intervals do not cover the true value $\theta = 1$ when $m_{\mathcal{A}} \leq 20$. In contrast, when using the PMM there is a clear widening in the posterior for small $m_{\mathcal{A}}$, and in general the true value of θ is within a standard deviation of the posterior mode.

Note also that convergence is faster when using the natural kernel k, compared with the integral kernel $\hat{k}$; this is to be expected considering the faster convergence exhibited in Sec. 3.2.1.

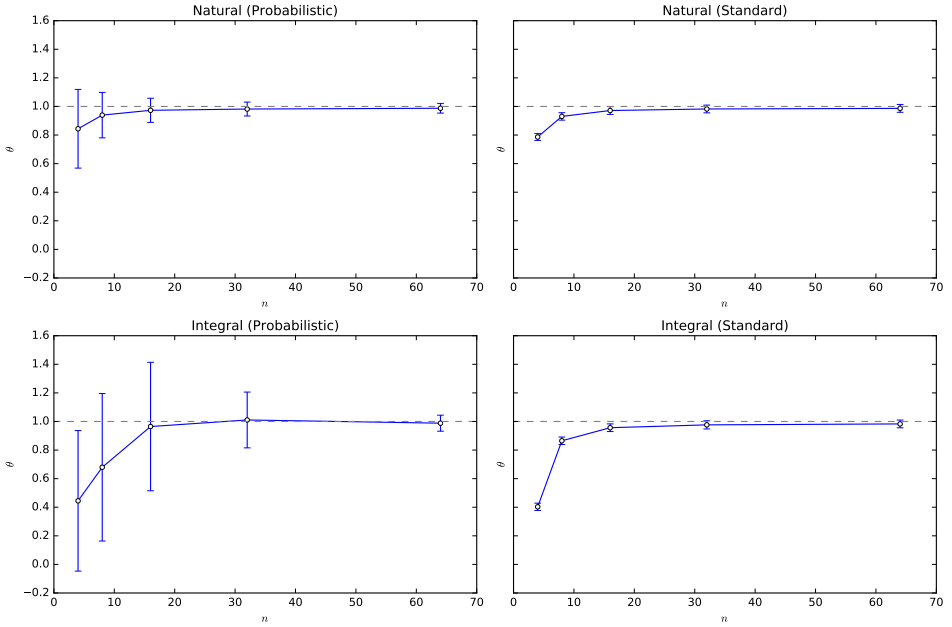
3.3.2 Calibration of Kernel Parameters

The requirement to posit a kernel $\tilde{k}$ typically introduces nuisance parameters. This issue has so far received little attention in the literature on meshless methods, but is crucial to this work since the choice of parameters directly influences the spread of the probability model for numerical error.

The problem of selecting kernel parameters appears regularly in PN methods [Conrad et al., 2015, Briol et al., 2016, Kersting and Hennig, 2016]. One principled approach to selection of kernel parameters would simply be to maximise the likelihood of the data, i.e. to maximise Eqn. 13 over all nuisance parameters. This is known in statistics as "empirical Bayes", but other approaches are possible, including marginalisation and cross-validation; see the discussion in Briol et al. [2016]. An alternative approach is to consider the nuisance parameters of the kernel as additional parameters to be inferred; this allows marginalisation of those parameters so that inferences do not depend upon point estimates. Both solutions are explored in the applications in Sec. 7.



(a) Posterior distributions $\Pi_\theta^{\boldsymbol{y}}$



(b) Posterior credible intervals

Abbildung 4: Probabilistic meshless methods: Solutions to the inverse problem are contrasted with the standard approach that plugs a discrete approximation to the exact PDE solution into the data-likelihood, leading to over-confidence. (a) Posterior distributions $\Pi_\theta^{\boldsymbol{y}}$ as a function of the number $m_{\mathcal{A}}$ of design points. (b) One standard deviation posterior credible intervals for θ, again as a function of the number $m_{\mathcal{A}}$ of design points. In each case the natural kernel k is contrasted with the integral-type kernel $\hat{k}$, as explained in the main text.

4 Error Analysis for the Forward Problem

This section presents error analysis for the forward problem where $\omega_0 \in \Omega$ is fixed. Denote by $u_0 = u(\cdot, \omega_0, \omega_0')$ the true solution to the PDE.

Two Hilbert spaces H, H' are said to be "(norm-) equivalent" when each is embedded in the other and we write $H \equiv H'$. The results restrict to the following situation:

(A1) Suppose that $H(D)$ is norm-equivalent to the Sobolev space $\mathbb{H}^\beta(D)$ of order $\beta > d/2$, with norm denoted by $\|\cdot\|_{\mathbb{H}^\beta(D)}$.

(A1) can be satisfied by construction since we are free to select the kernel $\tilde{k}$ and $H(D) = H_{\tilde{k}}(D)$. It is implicitly assumed that the orders (i.e. number of derivatives) of $\mathcal{A}$ and $\mathcal{B}$ are $O(\mathcal{A}), O(\mathcal{B}) < \beta - d/2$, so that the stochastic processes $\mathcal{A}u$ and $\mathcal{B}u$ are well-defined.

The analysis is rooted in a dual relationship between the posterior variance and the worst-case error:

Proposition 6 (Local accuracy)**.** *For all $\boldsymbol{x} \in D$ we have $|\mu(\boldsymbol{x}) - u_0(\boldsymbol{x})| \leq \sigma(\boldsymbol{x}) \|u_0\|_{\hat{k}}$.*

Prop. 6 shows that minimising $\sigma(\boldsymbol{x})$ leads to accurate estimates $\mu(\boldsymbol{x})$. This reassures us that the conditional measure $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ over the solution space is sensible in a local sense. Moreover it reveals a minimax characterisation, since $\sigma(\boldsymbol{x})$ is in fact equal to the supremum of $|\mu(\boldsymbol{x}) - u(\boldsymbol{x})|$ over all $u \in H_{\hat{k}}(D)$ with $\|u\|_{\hat{k}} = 1$.

To make precise the notion of minimising $\sigma(\boldsymbol{x})$, define the "fill distance" of the design X_0 as

$$h := \sup_{\boldsymbol{x} \in D} \min_{\boldsymbol{x}' \in X_0} \|\boldsymbol{x} - \boldsymbol{x}'\|_2.$$

The following is Lemma 3.4 of Cialenco et al. [2012]; see also Secs. 11.3 and 16.3 of Wendland [2005]:

Proposition 7. *For all $\boldsymbol{x} \in D$ and all $h > 0$ sufficiently small, we have $\sigma(\boldsymbol{x}) \leq Ch^{\beta-\rho-d/2}$ where $\rho = \max\{O(\mathcal{A}), O(\mathcal{B})\}$ and C denotes a generic constant.*

Denote $\|u\|_2^2 = \int_D u(\boldsymbol{x})^2 \mathrm{d}\boldsymbol{x}$. Prop. 7 is used below to establish contraction of the conditional measure $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ over $H(D)$ to the true solution u_0 as the fill distance tends to zero:

Theorem 8 (Contraction of conditional measure to u_0)**.** *Fix $\epsilon > 0$ and consider the limit $h \to 0$. For a ball $B_\epsilon(u_0) := \{u \in H(D) : \|u - u_0\|_2^2 \leq \epsilon\}$ centred on the true solution u_0 of the PDE, we have*

$$1 - \Pi_u^{\boldsymbol{g},\boldsymbol{b}}[B_\epsilon(u_0)] = \mathcal{O}\left(\frac{h^{2\beta-2\rho-d}}{\epsilon}\right).$$

A similar result is presented as Lemma 3.5 in Cialenco et al. [2012].

Thm. 8 is an important and reassuring result, showing that the conditional measure $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ provides sensible uncertainty quantification in a global sense. However it is not the full story; our ultimate goal is to make accurate inferences on θ. This introduces several considerations that go well beyond analysis of the forward problem. For example, given noisy observations $\boldsymbol{y}$ of the solution, it makes sense that the design points X_0 should also lie "close" to where these data are obtained above. For non-uniform observation locations, this objective does not coincide with the fill distance objective. The proposed resolution involves the use of sophisticated techniques from experimental design, that are presented in Sec. 6.

Furthermore, in the inverse problem setting, the solution $u = u(\cdot, \omega, \omega')$ depends on $\omega \in \Omega$. Thus the term $\|u\|_{\hat{k}}$ in Prop. 6 will be a random variable. In fact, from a broader perspective we must examine whether, and in what sense, the solution $u(\cdot, \omega, \omega')$ exists as a random object.

These points are addressed in the next sections.

5 Error Analysis for the Inverse Problem

The aim of Sec. 5.1 is to establish the existence of $u(\cdot, \omega, \omega')$ as a random variable. Sec. 5.2 shows that consistent estimation of θ is possible provided that the fill distance h, that controls the accuracy of the PMM, decreases at an appropriate rate relative to the number n of observational data.

5.1 Existence of the Doubly Stochastic Solution

This section makes precise the sense in which the doubly stochastic solution $u(\cdot, \omega, \omega')$ exists as a random variable. Let $\mathbb{E}$ and $\mathbb{E}'$ denote expectations with respect to $\mathbb{P}$ and $\mathbb{P}'$. First recall the notion of a "Hilbert scale" of spaces.

Since D is not compact we require a generalization of Mercer's theorem for continuous kernels as given in Steinwart and Scovel [2012]. Since k is continuous and, by virtue of (A2), below, its RKHS $H_k(D)$ is embedded in $L^2(D)$, there exists an orthonormal basis for $H_k(D)$ such that a generic element $u \in H_k(D)$ can be written as $u = \sum_{i=1}^{\infty} c_i h_i$, $h_i = \sqrt{\lambda_i} e_i$ where $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ are eigenvalues and e_i are associated eigenvectors of the integral operator $u(\cdot) \mapsto \int_D u(\boldsymbol{x}) k(\boldsymbol{x}, \cdot) \mathrm{d}\boldsymbol{x}$. The norm for this space is characterised by $\|u\|_k^2 = \sum_{i=1}^{\infty} c_i^2$. Define the Hilbert scale of spaces $H^t = \{h = \sum_i c_i h_i \text{ s.t. } \|h\|_{k,t}^2 := \sum_i \lambda_i^{-t} c_i^2 < \infty\}$ for $t \in \mathbb{R}$ [Dashti and Stuart, 2015, Section 7.1.3]. For a generic RKHS H we have that $H^0 = H$, while $H^s \supseteq H^t$ whenever $s \leq t$. The intuition here is that H^t, $t < 0$, is a relaxation of H. The integral-type kernel from Sec. 3.2 is constructed such that $H_{\hat{k}}^{-1}(D) = H(D)$.

An assumption is now made on the regularity of the inverse problem, as captured by the regularity of the natural solution space $H_k(D)$ from Sec. 3.1.1. Recall that $H_k(D)$ is a random space, depending on $\omega \in \Omega$ through the value of the parameter $\theta \in \Theta$. Write $\lambda_i^{(k)}$ for the eigenvalues associated with k. Similarly write $\lambda_i^{(\alpha)}$ for the eigenvalues associated with the Sobolev space $\mathbb{H}^\alpha(D)$.

(A2) For some $\alpha \geq \beta$, all $-1 < t < -d/2\alpha$ and $\mathbb{P}$- almost all $\omega \in \Omega$, there exist constants $0 < C_\omega$ and $C_{\omega,t} < \infty$ such that, for all $v \in H_k(D)$, $i \in \mathbb{N}$,

$$\|v\|_{\mathbb{H}^\alpha(D),t}^2 \;\leq\; C_{\omega,t} \|v\|_{k,t}^2, \quad \lambda_i^{(\alpha)} \;\leq\; C_\omega \lambda_i^{(k)}$$

and $\mathbb{E}[\, C_\omega^t \, C_{\omega,t} \,] \;<\; \infty$.

This implies, in particular, that $H_k^t(D)$ is embedded in $[\mathbb{H}^\alpha(D)]^t \equiv \mathbb{H}^{(1+t)\alpha}(D)$ for Π_θ- almost all values of the parameter $\theta \in \Theta$. Note that α and β assume distinct roles in the analysis; α captures the regularity of the unavailable natural solution space $H_k(D)$, while β captures the regularity of the larger space $H(D)$ in which the numerical solver operates. In general we must have $\alpha \geq \beta$, although in Prop. 10 below it is required that $\alpha > 2\beta + d/2$.

Theorem 9 (Existence)**.** *For all $0 < s < \alpha - d/2$, the function u exists as a random variable in*

$$L^2_{\mathbb{P}\mathbb{P}'}(\Omega, \Omega'; \mathbb{H}^s(D)) := \{v : D \times \Omega \times \Omega' \to \mathbb{R} \; s.t. \; \mathbb{E}\mathbb{E}' \|v\|^2_{\mathbb{H}^s(D)} < \infty\}$$

and takes values $(\mathbb{P}, \mathbb{P}')$- almost surely in $\mathbb{H}^s(D)$.

The sense of existence used in Theorem 9 is precisely the same sense in which a Gaussian process exists, where the covariance function forms a kernel for $\mathbb{H}^\alpha(D)$ [Dashti and Stuart, 2015, Theorem 2.10]. As an aside, note the following bound on the average-case accuracy of the point estimate μ provided by the meshless method:

Proposition 10. *Suppose $\alpha > 2\beta + d/2$. Then $\|\mu(\cdot, \omega, \omega') - u(\cdot, \omega, \omega')\|_\infty = \mathcal{O}_{\mathbb{P}\mathbb{P}'}(h^{\beta - \rho - d/2})$.*

Prop. 10 shows that, when both the parameter and the forcing function are drawn at random from their respective distributions, the error of the point estimate μ is controlled by the fill distance. As noted by Novak [1988], such L_∞ error bounds tend to be inefficient by a factor of $\sqrt{n}$, suggesting that a sharper bound could be obtained.

5.2 Posterior Contraction

Let $X = \{\boldsymbol{x}_j\}_{j=1}^n \subset D$ be the locations at which data $\boldsymbol{y} \sim N(\boldsymbol{u}, \boldsymbol{\Gamma})$ are obtained. Define the potential $\Phi_h(\boldsymbol{y}, \theta) = -\log \pi(\boldsymbol{y}|\theta, \boldsymbol{g}, \boldsymbol{b})$ corresponding to the PMM, where we have chosen notation that emphasises the role of the fill distance h. Note that h does not uniquely define Φ_h since several different sets X_0 can give rise to the same fill distance. Then $\Phi_h(\boldsymbol{y}, \cdot) : \Theta \to \mathbb{R}$ is Π_θ-measurable. From Bayes' theorem [Theorem 1.1 Dashti and Stuart, 2015], the posterior distribution of $\theta|\boldsymbol{y}$, denoted by $\Pi_\theta^{\boldsymbol{y},h}$, exists and is absolutely continuous with respect to Π_θ, with Radon–Nikodým derivative

$$\frac{\mathrm{d}\Pi_\theta^{\boldsymbol{y},h}}{\mathrm{d}\Pi_\theta}(\theta) = \frac{1}{Z_h} \exp(-\Phi_h(\boldsymbol{y}, \theta)). \tag{14}$$

where

$$Z_h := \int_\Theta \exp(-\Phi_h(\boldsymbol{y}, \theta))\Pi_\theta(\mathrm{d}\theta).$$

Following similar arguments to Dashti and Stuart [2015] (Section 3.4.3), $Z_h > 0$. This requires the additional assumption that $\mathcal{A}$, $\mathcal{B}$ are non-degenerate, so that the conditional covariance $\Sigma(\theta)$ is finite and positive-definite for all $\theta \in \Theta$.

This section elaborates on the sense in which $\Pi_\theta^{\boldsymbol{y},h}$ approximates the idealised posterior $\Pi_\theta^{\boldsymbol{y}}$ in Eqn. 3 as the fill distance h is decreased to zero.

Firstly we define what it means for the posterior $\Pi_\theta^{\boldsymbol{y},h}$ to contract. Denote $B_\epsilon(\theta_0) := \{\theta \in \Theta : \|\theta - \theta_0\|_\Theta < \epsilon\}$. The posterior $\Pi_\theta^{\boldsymbol{y},h}$ is said to "contract" to the true value θ_0 of the parameter if, for any $\epsilon > 0$, $\ell < 1$, we have $\Pi_{\boldsymbol{y}}[\Pi_\theta^{\boldsymbol{y},h}[B_\epsilon(\theta_0)] \geq l] \to 1$. Some authors also refer to this property as "consistency" [e.g. Vollmer, 2013], while others use stronger definitions [e.g. Ghosal et al., 2000]. A suitably chosen pair (Π_θ, Φ), consisting of prior Π_θ and potential Φ, lead to contraction of the exact posterior $\Pi_\theta^{\boldsymbol{y}}$ under conditions that are now well-studied, see Vollmer [2013] for a comprehensive review.

The question we now ask how fast h must decrease relative to number n of data in order for the posterior $\Pi_\theta^{\boldsymbol{y},h}$ to be asymptotically equal to the exact posterior $\Pi_\theta^{\boldsymbol{y}}$? [A similar question was asked in a non-PN context by Xue et al., 2010]. This is interesting because asymptotic equality implies that the posterior $\Pi_\theta^{\boldsymbol{y},h}$ contracts under the same conditions that are sufficient for contraction in the idealised problem. The result is provided in the following:

Theorem 11 (Contraction of posterior measure to θ_0)**.** *Suppose that the pair (Π_θ, Φ) is such that the exact posterior $\Pi_\theta^{\boldsymbol{y}}$ contracts to the true parameter θ_0. For simplicity, assume independent errors in the observation process; $\boldsymbol{\Gamma} = \gamma^2 \boldsymbol{I}$. Then, if the fill distance scales as $h = o(n^{-1/(\beta-\rho-d/2)})$, the pair (Π_θ, Φ_h) is such that the PN posterior $\Pi_\theta^{\boldsymbol{y},h}$ also contracts to θ_0.*

It is seen that when the true solution u is smooth, β can be taken to be large. Then, the rate at which design points must be added to achieve consistency is slow, and in particular can be orders of magnitude slower than the rate at which data are obtained. This gives us confidence that it is possible, at least in principle, to consider tackling regular problems with relatively small sets of design points. Conversely, when the true solution u is not smooth, $\beta - \rho$ can be arbitrarily close to $d/2$. Then the rate at which design points must be added to achieve consistency can be arbitrarily large. This reflects the intrinsic complexity associated with such non-smooth function spaces.

6 Implementation Details

Here the practical question of implementation is addressed. In Sec. 6.1 we outline a Markov chain Monte Carlo (MCMC) scheme that operates in either finite or infinite dimensional parameter spaces Θ and targets the posterior $\Pi_\theta^{\boldsymbol{y},h}$. In Sec. 6.2 we cast the problem of choosing the locations X_0 as a problem of statistical experimental design. Then Sec. 6.3 extends PMMs to operate in the presence of multiple solutions to the forward problem. The reader may prefer to jump ahead to Sec. 7 and return to Sec. 6 later.

6.1 Infinite Dimensional MCMC

The goal here is to obtain samples from $\Pi_\theta^{\boldsymbol{y},h}$. Intractability of Z_h motivates the use of MCMC techniques; these construct a measure-preserving Markov transition kernel over Θ that can be used to obtain approximately independent samples from $\Pi_\theta^{\boldsymbol{y},h}$. Crucially, MCMC requires knowledge of the Radon–Nikodým derivative only up to a multiplicative constant, avoiding the need to calculate Z_h.

For infinite-dimensional parameter inference problems we propose to use the preconditioned Crank–Nicolson (pCN) algorithm [Cotter et al., 2013, Dashti and Stuart, 2015]. This proceeds as follows: Assume that Θ is a Hilbert space and a Gaussian prior $\Pi_\theta = N(0, C)$ is assigned over Θ. Given potential function Φ_h, Algorithm 1 details the pCN method for constructing a Markov chain θ^i, $i = 1, \ldots, I$, that targets the posterior $\Pi_\theta^{\boldsymbol{y},h}$:

Algorithm 1 (pCN Method)**.**

Pick $\theta^0 \in \Theta$

For $i = 1 \dots I$:

1. Draw $\xi^i \sim N(0, C)$
2. Propose $\theta^* \leftarrow \sqrt{1-\lambda^2}\theta^i + \lambda\xi^i$
3. Compute
$$\alpha \leftarrow \min\left\{1, \frac{\exp(-\Phi_h(\boldsymbol{y}, \theta^*))}{\exp(-\Phi_h(\boldsymbol{y}, \theta^i))}\right\}$$
4. Set $\theta^{i+1} \leftarrow \theta^*$ with probability α; otherwise set $\theta^{i+1} \leftarrow \theta^i$

End.

The parameter λ governs the trade-off between local and global proposals, and should be tuned to achieve a reasonable acceptance probability.

6.2 Experimental Design

The inferences drawn with PMMs are valid from a statistical perspective, regardless of the locations X_0 that are used to implement the method. However the informativeness of the inference will depend upon choice of X_0. This section provides a principled approach to selecting X_0 based on experimental design.

6.2.1 A-Optimal Designs X_0

The selection of the locations X_0 that define the meshless method is subject to competing considerations. On one hand, the locations should be close to the locations X where data are obtained, so that uncertainty in the solution is reduced at these locations. On the other hand, the locations must be sufficiently well spread in the domain $D \cup \partial D$ in order to properly encode the governing equation(s) and the boundary condition(s). Two additional requirements arise from a practical perspective: Firstly, we require sparsity; i.e. that $|X_0|$ is not too large. Recall $X_0 = X_0^{\mathcal{A}} \cup X_0^{\mathcal{B}}$, where $X_0^{\mathcal{A}} \subset D$ contains $m_{\mathcal{A}}$ elements and $X_0^{\mathcal{B}} \subset \partial D$ contains $m_{\mathcal{B}}$ elements. For efficient computer code it is desirable that the numbers $m_{\mathcal{A}}$ and $m_{\mathcal{B}}$ of locations be determined *a priori* and then held fixed, so that matrix dimensions do not need to be varied. Secondly, the method used to select locations must not be computationally intensive.

The approach pursued is to cast the selection of locations as a problem of statistical experimental design. This is made possible by the probabilistic formulation of the meshless method. Below we write $\boldsymbol{\Sigma}(\theta, X_0)$ to emphasise that the posterior covariance matrix for the solution vector $\boldsymbol{u}$ depends on both the parameter θ and the choice of design X_0. For different values of the parameter θ, different designs X_0 will in general be required. Indeed, it is natural that X_0 should depend on the parameter θ, since both the governing equation(s) and the boundary condition(s) can depend on θ.

In some sense we want the entries of $\boldsymbol{\Sigma}(\theta, X_0)$ to be small; write $L[\boldsymbol{\Sigma}]$ for a generic loss function that is a function of $\boldsymbol{\Sigma}$ only. Define an optimal design $X_0^*(\theta)$ for fixed parameter θ to satisfy:

$$X_0^*(\theta) \in \underset{\substack{X_0^{\mathcal{A}} \subset D,\ X_0^{\mathcal{B}} \subset \partial D \\ |X_0^{\mathcal{A}}| = m_{\mathcal{A}},\ |X_0^{\mathcal{B}}| = m_{\mathcal{B}}}}{\arg\min} L[\boldsymbol{\Sigma}(\theta, X_0)]. \tag{15}$$

Particular choices of L are suggested from Bayesian decision theory. For example, the standard optimality criteria "A-optimality" seeks to minimise $L[\boldsymbol{\Sigma}] = \mathrm{Tr}[\boldsymbol{\Sigma}]$, while another standard criteria, "D-optimality", takes $L(\boldsymbol{\Sigma}) = \det[\boldsymbol{\Sigma}]$. Both A- and D-optimality can be motivated from a Hilbert-space perspective on posterior uncertainty. In particular, the A-optimality criteria above is equivalent to minimising the trace of the posterior covariance operator $\int_D \sigma(\boldsymbol{x})^2 \mathrm{d}\boldsymbol{x}$, while D-optimality minimises the volume of the uncertainty ellipsoid [Alexanderian et al., 2016a].

6.2.2 Related Literature

In this paper A-optimal experimental designs are pursued, to limit scope. In this case the solution to Eqn. 15 is in general unavailable and numerical minimisation is required. A hybrid strategy, that combines the Approximate Coordinate Exchange (ACE) algorithm of Overstall et al. [2015] with Bayesian Optimisation [Mockus et al., 1978] performed well at this task. In particular, the ACE approach elegantly interlaces with MCMC: For one iteration of the MCMC the difference between θ^{i+1} and θ^i will usually be small. It is therefore reasonable to expect that $X_0^*(\theta^i)$ will be a sensible approximation to $X_0^*(\theta^{i+1})$. The ACE algorithm. at iteration i of the MCMC, takes a numerical approximation of $X_0^*(\theta^i)$ as a starting point and performs a local search in design space in order to locate an approximate $X_0^*(\theta^{i+1})$. This approach avoids the need to repeatedly solve challenging multi-variate optimisation problems *de novo* at each iteration of the MCMC. Computational overhead can be further reduced by operating on approximations to the loss function and employing efficient optimisation routines based on Bayesian Optimisation.

Related work on experimental design in inverse problems is more classical, in the sense that numerical error is assumed to be negligible and instead one aims to select the locations X of sensors, that will be used to obtain the data $\boldsymbol{y}$, in order to minimise expected posterior uncertainty over the parameter θ. In this context, recent work includes a series of papers by Alexanderian et al. [2016a, 2014, 2016b]. In addition, recent work by Gorodetsky and Marzouk [2015] applies these techniques to Gaussian process regression using the integrated variance criterion to determine optimal sensor locations; this is equivalent to the A-optimal approach which we pursue here. The method used to attain the optimal design is to minimise a Monte-Carlo estimate of the objective function using gradient-based optimizers, rather than ACE.

From the meshless methods literature, Ling et al. [2006] and Ling and Schaback [2008] both consider the use of greedy algorithm. to select locations X in order to minimise a criterion relating to numerical error in the forward problem. These papers differ to ours in several respects. Firstly, the context is asymmetric collocation, rather than symmetric collocation; secondly the formulation is not probabilistic and does not have the associated interpretation as a problem in experimental design. Finally inverse problems are not considered.

To conclude this section, note that the direct approach of minimising uncertainty over the parameter θ appears to be challenging in this framework. The present proposal, to minimise uncertainty over the solution vector $\boldsymbol{u}$ at each value of the parameter θ, provides a practical approach that acts as a proxy for uncertainty in the parameter.

6.3 Extension to a Class of Semi-linear PDEs

An important motivation for probabilistic numerical solvers comes from inverse problems that involve a non-linear forward model. At present little is known about the performance of meshless methods in this setting, and in a wider context such inverse problems are not widely studied. Non-linear problems abound in the applied sciences and FEA methods for these problems require substantially more computational effort compared to the linear case. There is thus a strong computational motivation for exploiting meshless methods in many non-linear inverse problems. However, for inferences to be valid a more detailed characterisation of numerical error is required than is currently available in the literature. In this section the framework of PMM is extended to the case when case when the underlying PDE model is non-linear. To limit scope, the focus here is on a particular class of non-linear PDEs, known as "semi-linear" PDEs, that are rich enough to exhibit canonical non-linear behaviour (e.g. multiple solutions), whilst also permitting tractable computation.

6.3.1 A Latent Variables Approach For Semi-Linear PDEs

Previous sections considered the case where the operator $\mathcal{A}$ is linear. Here we generalise to operators of the form $\mathcal{A} = \mathcal{A}_1 + \cdots + \mathcal{A}_m$, where each of the $\mathcal{A}_j$ is either (i) a linear differential operator, or (ii) a (possibly non-linear) monotonic operator. This class is motivated by the observation that monotonic operators are invertible. Below this invertibility is exploited to reduce the system to a linear system, to which above methods can be applied.

The approach is best described by example; Consider the steady-state Allen-Cahn equation, which

has been extensively used to model the boundaries between phases in alloys.

$$-\theta\nabla^2 u(\boldsymbol{x}) + \theta^{-1}(u(\boldsymbol{x})^3 - u(\boldsymbol{x})) = g(\boldsymbol{x}).$$

This is a semi-linear PDE with linear differential operator $\mathcal{A}_1 u = -\theta\nabla^2 u - \theta^{-1}u$ and monotonic operator $\mathcal{A}_2 u = \theta^{-1}u^3$.

In the case of $m = 2$ we have $\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2$ and the indirect, non-linear observations $\mathcal{A}u(\boldsymbol{x}_i) = g(\boldsymbol{x}_i)$ can be decomposed into direct observations by introducing a latent function $z(\boldsymbol{x}_i)$ such that $\mathcal{A}_1 u(\boldsymbol{x}_i) = z(\boldsymbol{x}_i)$ and $\mathcal{A}_2 u(\boldsymbol{x}_i) = g(\boldsymbol{x}_i) - z(\boldsymbol{x}_i)$. As a concrete example, for the Allen-Cahn system,

$$\begin{aligned} -\theta\nabla^2 u(\boldsymbol{x}_i) - \theta^{-1}u(\boldsymbol{x}_i) &= z(\boldsymbol{x}_i) \\ \theta^{-1}u(\boldsymbol{x}_i)^3 &= g(\boldsymbol{x}_i) - z(\boldsymbol{x}_i). \end{aligned}$$

The final equation can be inverted to produce $u(\boldsymbol{x}) = (\theta(g(\boldsymbol{x}_i) - z(\boldsymbol{x}_i)))^{1/3}$, which leads to a system of equations that is linear in the solution u, depending on the unknown function z, and can be solved using the methods introduced in previous sections. However, to make inferences on both the actual solution u in the forward problem, and θ in the inverse problem, we must be able to efficiently marginalise the unknown latent function z. The proposed computational approach is discussed in Sec. 6.3.3, but first in Sec. 6.3.2 the arguments of this section are made formal.

6.3.2 Conditional Measure With Latent Variables

To keep the notation under control, details are presented for the simplest case where $\mathcal{A}_1$ is a linear differential operator and $\mathcal{A}_2$ is a monotonic function such that $\mathcal{A}_2^{-1}$ is known. In this case the previous notation is extended as follows:

$$\boldsymbol{z} = \mathcal{A}_1\boldsymbol{u}, \quad \mathcal{L} = \begin{bmatrix} \mathcal{A}_1 \\ \mathcal{I} \\ \mathcal{B} \end{bmatrix}, \quad \bar{\mathcal{L}} = \begin{bmatrix} \bar{\mathcal{A}}_1 & \mathcal{I} & \bar{\mathcal{B}} \end{bmatrix}$$

where $\mathcal{I} : H(D) \to H(D)$ is the identity operator. Here we have written $\boldsymbol{z}$ for the $m_{\mathcal{A}} \times 1$ vector with jth element z_i. To simplify the notation in this section, dependence on the parameter θ is suppressed; this will be re-introduced in Sec. 6.3.3. For non-linear $\mathcal{A}_2$, the marginal probability distribution $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ will no longer be Gaussian. However, when $\boldsymbol{z}$ is included, $\Pi_u^{\boldsymbol{z},\boldsymbol{g},\boldsymbol{b}}$, representing the conditional distribution of $u|\boldsymbol{z},\boldsymbol{g},\boldsymbol{b}$, is Gaussian and its finite dimensional distribution at the test points X takes the form

$$\boldsymbol{u}|\boldsymbol{z},\boldsymbol{g},\boldsymbol{b} \sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$$

where

$$\begin{aligned} \boldsymbol{\mu} &= \bar{\mathcal{L}}\hat{\boldsymbol{K}}(X,X_0)[\mathcal{L}\bar{\mathcal{L}}\hat{\boldsymbol{K}}(X_0)]^{-1}[\boldsymbol{z}^T\ \mathcal{A}_2^{-1}(\boldsymbol{g}-\boldsymbol{z})^T\ \boldsymbol{b}^T]^T \\ \boldsymbol{\Sigma} &= \hat{\boldsymbol{K}}(X) - \bar{\mathcal{L}}\hat{\boldsymbol{K}}(X,X_0)[\mathcal{L}\bar{\mathcal{L}}\hat{\boldsymbol{K}}(X_0)]^{-1}\mathcal{L}\hat{\boldsymbol{K}}(X_0,X). \end{aligned}$$

This observation suggests an efficient sampling scheme for $\Pi_u^{\boldsymbol{z},\boldsymbol{g},\boldsymbol{b}}$ can be constructed, with samples from $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ appearing as a sample marginal. Details are provided in Sec. 6.3.3 below.

6.3.3 Pseudo-Marginal MCMC

Turning to the inverse problem, we now present an MCMC scheme for sampling the joint distribution of the solution vector $\boldsymbol{u}$, the latent states $\boldsymbol{z}$ and the parameter θ. Our primary interest is in θ, so a natural approach is to focus sampling effort on θ via Pseudo-Marginal MCMC [Andrieu and Roberts, 2009]. For simplicity we restrict to finite dimensional $\Theta \subseteq \mathbb{R}^M$. This is to avoid technical issues associated with infinite-dimensional parameters in Pseudo-Marginal MCMC.

When the forward model is non-linear, the data-likelihood is an intractable integral

$$\pi(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{b},\theta) = \int \pi(\boldsymbol{z}|\theta) \underbrace{\int \pi(\boldsymbol{y}|\boldsymbol{u})\pi(\boldsymbol{u}|\boldsymbol{z},\boldsymbol{g},\boldsymbol{b},\theta)\mathrm{d}\boldsymbol{u}}_{(*)} \mathrm{d}\boldsymbol{z} \tag{16}$$

where we note that the interior integral $(*)$ is available in closed form as before in Sec. 3.3.2. An improper uniform measure $\pi(\boldsymbol{z}|\theta) = 1$ was taken; note that the impropriety of $\pi(\boldsymbol{z}|\theta)$ implies that Eqn. 16 is only defined up to a multiplicative constant.

To construct an estimate to Eqn. 16, an importance density $r(\boldsymbol{z}|\boldsymbol{y}, \theta)$ was constructed, defined in Sec. 6.3.4, and rewrite Eqn. 16 as

$$\pi(\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{b}, \theta) = \int \frac{1}{r(\boldsymbol{z}|\boldsymbol{y}, \theta)} \int \pi(\boldsymbol{y}|\boldsymbol{u})\pi(\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{g}, \boldsymbol{b}, \theta)\mathrm{d}\boldsymbol{u}\; r(\boldsymbol{z}|\boldsymbol{y}, \theta)\mathrm{d}\boldsymbol{z}.$$

Then an explicit, almost-surely positive, unbiased estimator $\hat{\pi}(\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{b}, \theta)$ for the likelihood $\pi(\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{b}, \theta)$ can be constructed as follows:

1. $\boldsymbol{z}^* \sim r(\boldsymbol{z}|\boldsymbol{y}, \theta)$
2. $\hat{\pi}(\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{b}, \theta) \leftarrow \frac{1}{r(\boldsymbol{z}^*|\boldsymbol{y}, \theta)} \int \pi(\boldsymbol{y}|\boldsymbol{u})\pi(\boldsymbol{u}|\boldsymbol{z}^*, \boldsymbol{g}, \boldsymbol{b}, \theta)d\boldsymbol{u}$

The Pseudo-Marginal approach constructs a Markov chain θ^i, $i = 1, \ldots, I$, as follows: Specify a position-dependent proposal density $q(\theta^*|\boldsymbol{y}, \theta)$, for example a random walk.

Algorithm 2 (Pseudo-Marginal MCMC)**.**
Pick $\theta^0 \in \Theta$ and simulate $\hat{\pi}^0 = \hat{\pi}(\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{b}, \theta^0)$.
For $i = 1, \ldots, I$:

1. Propose $\theta^* \sim q(\theta^*|\boldsymbol{y}, \theta^i)$
2. Simulate $\hat{\pi}^* \leftarrow \hat{\pi}(\boldsymbol{y}|\boldsymbol{g}, \boldsymbol{b}, \theta^i)$
3. Compute
$$\alpha \leftarrow \min\left\{1, \frac{\pi(\theta^*)}{\pi(\theta^i)} \cdot \frac{\hat{\pi}^*}{\hat{\pi}^i} \cdot \frac{q(\theta^i|\boldsymbol{y}, \theta^*)}{q(\theta^*|\boldsymbol{y}, \theta^i)}\right\}$$
4. Set $\theta^{i+1} \leftarrow \theta^*$, $\hat{\pi}^{i+1} \leftarrow \hat{\pi}^*$ with probability α; else set $\theta^{i+1} \leftarrow \theta^i$, $\hat{\pi}^{i+1} \leftarrow \hat{\pi}^i$

End.

6.3.4 Multiple Solutions

The performance of Algorithm 2 depends on how close $r(\boldsymbol{z}|\boldsymbol{y}, \theta)$ can be made to the latent variable posterior $\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{g}, \boldsymbol{b}, \theta)$. Our preference is to employ adaptive proposals that automatically take into account the varying nature of the parameter θ. The construction is complicated by the fact that non-linear PDEs are not guaranteed a unique solution, leading to multiple values of $\boldsymbol{z}$ which are each consistent with some solution of the PDE. Here details are provided for the choice of importance density.

In this work a known, fixed number of solutions are assumed to exist for the non-linear forward problem and that these each vary smoothly with θ. Then our strategy is to augment the MCMC procedure with an additional parameter, i, describing the solution index. A joint proposal over (θ, i) then operates in a Metropolis-within-Gibbs sampler. The proposal distribution for i was taken to be uniform for simplicity. To be specific, consider a semi-linear PDE of the form $\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2$ as in Sec. 6.3.2. For fixed (θ, i), a crude approximation $\hat{\boldsymbol{u}}_i$ to the solution $\boldsymbol{u}_i$, indexed by i of the PDE, was obtained. These approximations need only be crude, and so for compuational efficiency can be obtained using (e.g.) FEA on a coarse mesh. In this work the recent "deflation" approach of Farrell et al. [2015] was applied. Finally we take $r(\boldsymbol{z}|\boldsymbol{y}, \theta, i) = N(\mathcal{A}_2^{-1}\hat{\boldsymbol{u}}_i, \mathcal{C})$ for an appropriate covariance $\mathcal{C}$. We emphasise that these choices affect only the mixing properties of the MCMC, not the posterior distributions that are the central focus of this work.

The total computational cost of this method is equivalent to a single application of the deflation technique, following by the cost of sampling from the importance distribution to obtain an unbiased estimate of the likelihood. This latter cost is minimal, as the matrices required to compute $\boldsymbol{u}|\theta, \boldsymbol{z}$ are predominantly independent of $\boldsymbol{z}$; as a result, using many samples from $\boldsymbol{z}$ to approximate the data-likelihood is computationally inexpensive.

7 Experiments

This section deploys PMMs to solve two relevant and challenging inverse problems. The first is an application to Electrical Impedance Tomography (EIT), an infinite-dimensional inverse problem with linear governing equations. The second is a more challenging application to the steady-state Allen–Cahn equation, a finite-dimensional inverse problem with non-linear governing equations. These two applications combine to illustrate the salient properties of the method.

These experiments can be reproduced using the Python library hosted at `https://github.com/jcockayne/bayesian_pdes`.

7.1 Application to Electrical Impedance Tomography

EIT is a technique used for medical imaging in which an electrical current is passed through electrodes attached to a patient. The statistical challenge is to use measurements taken on the exterior to determine interior conductivity, for example for the purposes of detecting brain tumours. Recent work by Dunlop and Stuart [2015] shows that EIT is well-posed as a Bayesian inverse problem, while Iglesias [2016] proposes a fast, derivative-free method for finding the posterior mean for the conductivity field. EIT is posed as an elliptic PDE which is linear in the solution $u(\boldsymbol{x})$:

$$-\nabla \cdot (a(\boldsymbol{x})\nabla u(\boldsymbol{x})) = 0 \qquad \boldsymbol{x} \in D \tag{17}$$

$$c_i(\boldsymbol{x})\frac{\partial u}{\partial n}(\boldsymbol{x}) = c(\boldsymbol{x}) \quad \boldsymbol{x} \in \partial D \tag{18}$$

The formulation above is the simplified continuum model as originally posed in Calderón [1980]. In this setting we assume that a continuous stimulating current is applied on ∂D. As noted in Dunlop and Stuart [2015], stimulation patterns must be linearly independent to assist in inferring the true conductivity field; these patterns are modelled by $c_i(\boldsymbol{x})$, $i = 1 \dots n$. For each of these we measure the value of the voltage $u(\boldsymbol{x})$ on the boundary, and label this voltage $u_i(\boldsymbol{x})$, $\boldsymbol{x} \in \partial D$. The inverse problem is to infer $a(\boldsymbol{x})$ from the observations $(c_i(\boldsymbol{x}), u_i(\boldsymbol{x}))$, $i = 1 \dots n$. In practice we discretise the functions c_i, u_i so that observations of each are made at point electrode locations $\{\boldsymbol{x}_j^{\mathcal{B}}\}_{j=1}^{m_{\mathcal{B}}} \subset \partial D$.

To pose this as a Bayesian inverse problem, consider the unknown parameter of interest to be $\theta(\boldsymbol{x}) = \log a(\boldsymbol{x})$ and endow θ with a Gaussian prior. The nature of the observations, coupled with the complexity of the conductivity field $a(\boldsymbol{x})$, makes sampling from the posterior challenging. The quantification of uncertainty provided by PMMs here is attractive. Indeed, these methods permit a cheap, coarse discretisation to be used while still providing a rigorous statistical inference for the conductivity field. Below it is shown that when the conductivity field is recovered using a coarse probabilistic solver for the forward problem, the posterior variance of the field is appropriately inflated.

Data were generated on the domain D, here a circle with unit radius centered at the origin. The diffusivity field used as ground truth is shown in Fig. 5, on a fine grid of 316 points. A number $n = 31$ of stimulation patterns were generated by applying currents of 0.1 to adjacent electrodes, uniformly spaced on ∂D:

$$c_i(x_j^{\mathcal{B}}) = \begin{cases} 0.1 & \text{if } i = j+1 \\ -0.1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The observations obtained by solving this model with the true diffusivity field are then corrupted with independent Gaussian noise, with covariance $\boldsymbol{\Gamma} = 0.0001^2\boldsymbol{I}$.

The PMM was applied based on the exponentiated quadratic kernel

$$\hat{k}(\boldsymbol{x}, \boldsymbol{x}'; \ell) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2}{2\ell^2}\right). \tag{19}$$

The length-scale parameter ℓ was set by maximising the data-likelihood rather than sampled from, as in the previous example; this is for performance reasons as, for fixed ℓ, sampling a new $\theta(\boldsymbol{x})$ does not require costly recomputation of the large matrices involved in the posterior mean and covariance; only multiplying them by the new values in $\theta(\boldsymbol{x})$.

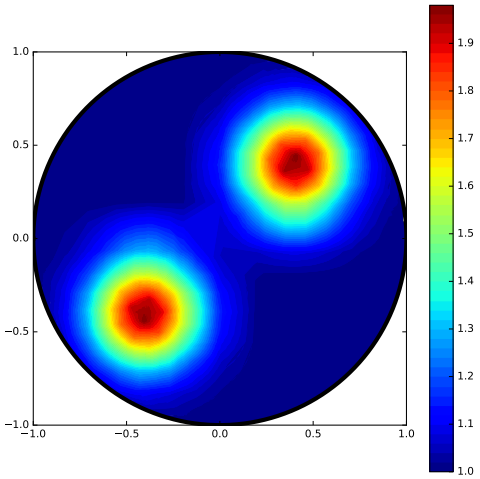


Abbildung 5: Application to EIT: True conductivity field $a_0(\boldsymbol{x}) = \log \theta_0(\boldsymbol{x})$ for the simulated data.

An initial design X_0 was created by placing $m_{\mathcal{A}}$ points at random throughout the domain, in addition to the boundary points $\{\boldsymbol{x}_j^{\mathcal{B}}\}_{j=1}^{m_{\mathcal{B}}}$. The algorithm described in Sec. 6.2 was then applied to approximate an optimal design. Initial and optimal designs for $m_{\mathcal{A}} = 80$ points are depicted in Fig. 6. Note that this design appears approximately space-filling; we conjecture that this is to some extent an artefact of using an isotropic kernel rather than the (unavailable) natural kernel.

To sample from the posterior conductivity we apply the pCN algorithm as given in Sec. 6.1, with $\lambda = 0.01$. The recovered conductivity field is depicted in Fig. 7. This shows how, as the number of design points $m_{\mathcal{A}}$ is increased, the recovered field increasingly resembles the original field from which simulated data was constructed as error due to discretisation vanishes and converges on the exact Bayesian estimate (which for this example was a good approximation to the true conductivity field).

A comparison between the posterior variance when using the PMM, compared with using standard collocation, is shown in Fig. 8. This shows inflated posterior variance when using the probabilistic method, as expected.

7.2 Application to the steady-state Allen–Cahn Equation

The second application considered the steady-state Allen–Cahn equation

$$-\theta \, \nabla^2 u + \theta^{-1}(u^3 - u) = 0$$

[Allen and Cahn, 1979] on $D = (0,1)^2$ with boundary conditions

$$\begin{aligned} u &= +1 \quad \text{on } x_1 \in \{0,1\},\ 0 < x_2 < 1 \\ u &= -1 \quad \text{on } x_2 \in \{0,1\},\ 0 < x_1 < 1. \end{aligned}$$

The data-generating value $\theta = 0.04$ was used for this simulation study. This specific choice is interesting because it leads to three distinct solutions of the PDE. This set-up was recently considered in Farrell et al. [2015] where the deflation technique was used to determine all solutions u_1 ("negative stable"), u_2 ("unstable") and u_3 ("positive stable"), which are shown in Figure 9. The existence of multiple solutions provides strong motivation for the quantitative description of solver error that is provided by PMMs.

Data were generated from the "unstable" solution u_2 to this system; in total $n = 16$ observations were taken on a 4×4 grid in the interior of the domain and each observation was corrupted with Gaussian noise with covariance $\boldsymbol{\Gamma} = 0.1^2 \boldsymbol{I}$. For the inverse problem a uniform prior for θ over $(0.02, 0.15)$ was used, which ensures three distinct solutions to the PDE for each value of θ.

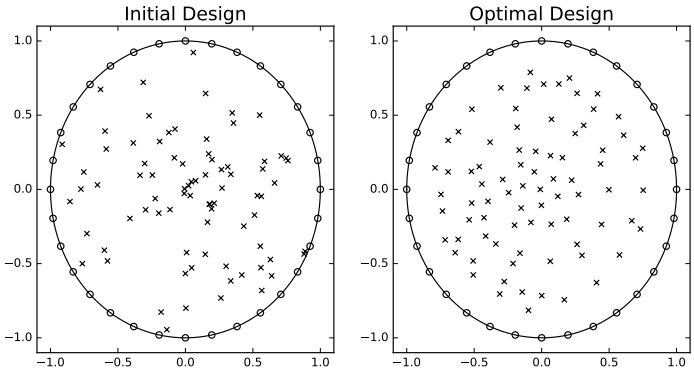
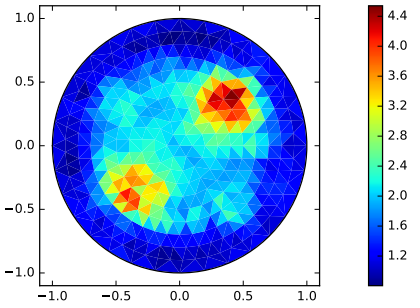
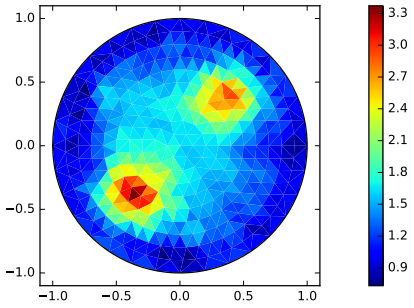


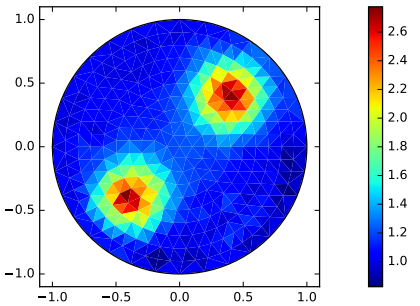
Abbildung 6: Application to EIT: Initial vs. optimal design, $m_{\mathcal{A}} = 80$ points. Black crosses represent design points, which are allowed to vary, while circles on the boundary represent the fixed design points, which in this case coincided with the location of the observations.



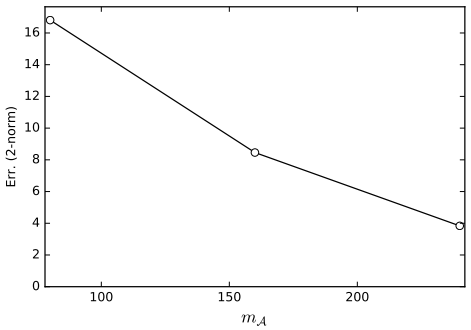
(a) $m_{\mathcal{A}} = 80$



(b) $m_{\mathcal{A}} = 160$

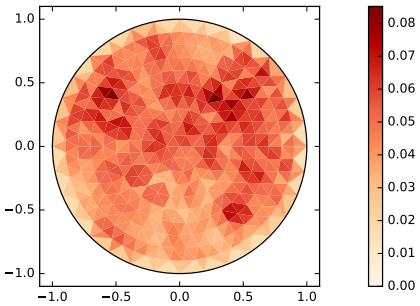


(c) $m_{\mathcal{A}} = 240$

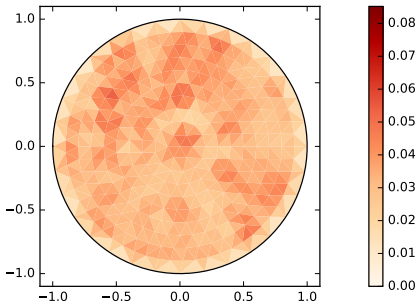


(d) L_2-accuracy of the posterior mean.

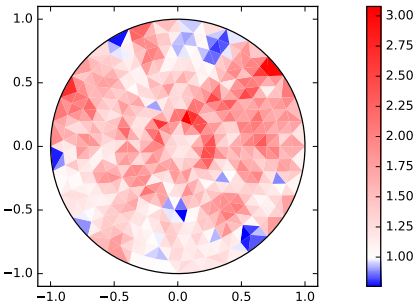
Abbildung 7: Application to EIT: Recovered conductivity fields (posterior mean μ) for varying number $m_{\mathcal{A}}$ of design points. (a) $m_{\mathcal{A}} = 80$, (b) $m_{\mathcal{A}} = 160$, (c) $m_{\mathcal{A}} = 240$. Panel (d) plots the error $\|\mu - u_0\|_2$ against $m_{\mathcal{A}}$.



(a) Probabilistic meshless method



(b) Standard meshless method



(c) Ratio of 8a to 8b.

Abbildung 8: Application to EIT: Recovered conductivity fields (posterior variance σ^2) for $m_{\mathcal{A}} = 80$. In (c), blue tinted areas indicate lower variance in the probabilistic meshless method, while red indicates larger.

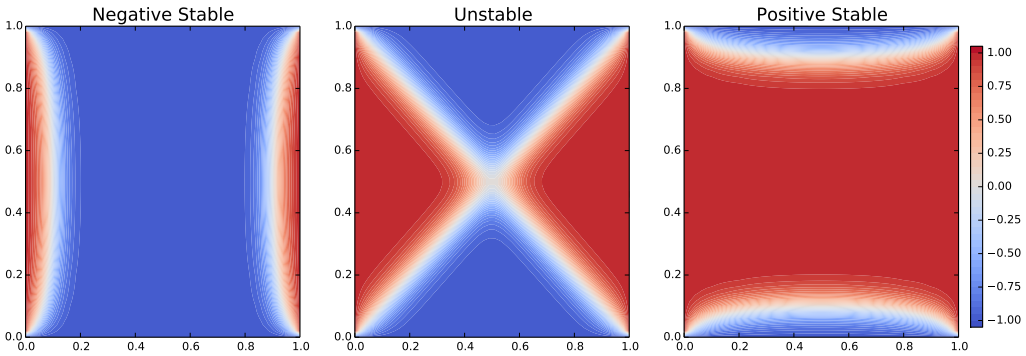


Abbildung 9: Application to Allen–Cahn: Three distinct solutions of the PDE, here shown at the parameter value $\theta = 0.04$.

For the PMM, the reference measure was based on the exponentiated quadratic kernel given in Eqn. 19. The length scale ℓ was assigned a standard half-Cauchy prior and was marginalised out, following the recommendation of Gelman [2006]. MCMC trace plots for (θ, ℓ) are shown in Fig. 10.

Experimental designs were computed as discussed in Sec. 6.2. In practice it was sufficient to fix one design for all θ, in this case based on the the solution at $\theta = 0.04$, which was essentially space-filling and shown in Fig. 11. This highlights an inefficiency in the assumption of isotropic covariance; in the case of the Allen–Cahn system, it is clear that the three solutions are highly non-uniform throughout the domain, and that the area of highest variation is distinct for each solution. This suggests that more efficient methods might be found by use of a non-isotropic covariance.

Posterior distributions, generated using the methodology, are shown in Fig. 12 based on $m_{\mathcal{A}} \in \{5, 10, 20, 40, 80\}$. Posteriors generated by using FEA are included for comparsion.

Results showed that both the PMM and the FEA method posess similar bias for smaller numbers of design points or coarser meshes, respectively. However while the posteriors generated using FEA are sharply peaked around incorrect values for coarse meshes, a larger, more appropriate variance is reported by the PMM.

8 Discussion

In this paper the notion of a probabilistic meshless method was introduced. While standard numerical solvers return a single element from the solution space, a probabilistic solver instead returns a full distribution over the solution space, with a view to capturing epistemic uncertainty due to discretisation error. These results provide theoretical support to this approach and demonstrate that the quantification of numerical error provided by these methods is sensible, ensuring posterior contraction in an appropriate limit. Moreover, through two applications, the integration of a probabilistic model for numerical error into the inverse problem was shown to enable valid statistical inferences to be drawn, overcoming the problems of bias and/or over-confidence that result from the standard approach of simply neglecting numerical error.

Attention was restricted to strong-form solution of stationary PDEs. In future we seek to relax this restriction to examine parabolic, time-evolving PDEs. To restrict the presentation we did not consider combining probabilistic meshless methods with emulation of the data-likelihood [Stuart and Teckentrup, 2016] or reduced-model approached [Cohen and DeVore, 2015] in solution of the inverse problem. This would provide an obvious and immediate reduction in computational cost in the examples studied in this paper.

The method proposed here shares similarities with the meshless construction recently presented in Owhadi [2016]. In that work, Owhadi shows how the meshless approach can be made to operate locally in space (using a construction called "gamblets"). This is shown to speed up computation and permit sophisticated multi-level schemes. However improved computational performance requires truncation of the gamblets to produce a local basis; the fully probabilistic interpretation is of comparable computational

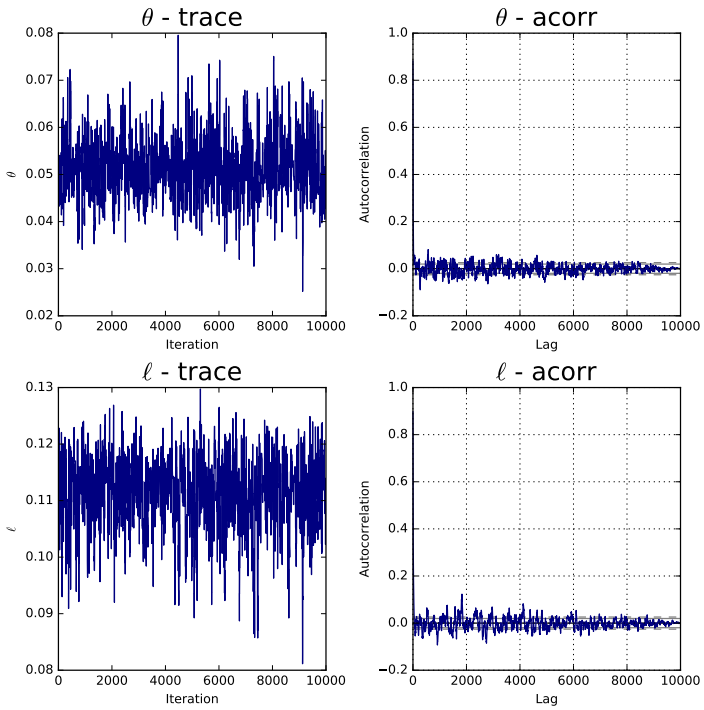


Abbildung 10: Application to Allen–Cahn: MCMC trace and autocorrelation plots for the unknown parameter θ and the kernel length scale ℓ, based on a probabilistic meshless method with $m_{\mathcal{A}} = 20$ design points.

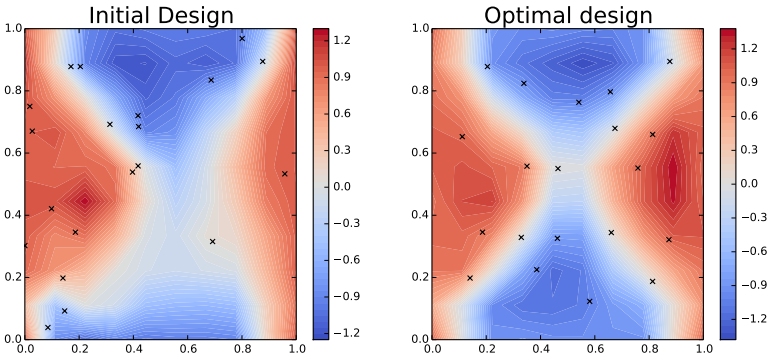
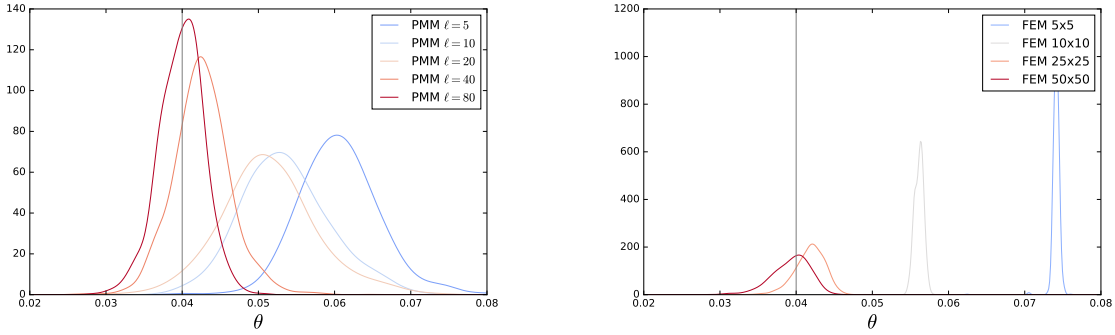


Abbildung 11: Application to Allen–Cahn: Initial vs. optimal design, $m_{\mathcal{A}} = 20$ points. The heat map shown is the mean function of the conditional measure $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$ for the unstable solution, the accuracy of which is controlled by the quality of the design. Computation was based on 500 samples of the latent variable $\boldsymbol{z}$.



(a) Probabilistic.

(b) Standard

Abbildung 12: Application to Allen–Cahn: Posterior distributions for θ, computed with (a) a probabilistic meshless method (PMM), (b) a finite element method (FEM). In (a) the legend denotes the number $m_{\mathcal{A}}$ of design points, while in (b) the legend denotes the size of the mesh used by the FEM.

complexity to the approach described herein.

A fundamental distinction of the present paper is that uncertainty due to numerical error was propagated into the inverse problem, whereas Owhadi [2016] does not make use of the probability model beyond observing its Bayesian interpretation. A second, less fundamental, distinction is that, although Owhadi's method is in principle meshless, computations are performed on a grid, so that there is no analogue of the experimental design approach pursued in this paper. Future work will aim to leverage the strengths of both approaches.

Appendices

Proof of Prop. 1. First, we claim that Λ admits the following characterisation:

$$\Lambda(\boldsymbol{x}', \boldsymbol{x}) = \int_D G_\Lambda(\boldsymbol{x}', \boldsymbol{z}) G_\Lambda(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z}$$

where G_Λ is the Green's function for the system $(\mathcal{A}_\Lambda, \mathcal{B}_\Lambda)$ defined in Eqn. 5. Indeed, since the reproducing kernel must be unique, it suffices to verify that Eqn. 8 is reproducing in $H_\Lambda(D)$:

$$\begin{aligned} \left\langle g, \int_D G_\Lambda(\cdot, \boldsymbol{z}) G_\Lambda(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z} \right\rangle_\Lambda &= \left\langle \mathcal{A}_\Lambda g, \mathcal{A}_\Lambda \int_D G_\Lambda(\cdot, \boldsymbol{z}) G_\Lambda(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z} \right\rangle_{L^2(D)} \\ &= \langle \mathcal{A}_\Lambda g, G_\Lambda(\boldsymbol{x}, \cdot) \rangle_{L^2(D)} \;=\; g(\boldsymbol{x}) \end{aligned}$$

where the final line follows from the definition of the Green's function G_Λ and the fact that

$$\begin{aligned} \mathcal{A}_\Lambda \int_D G_\Lambda(\cdot, \boldsymbol{z}) G_\Lambda(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z} &= \int_D [\mathcal{A}_\Lambda G_\Lambda(\cdot, \boldsymbol{z})] G_\Lambda(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\ &= \int_D [\delta(\cdot - \boldsymbol{z})] G_\Lambda(\boldsymbol{x}, \boldsymbol{z}) \mathrm{d}\boldsymbol{z} \;=\; G_\Lambda(\boldsymbol{x}, \cdot). \end{aligned}$$

Second, using the definition of the Green's functions G and G_Λ,

$$\begin{aligned}
\langle v, k(\cdot,\boldsymbol{x})\rangle_k &= \int_D [\mathcal{A}_\Lambda \mathcal{A} v(\boldsymbol{z})][\mathcal{A}_\Lambda \mathcal{A} k(\boldsymbol{z},\boldsymbol{x})]\mathrm{d}\boldsymbol{z} \\
&= \int_D \int_D \int_D [\mathcal{A}_\Lambda \mathcal{A} v(\boldsymbol{z})][\mathcal{A}_\Lambda \underbrace{\mathcal{A} G(\boldsymbol{z},\boldsymbol{z}')}_{\delta(\boldsymbol{z}-\boldsymbol{z}')} G(\boldsymbol{x},\boldsymbol{z}'')\Lambda(\boldsymbol{z}',\boldsymbol{z}'')]\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\mathrm{d}\boldsymbol{z}'' \\
&= \int_D \int_D [\mathcal{A}_\Lambda \mathcal{A} v(\boldsymbol{z})][\underbrace{\mathcal{A}_\Lambda \Lambda(\boldsymbol{z},\boldsymbol{z}'')}_{G_\Lambda(\boldsymbol{z},\boldsymbol{z}'')} G(\boldsymbol{x},\boldsymbol{z}'')]\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'' \\
&= \int_D \underbrace{\int_D [\mathcal{A}_\Lambda \mathcal{A} v(\boldsymbol{z})] G_\Lambda(\boldsymbol{z},\boldsymbol{z}'')\mathrm{d}\boldsymbol{z}}_{\mathcal{A} v(\boldsymbol{z}'')} G(\boldsymbol{x},\boldsymbol{z}'')\mathrm{d}\boldsymbol{z}'' \;=\; v(\boldsymbol{x}).
\end{aligned}$$

This proves the reproducing property in $H_k(D)$. We also need to check boundedness of the evaluation function:

$$|v(\boldsymbol{x})| = |\langle v, k(\cdot,\boldsymbol{x})\rangle_k| \le \langle v, v\rangle_k^{1/2} \langle k(\cdot,\boldsymbol{x}), k(\cdot,\boldsymbol{x})\rangle_k^{1/2} = \langle v, v\rangle_k^{1/2} k(\boldsymbol{x},\boldsymbol{x})^{1/2},$$

which follows from the Cauchy–Schwarz inequality. □

Proof of Prop. 2. This is essentially Prop. 3.1 of Owhadi [2015]. Noting that

$$u(\boldsymbol{x}) = \int_D G(\boldsymbol{x},\boldsymbol{z}) g(\boldsymbol{z})\mathrm{d}\boldsymbol{z},$$

we have

$$\begin{aligned}
\mathbb{E}[u(\boldsymbol{x})u(\boldsymbol{x}')] &= \mathbb{E}\left[\int_D \int_D G(\boldsymbol{x},\boldsymbol{z}) g(\boldsymbol{z}) G(\boldsymbol{x}',\boldsymbol{z}') g(\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\right] \\
&= \int_D \int_D G(\boldsymbol{x},\boldsymbol{z}) G(\boldsymbol{x}',\boldsymbol{z}')\mathbb{E}[g(\boldsymbol{z})g(\boldsymbol{z}')]\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \\
&= \int_D \int_D G(\boldsymbol{x},\boldsymbol{z}) G(\boldsymbol{x}',\boldsymbol{z}')\Lambda(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' = k(\boldsymbol{x},\boldsymbol{x}').
\end{aligned}$$

Moreover, the stochastic process is well-defined since, from Prop. 1, the covariance function k is positive definite over $D \times D$. □

Proof of Prop. 4. Since D is open and bounded and $\tilde{k}$ is symmetric positive definite, Mercer's theorem [Steinwart and Scovel, 2012] guarantees the existence of a countable set of eigenvalues and eigenvectors $\{\lambda_i\}$ and $\{e_i\}$ such that $\lambda_1 \ge \lambda_2 \ge \cdots > 0$, $\sum \lambda_i < \infty$, $\{e_i\}$ are an orthonormal basis of $L_2(D)$ and

$$\tilde{k}(\boldsymbol{x},\boldsymbol{x}') = \sum \lambda_i e_i(\boldsymbol{x}) e_i(\boldsymbol{x}').$$

Moreover

$$\left\|\sum c_i \sqrt{\lambda_i} e_i\right\|_{\tilde{k}}^2 = \sum c_i^2.$$

Then the integral-type kernel $\hat{k}$ can be checked to have eigenvalues and eigenvectors $\{\lambda_i^2\}$ and $\{e_i\}$. To see that (i) is satisfied, define a stochastic process $S = \sum \xi_i \lambda_i e_i$ with $\xi_i \sim N(0,1)$ independent, corresponding to a generic sample from Π_u. Then S lies in $H(D)$ with probability one since

$$\mathbb{E}\left(\|S\|_{\tilde{k}}^2\right) = \mathbb{E}\left(\sum \xi_i^2 \lambda_i\right) = \sum \lambda_i < \infty.$$

To see that (ii) is satisfied, given an element $c = \sum c_i \sqrt{\lambda_i} e_i$ of $H(D)$ we have $\sum c_i^2 < \infty$ and so the partial sums $c^{(N)} = \sum_{i=1}^N c_i \sqrt{\lambda_i} e_i$ converge to c under the norm $\|\cdot\|_{\tilde{k}}$. Since each $c^{(N)}$ also belongs to $H_{\hat{k}}(D)$, it follows that the set $H_{\hat{k}}(D)$ is dense in the space $(H(D), \|\cdot\|_{\tilde{k}})$. □

Proof of Prop. 6. First note that

$$\mu(\boldsymbol{x}) = \sum_{\mathcal{L}\in\{\mathcal{A},\mathcal{B}\}} \sum_{j=1}^{n_{\mathcal{L}}} w_j^{\mathcal{L}} \mathcal{L}u(\boldsymbol{x}_{0,j}^{\mathcal{L}})$$

where the weights are

$$\begin{bmatrix} \boldsymbol{w}^{\mathcal{A}} \\ \boldsymbol{w}^{\mathcal{B}} \end{bmatrix}^T = \bar{\mathcal{L}}\hat{\boldsymbol{K}}(\boldsymbol{x}, X_0)[\mathcal{L}\bar{\mathcal{L}}\hat{\boldsymbol{K}}(X_0)]^{-1}. \tag{20}$$

Now, from the reproducing property we have

$$\mu(\boldsymbol{x}) = \sum_{\mathcal{L}\in\{\mathcal{A},\mathcal{B}\}} \sum_{j=1}^{n_{\mathcal{L}}} w_j^{\mathcal{L}} \mathcal{L} \left\langle u, \hat{k}(\cdot, \boldsymbol{x}_{0,j}^{\mathcal{L}}) \right\rangle_{\hat{k}} = \left\langle u, \sum_{\mathcal{L}\in\{\mathcal{A},\mathcal{B}\}} \sum_{j=1}^{n_{\mathcal{L}}} w_j^{\mathcal{L}} \bar{\mathcal{L}}\hat{k}(\cdot, \boldsymbol{x}_{0,j}^{\mathcal{L}}) \right\rangle_{\hat{k}}$$

and hence, using the reproducing property again,

$$u(\boldsymbol{x}) - \mu(\boldsymbol{x}) = \left\langle u, \hat{k}(\cdot, \boldsymbol{x}) - \sum_{\mathcal{L}\in\{\mathcal{A},\mathcal{B}\}} \sum_{j=1}^{n_{\mathcal{L}}} w_j^{\mathcal{L}} \bar{\mathcal{L}}\hat{k}(\cdot, \boldsymbol{x}_{0,j}^{\mathcal{L}}) \right\rangle_{\hat{k}}.$$

Finally, using Cauchy–Schwarz produces

$$|u(\boldsymbol{x}) - \mu(\boldsymbol{x})| \leq \|u\|_{\hat{k}} \left\| \hat{k}(\cdot, \boldsymbol{x}) - \sum_{\mathcal{L}\in\{\mathcal{A},\mathcal{B}\}} \sum_{j=1}^{n_{\mathcal{L}}} w_j^{\mathcal{L}} \bar{\mathcal{L}}\hat{k}(\cdot, \boldsymbol{x}_{0,j}^{\mathcal{L}}) \right\|_{\hat{k}}.$$

Upon substitution of the expression for the weights $w_j^{\mathcal{L}}$ provided in Eqn. 20, the second term is recognised as $\sigma(\boldsymbol{x})$. Taking $u = u_0$ completes the proof. □

Proof of Thm. 8. Suppose u is a random variable with distribution $\Pi_u^{\boldsymbol{g},\boldsymbol{b}}$. Then we have

$$\begin{aligned} \int_{\Omega'} \|u - u_0\|_2^2 \mathrm{d}\Pi_u^{\boldsymbol{g},\boldsymbol{b}} &\leq \int_{\Omega'} \|u - \mu\|_2^2 \mathrm{d}\Pi_u^{\boldsymbol{g},\boldsymbol{b}} + \underbrace{\int_{\Omega'} \|\mu - u_0\|_2^2 \mathrm{d}\Pi_u^{\boldsymbol{g},\boldsymbol{b}}}_{\text{indep. of } u} \\ &= \int_D \int_{\Omega'} (u(\boldsymbol{x}) - \mu(\boldsymbol{x}))^2 \mathrm{d}\Pi_u^{\boldsymbol{g},\boldsymbol{b}} \mathrm{d}\boldsymbol{x} + \int_D (\mu(\boldsymbol{x}) - u_0(\boldsymbol{x}))^2 \mathrm{d}\boldsymbol{x} \\ &\leq \int_D \sigma(\boldsymbol{x})^2 \mathrm{d}\boldsymbol{x} + \|u_0\|_{\hat{k}}^2 \int_D \sigma(\boldsymbol{x})^2 \mathrm{d}\boldsymbol{x} \end{aligned}$$

where the second line uses Fubini's theorem to interchange the order of integration and the final line makes use of Prop. 6. Since the domain D is bounded, we have from Prop. 7 that there exists a generic constant C for which $\int_D \sigma(\boldsymbol{x})^2 \mathrm{d}\boldsymbol{x} \leq Ch^{2\beta-2\rho-d}$ and therefore

$$\int_{\Omega'} \|u - u_0\|_2^2 \mathrm{d}\Pi_u^{\boldsymbol{g},\boldsymbol{b}} \leq C(1 + \|u_0\|_{\hat{k}}^2) h^{2\beta-2\rho-d}.$$

The result follows from Markov's inequality: For fixed $\epsilon > 0$ we have

$$1 - \Pi_u^{\boldsymbol{g},\boldsymbol{b}}[B_\epsilon(u_0)] \leq \frac{\int \|u - u_0\|_2^2 \, \mathrm{d}\Pi_u^{\boldsymbol{g},\boldsymbol{b}}}{\epsilon} \leq \frac{C(1 + \|u_0\|_{\hat{k}}^2) h^{2\beta-2\rho-d}}{\epsilon}$$

as required. □

Proof of Thm. 9. The proof below consists of three discrete steps.

Step #1: Fix $\omega \in \Omega$ and $t \leq 0$. Consider an element u in the Hilbert scale of spaces $H_k^t(D)$. Using the fact that $\sqrt{\lambda_i} e_i$ are an orthonormal basis for $H_k^t(D)$, we have

$$\|u\|_{k,t}^2 = \sum_{i=1}^{\infty} \lambda_i^{-t} \langle u, \sqrt{\lambda_i} e_i \rangle_k^2.$$

By construction, a generic element $g \in H_\Lambda(D)$ can be written as $g = \sum_{i=1}^\infty c_i g_i$ where $g_i = \mathcal{A}\sqrt{\lambda_i} e_i$ form an orthonormal basis $\{g_i\}_{i=1}^\infty$ for $H_\Lambda(D)$ and $\|g\|_\Lambda^2 = \sum_{i=1}^\infty c_i^2 < \infty$. Thus we have

$$\|u\|_{k,t}^2 \;=\; \sum_{i=1}^\infty \lambda_i^{-t} \langle \mathcal{A}u, \mathcal{A}\sqrt{\lambda_i} e_i\rangle_\Lambda^2 \;=\; \sum_{i=1}^\infty \lambda_i^{-t} \langle \mathcal{A}u, g_i\rangle_\Lambda^2 \;=\; \|\mathcal{A}u\|_{\Lambda,t}^2.$$

Step #2: The stochastic process $g(\cdot,\omega')$ with kernel Λ can be expressed as a Karhunen-Loève expansion as $g(\cdot,\omega') = \sum_{i=1}^\infty \xi_i(\omega') g_i(\cdot)$, where the $\xi_i(\omega')$ are independent standard normal random variables under $\mathbb{P}'$ and the $g_i(\cdot) = \mathcal{A}\sqrt{\lambda_i} e_i(\cdot)$ were defined in Step #1. From sub-additivity of measure, we have that

$$\begin{aligned}
\mathbb{E}'\|g(\cdot,\omega')\|_{\Lambda,t}^2 &= \mathbb{E}' \sum_{i=1}^\infty \lambda_i^{-t} \xi_i(\omega')^2 \\
&\le \sum_{i=1}^\infty \lambda_i^{-t} \mathbb{E}' \xi_i(\omega')^2 \;=\; \sum_{i=1}^\infty \lambda_i^{-t}.
\end{aligned}$$

From (A2) we obtain (recall $t < -d/2\alpha < 0$)

$$\mathbb{E}'\|g(\cdot,\omega')\|_{\Lambda,t}^2 \;\le\; C_\omega^t \sum_{i=1}^\infty [\lambda_i^{(\alpha)}]^{-t}.$$

Combining this with the result of Step #1 implies that

$$\begin{aligned}
\mathbb{E}'\|u(\cdot,\omega,\omega')\|_{k,t}^2 &= \mathbb{E}'\|\mathcal{A}u(\cdot,\omega,\omega')\|_{\Lambda,t}^2 \\
&= \mathbb{E}'\|g(\cdot,\omega')\|_{\Lambda,t}^2 \;\le\; C_\omega^t \sum_{i=1}^\infty [\lambda_i^{(\alpha)}]^{-t}.
\end{aligned}$$

Step #3: Consider the doubly stochastic process $u(\cdot,\omega,\omega')$ and a double expectation $\mathbb{E}\mathbb{E}'$ over ω and ω'. From (A2) we have

$$\mathbb{E}\mathbb{E}'\|u(\cdot,\omega,\omega')\|_{\mathbb{H}^\alpha(D),t}^2 \;\le\; \mathbb{E}\, C_{\omega,t}\, \mathbb{E}'\|u(\cdot,\omega,\omega')\|_{k,t}^2.$$

The output of Step #2 then implies that

$$\mathbb{E}\mathbb{E}'\|u(\cdot,\omega,\omega')\|_{\mathbb{H}^\alpha(D),t}^2 \;\le\; \underbrace{\mathbb{E}[\, C_\omega^t\, C_{\omega,t}\,]}_{(*)} \sum_{i=1}^\infty [\lambda_i^{(\alpha)}]^{-t}.$$

Under (A2), the term $(*)$ is finite when $t < -d/2\alpha$. Observe that, since $\lambda_i^{(\alpha)} \asymp i^{-2\alpha/d}$, the right hand side is finite for all values of $t < -d/2\alpha$. This implies that $u \in L^2_{\mathbb{P}\mathbb{P}'}(\Omega,\Omega';[\mathbb{H}^\alpha(D)]^t)$ for all $-1 < t < -d/2\alpha$ and hence $u \in L^2_{\mathbb{P}\mathbb{P}'}(\Omega,\Omega';\mathbb{H}^s(D))$ for all $0 < s < \alpha - d/2$. □

Proof of Prop. 10. The conclusion of Props. 6 and 7 show that, for h sufficiently small,

$$\begin{aligned}
|\mu(\boldsymbol{x}) - u_0(\boldsymbol{x})| &\le \sigma(\boldsymbol{x}) \|u_0\|_{\hat{k}} \\
&\le C h^{\beta-\rho-d/2} \|u_0\|_{\hat{k}}
\end{aligned}$$

where C is independent of $\boldsymbol{x}$, ω and ω'. Hence

$$\|\mu - u_0\|_\infty \;\le\; C h^{\beta-\rho-d/2} \|u_0\|_{\hat{k}}.$$

Squaring both sides and taking a double expectation over $\omega_0 \sim \mathbb{P}$, $\omega_0' \sim \mathbb{P}'$ leads to

$$\mathbb{E}\mathbb{E}'\|\mu - u\|_\infty^2 \;\le\; C^2 h^{2\beta-2\rho-d} \mathbb{E}\mathbb{E}'\|u\|_{\hat{k}}^2.$$

Now (A1) implies that

$$\begin{aligned}
H_{\hat{k}}(D) &\equiv [H(D)]^1 \\
&\equiv [\mathbb{H}^\beta(D)]^1 \;\equiv\; \mathbb{H}^{(1+1)\beta}(D) \;=\; \mathbb{H}^{2\beta}(D).
\end{aligned}$$

Thus, for $s := 2\beta < \alpha - d/2$, Thm. 9 shows that $\mathbb{E}\mathbb{E}'\|u\|_{\hat{k}}^2 < \infty$ and hence $\mathbb{E}\mathbb{E}'\|\mu - u\|_\infty^2 = \mathcal{O}(h^{2\beta-2\rho-d})$. It follows that $\|u - \mu\|_\infty = \mathcal{O}_{\mathbb{P}\mathbb{P}'}(h^{\beta-\rho-d/2})$. □

Proof of Thm. 11. We show that, when the fill distance h (that determines the error of the numerical solver) is appropriately coupled to the number n of data, $\Pi_\theta^{\boldsymbol{y},h}$ and $\Pi_\theta^{\boldsymbol{y}}$ are asymptotically identical. In the remainder of the proof $\omega_0 \in \Omega$ and $\omega_0' \in \Omega'$ are each fixed.

From Thm. 4.9 of Dashti and Stuart [2015], it is sufficient to establish that the two potentials Φ_h and Φ are asymptotically identical. Recall the identity $(\boldsymbol{I}-\boldsymbol{M})^{-1} = \sum_{m=0}^{\infty} \boldsymbol{M}^m$, valid when the largest eigenvalue $\lambda_{\max}[\boldsymbol{M}] < 1$. Assume h is sufficiently small that $\lambda_{\max}[\boldsymbol{\Sigma}] < \gamma^2$. Then, using the identity with $\boldsymbol{M} = -\gamma^{-2}\boldsymbol{\Sigma}$, we have

$$\begin{aligned}
\Phi_h &= (\boldsymbol{y}-\boldsymbol{\mu})^T(\boldsymbol{\Sigma}+\gamma^2\boldsymbol{I})^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) \\
&= (\boldsymbol{y}-\boldsymbol{\mu})^T(\gamma^2\boldsymbol{I})^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) + \gamma^{-2}\sum_{m=1}^{\infty}(\boldsymbol{y}-\boldsymbol{\mu})^T(-\gamma^{-2}\boldsymbol{\Sigma})^m(\boldsymbol{y}-\boldsymbol{\mu}) \\
&= (\boldsymbol{y}-\boldsymbol{u})^T(\gamma^2\boldsymbol{I})^{-1}(\boldsymbol{y}-\boldsymbol{u}) + \underbrace{2(\boldsymbol{y}-\boldsymbol{u})^T(\gamma^2\boldsymbol{I})^{-1}(\boldsymbol{u}-\boldsymbol{\mu})}_{(*)} \\
&\quad + \underbrace{(\boldsymbol{u}-\boldsymbol{\mu})^T(\gamma^2\boldsymbol{I})^{-1}(\boldsymbol{u}-\boldsymbol{\mu})}_{(**)} + \underbrace{\gamma^{-2}\sum_{m=1}^{\infty}(\boldsymbol{y}-\boldsymbol{\mu})^T(-\gamma^{-2}\boldsymbol{\Sigma})^m(\boldsymbol{y}-\boldsymbol{\mu})}_{(***)} \\
&= \Phi + (*) + (**) + (***).
\end{aligned}$$

Each of the terms $(*), (**), (***)$ can be bound using Props. 6 and 7:

$$\begin{aligned}
(*) \,:\, |2(\boldsymbol{y}-\boldsymbol{u})^T(\gamma^2\boldsymbol{I})^{-1}(\boldsymbol{u}-\boldsymbol{\mu})| &\leq 2\gamma^{-2}\|\boldsymbol{y}-\boldsymbol{\mu}\|_2\|\boldsymbol{u}-\boldsymbol{\mu}\|_2 \\
&\leq 2\gamma^{-2}[\|\boldsymbol{y}-\boldsymbol{u}\|_2 + \|\boldsymbol{u}-\boldsymbol{\mu}\|_2]\|\boldsymbol{u}-\boldsymbol{\mu}\|_2 \\
&\leq 2\gamma^{-2}[\underbrace{\|\boldsymbol{y}-\boldsymbol{u}\|_2}_{\mathcal{O}_P(n^{1/2})} + n^{1/2}\underbrace{\|\boldsymbol{u}-\boldsymbol{\mu}\|_\infty}_{\mathcal{O}(h^{\beta-\rho-d/2})}]n^{1/2}\underbrace{\|\boldsymbol{u}-\boldsymbol{\mu}\|_\infty}_{\mathcal{O}(h^{\beta-\rho-d/2})}.
\end{aligned}$$

Here the first inequality here is Cauchy–Schwarz, the second is the triangle inequality and the third is the bound $\|\cdot\|_2 \leq \sqrt{n}\|\cdot\|_\infty$. The notation P is used to reflect randomness in the observation process for $\boldsymbol{y}|\boldsymbol{u}$, distinguishing this source of randomness from $\mathbb{P}$ and $\mathbb{P}'$. Similarly

$$\begin{aligned}
(**) \,:\, |(\boldsymbol{u}-\boldsymbol{\mu})^T(\gamma^2\boldsymbol{I})^{-1}(\boldsymbol{u}-\boldsymbol{\mu})| &= \gamma^{-2}\|\boldsymbol{u}-\boldsymbol{\mu}\|_2^2 \\
&\leq \gamma^{-2}n\underbrace{\|\boldsymbol{u}-\boldsymbol{\mu}\|_\infty^2}_{\mathcal{O}(h^{2\beta-2\rho-d})}
\end{aligned}$$

and also

$$\begin{aligned}
(***) \,:\, \left|\gamma^{-2}\sum_{m=1}^{\infty}(\boldsymbol{y}-\boldsymbol{\mu})^T(-\gamma^{-2}\boldsymbol{\Sigma})^m(\boldsymbol{y}-\boldsymbol{\mu})\right| &\leq \gamma^{-2}\left|\sum_{m=1}^{\infty}(-\gamma^{-2})^m\lambda_{\max}[\boldsymbol{\Sigma}^m](\boldsymbol{y}-\boldsymbol{\mu})^T(\boldsymbol{y}-\boldsymbol{\mu})\right| \\
&= \gamma^{-2}\|\boldsymbol{y}-\boldsymbol{\mu}\|_2^2\left|\sum_{m=1}^{\infty}(-\gamma^{-2})^m\lambda_{\max}[\boldsymbol{\Sigma}]^m\right| \\
&= \gamma^{-2}\|\boldsymbol{y}-\boldsymbol{\mu}\|_2^2\frac{\lambda_{\max}[\boldsymbol{\Sigma}]}{\lambda_{\max}[\boldsymbol{\Sigma}]+\gamma^2} \\
&\leq \gamma^{-2}\|\boldsymbol{y}-\boldsymbol{\mu}\|_2^2\lambda_{\max}[\boldsymbol{\Sigma}].
\end{aligned}$$

A bound $\lambda_{\max}[\boldsymbol{\Sigma}] \leq \mathrm{tr}[\boldsymbol{\Sigma}]$ then produces

$$\begin{aligned}
(***) &\leq \gamma^{-2}\|\boldsymbol{y}-\boldsymbol{\mu}\|_2^2\,\mathrm{tr}[\boldsymbol{\Sigma}] \\
&\leq \gamma^{-2}[\|\boldsymbol{y}-\boldsymbol{u}\|_2 + \|\boldsymbol{u}-\boldsymbol{\mu}\|_2]^2 \times n\max_{j=1,\ldots,n}\sigma(\boldsymbol{x}_j)^2 \\
&\leq \gamma^{-2}[\underbrace{\|\boldsymbol{y}-\boldsymbol{u}\|_2}_{\mathcal{O}_P(n^{1/2})} + n^{1/2}\underbrace{\|\boldsymbol{u}-\boldsymbol{\mu}\|_\infty}_{\mathcal{O}(h^{\beta-\rho-d/2})}]^2 \times n\underbrace{\max_{j=1,\ldots,n}\sigma(\boldsymbol{x}_j)^2}_{\mathcal{O}(h^{2\beta-2\rho-d})}
\end{aligned}$$

where Prop. 7 is used to compute the asymptotic order of the final term. Notably the bound $\lambda_{\max}[\mathbf{\Sigma}] \leq \mathrm{tr}[\mathbf{\Sigma}]$ could be further refined by relating the spectral gap to the fill distance.

Combining the bounds for $(*)$, $(**)$ and $(***)$ produces $\Phi_h = \Phi + \mathcal{O}_P(\xi^2)$ where $\xi := nh^{\beta-\rho-d/2}$. This shows that the two potentials Φ_h and Φ are asymptotically identical provided $\xi = o(1)$; that is, provided $h = o(n^{-1/(\beta-\rho-d/2)})$. This completes the proof. □

Acknowledgement

TJS was supported by the Free University of Berlin within the Excellence Initiative of the German Research Foundation (DFG). MG was supported by EPSRC [EP/J016934/1, EP/K034154/1], an EPSRC Established Career Fellowship, the EU grant [EU/259348] and a Royal Society Wolfson Research Merit Award.

The authors would like to thank John Skilling for useful discussion, Patrick Farrell for providing code used in generating these results and François-Xavier Briol for helpful feedback. In addition they express gratitude to the developers of the Python libraries Autograd and GPyOpt.

Literatur

Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification. *SIAM J. Sci. Comput.*, 36(5):A2122–A2148, 2014. ISSN 1064-8275. doi: 10.1137/130933381.

Alen Alexanderian, Philip J. Gloor, and Omar Ghattas. On Bayesian A- and D-optimal experimental designs in infinite dimensions. *Bayesian Anal.*, 11(3):671–695, 2016a. ISSN 1936-0975. doi: 10.1214/15-BA969.

Alen Alexanderian, Noemi Petra, Georg Stadler, and Omar Ghattas. A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems. *SIAM J. Sci. Comput.*, 38(1):A243–A272, 2016b. ISSN 1064-8275. doi: 10.1137/140992564.

Samuel M. Allen and John W. Cahn. A microscopic theory for antiphase domain boundary motion and its application to antiphase domain coarsening. *Acta Met.*, 27(6):1085–1095, 1979. doi: 10.1016/0001-6160(79)90196-2.

Christophe Andrieu and Gareth O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2):697–725, 2009. ISSN 0090-5364. doi: 10.1214/07-AOS574.

Andrea Arnold, Daniela Calvetti, and Erkki Somersalo. Linear multistep methods, particle filtering and sequential Monte Carlo. *Inverse Prob.*, 29(8):085007, 25, 2013. ISSN 0266-5611. doi: 10.1088/0266-5611/29/8/085007.

David Barber and Yali Wang. Gaussian processes for Bayesian estimation in ordinary differential equations. In Tony Jebara and Eric P. Xing, editors, *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1485–1493. JMLR Workshop and Conference Proceedings, 2014.

Jörn Behrens and Armin Iske. Grid-free adaptive semi-Lagrangian advection using radial basis functions. *Comput. Math. Appl.*, 43(3-5):319–327, 2002. ISSN 0898-1221. doi: 10.1016/S0898-1221(01)00289-9. Radial basis functions and partial differential equations.

T. Belytschko, N. Mo‘s, S. Usui, and C. Parimi. Arbitrary discontinuities in finite elements. *Int. J. Numer. Meth. Engng*, 50(4):993–1013, 2001. ISSN 1097-0207. doi: 10.1002/1097-0207(20010210)50:4⟨993::AID-NME164⟩3.0.CO;2-M. URL `http://dx.doi.org/10.1002/1097-0207(20010210)50:4<993::AID-NME164>3.0.CO;2-M`.

Alain Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic Publishers, Boston, MA, 2004. ISBN 1-4020-7679-7. doi: 10.1007/978-1-4419-9096-9.

F.-X. Briol, C.J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic. Probabilistic integration: A role for statisticians in numerical analysis?, 2016. arXiv:1512.00933v4.

Ben Calderhead, Mark Girolami, and Neil D. Lawrence. Accelerating Bayesian inference over nonlinear differential equations with Gaussian processes. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 217–224. Curran Associates, Inc., 2009. URL `http://papers.nips.cc/paper/3497-accelerating-bayesian-inference-over-nonlinear-differential-equations-with-gaussian-processes.pdf`.

Alberto-P. Calderón. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980)*, pages 65–73. Soc. Brasil. Mat., Rio de Janeiro, 1980.

Dave Campbell. *Bayesian Collocation Tempering and Generalized Profiling for Estimation of Parameters from Differential Equation Models*. PhD thesis, McGill University, 2007.

Marcos Capistrán, J. Andrés Christen, and Sophie Donnet. Bayesian analysis of ODEs: solver optimal accuracy and Bayes factors, 2013. arXiv:1311.2281v1.

S. W. Chi, J. S. Chen, H. Luo, H. Y. Hu, and L. Wang. Dispersion and stability properties of radial basis collocation method for elastodynamics. *Numer. Methods Partial Differential Equations*, 29(3): 818–842, 2013. ISSN 0749-159X. doi: 10.1002/num.21732.

Igor Cialenco, Gregory E. Fasshauer, and Qi Ye. Approximation of stochastic partial differential equations by a kernel-based collocation method. *Int. J. Comput. Math.*, 89(18):2543–2561, 2012. ISSN 0020-7160. doi: 10.1080/00207160.2012.688111.

Albert Cohen and Ronald DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numer.*, 24:1–159, 2015. ISSN 0962-4929. doi: 10.1017/S0962492915000033.

Patrick R. Conrad, Mark Girolami, Simo Särkkä, Andrew Stuart, and Konstantinos Zygalakis. Probability measures for numerical solutions of differential equations, 2015. arXiv:1506.04592v1.

Simon L. Cotter, Gareth O. Roberts, Andrew M. Stuart, and David. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013. ISSN 0883-4237. doi: 10.1214/13-STS421.

Masoumeh Dashti and Andrew M. Stuart. The Bayesian approach to inverse problems, 2015. arXiv:1302.6989v4.

Itai Dattner and Chris A. J. Klaassen. Optimal rate of direct estimators in systems of ordinary differential equations linear in functions of the parameters. *Electron. J. Statist.*, 9(2):1939–1973, 2015. doi: 10.1214/15-EJS1053.

Persi Diaconis. Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, Ind., 1986)*, pages 163–175. Springer, New York, 1988.

F. Dondelinger, S. Rogers, and D. Husmeier. ODE parameter inference using adaptive gradient matching with Gaussian processes. In *Sixteenth International Conference on Artificial Intelligence and Statistics; AISTATS*. 2013.

Matthew M. Dunlop and Andrew M. Stuart. The Bayesian formulation of EIT: Analysis and algorithms, 2015. arXiv:1508.04106v1.

Patrick E. Farrell, Ásgeir Birkisson, and Simon W. Funke. Deflation techniques for finding distinct solutions of nonlinear partial differential equations. *SIAM J. Sci. Comput.*, 37(4):A2026–A2045, 2015. ISSN 1064-8275. doi: 10.1137/140984798.

Gregory E Fasshauer. Solving partial differential equations by collocation with radial basis functions. In Alain Le Méhauté, Christophe Rabut, and Larry L. Schumaker, editors, *Surface Fitting and Multiresolution Methods. Vol. 2 of the Proceedings of the 3rd International Conference on Curves and Surfaces held in Chamonix–Mont-Blanc, June 27–July 3, 1996*, pages 131–178. Vanderbilt University Press, Nashville, TN, 1997. ISBN 0-8265-1294-1.

Gregory E. Fasshauer. Solving differential equations with radial basis functions: multilevel methods and smoothing. *Adv. Comput. Math.*, 11(2-3):139–159, 1999. ISSN 1019-7168. doi: 10.1023/A:1018919824891. Radial basis functions and their applications.

Gregory E. Fasshauer and Qi Ye. Reproducing kernels of generalized Sobolev spaces via a Green function approach with distributional operators. *Numer. Math.*, 119(3):585–611, 2011. ISSN 0029-599X. doi: 10.1007/s00211-011-0391-2.

Gregory E. Fasshauer and Qi Ye. A kernel-based collocation method for elliptic partial differential equations with random coefficients. In *Monte Carlo and quasi-Monte Carlo methods 2012*, volume 65 of *Springer Proc. Math. Stat.*, pages 331–347. Springer, Heidelberg, 2013. doi: 10.1007/978-3-642-41095-6_14.

Carsten Franke and Robert Schaback. Convergence order estimates of meshless collocation methods using radial basis functions. *Adv. Comput. Math.*, 8(4):381–399, 1998. ISSN 1019-7168. doi: 10.1023/A:1018916902176.

Andrew Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.*, 1(3):515–533 (electronic), 2006.

Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000. ISSN 0090-5364. doi: 10.1214/aos/1016218228.

Alex a. Gorodetsky and Youssef M. Marzouk. Mercer kernels and integrated variance experimental design: connections between Gaussian process regression and polynomial approximation. *SIAM J. Sci. Comput.*, xx:1–32, 2015.

Thore Graepel. Solving noisy linear operator equations by Gaussian processes: Application to ordinary and partial differential equations. In Nina Mishra and Tom Fawcett, editors, *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 234–241, 2003. URL `http://www.aaai.org/Papers/ICML/2003/ICML03-033.pdf`.

E. Hairer, R. I. McLachlan, and A. Razakarivony. Achieving Brouwer's law with implicit Runge-Kutta methods. *BIT*, 48(2):231–243, 2008. ISSN 0006-3835. doi: 10.1007/s10543-008-0170-3.

Markus Heinonen and Florence d'Alché Buc. Learning nonparametric differential equations with operator-valued kernels and gradient matching. 2014. arXiv:1411.5172v1.

Philipp Hennig. Probabilistic interpretation of linear solvers. *SIAM J. Optim.*, 25(1):234–260, 2015. ISSN 1052-6234. doi: 10.1137/140955501.

Philipp Hennig, Michael A. Osborne, and Mark Girolami. Probabilistic numerics and uncertainty in computations. volume 471, pages 20150142, 17, 2015. doi: 10.1098/rspa.2015.0142.

Y. C. Hon and Robert Schaback. Solvability of partial differential equations by meshless kernel methods. *Adv. Comput. Math.*, 28(3):283–299, 2008. ISSN 1019-7168. doi: 10.1007/s10444-006-9023-2.

T. E. Hull and J. R. Swenson. Tests of probabilistic models for the propagation of roundoff errors. *Comm. ACM*, 9:108–113, 1966. ISSN 0001-0782.

Marco A Iglesias. A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Prob.*, 32(2):025002, 2016. ISSN 0266-5611. doi: 10.1088/0266-5611/32/2/025002. URL `http://stacks.iop.org/0266-5611/32/i=2/a=025002?key=crossref.b2f1f1507b3a961ed3947d02a40c9002`.

Joseph B. Kadane. Parallel and sequential computation: a statistician's view. *J. Complexity*, 1(2):256–263, 1985. ISSN 0885-064X. doi: 10.1016/0885-064X(85)90014-7. Complexity of approximately solved problems (Morningside Heights, N.Y., 1985).

Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

Jari Kaipio and Erkki Somersalo. Statistical inverse problems: discretization, model reduction and inverse crimes. *J. Comput. Appl. Math.*, 198(2):493–504, 2007. ISSN 0377-0427. doi: 10.1016/j.cam.2005.09.027.

Hans Kersting and Philipp Hennig. Active uncertainty calibration in Bayesian ODE solvers. In Ihler Janzing, editor, *Uncertainty in Artificial Intelligence (UAI)*, volume 32. 2016.

B. Li, F. Habbal, and M. Ortiz. Optimal transportation meshfree approximation schemes for fluid and plastic flows. *Internat. J. Numer. Methods Engrg.*, 83(12):1541–1579, 2010. ISSN 0029-5981. doi: 10.1002/nme.2869.

B. Li, A. Kidane, G. Ravichandran, and M. Ortiz. Verification and validation of the Optimal Transportation Meshfree (OTM) simulation of terminal ballistics. *Int. J. Impact. Engng*, 42:25–36, 2012. ISSN 0734-743X. doi: 10.1016/j.ijimpeng.2011.11.003.

Leevan Ling and Robert Schaback. Stable and convergent unsymmetric meshless collocation methods. *SIAM J. Numer. Anal.*, 46(3):1097–1115, 2008. ISSN 0036-1429. doi: 10.1137/06067300X.

Leevan Ling, Roland Opfer, and Robert Schaback. Results on meshless collocation techniques. *Eng. Anal. Boundary Elements*, 30(4):247–253, 2006. doi: 10.1016/j.enganabound.2005.08.008.

R. A. Lorentz, F. J. Narcowich, and J. D. Ward. Collocation discretizations of the transport equation with radial basis functions. *Appl. Math. Comput.*, 145(1):97–116, 2003. ISSN 0096-3003. doi: 10.1016/S0096-3003(02)00472-1.

Xiang Ma and Nicholas Zabaras. An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method. *Inverse Prob.*, 25(3):035013, 27, 2009. ISSN 0266-5611. doi: 10.1088/0266-5611/25/3/035013.

Benn Macdonald, Catherine Higham, and Dirk Husmeier. Controversy in mechanistic modelling with Gaussian processes. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 1539–1547, 2015. URL `http://jmlr.org/proceedings/papers/v37/macdonald15.html`.

Youssef Marzouk and Dongbin Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. *Commun. Comput. Phys.*, 6(4):826–847, 2009. ISSN 1815-2406. doi: 10.4208/cicp.2009.v6.p826.

Jonas Mockus, Vytautas Tiesis, and Antanas Zilinskas. The application of Bayesian methods for seeking the extremum. *Towards Global Optimization*, 2(2):117–129, 1978.

K.-S. Moon, E. von Schwerin, A. Szepessy, and R. Tempone. Convergence rates for an adaptive dual weighted residual finite element algorithm. *BIT*, 46(2):367–407, 2006. ISSN 0006-3835. doi: 10.1007/s10543-006-0058-z.

Kyoung-Sook Moon, Erik von Schwerin, Anders Szepessy, and Raúl Tempone. An adaptive algorithm for ordinary, stochastic and partial differential equations. 383:325–343, 2005. doi: 10.1090/conm/383/07176.

Sebastian Mosbach and Amanda G. Turner. A quantitative probabilistic investigation into the accumulation of rounding errors in numerical ODE solution. *Comput. Math. Appl.*, 57(7):1157–1167, 2009. ISSN 0898-1221. doi: 10.1016/j.camwa.2009.01.020.

Erich Novak. *Deterministic and stochastic error bounds in numerical analysis*, volume 1349. Springer Berlin, 1988.

A. O'Hagan. Bayes–Hermite quadrature. *J. Statist. Plann. Inference*, 29(3):245–260, 1991. ISSN 0378-3758. doi: 10.1016/0378-3758(91)90002-V.

Antony Overstall, David Woods, and Ben Parker. Bayesian optimal design for ordinary differential equation models. 2015. arXiv:1509.04099v2.

Houman Owhadi. Bayesian numerical homogenization. *Multiscale Model. Simul.*, 13(3):812–828, 2015. ISSN 1540-3459. doi: 10.1137/140974596.

Houman Owhadi. Multi-grid with rough coefficients and multiresolution operator decomposition from hierarchical information games. *SIAM Rev.*, 2016. To appear.

Natesh S. Pillai, Qiang Wu, Feng Liang, Sayan Mukherjee, and Robert L. Wolpert. Characterizing the function space for Bayesian kernel models. *J. Mach. Learn. Res.*, 8:1769–1797, 2007. ISSN 1532-4435.

Friedrich Pukelsheim. *Optimal Design of Experiments*, volume 50 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. ISBN 0-89871-604-7. doi: 10.1137/1.9780898719109. Reprint of the 1993 original.

J. O. Ramsay, G. Hooker, D. Campbell, and J. Cao. Parameter estimation for differential equations: a generalized smoothing approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69(5):741–796, 2007. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2007.00610.x. With discussions and a reply by the authors.

Simo Särkkä. Linear operators and stochastic partial differential equations in Gaussian process regression. In Timo Honkela, Włodzisław Duch, Mark Girolami, and Samuel Kaski, editors, *Artificial Neural Networks and Machine Learning – ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part II*, pages 151–158. Springer, Berlin, Heidelberg, 2011. ISBN 978-3-642-21738-8. doi: 10.1007/978-3-642-21738-8_20.

Michael Schober, David K Duvenaud, and Philipp Hennig. Probabilistic ODE solvers with Runge–Kutta means. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 739–747. Curran Associates, Inc., 2014. URL `http://papers.nips.cc/paper/5451-probabilistic-ode-solvers-with-runge-kutta-means.pdf`.

John Skilling. Bayesian solution of ordinary differential equations. In C. Ray Smith, Gary J. Erickson, and Paul O. Neudorfer, editors, *Maximum Entropy and Bayesian Methods: Seattle, 1991*, pages 23–37. Springer, Dordrecht, 1992. ISBN 978-94-017-2219-3. doi: 10.1007/978-94-017-2219-3_2.

Ingo Steinwart and Clint Scovel. Mercer's Theorem on General Domains: On the Interaction between Measures, Kernels, and RKHSs. *Constructive Approximation*, 35(3):363–417, 2012. ISSN 01764276. doi: 10.1007/s00365-012-9153-3.

A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010. ISSN 0962-4929. doi: 10.1017/S0962492910000061.

Andrew M. Stuart and Aretha L. Teckentrup. Posterior consistency for Gaussian process approximations of Bayesian posterior distributions, 2016. arXiv:1603.02004v1.

T. J. Sullivan. *Introduction to Uncertainty Quantification*, volume 63 of *Texts in Applied Mathematics*. Springer, Cham, 2015. ISBN 978-3-319-23394-9; 978-3-319-23395-6. doi: 10.1007/978-3-319-23395-6.

Rüdiger Verfürth. *A Posteriori Error Estimation Techniques for Finite Element Methods*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2013. ISBN 978-0-19-967942-3. doi: 10.1093/acprof:oso/9780199679423.001.0001.

Sebastian J. Vollmer. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Prob.*, 29(12):125011, 32, 2013. ISSN 0266-5611. doi: 10.1088/0266-5611/29/12/125011.

Mort David Webster, Menner A. Tatang, and Gregory J. McRae. Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model. In *MIT Joint Program on the Science and Policy of Global Change*. 1996.

Holger Wendland. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.*, 4(4):389–396, 1995. ISSN 1019-7168. doi: 10.1007/BF02123482.

Holger Wendland. *Scattered Data Approximation*, volume 17 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2005. ISBN 978-0521-84335-5; 0-521-84335-9.

Henryk Woźniakowski. What is information-based complexity? pages 89–95, 2009. doi: 10.4171/069-1/5.

Hongqi Xue, Hongyu Miao, and Hulin Wu. Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Ann. Statist.*, 38(4):2351–2387, 2010. ISSN 0090-5364. doi: 10.1214/09-AOS784.

Supplementary Material

Computation of the Natural Kernel

Here we detail the computations required to find the natural kernel in closed form for the illustrative example of Sec. 3.2.1. We must solve:

$$k(x,x') = \int_0^1 \int_0^1 G(x,z)G(x',z')\Lambda(z,z')\mathrm{d}z\mathrm{d}z'$$

where

$$G(x,x') = \begin{cases} x'(x-1) & x' < x \\ x(x'-1) & x' > x \end{cases}$$
$$\Lambda(x,x') = \begin{cases} (1-\epsilon(x-x'))_+^2 & x > x' \\ (1-\epsilon(x'-x))_+^2 & x < x' \end{cases}$$

Note that each of G, Λ is continuous and piecewise polynomial, and that each has a discontinuous derivative at $x = x'$. $\Lambda(x,x')$ has an additional discontinuity in its derivative at $x = x' - \epsilon^{-1}$ and $x = x' + \epsilon^{-1}$. We are aided, however, by the fact that $\Lambda(x,x') = 0$ wherever $|x - x'| > \epsilon^{-1}$. Further note that, as kernels are symmetric in x and x', we may without loss of generality assume $x' > x$.

Symbolic integration packages tend to have difficulty with piecewise expressions such as these, so we are forced to divide the integrals manually before passing to such integrators. That is the purpose of the derivation below.

Start by dividing the integrals according to the discontinuities in G:

$$\begin{aligned} k(x,x') =& (x-1)(x'-1)\int_0^x z \int_0^{x'} z'\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \\ &+ (x-1)x'\int_0^x z\int_{x'}^1 (z'-1)\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \\ &+ x(x'-1)\int_x^1 (z-1)\int_0^{x'} z'\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \\ &+ xx'\int_x^1 (z-1)\int_{x'}^1 (z'-1)\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \end{aligned}$$

so that

$$\begin{aligned} I_1(x,x') &= \int_0^x z\int_0^{x'} z'\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \\ I_2(x,x') &= \int_0^x z\int_{x'}^1 (z'-1)\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \\ I_3(x,x') &= \int_x^1 (z-1)\int_0^{x'} z'\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \\ I_4(x,x') &= \int_x^1 (z-1)\int_{x'}^1 (z'-1)\Lambda(z,z')\mathrm{d}z'\mathrm{d}z \end{aligned}$$

and

$$\begin{aligned} k(x,x') =& (x-1)(x'-1)I_1(x,x') + (x-1)x'I_2(x,x') \\ &+ x(x'-1)I_3(x,x') + xx'I_4(x,x'). \end{aligned}$$

The only change-points remaining in each integrand are in the $\Lambda(z,z')$ term. Our general strategy for reducing these integrals to those which can be solved using symbolic integration packages is to compute

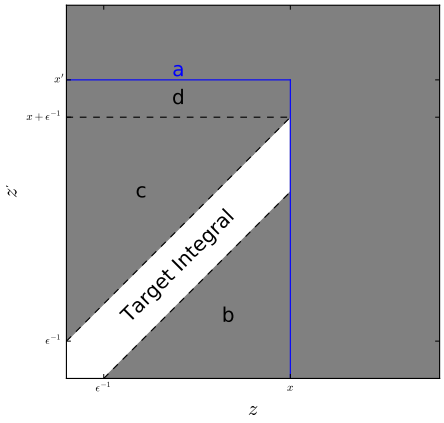


Abbildung 13: A sketch of the procedure for computing I_1. We require the region in white, which we compute by first determining integral a, then subtracting b, c and d.

the integral ignoring the $(\cdot)_+$ term in Λ, and then subtract off the terms which are outside this support. A schematic for I_1 is included in Fig. 13; the other integrals are analogous. When decomposed as such we have found that the integrals in the decomposition can be solved by the software.

$$
\begin{aligned}
I_1(x,x') = &\int_0^x \int_0^{x'} zz'(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z \\
&- \mathbb{I}(x > \epsilon^{-1}) \int_{\epsilon^{-1}}^x \int_0^{z-\epsilon^{-1}} zz'(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z \\
&- \mathbb{I}(x' > \epsilon^{-1}) \int_{\epsilon^{-1}}^{x+\epsilon^{-1}\vee x'} \int_0^{z'-\epsilon^{-1}} zz'(1-\epsilon(x-x'))^2 \mathrm{d}z\mathrm{d}z' \\
&- \mathbb{I}(x' > x+\epsilon^{-1}) \int_{x+\epsilon^{-1}}^{x'} \int_0^{x} zz'(1-\epsilon(x-x'))^2 \mathrm{d}z\mathrm{d}z'
\end{aligned}
$$

where $\mathbb{I}(\cdot)$ is an indicator function. Similarly:

$$
I_2(x,x') = \mathbb{I}(x' < x+\epsilon^{-1}) \int_{0\wedge(x'-\epsilon^{-1})}^x \int_{x'}^{1\vee(x+\epsilon^{-1})} z(z'-1)(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z
$$

$$
\begin{aligned}
I_3(x,x') = * &\int_x^{x'-\epsilon^{-1}} \int_{z-\epsilon^{-1}}^{z+\epsilon^{-1}} (z-1)z'(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z \\
&+ \int_{x\wedge x'-\epsilon^{-1}}^{x'} \int_{z-\epsilon^{-1}}^{z+\epsilon^{-1}} (z-1)z'(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z \\
&+ \mathbb{I}(x'-\epsilon^{-1} > x) \int_{x'}^{x'+\epsilon^{-1}\vee 1} \int_{z-\epsilon^{-1}}^{x'} (z-1)z'(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z \\
&- \mathbb{I}(x < \epsilon^{-1}) \int_x^{\epsilon^{-1}} \int_{z-\epsilon^{-1}}^{0} (z-1)z'(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z
\end{aligned}
$$

$$
\begin{aligned}
I_4(x,x') = &\int_x^1 \int_{x'}^1 (z-1)(z'-1)(1-\epsilon(x-x'))^2 \mathrm{d}z\mathrm{d}z' \\
&- \mathbb{I}(x' < 1-\epsilon^{-1}) \int_{x'+\epsilon^{-1}}^1 \int_{x'}^{z-\epsilon^{-1}} (z-1)(z'-1)(1-\epsilon(x-x'))^2 \mathrm{d}z\mathrm{d}z' \\
&- \mathbb{I}(x < 1-\epsilon^{-1}) \int_{x'\wedge x+\epsilon^{-1}}^1 \int_{x'-\epsilon^{-1}\wedge x}^{z'-\epsilon^{-1}} (z-1)(z'-1)(1-\epsilon(x-x'))^2 \mathrm{d}z'\mathrm{d}z \\
&- \mathbb{I}(x' < x-\epsilon^{-1}) \int_{x'}^1 \int_x^{x'-\epsilon^{-1}} (z-1)(z'-1)(1-\epsilon(x-x'))^2 \mathrm{d}z\mathrm{d}z'
\end{aligned}
$$