Journal of Computational Physics 520 (2025) 113488

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Error analysis of kernel/GP methods for nonlinear and parametric PDEs

Pau Batlle [a], Yifan Chen [a], Bamdad Hosseini [b,*], Houman Owhadi [a], Andrew M. Stuart [a]

[a] Computing and Mathematical Sciences, Caltech, Pasadena, CA, United States of America
[b] Department of Applied Mathematics, University of Washington, Seattle, WA, United States of America

ARTICLE INFO

MSC:
60G15
65M75
65N75
65N35
47B34
41A15
35R30
34B15

Keywords:
Kernel methods
Gaussian processes
Optimal recovery
Nonlinear PDEs
High-dimensional PDEs
Parametric PDEs

ABSTRACT

We introduce a priori Sobolev-space error estimates for the solution of arbitrary nonlinear, and possibly parametric, PDEs that are defined in the strong sense, using Gaussian process and kernel based methods. The primary assumptions are: (1) a continuous embedding of the reproducing kernel Hilbert space of the kernel into a Sobolev space of sufficient regularity; and (2) the stability of the differential operator and the solution map of the PDE between corresponding Sobolev spaces. The proof is articulated around Sobolev norm error estimates for kernel interpolants and relies on the minimizing norm property of the solution. The error estimates demonstrate dimension-benign convergence rates if the solution space of the PDE is smooth enough. We illustrate these points with applications to high-dimensional nonlinear elliptic PDEs and parametric PDEs. Although some recent machine learning methods have been presented as breaking the curse of dimensionality in solving high-dimensional PDEs, our analysis suggests a more nuanced picture: there is a trade-off between the regularity of the solution and the presence of the curse of dimensionality. Therefore, our results are in line with the understanding that the curse is absent when the solution is regular enough.

1. Introduction

In recent years the adoption of machine learning in the natural sciences and engineering has led to the development of new methods for solving PDEs [70,86,87,47,52]. The majority of these methods rely on the approximation power of artificial neural networks (ANNs) either as a function class to approximate the solution of the PDE or as a high-dimensional function class to approximate the solution map of the PDE. Despite the empirical success of the aforementioned ANN based methods, current theoretical understanding of these PDE solvers is scarce and, beyond particular PDEs (e.g., [81,53,25]), results are oftentimes limited to existence results rather than convergence guarantees or rates.

Similar to ANNs, kernel methods and Gaussian processes (GPs) have been very effective in scientific computing and machine learning [80,59,7,89] and at the same time they are supported by rigorous theoretical foundation [7,88,66]. Recently in [12], the authors introduced a kernel collocation method for solving arbitrary nonlinear PDEs with a rigorous convergence guarantee. The

* Corresponding author.
E-mail addresses: pbatllef@caltech.edu (P. Batlle), yifanc@caltech.edu (Y. Chen), bamdadh@uw.edu (B. Hosseini), owhadi@caltech.edu (H. Owhadi), astuart@caltech.edu (A.M. Stuart).

https://doi.org/10.1016/j.jcp.2024.113488
Received 1 May 2024; Received in revised form 26 September 2024; Accepted 4 October 2024

Available online 10 October 2024

theory presented in that work was based on the assumptions that (1) the solution belongs to the reproducing kernel Hilbert space (RKHS) defined by the underlying kernel which in turn is embedded in the Sobolev space H^s for $s > d/2 +$ "order of the PDE" (where d is the dimension of the domain of the PDE) and (2) the fill-distance between collocation points goes to zero. Convergence was proved via a compactness argument but no convergence rates were provided.

The goal of this article is to provide quantitative convergence rates for the PDE solver introduced in [12]. Our quantitative rates also reveal the interplay between the regularity of the solution of the PDE and the dimension d of the problem. At the same time we make improvements to the methodology of [12] and extend it to the case of parametric PDEs. In the rest of this section we summarize our main contributions in Subsection 1.2 followed by a review of the relevant literature in Subsection 1.3, and an outline of the article in Subsection 1.5.

1.1. Significance of contributions

The use of meshless, collocation methods with radial basis functions for solving PDEs dates back to the 1990s [39,40,30,26]. Typical approaches include symmetric collocation [30,26] and unsymmetric collocation methods (also known as the Kansa method [39,40]); see [77,28] for reviews. It is recognized that the unsymmetric collocation approach may encounter instability issues and require additional techniques [48,74,49,15], whereas the symmetric collocation approach always yields positive definite symmetric matrices and is stable. The reason is that the symmetric collocation approach includes higher-order derivatives of the kernel as basis functions and leads to solutions that can be identified as *optimal recovery* (worst case minimax optimal) solutions. This optimal recovery property makes the algorithm generally applicable for *any well-posed linear PDEs* [38,75,76]. Furthermore the analysis becomes straightforward because of this optimality; see [29,30] for linear PDEs and [9,10] for some degree of generalization to quasi-linear/nonlinear problems. Note that the optimality has long been recognized, however not extensively acknowledged, as highlighted in [75]: "This technique has been around since at least 1998, but its optimality properties went unnoticed." The GP-PDE methodology proposed in [12] can be seen as a nonlinear generalization of the optimal recovery approach to solving (and learning) *arbitrary classically/strongly defined nonlinear PDEs*. This paper aims to offer a simple and transparent theoretical error analysis of this nonlinear optimal recovery method. This analysis extends the linear setting [33] and shares conceptual steps (in terms of the role of stability and sampling inequalities) with [9] while explicitly focusing on optimal recovery solutions rather than solutions obtained from finite-dimensional trial spaces and residual minimization. Such theoretical analyses are notably rare within the sphere of machine learning-based PDE solutions (where theoretical guarantees are typically limited to existence results). Furthermore, while there is no general theory for strongly defined arbitrary nonlinear PDEs, the optimal recovery approach provides a way of obtaining general theoretical guarantee for the numerical approximation of such PDEs. Beyond its wide scope, the proposed analysis also lays the groundwork for developing rigorous, efficient, and scalable (near-linear complexity) learning-based methods for arbitrary nonlinear PDEs. This can be achieved by integrating the proposed error estimates with the fast algorithms developed in [13] for kernel matrices whose entries contain higher-order derivatives of the kernel, a setting well suited for the optimal recovery approach. We note that the analysis in the paper is focused on the minimizer of a loss function induced by kernel and GP methods. Understanding theoretically how iterative algorithms are able to achieve this minimizer represents another crucial stride towards our ultimate objective; this could be potentially done by combining analysis results for iterative linearization, for example the work in [6].

1.2. Summary of contributions

Throughout the article we consider parametric PDEs of the form

$$\begin{cases} \mathcal{P}(u^\star)(\mathbf{x};\boldsymbol{\theta}) = f(\mathbf{x};\boldsymbol{\theta}), & (\mathbf{x},\boldsymbol{\theta}) \in \Omega \times \Theta, \\ \mathcal{B}(u^\star)(\mathbf{x};\boldsymbol{\theta}) = g(\mathbf{x};\boldsymbol{\theta}), & (\mathbf{x},\boldsymbol{\theta}) \in \partial\Omega \times \Theta, \end{cases} \tag{1.1}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded connected domain with an appropriately smooth boundary $\partial\Omega$, $\mathcal{P}$ and $\mathcal{B}$ are the interior and boundary differential operators that define the PDE and f, g are the source and boundary data. $\mathbf{x}$ denotes the spatial variable with $\boldsymbol{\theta}$ denoting a parameter belonging to a compact set $\Theta \subset \mathbb{R}^p$. The function $u^\star$ denotes the exact, strong solution of this PDE.

We view the solution $u^\star$ as a function on $\overline{\Omega} \times \Theta$ and approximate it in an appropriate RKHS by imposing the PDE as a constraint on a set of collocation points in the product space $\overline{\Omega} \times \Theta$. Our main contributions are four-fold as summarized below:

1. We extend the kernel PDE solver of [12] to the case of the parametric PDE (1.1). This extension follows by viewing the solution $u^\star(\mathbf{x};\theta)$ as a continuous function defined on $\overline{\Omega} \times \Theta$ and approximating it with a function $u^\dagger$ in an appropriate RKHS $\mathcal{U}$ after imposing the PDE as a constraint on a set of collocation points. At the same time we improve the efficacy and performance of the Gauss-Newton (GN) algorithm of [12] through an approach of "linearize first then apply the kernel solver". For many prototypical PDEs this new approach leads to smaller kernel matrices that can be factored or inverted more efficiently. These numerical strategies are outlined in Section 2, and our proposed methodology is summarized in Algorithms 2.1 and 2.2.
2. We provide explicit a priori convergence rates for the kernel estimator $u^\dagger \in \mathcal{U}$ to the true solution $u^\star$. Our proof relies on three assumptions: (1) the RKHS $\mathcal{U} \subset H^s(\Omega)$ for $s > (d+p)/2 +$ "order of the PDE"; (2) the true unique solution $u^\star \in \mathcal{U}$; (3) the forward PDE operator and the associated solution map of the PDE are Lipschitz stable. Our error estimates are of the general form

$$\|u^\dagger - u^\star\|_{L^2(\Omega)} \lesssim h^s \|u^\star\|_{\mathcal{U}}, \tag{1.2}$$

$$
\begin{aligned}
\text{Solution operator is stable} &\longrightarrow \|u^\dagger - u_N\|_1 \le C\|\mathcal{P}(u^\dagger) - \mathcal{P}(u_N)\|_2 \\
\text{Sampling error estimates} &\longrightarrow \|\mathcal{P}(u^\dagger) - \mathcal{P}(u_N)\|_2 \le Ch^\gamma \|\mathcal{P}(u^\dagger) - \mathcal{P}(u_N)\|_3 \\
\text{Forward operator is stable} &\longrightarrow \|\mathcal{P}(u^\dagger) - \mathcal{P}(u_N)\|_3 \le C\|u^\dagger - u_N\|_4 \\
\text{Continuous embedding of RKHS} &\longrightarrow \|u^\dagger - u_N\|_4 \le C\|u^\dagger - u_N\|_{\mathcal{U}} \\
\|u_N\|_{\mathcal{U}} \le \|u^\dagger\|_{\mathcal{U}} &\longrightarrow \|u^\dagger - u_N\|_{\mathcal{U}} \le 2\|u^\dagger\|_{\mathcal{U}} \\
&\longrightarrow \|u^\dagger - u_N\|_1 \le Ch^\gamma \|u^\dagger\|_{\mathcal{U}}
\end{aligned}
$$

Fig. 1. A summary of the main steps in our proof of convergence rates outlined in Theorems 3.1, 3.3, 3.8, and 3.12. The 1–4 norms denote arbitrary norms on appropriate Banach spaces while the $\|\cdot\|_{\mathcal{U}}$-norm can be chosen as an RKHS norm or another desired norm with respect to which the numerical algorithm is stable.

where h is the fill-distance (mesh-norm) of our collocation points. Indeed, if expressed in terms of N, the number of collocation points, the rate will read as $\mathcal{O}(N^{-s/(d+p)})$. The above rate indicates a trade-off between the regularity of the solution space and the dimension $d+p$ of $\Omega \times \Theta$ stating that the convergence rate is dimension-benign so long as the solution $u^\star$ is sufficiently regular; these results are outlined in Section 3.

3. In fact, our method for proving the rate (1.2) is more general than the case of PDEs (see Fig. 1 for the road map of the proof technique). The proof can be viewed as a recipe for convergence analysis of solutions to nonlinear functional equations of the form $\mathcal{P}(u) = f$ where u, f belong to sufficiently regular function spaces and $\mathcal{P}$ is invertible (at least locally). Then the Lipschitz stability of $\mathcal{P}$ and $\mathcal{P}^{-1}$ plus RKHS interpolation bounds on f yield convergence rates for u. Results at this level of generality are presented in [76] for linear maps $\mathcal{P}$ which are then extended to nonlinear problems in [9,10]. In these works, the stability of the discretization method is furthermore assumed. In the GP methodology, this property is guaranteed, due to the minimal RKHS norm and optimal recovery property of the solution. This has been pointed out in [76, Sec. 10] for linear PDEs. Our theory can be seen as a generalization of the result in [76] to the nonlinear case.
4. We present a suite of numerical experiments that elucidate and extend our theoretical analysis in item 2. We present an example of a nonlinear elliptic PDE with a prescribed solution of varying regularity in various dimensions. We then explore the interplay between regularity and dimensionality as well as the rate in (1.2). We further verify our result for a one dimensional parametric PDE by varying p, the dimension of the parameter space Θ. Because of this trade-off between regularity and dimensionality, showing that a numerical method remains accurate for a high dimensional PDE may not be an indication that it is breaking the curse of dimensionality but simply an indication that the problem being solved is very regular; see our experiments in Section 4 and in particular Subsection 4.3.

1.3. Literature review

Below we present a brief review of the literature relevant to the current work.

1.3.1. Kernel and Gaussian process solvers for PDEs

As mentioned earlier our algorithmic and theoretical developments are focused on the kernel method introduced in [12] and extending that approach to parametric PDEs. Further extensions and applications of the aforementioned framework can also be found in [58,57,50,13]. When applied to linear PDEs our kernel method coincides with the so-called symmetric collocation method [77, Sec. 14] and is closely associated with radial basis function (RBF) PDE solvers [28,26,30]. Various error analyses for RBF collocation methods can be found in [29,30]. In particular, the article [33] is the closest to our work and their rates coincide with ours in the linear PDE setting. The articles [48,74,49,15] present similar bounds for the so-called Kansa method [39,40], a non-symmetric RBF collocation PDE solver. Finally, [76] presents an abstract set of convergence rates for RBF interpolation of "well-posed" linear maps between regular function spaces that includes RBF PDE solvers as a special case. All of the aforementioned analyses consider linear PDEs and some generalizations to nonlinear problems are studied in [9,10].

The deep connection of Kernels and RKHSs to the theory of GPs [8,43,89,84] suggests that kernel PDE solvers can be viewed from lens of probability theory as a conditioning problem for GPs. While not as extensively developed as the kernel solvers mentioned earlier, this direction has been explored for the solution of linear PDEs as well as nonlinear ODEs [18,21,64,73,82] and recent works have extended this idea to some nonlinear and time-dependent PDEs [20,69,85]. The GP interpretation is attractive due to the ability to provide rigorous uncertainty estimates along with the solution to the PDE. The idea here is that the uncertainties can serve as a posterior or a priori error indicators for the PDE solver. Some ideas related to this direction were discussed in [12,20]. A fully probabilistic GP interpretation of our kernel framework for linear PDEs can be found in [64,20,66] but the case of nonlinear PDEs remains partially investigated [12,57,58,51]. Moreover we note that in the GP framework, hierarchical Bayes learning can be used to select kernels to get better convergence rates [14,90,67,24].

Table 1
A qualitative comparison of the properties of traditional PDE solvers (such as FEM, FVM, FDM, spectral methods, etc) against kernel methods and ANNs.

Method	Ease of implementation in high-dimensions	Provable guarantees	Near linear complexity	Occam's razor	Transparent	Ease of reproducibility	Built-in UQ	Software support
Trad.	✗	✓	✓	✓	✓	✓	✗	✓
Kernel	✓	✓	✓	✓	✓	✓	✓	Limited
ANN	✓	Limited	✗	✗	✗	Limited	✗	✓

1.3.2. Parametric and high-dimensional PDEs

Parametric PDEs are ubiquitous in physical sciences and engineering and in particular in the context of uncertainty quantification (UQ) and solution of stochastic PDEs (SPDEs) [19,32,44,92]. A vast literature exists on the subject, connecting it to reduced basis models [1,63], emulation of computer codes [42], reduced order models [54], and numerical homogenization [66]; for settings that most closely resemble our problems we refer the reader to [32,91,22] for a general overview. Broadly speaking, the dominant approaches for approximation of high-dimensional and parametric solution maps include polynomial/Taylor approximation methods [5,16,17,62,61]; Galerkin methods [36,23]; reduced basis methods [37]; and more recently ANN operator learning techniques such as [47,52]. In comparison to the aforementioned works we propose to directly approximate the solution of the parametric PDE as a function on the tensor product space of the physical and parameter domains in a similar spirit as [41]. The recent article [4] also presents a kernel based operator learning approach to various PDE problems including parametric PDEs.

1.3.3. The curse of dimensionality

Although the trade-off between regularity and accuracy is well understood in numerical approximation/integration, where it has led to the development of the Kolmogorov N-width and stress tests for finite-element methods [68,56,3], its impact is oftentimes overlooked when communicating the convergence of Machine Learning and Deep Learning methods for high-dimensional PDEs. In particular, since artificial neural networks (ANNs) can be interpreted as kernel methods [60,45,65] with data-dependent parameterized kernels, our results raise the further question of understanding whether the (empirically observed) convergence of ANN-based methods for high-dimensional PDEs is an indication of the absence of the curse (i.e., the regularity of the solution in selected numerical experiments is high) or the breaking of that curse. In particular, empirically observing numerical accuracy for an algorithm and particular solutions is insufficient to prove that the curse of dimensionality is broken, and one must also show that the underlying problem and those solutions are not too regular. We emphasize that the curse of dimensionality referred to here, is the one associated with the worsening of the accuracy of a numerical approximation algorithm as a function of the dimension of the domain of the PDE as opposed to the impact of the curse on the number of degrees of freedom in the implementation level (e.g., finite difference methods suffer from that second curse but ANN/kernel based methods do not).

1.4. The potential value of kernel/GP methods

The proposed work aims to further develop Gaussian Process (GP) and kernel methods for solving PDEs. We are motivated to do so because GP methods have the potential to offer the best of both worlds by combining the profound theory underlying traditional methods (and in particular finite element methods) with the ease of implementation of emerging Deep Learning (DL) methods. They also come equipped with automatic uncertainty quantification (UQ) capabilities, not readily available in either traditional or deep-learning based methods. And finally they provide easily implementable meshless methods that can be used to benchmark other machine learning based algorithms such as PINNs [70].

Compared to traditional methods (such as finite element methods (FEM), finite volume methods (FVM), finite difference methods (FDM), spectral methods, etc), GP methods generalize meshless, RBF, optimal recovery methods and are flexible and applicable in high dimensions. Compared to DL methods that use an expressive neural network representation, GPs offer transparent methods that are easy to reproduce and analyze. Furthermore the natural probabilistic interpretation of GPs enables convenient UQ and also facilitates the process of scientific discovery itself [66]; FEM and DL methods do not interface so cleanly with UQ. Moreover, with hierarchical kernel learning [14,90,67,24], GP methods can also be made highly expressive. In fact, as shown in the table below, GP methods can offer many advantages over traditional and DL methods. In the context of PDEs these advantages include: greater flexibility, applicability in high dimensions, provable guarantees, near-linear complexity computation, Occam's razor principle in the design of statistical models, mathematical transparency and interpretability, and ease of reproducibility; see Table 1. Although software support for GPs is currently not as advanced as that for DL and traditional methods, GPs are still easy to program and can be seamlessly integrated into an engineering pipeline. Table 1 should be interpreted in this light: as an argument for further deployment and development of software infrastructure for GP-PDE based methods.

Given their long training times, ANN-based methods may not be competitive with FEM in low dimensions [35]. In contrast, GP-based methods can achieve near-linear complexity when combined with fast algorithms for kernel methods such as the sparse Cholesky factorization [78,79,13]. In some applications, these algorithms can be competitive (both in terms of complexity and accuracy) even when compared to highly optimized algebraic multigrid solvers such as AMGCL and Trilinos [11]. GP methods are naturally amenable to analysis and come with simple provable guarantees, while ANN-based methods involve complicated optimizations and many heuristics, which can make them hard to understand. GP methods fit Occam's razor, offering a clarity of purpose in their structure.

We can understand why and when they work, which is of scientific importance [27]. Therefore, it is of potential value to benchmark deep learning methods against kernel-based methods to ensure that the deep part of a DL method serves a significant purpose beyond adding complexity.

1.5. Outline of the article

The rest of the article is organized as follows: We present a brief overview of our GP and kernel approach for solving nonlinear and parametric PDEs in Section 2; our error analysis is outlined in Section 3; followed by numerical experiments in Section 4 and conclusions in Section 5. Auxiliary results are collected in Appendices A to C.

2. Kernel methods for parametric PDEs

In this section we extend the kernel methodology of [12] to the case of parametric PDEs as outlined in Subsection 2.1. Some numerical strategies and ideas for improving the efficiency of the solver are discussed in Subsection 2.2.

2.1. Solving parametric PDEs

Let us consider bounded connected domains $\Omega \in \mathbb{R}^d$ with a Lipschitz boundary for $d \geq 1$ and $\Theta \subset \mathbb{R}^p$ for $p \geq 1$. We consider nonlinear and parametric PDEs of the form

$$\begin{cases} \mathcal{P}(u^\star)(\mathbf{x};\boldsymbol{\theta}) = f(\mathbf{x};\boldsymbol{\theta}), & (\mathbf{x},\boldsymbol{\theta}) \in \Omega \times \Theta, \\ \mathcal{B}(u^\star)(\mathbf{x};\boldsymbol{\theta}) = g(\mathbf{x};\boldsymbol{\theta}), & (\mathbf{x},\boldsymbol{\theta}) \in \partial\Omega \times \Theta, \end{cases} \tag{2.1}$$

where $\mathcal{P}, \mathcal{B}$ are nonlinear differential operators in the interior and boundary of Ω, $\boldsymbol{\theta}$ is a parameter, and f, g are the PDE source and boundary data. For now we assume that the above PDE is well-posed and has a unique solution $u^\star(\mathbf{x};\boldsymbol{\theta})$ which is assumed to exist in the strong sense over $\overline{\Omega}$ and for all values of $\boldsymbol{\theta} \in \Theta$. In [12] the authors introduced a GP/kernel method for solving nonlinear PDEs of the form (2.1) without parametric dependence. Here we extend that approach to the parametric case.

Let $\Upsilon := \Omega \times \Theta$ and $\partial\Upsilon := \partial\Omega \times \Theta$ and write $\mathbf{s} := (\mathbf{x}, \boldsymbol{\theta})$. Choose $M \geq 1$ collocation points $\{\mathbf{s}_m\}_{m=1}^M \in \overline{\Upsilon}$ such that $\{\mathbf{s}_m\}_{m=1}^{M_\Omega} \in \Upsilon$ and $\{\mathbf{s}_m\}_{m=M_\Omega+1}^{M} \in \partial\Upsilon$[1] and consider a kernel $K : \overline{\Upsilon} \times \overline{\Upsilon} \to \mathbb{R}$ with its corresponding RKHS denoted by $\mathcal{U}$ and norm $\|\cdot\|_{\mathcal{U}}$. We then propose to approximate $u^\star(\mathbf{s})$ by solving the optimization problem

$$\begin{cases} \underset{u \in \mathcal{U}}{\text{minimize}} & \|u\|_{\mathcal{U}} \\ \text{s.t.} & \mathcal{P}(u)(\mathbf{s}_m) = f(\mathbf{s}_m), \quad m = 1, \dots, M_\Omega, \\ & \mathcal{B}(u)(\mathbf{s}_m) = g(\mathbf{s}_m), \quad m = M_\Omega + 1, \dots, M. \end{cases} \tag{2.2}$$

Observe that our approach above approximates the solution u as a function defined on the set product set Υ which is different from previous works [5,16,17,62,61] where the solution map $\boldsymbol{\theta} \to u^\star(\cdot;\boldsymbol{\theta})$, as a mapping from Θ to an appropriate function space, is characterized and approximated. The latter approach requires different discretization methods for the $\boldsymbol{\theta}$ parameter and the functions $u^\star(\cdot;\boldsymbol{\theta})$ while our approach leads to a meshless collocation method on the product space which is desirable and convenient at the level of implementation, following [12, Sec. 3.1] (see also [33]).

We make the following assumption on the differential operators $\mathcal{P}, \mathcal{B}$.

Assumption 2.1. There exist bounded and linear operators $L_1, \dots, L_{Q_\Omega} \in \mathcal{L}\big(\mathcal{U}; C(\Upsilon)\big)$ and $L_{Q_\Omega+1}, \dots, L_Q \in \mathcal{L}\big(\mathcal{U}; C(\partial\Upsilon)\big)$ for some $1 \leq Q_\Omega < Q$ together with maps $P : \mathbb{R}^{Q_\Omega} \to \mathbb{R}$ and $B : \mathbb{R}^{Q-Q_\Omega} \to \mathbb{R}$, which may be nonlinear, so that $\mathcal{P}, \mathcal{B}$ can be written as

$$\begin{aligned} \mathcal{P}(u)(\mathbf{s}) &= P\Big(L_1(u)(\mathbf{s}), \dots, L_{Q_\Omega}(u)(\mathbf{s})\Big) \qquad \forall \mathbf{s} \in \Upsilon, \\ \mathcal{B}(u)(\mathbf{s}) &= B\Big(L_{Q_\Omega+1}(u)(\mathbf{s}), \dots, L_Q(u)(\mathbf{s})\Big) \quad \forall \mathbf{s} \in \partial\Upsilon. \quad \diamondsuit \end{aligned} \tag{2.3}$$

We briefly introduce a running example of a parametric PDE for which the above assumptions can be verified easily.

Example 2.2 *(Nonlinear Darcy flow).* Consider the nonlinear Darcy flow PDE

$$\begin{cases} -\mathrm{div}_{\mathbf{x}}\big(\exp(a(\mathbf{x},\boldsymbol{\theta}))\nabla u\big)(\mathbf{x}) + \tau(u(\mathbf{x})) = 1, & \mathbf{x} \in \Omega, \\ u(\mathbf{x}) = 0, & \mathbf{x} \in \partial\Omega, \end{cases} \tag{2.4}$$

[1] Note that we do not specifically ask for collocation points on $\partial\Omega \times \partial\Theta$ since we may not have boundary data on the $\boldsymbol{\theta}$ parameter.

where $\Omega \subset \mathbb{R}^d$ is a bounded domain with a Lipschitz boundary and $\tau : \mathbb{R} \to \mathbb{R}$ is a continuous and nonlinear map. We assume that the permeability field a is parameterized as

$$a(\mathbf{x}, \boldsymbol{\theta}) = \sum_{j=1}^{p} \theta_j \psi_j(\mathbf{x}), \tag{2.5}$$

where $\theta_j \in (0,1)$ so that $\Theta = (0,1)^p$ and $\psi_j \in C(\overline{\Omega})$. Substituting into the PDE and expanding the differential operator we can rewrite our nonlinear PDE as

$$\begin{cases} -\exp\left(\sum_{j=1}^{p} \theta_j \psi_j(\mathbf{x})\right) \sum_{j=1}^{p} \theta_j \nabla_{\mathbf{x}} \psi_j(\mathbf{x}) \cdot \nabla_{\mathbf{x}} u(\mathbf{x};\boldsymbol{\theta}) \\ \qquad -\exp\left(\sum_{j=1}^{p} \theta_j \psi_j(\mathbf{x})\right) \Delta_{\mathbf{x}} u(\mathbf{x};\boldsymbol{\theta}) + \tau(u(\mathbf{x};\boldsymbol{\theta})) = 1, \quad (\mathbf{x},\boldsymbol{\theta}) \in \Omega \times \Theta, \\ \qquad\qquad u(\mathbf{x};\boldsymbol{\theta}) = 0, \quad (\mathbf{x},\boldsymbol{\theta}) \in \partial\Omega \times \Theta, \end{cases}$$

where we used subscripts on the differential operators to highlight that derivatives are computed for the $\mathbf{x}$ variable only and not $\boldsymbol{\theta}$. We also did not use the compact notation $\mathbf{s} \equiv (\mathbf{x}, \boldsymbol{\theta})$ since it is more helpful to be able to distinguish between the $\mathbf{x}$ and $\boldsymbol{\theta}$ variables in this example. We can directly verify Assumption 2.1 with the bounded and linear operators

$$\begin{aligned} &L_1 : u(\mathbf{x};\boldsymbol{\theta}) \mapsto u(\mathbf{x};\boldsymbol{\theta}), \\ &L_2 : u(\mathbf{x};\boldsymbol{\theta}) \mapsto \exp\left(\sum_{j=1}^{p} \theta_j \psi_j(\mathbf{x})\right) \sum_{j=1}^{p} \theta_j \nabla_{\mathbf{x}} \psi_j(\mathbf{x}) \cdot \nabla_{\mathbf{x}} u(\mathbf{x};\boldsymbol{\theta}), \\ &L_3 : u(\mathbf{x};\boldsymbol{\theta}) \mapsto \exp\left(\sum_{j=1}^{p} \theta_j \psi_j(\mathbf{x})\right) \Delta_{\mathbf{x}} u(\mathbf{x};\boldsymbol{\theta}) \\ &L_4 : u(\mathbf{x};\boldsymbol{\theta}) \mapsto u(\mathbf{x};\boldsymbol{\theta}). \end{aligned}$$

Note that operators L_1, L_4 are the same here since the point values of u appear in both the interior and boundary conditions. Thus we have $Q_\Omega = 3$ and $Q = 4$ and the maps

$$P(t_1,t_2,t_3) = -t_2 - t_3 + \tau(t_1), \qquad B(t_1) = t_1. \quad \diamondsuit$$

If $\mathcal{U}$ is sufficiently regular and Assumption 2.1 holds, then we can define the functionals $\phi_m^q \in \mathcal{U}^*$ for $1 \le q \le Q$ as

$$\phi_m^q := \delta_{(\mathbf{s}_m)} \circ L_q, \quad \text{where} \quad \begin{cases} 1 \le m \le M_\Omega & \text{if} \quad 1 \le q \le Q_\Omega \\ M_\Omega + 1 \le m \le M & \text{if} \quad Q_{\Omega+1} \le q \le Q. \end{cases} \tag{2.6}$$

In what follows we write $[\phi, u]$ to denote the duality pairing between $\mathcal{U}$ and $\mathcal{U}^*$ and further use the shorthand notation $\boldsymbol{\phi}^{(q)}$ to denote the vector of dual elements ϕ_m^q for a fixed index q. Note that $\boldsymbol{\phi}^{(q)} \in (\mathcal{U}^*)^{\otimes M_\Omega}$ if $q \le Q_\Omega$ but $\boldsymbol{\phi}^{(q)} \in (\mathcal{U}^*)^{\otimes (M - M_\Omega)}$ if $q > Q_\Omega$ in order to accommodate different differential operators defining the PDE and the boundary conditions. We further write $N = M_\Omega Q_\Omega + (M - M_\Omega)(Q - Q_\Omega)$ and define

$$\boldsymbol{\phi} = \left(\boldsymbol{\phi}^{(1)}, \dots, \boldsymbol{\phi}^{(Q)}\right) \in (\mathcal{U}^*)^{\otimes N}.$$

Henceforth we write ϕ_n for $n = 1, \dots, N$ to denote the entries of the vector $\boldsymbol{\phi}$ and write $[\boldsymbol{\phi}, u] = ([\phi_1, u], \dots, [\phi_N, u]) \in \mathbb{R}^N$. With this notation we rewrite problem (2.2) as

$$\begin{cases} \text{minimize} & \|u\|_{\mathcal{U}} \\ \text{s.t.} & F([\boldsymbol{\phi}, u]) = \mathbf{y}, \end{cases}$$

where the data vector $\mathbf{y} \in \mathbb{R}^M$ has entries

$$y_m := \begin{cases} f(\mathbf{s}_m), & \text{if} \quad 1 \le m \le M_\Omega, \\ g(\mathbf{s}_m), & \text{if} \quad M_\Omega + 1 \le m \le M, \end{cases}$$

and $F : \mathbb{R}^N \to \mathbb{R}^M$ is a nonlinear map whose output components are defined as

$$\left(F([\boldsymbol{\phi}, u])\right)_m := \begin{cases} P\left([\phi_m^1, u], \dots, [\phi_m^{Q_\Omega}, u]\right) & \text{if } 1 \le m \le M_\Omega, \\ B\left([\phi_m^{Q_\Omega+1}, u], \dots, [\phi_m^{Q}, u]\right) & \text{if } M_\Omega + 1 \le m \le M. \end{cases} \tag{2.7}$$

Further define the kernel vector field

$$K(\cdot,\boldsymbol{\phi}):\overline{\Upsilon}\to\mathcal{U}^N,\qquad K(\mathbf{s},\boldsymbol{\phi})_k:=[\phi_k,K(\mathbf{s},\cdot)] \tag{2.8}$$

and the kernel matrix

$$K(\boldsymbol{\phi},\boldsymbol{\phi})\in\mathbb{R}^{N\times N},\qquad K(\boldsymbol{\phi},\boldsymbol{\phi})_{nk}:=[\phi_n,K(\cdot,\boldsymbol{\phi})_k]. \tag{2.9}$$

We can then characterize the minimizers of (2.2) via the following representer theorem which is a direct consequence of [12, Prop. 2.3]:

Proposition 2.3. *Suppose Assumption 2.1 holds and $K(\boldsymbol{\phi},\boldsymbol{\phi})$ is invertible. Then a function $u^\dagger:\overline{\Upsilon}\to\mathbb{R}$ is a minimizer of (2.2) if and only if*

$$u^\dagger(\mathbf{s})=K(\mathbf{s},\boldsymbol{\phi})K(\boldsymbol{\phi},\boldsymbol{\phi})^{-1}\mathbf{z}^\dagger, \tag{2.10}$$

where $\mathbf{z}^\dagger\in\mathbb{R}^N$ solves

$$\begin{cases}\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} & \mathbf{z}^T K(\boldsymbol{\phi},\boldsymbol{\phi})^{-1}\mathbf{z},\\ \text{s.t.} & F(\mathbf{z})=\mathbf{y},\end{cases} \tag{2.11}$$

with the nonlinear map F defined in (2.7).

This result allows us to reduce the infinite-dimensional optimization problem (2.2) to a finite-dimensional optimization problem without incurring any approximation errors; it is an instance of the well-known family of representer theorems [80, Sec. 4.2]. Thus, to find an approximation to $u^\star$ we simply need to solve (2.11) and apply the formula (2.10); algorithms for this task are discussed next.

2.2. Numerical strategies

We now summarize various numerical strategies for solution of (2.11). These strategies are naturally applicable to non-parametric PDEs as they can be viewed as a special case of (2.1) with a fixed parameter. In Subsection 2.2.1 we summarize a Gauss-Newton algorithm that was introduced in [12] followed by a new and, often, more efficient strategy that linearizes the PDE first before formulating the optimization problem in Subsection 2.2.2.

2.2.1. Gauss-Newton

To solve the optimization problem (2.11), a Gauss-Newton algorithm was proposed in [12] which we recall briefly. The equality constraints can be dealt with either by elimination or relaxation. Suppose that there exists a map $\overline{F}:\mathbb{R}^{N-M}\times\mathbb{R}^M\to\mathbb{R}^N$ so that

$$F(\mathbf{z})=\mathbf{y}\quad\text{if and only if}\quad\mathbf{z}=\overline{F}(\mathbf{w},\mathbf{y}),\quad\text{for a unique }\mathbf{w}\in\mathbb{R}^{N-M}.$$

Then, we rewrite (2.11) as the unconstrained optimization problem

$$\underset{\mathbf{w}\in\mathbb{R}^{N-M}}{\text{minimize}}\quad\overline{F}(\mathbf{w},\mathbf{y})^T K(\boldsymbol{\phi},\boldsymbol{\phi})^{-1}\overline{F}(\mathbf{w},\mathbf{y}). \tag{2.12}$$

Then a minimizer $\mathbf{w}^\dagger$ of (2.12) can be approximated with a sequence of elements $\mathbf{w}^\ell$ defined iteratively via $\mathbf{w}^{\ell+1}=\mathbf{w}^\ell+\alpha^\ell\delta\mathbf{w}^\ell$, where $\alpha^\ell>0$ is an appropriate step size while $\delta\mathbf{w}^\ell$ is the minimizer of the optimization problem

$$\underset{\delta\mathbf{w}\in\mathbb{R}^{N-M}}{\text{minimize}}\quad\left(\overline{F}(\mathbf{w}^\ell,\mathbf{y})+\nabla_{\mathbf{w}}\overline{F}(\mathbf{w}^\ell,\mathbf{y})\delta\mathbf{w}\right)^T K(\boldsymbol{\phi},\boldsymbol{\phi})^{-1}\left(\overline{F}(\mathbf{w}^\ell,\mathbf{y})+\nabla_{\mathbf{w}}\overline{F}(\mathbf{w}^\ell,\mathbf{y})\delta\mathbf{w}\right).$$

Alternatively, if the map $\overline{F}$ does not exist or is hard to compute, i.e., eliminating the constraints is not feasible, then we consider the relaxed problem

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\quad\frac{1}{2}\mathbf{z}^T K(\boldsymbol{\phi},\boldsymbol{\phi})^{-1}\mathbf{z}+\frac{1}{2\beta^2}|F(\mathbf{z})-\mathbf{y}|^2,$$

for a sufficiently small parameter $\beta>0$. Here $\|\cdot\|$ is the L^2 norm of the vector. A minimizer $\mathbf{z}^\dagger_\beta$ of the above problem can be approximated with a sequence $\mathbf{z}^\ell$ where $\mathbf{z}^{\ell+1}=\mathbf{z}^\ell+\alpha^\ell\delta\mathbf{z}^\ell$ where $\delta\mathbf{z}^\ell$ is the minimizer of

$$\underset{\delta\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\quad\delta\mathbf{z}^T K(\boldsymbol{\phi},\boldsymbol{\phi})^{-1}\mathbf{z}^\ell+\frac{1}{2\beta^2}|F(\mathbf{z}^\ell)+\nabla F(\mathbf{z}^\ell)\delta\mathbf{z}-\mathbf{y}|^2.$$

We summarize the proposed Gauss-Newton algorithm for solution of parametric PDEs in Algorithm 2.1.

2.2.2. Linearize then optimize

The Gauss-Newton approach of Subsection 2.2.1 is applicable to wide families of nonlinear PDEs. The primary computational bottleneck of that approach is the construction and factorization of the kernel matrix $K(\boldsymbol{\phi},\boldsymbol{\phi})$ which for some PDEs can be prohibitively large. To get around this difficulty we propose an alternative approach to approximating the solution of (2.2) by first linearizing the

Algorithm 2.1 Kernel Methods for Parametric PDEs using Gauss-Newton (Subsection 2.2.1).

Input: PDE of the form (2.1) defined on $\Upsilon = \Omega \times \Theta$ with boundary condition on $\partial\Upsilon = \partial\Omega \times \Theta$, $M \geq 1$ collocation points in $\overline{\Upsilon}$, and kernel $K : \overline{\Upsilon} \times \overline{\Upsilon} \to \mathbb{R}$

Output: Approximation $u^\dagger(s)$ to exact solution $u^*(s)$

$N := M_\Omega Q_\Omega + (M - M_\Omega)(Q - Q_\Omega)$
for $i = 1 : N$ **do** ▷ Build Kernel Matrix $K(\boldsymbol{\phi}, \boldsymbol{\phi})$ (2.9)
  for $j = 1 : N$ **do**
    $K(\boldsymbol{\phi}, \boldsymbol{\phi})_{i,j} = [\phi_i, K(\cdot, \boldsymbol{\phi})_j]$
  end for
end for
while not converged **do**
  $\delta\mathbf{w}^\ell := \arg\min_{\delta\mathbf{w}} \left(\overline{F}(\mathbf{w}^\ell, \mathbf{y}) + \nabla_{\mathbf{w}}\overline{F}(\mathbf{w}^\ell, \mathbf{y})\delta\mathbf{w}\right)^T K(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \left(\overline{F}(\mathbf{w}^\ell, \mathbf{y}) + \nabla_{\mathbf{w}}\overline{F}(\mathbf{w}^\ell, \mathbf{y})\delta\mathbf{w}\right)$
  $\mathbf{w}^{\ell+1} = \mathbf{w}^\ell + \alpha^\ell \delta\mathbf{w}^\ell$,
end while
$\mathbf{z}^\dagger := \overline{F}(\mathbf{w}, \mathbf{y})$
$u^\dagger(\mathbf{s}) := K(\mathbf{s}, \boldsymbol{\phi}) K(\boldsymbol{\phi}, \boldsymbol{\phi})^{-1} \mathbf{z}^\dagger$ ▷ Apply (2.10)

PDE operators before applying Proposition 2.3. The resulting approach is more intrusive in comparison to the Gauss-Newton method as it requires explicit calculations involving the PDE but often leads to smaller kernel matrices and better performance. This method can also be viewed as applying the methodology of [12,33] to discretize successive linearizations of the PDE.

Let $u^\dagger$ denote the minimizer of (2.2) as before. Assuming that the operators $\mathcal{P}$ and $\mathcal{B}$ are Fréchet differentiable we then approximate $u^\dagger$ with a sequence of elements u^ℓ obtained by solving the problem

$$\begin{cases} \underset{u \in \mathcal{U}}{\text{minimize}} & \|u\|_{\mathcal{U}} \\ \text{s.t.} & \left(\mathcal{P}(u^{\ell-1}) + \mathcal{P}'(u^{\ell-1})(u - u^{\ell-1})\right)|_{\mathbf{s}_m} = f(\mathbf{s}_m), \quad m = 1, \dots, M_\Omega, \\ & \left(\mathcal{B}(u^{\ell-1}) + \mathcal{B}'(u^{\ell-1})(u - u^{\ell-1})\right)|_{\mathbf{s}_m} = g(\mathbf{s}_m), \quad m = M_\Omega + 1, \dots, M, \end{cases} \tag{2.13}$$

where, $\mathcal{P}'$ and $\mathcal{B}'$ are the Fréchet derivatives of $\mathcal{P}$ and $\mathcal{B}$.

Let us further suppose that Assumption 2.1 holds. Observing that the constraints in (2.13) are linear in u we obtain an explicit formula for u^ℓ by [12, Prop. 2.2]:

$$u^\ell(\mathbf{s}) = K(\mathbf{s}, \tilde{\boldsymbol{\phi}}^{\ell-1}) K(\tilde{\boldsymbol{\phi}}^{\ell-1}, \tilde{\boldsymbol{\phi}}^{\ell-1})^{-1} \mathbf{z}^{\ell-1} \tag{2.14}$$

where $\mathbf{z}^{\ell-1} = (z_1^{\ell-1}, \dots, z_M^{\ell-1})^T$ has entries

$$z_m^{\ell-1} = \begin{cases} \left(f - \mathcal{P}(u^{\ell-1}) + \mathcal{P}'(u^{\ell-1})u^{\ell-1}\right)|_{\mathbf{s}_m}, & \text{if } 1 \leq m \leq M_\Omega, \\ \left(f - \mathcal{P}(u^{\ell-1}) + \mathcal{B}'(u^{\ell-1})u^{\ell-1}\right)|_{\mathbf{s}_m}, & \text{if } M_\Omega + 1 \leq m \leq M. \end{cases} \tag{2.15}$$

The vectors $\tilde{\boldsymbol{\phi}}^{\ell-1} \in (\mathcal{U}^*)^{\otimes M}$ are obtained by concatenating the dual elements

$$\tilde{\phi}_m^{\ell-1} := \begin{cases} \delta_{(\mathbf{s}_m)} \circ \mathcal{P}'(u^{\ell-1}), & \text{if } 1 \leq m \leq M_\Omega, \\ \delta_{(\mathbf{s}_m)} \circ \mathcal{B}'(u^{\ell-1}), & \text{if } M_\Omega + 1 \leq m \leq M. \end{cases} \tag{2.16}$$

We note that the above scheme implicitly assumes that $\mathcal{U}$ is sufficiently regular so that the derivatives $\mathcal{P}'(u^{\ell-1})$ and $\mathcal{B}'(u^{\ell-1})$ can be regarded as linear operators mapping $\mathcal{U}$ to $C(\overline{\Upsilon})$ where pointwise evaluation is well-defined. The most important feature of the linearize-then-optimize approach is that the kernel matrices $K(\tilde{\boldsymbol{\phi}}^{\ell-1}, \tilde{\boldsymbol{\phi}}^{\ell-1})$ are of size $M \times M$ while the kernel matrix $K(\boldsymbol{\phi}, \boldsymbol{\phi})$, used in the Gauss-Newton approach of Subsection 2.2.1, is of size $N \times N$. Note that $N/M \approx Q_\Omega$ and for instance, in Example 2.2, $Q_\Omega = 3$ so there is an approximately $3\times$ reduction in the kernel matrix size. Thus, the linearize-then-optimize approach requires the inversion of much smaller kernel matrices at each iteration but these matrices need to be updated successively since the $\tilde{\boldsymbol{\phi}}^\ell$ depend on the previous solution $u^{\ell-1}$. In the case where Assumption 2.1 holds and P, B are differentiable, then the $\mathcal{P}'(u)$ and $\mathcal{B}'(u)$ operators can be written explicitly as

$$\mathcal{P}'(u^{\ell-1}) : u \mapsto \nabla P\left(L_1(u^{\ell-1}), \dots, L_{Q_\Omega}(u^{\ell-1})\right)^T \begin{bmatrix} L_1(u) \\ \vdots \\ L_{Q_\Omega}(u) \end{bmatrix},$$

$$\mathcal{B}'(u^{\ell-1}) : u \mapsto \nabla B\left(L_{Q_\Omega+1}(u^{\ell-1}), \dots, L_Q(u^{\ell-1})\right)^T \begin{bmatrix} L_{Q_\Omega+1}(u) \\ \vdots \\ L_Q(u) \end{bmatrix}.$$

Algorithm 2.2 The linearize-then-optimize approach to parametric PDEs.

Input: PDE of the form (2.1) defined on $\Upsilon = \Omega \times \Theta$ with boundary condition on $\partial\Upsilon = \partial\Omega \times \Theta$, $M \geq 1$ collocation points in $\overline{\Upsilon}$, and kernel $K : \overline{\Upsilon} \times \overline{\Upsilon} \to \mathbb{R}$

Output: Approximation $u^\dagger(s)$ to exact solution $u^*(s)$

1: **while** not converged **do** ▷ Iteratively approximate $u^\dagger$

2: $z_m^{\ell-1} := \begin{cases} \left(f - \mathcal{P}(u^{\ell-1}) + \mathcal{P}'(u^{\ell-1})u^{\ell-1}\right)|_{s_m}, & \text{if } m = 1, \dots, M_\Omega, \\ \left(f - \mathcal{P}(u^{\ell-1}) + \mathcal{B}'(u^{\ell-1})u^{\ell-1}\right)|_{s_m}, & \text{if } m = M_\Omega + 1, \dots, M. \end{cases}$ ▷ Build $\mathbf{z}^{\ell-1}$

3: $\tilde{\boldsymbol{\phi}}_m^{\ell-1} := \begin{cases} \delta_{(s_m)} \circ \mathcal{P}'(u^{\ell-1}), & \text{if } m = 1, \dots, M_\Omega, \\ \delta_{(s_m)} \circ \mathcal{B}'(u^{\ell-1}), & \text{if } m = M_\Omega + 1, \dots, M. \end{cases}$ ▷ Build $\tilde{\boldsymbol{\phi}}^{\ell-1}$

4: $u^\ell(\mathbf{s}) = K(\mathbf{s}, \tilde{\boldsymbol{\phi}}^{\ell-1}) K(\tilde{\boldsymbol{\phi}}^{\ell-1}, \tilde{\boldsymbol{\phi}}^{\ell-1})^{-1} \mathbf{z}^{\ell-1}$

5: **end while**

We note that while the "linearize then optimize" approach can reduce the size of kernel matrices by a constant factor, which is significant in practice, the size still scales with the number of collocation points. For very large-scale problems requiring many collocation points, we can further employ fast algorithms for these kernel matrices; see for example [13].

Remark 2.4. The "linearize then optimize" approach performs linearization first at the continuous level, while the Gauss-Newton iteration linearizes at the discrete level after applying the representer theorem and transforming the optimization problem into an unconstrained form, either through elimination or relaxation. The "linearize then optimize" approach and the Gauss-Newton iteration are mathematically equivalent if the latter is implemented using elimination with a specific choice of $\overline{F}$. This equivalence is demonstrated in [13, Sec. 5.1] for a nonlinear elliptic example, where the algorithm is also shown to be equivalent to a sequential quadratic programming approach for solving (2.11). In general, these approaches may differ in how the nonlinear operators $\mathcal{P}, \mathcal{B}$, or the nonlinear map F are represented. ◊

Finally, we summarize our linearize-then-optimize approach as a pseudo-algorithm within Algorithm 2.2.

3. Error analyses

We now present our main theoretical results concerning convergence rates for the minimizers $u^\dagger$ of (2.2) to the respective true solutions $u^\star$. We start in Subsection 3.1 by articulating the abstract framework, main theorem and proof. We then consider the simple setting of a nonlinear PDE in Subsection 3.2 where the RKHS $\mathcal{U}$ already satisfies the boundary conditions of the PDE to convey the main ideas of the proof in a simple setting. Non-trivial boundary conditions are then considered in Subsection 3.3 followed by the case of parametric PDEs in Subsection 3.4. Our proof technique is a generalization of the results of [33,76] to the case of nonlinear and parametric PDEs that are Lipschitz stable and well-posed.

3.1. An abstract framework for obtaining convergence rates

We present here an abstract theoretical result that allows us to obtain convergence rates for nonlinear operator equations. Our error analyses concerning the numerical solutions $u^\dagger$ and the true solution to the PDE $u^\star$ then follow as applications of this abstract result. Our main result here can also be viewed as a generalization of the results of [76, Sec. 10], which focused on linear operators, to the nonlinear case.

Let us consider operator equations of the form

$$\mathcal{T}(v^\star) = w^\star, \tag{3.1}$$

where $v^\star, w^\star$ are elements of appropriate Banach spaces and $\mathcal{T}$ is a nonlinear map. In the setting of PDEs the map $\mathcal{T}$ is defined by the differential operator of the PDE, $v^\star$ coincides with the solution and $w^\star$ is the source/boundary data. Broadly speaking our goal is to approximate the solution $v^\star$ under assumptions on its regularity and the stability properties of the map $\mathcal{T}$. To this end, we present a general result that allows us to control the error of approximating $v^\star$ given an appropriate candidate $v^\dagger$. Henceforth we write $B_r(V)$ to denote the ball of radius r centered at zero in a Banach space V.

Theorem 3.1. *Consider abstract Banach spaces $(V_i, \|\cdot\|_i)_{i=1}^4$ as well as $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$. Suppose the following conditions are satisfied for any choice of $r > 0$ (all the appeared constants $C(r)$ are non-decreasing regarding r):*

(A1) For any pair $v, v' \in B_r(V_1)$ there exists a constant $C = C(r) > 0$ so that

$$\|v - v'\|_1 \leq C\|\mathcal{T}(v) - \mathcal{T}(v')\|_2. \tag{3.2}$$

(A2) For any pair $v, v' \in B_r(V_4)$ there exists a constant $C = C(r) > 0$ so that

$$\|\mathcal{T}(v) - \mathcal{T}(v')\|_3 \leq C\|v - v'\|_4. \tag{3.3}$$

(A3) For any $v \in V_4$, there exists a constant $C > 0$ so that

$$\|v\|_4 \leq C\|v\|_{\mathcal{U}}.$$

(A4) There exists a set $\widetilde{V} \subset V_2 \cap V_3$ and a constant $\varepsilon > 0$, so that for all $w, w' \in \widetilde{V}$ it holds that

$$\|w - w'\|_2 \leq \varepsilon \|w - w'\|_3. \tag{3.4}$$

Suppose problem (3.1) is uniquely solvable with $v^\star \in \mathcal{U}$ and let $v^\dagger \in \mathcal{U}$ be any other function such that:

(A5) $\mathcal{T}(v^\star), \mathcal{T}(v^\dagger) \in \widetilde{V}$.
(A6) There exists a constant $C > 0$, independent of $v^\star$ and $v^\dagger$, so that

$$\|v^\dagger\|_{\mathcal{U}} \leq C\|v^\star\|_{\mathcal{U}}. \tag{3.5}$$

Then there exists a constant $C > 0$, depending only on $\|v^\star\|_{\mathcal{U}}$, such that

$$\|v^\dagger - v^\star\|_1 \leq C\varepsilon\|v^\star\|_{\mathcal{U}}.$$

Proof. By (A1) we have that

$$\|v^\dagger - v^\star\|_1 \leq C\|\mathcal{T}(v^\dagger) - \mathcal{T}(v^\star)\|_2. \tag{3.6}$$

Then (A4) and (A5) imply that $\|\mathcal{T}(v^\dagger) - \mathcal{T}(v^\star)\|_2 \leq C\varepsilon\|\mathcal{T}(v^\dagger) - \mathcal{T}(v^\star)\|_3$. By the triangle inequality we have $\|\mathcal{T}(v^\dagger) - \mathcal{T}(v^\star)\|_3 \leq \|\mathcal{T}(v^\dagger) - \mathcal{T}(0)\|_3 + \|\mathcal{T}(v^\star) - \mathcal{T}(0)\|_3$. Using (A2), (A3), and (A6) in that order, we get $\|\mathcal{T}(v^\dagger) - \mathcal{T}(0)\|_3 \leq C\|v^\dagger\|_4 \leq C\|v^\dagger\|_{\mathcal{U}} \leq C\|v^\star\|_{\mathcal{U}}$. Similarly, we have $\|\mathcal{T}(v^\star) - \mathcal{T}(0)\|_3 \leq C\|v^\star\|_{\mathcal{U}}$. Combining these bounds we obtain $\|\mathcal{T}(v^\dagger) - \mathcal{T}(v^\star)\|_3 \leq C\varepsilon\|v^\star\|_{\mathcal{U}}$ which yields the desired result due to (3.6). □

Let us provide some remarks regarding the assumptions of the theorem. In our PDE examples we often take the V_i spaces to be Sobolev spaces of appropriate smoothness while $\mathcal{U}$ is taken as an RKHS that is sufficiently smooth and so $v^\star \in \mathcal{U}$ amounts to an assumption on the regularity of the true solution to the problem. Conditions (A1) and (A2) amount to forward and inverse Lipschitz stability of the operator $\mathcal{T}$ while (A4) is often given by a sampling/Poincaré-type inequality for our numerical method. We treat the constant ε separately from the other constants in the theorem since in practice ε often coincides with some power of the resolution (fill-distance/meshnorm) of our numerical scheme, constituting the rate of convergence of the method. Assumption (A3) also concerns the regularity of the RKHS and the choice of the space V_4 (we simply ask for $\mathcal{U}$ to be continuously embedded in V_4) and is a matter of the setup of the problem. Condition (A6) is less natural as it requires the norm of the approximate solution $v^\dagger$ to be controlled by the norm of $v^\star$. While this condition does not hold for many numerical approximation schemes, we will see that it follows easily from the setup of our collocation/optimal recovery scheme.

In plain words, the most important message of Theorem 3.1 is that: *given Condition (3.5) and the Lipschitz-continuity of $\mathcal{T}$ and its inverse, it follows that the approximation error between $v^\dagger$ and $v^\star$ is bounded by the approximation error between $\mathcal{T}(v^\dagger)$ and $\mathcal{T}(v^\star)$*. This result can be applied to both GP/kernel and ANN based collocation methods, since both seek to minimize the error between $\mathcal{T}(v^\dagger)$ and $\mathcal{T}(v^\star)$ at collocation points. This Condition (3.5) is automatically satisfied for our GP/Kernel based methods that solve problems of the form

$$v^\dagger = \arg\min \|v\|_{\mathcal{U}} \qquad \text{s.t.} \qquad [\phi_i, \mathcal{T}(v)] = [\phi_i, \mathcal{T}(v^\star)], \quad i = 1, \dots, M,$$

with $\mathcal{T}$ denoting the differential operator of a PDE and ϕ_i denoting a set of dual elements (e.g. pointwise evaluations at collocation points). Then since the true solution $v^\star$ satisfies the PDE for an infinite collection of dual elements (e.g. pointwise within a set, or in a weak sense) then we immediately have that $\|v^\dagger\|_{\mathcal{U}} \leq \|v^\star\|_{\mathcal{U}}$. One can also take $\mathcal{U}$ to be a Barron space (indeed the V_i norms could be arbitrary) to obtain an analogous result for ANNs, but it is unclear if this setup coincides with (or leads to) any practical algorithms.

3.2. The case of second order nonlinear PDEs

We begin our error analysis in the case where (2.1) does not depend on the parameter θ and homogeneous Dirichlet boundary conditions are imposed, i.e., nonlinear second order PDEs of the form

$$\begin{cases} \mathcal{P}(u^\star)(\mathbf{x}) = f(\mathbf{x}), & \mathbf{x} \in \Omega, \\ u^\star(\mathbf{x}) = 0, & \mathbf{x} \in \partial\Omega. \end{cases} \tag{3.7}$$

The choice of Dirichlet boundary conditions is only made for simplicity here and can be replaced with other conditions of interest. We will also consider approximate boundary conditions in Subsection 3.3. We further assume that the kernel K is chosen so that the elements of $\mathcal{U}$ readily satisfy the boundary conditions of the PDE and consider optimization problems of the form

$$\begin{cases} \underset{u\in\mathcal{U}}{\text{minimize}} & \|u\|_{\mathcal{U}} \\ \text{s.t.} & \mathcal{P}(u)(\mathbf{x}_m) = f(\mathbf{x}_m), \quad m = 1,\dots,M_\Omega, \\ & u(\mathbf{x}) = 0, \qquad\qquad \mathbf{x}\in\partial\Omega, \end{cases} \tag{3.8}$$

where $X_\Omega := \{\mathbf{x}_m\}_{m=1}^{M_\Omega} \subset \Omega$ are a set of collocation points. We need to impose appropriate assumptions on the RKHS $\mathcal{U}$, the domain Ω and the PDE operator $\mathcal{P}$.

Assumption 3.2. The following conditions hold:

(B1) *(Regularity of the domain)* $\Omega \subset \mathbb{R}^d$ is a compact set with a Lipschitz boundary.

(B2) *(Stability of $\mathcal{P}$)* There exist indices $\gamma > 0$ and $k \in \mathbb{N}$ satisfying $d/2 < k + \gamma$ and $s \geq 1, \ell \in \mathbb{R}$, so that for any $r > 0$ it holds that

$$\|u_1 - u_2\|_{H^\ell(\Omega)} \leq C\|\mathcal{P}(u_1) - \mathcal{P}(u_2)\|_{H^k(\Omega)}, \qquad \forall u_1, u_2 \in B_r(H^\ell(\Omega)\cap H_0^1(\Omega)) \tag{3.9}$$

$$\|\mathcal{P}(u_1) - \mathcal{P}(u_2)\|_{H^{k+\gamma}(\Omega)} \leq C\|u_1 - u_2\|_{H^s(\Omega)}, \qquad \forall u_1, u_2 \in B_r(H^s(\Omega)\cap H_0^1(\Omega)) \tag{3.10}$$

where $C = C(r) > 0$ is independent of the u_i's. The space $H^s(\Omega)\cap H_0^1(\Omega)$ can be equipped with the norm $\|\cdot\|_{H^s(\Omega)}$, which is used to define the balls above.

(B3) $\mathcal{U}$ is continuously embedded in $H^s(\Omega)\cap H_0^1(\Omega)$. ♢

Item (B1) is standard while (B3) dictates the choice of the RKHS $\mathcal{U}$, and in turn the kernel, which should be made based on a priori knowledge about regularity of the strong solution $u^\star$. We highlight that, asking elements of $\mathcal{U}$ to satisfy the boundary conditions is only practical for simple domains and boundary conditions such as periodic, Dirichlet, or Neumann conditions on hypercubes or spheres. Assumption (B2) on the other hand is a question in the analysis of nonlinear PDEs and is independent of our numerical scheme; simply put we require the PDE to be Lipschitz well-posed with respect to the right hand side/source term.

We are now ready to present our first theoretical result characterizing the convergence of the minimizer $u^\dagger$ of (3.8) to $u^\star$ the strong solution of (3.7).

Theorem 3.3. *Suppose Assumption 3.2 is satisfied and let $u^\star \in \mathcal{U}$ denote the unique strong solution of (3.7). Let $u^\dagger$ be a minimizer of (3.8) with a set of collocation points $X_\Omega \subset \Omega$ and define their fill-distance*

$$h_\Omega := \sup_{\mathbf{x}'\in\Omega} \inf_{\mathbf{x}\in X_\Omega} |\mathbf{x}' - \mathbf{x}|.$$

Then there exists a constant $h_0 > 0$ so that if $h_\Omega < h_0$ then

$$\|u^\dagger - u^\star\|_{H^\ell(\Omega)} \leq C h_\Omega^\gamma \|u^\star\|_{\mathcal{U}},$$

where the constant $C > 0$ is independent of $u^\dagger$, and h_Ω.

Proof. We will obtain the result by applying Theorem 3.1 with the map $\mathcal{T} \equiv \mathcal{P}$ and the spaces $V_1 \equiv H^\ell(\Omega)$, $V_2 \equiv H^k(\Omega)$, $V_3 \equiv H^{k+\gamma}(\Omega)$, and $V_4 \equiv H^s(\Omega)$. With this setup we proceed to verify the conditions of Theorem 3.1: Condition (A1) follows from (3.9), (A2) follows from (3.10), (A3) follows from (B3).

Condition (A6) holds since $u^\dagger$ is a minimizer of (3.8) and so $\|u^\dagger\|_{\mathcal{U}} \leq \|u^\star\|_{\mathcal{U}}$, since $u^\star$ is feasible but satisfies additional constraints compared with $u^\dagger$, i.e., it solves the PDE over the entire set Ω. Thus, (3.5) is verified with constant $C = 1$.

It remains to verify (A4): Let $\bar{f} = \mathcal{P}(u^\dagger) - \mathcal{P}(u^\star)$ and observe that $\bar{f}(\mathbf{x}) = 0$ for all $\mathbf{x} \in X_\Omega$. Thus $\bar{f} \in H^{k+\gamma}(\Omega)$ is zero on X_Ω and an application of Proposition A.1 yields the existence of a constant $h_0 > 0$ so that whenever $h_\Omega < h_0$ then $\|\bar{f}\|_{H^k(\Omega)} \leq C h_\Omega^\gamma \|\bar{f}\|_{H^{k+\gamma}(\Omega)}$. This verifies (3.4) with $\varepsilon \equiv C h_\Omega^\gamma$. □

Remark 3.4. We note that Item (B2) and in turn Theorem 3.3 can easily be modified to a local version where the stability estimates (3.10) and (3.10) are stated for u_1, u_2 belonging to a ball of radius $r > 0$ around the true solution $u^\star$. Then one can obtain an asymptotic rate for $\|u^\dagger - u^\star\|_{H^\ell(\Omega)}$ under the additional assumption that $u^\dagger$ is sufficiently close to $u^\star$. ♢

Remark 3.5. The assumptions and results of Assumption 3.2 are analogous to the one used to obtain error estimates in numerical homogenization for elliptic PDEs [66]. In particular Theorem 3.3 can be extended to the setting where measurements on the PDE are not pointwise but involve integral operators and where the coefficients may be rough. ♢

We now present a brief example where Assumption 3.2 can be verified and so Theorem 3.3 is applicable to obtain convergence rates for our GP/kernel collocation solver.

Example 3.6 *(Nonlinear Darcy flow continued)*. Let us consider the nonlinear Darcy flow PDE (2.4) and assume that Ω has a smooth boundary and $a(\mathbf{x}) \in C^\infty(\Omega)$ is fixed and satisfies $a(\mathbf{x}) \geq 1$. Further suppose $\tau(z) = 1 + \tanh(\beta z)$ for a fixed constant $\beta > 0$ to be determined. Now pick $k = \lceil d/2 + \alpha \rceil$ from which it follows that $H^k(\Omega)$ is continuously embedded in $C^\alpha(\bar{\Omega})$ [34, Thm. 7.26] and fix an integer $\gamma > 0$.[2] It is then straightforward to verify (3.10) with $s = k + \gamma + 2$ as well as Assumption 3.2(iii) by choosing the kernel K to be the Green's function of the operator $(-\Delta)^s$ on the domain Ω, subject to homogeneous Dirichlet boundary conditions. We note that approximating this Green's function can be expensive in practice and in Subsection 3.3 we propose a way around this step by using collocation points on the boundary of Ω to impose the boundary conditions.

Furthermore our assumptions on a and τ imply that $\mathcal{P}$ is uniformly elliptic in $\bar{\Omega}$ (see [34, Part II] for definition of ellipticity for nonlinear elliptic PDEs). Since τ is smooth it follows from [34, Thm. 13.8] that, for any $\alpha \in (0,1)$ and $f \in C^\alpha(\bar{\Omega})$, the PDE

$$\begin{cases} -\mathrm{div}\big(\exp(a)\nabla u\big) + \tau(u) = f, & \mathbf{x} \in \Omega, \\ u = 0, & \mathbf{x} \in \partial\Omega, \end{cases} \tag{3.11}$$

has a solution $u \in C^2(\bar{\Omega})$. Now pick $f_1, f_2 \in H^k(\Omega)$ which, by the aforementioned Sobolev embedding result, belong to $C^\alpha(\bar{\Omega})$. Write $u_1, u_2 \in C^2(\bar{\Omega})$ for the solution of the PDE with both right hand sides and observe that the difference $w := u_1 - u_2$ solves the PDE

$$-\mathrm{div}\big(\exp(a)\nabla w\big) = f_1 - f_2 + \tau(u_2) - \tau(u_1).$$

Standard stability results for linear elliptic PDEs then imply the bound

$$\|w\|_{H^2(\Omega)} \leq B\left(\|f_1 - f_2\|_{L^2(\Omega)} + \|\tau(u_2) - \tau(u_1)\|_{L^2(\Omega)}\right),$$

for a constant $B > 0$ independent of w, f_1, f_2, u_1, u_2. Since τ is globally β-Lipschitz we infer that $\|\tau(u_1) - \tau(u_2)\|_{L^2(\Omega)} \leq \beta\|w\|_{L^2(\Omega)}$ which, together with the subsequent bound, yields

$$\|w\|_{H^2(\Omega)} \leq \frac{B}{1 - B\beta}\|f_1 - f_2\|_{L^2(\Omega)}.$$

Thus, assumption (3.9) is satisfied with $\ell = 2$ as long as $\beta B < 1$. ◊

3.3. Handling boundary conditions

We now turn our attention to the case where (2.1) is still independent of the $\boldsymbol{\theta}$ parameter but involves non-trivial boundary conditions, i.e.,

$$\begin{cases} \mathcal{P}(u^\star)(\mathbf{x}) = f(\mathbf{x}), & \mathbf{x} \in \Omega, \\ \mathcal{B}(u^\star)(\mathbf{x}) = g(\mathbf{x}), & \mathbf{x} \in \partial\Omega. \end{cases} \tag{3.12}$$

We will further assume that the elements of $\mathcal{U}$ do not satisfy the boundary conditions exactly and so boundary collocation points are utilized to approximate those conditions leading to the problem

$$\begin{cases} \underset{u \in \mathcal{U}}{\text{minimize}} & \|u\|_{\mathcal{U}} \\ \text{s.t.} & \mathcal{P}(u)(\mathbf{x}_m) = f(\mathbf{x}_m), \quad m = 1, \dots, M_\Omega, \\ & \mathcal{B}(u)(\mathbf{x}_m) = g(\mathbf{x}_m), \quad m = M_\Omega + 1, \dots, M, \end{cases} \tag{3.13}$$

where $X_\Omega := \{\mathbf{x}_m\}_{m=1}^{M_\Omega} \subset \Omega$ are the interior collocation points as before and $X_{\partial\Omega} := \{\mathbf{x}_m\}_{m=M_\Omega+1}^{M} \subset \partial\Omega$ are the boundary collocation points. We will state our assumptions and results for PDEs in $d > 1$ dimensions since in the 1D case we can, in principle, impose the boundary conditions exactly by placing some collocation points on boundary. The main difference, in comparison to Theorem 3.3, is that here we need to impose new assumptions on the PDE operators $\mathcal{P}$ and $\mathcal{B}$ and the boundary of Ω to be able to use Proposition A.1 (sampling inequality on manifolds) in the final step of the proof to obtain approximation rates for the boundary data.

Assumption 3.7. The following conditions hold:

(C1) *(Regularity of the domain and its boundary)* $\Omega \subset \mathbb{R}^d$ with $d > 1$ is a compact set and $\partial\Omega$ is a smooth connected Riemannian manifold of dimension $d - 1$ endowed with a geodesic distance $\rho_{\partial\Omega}$.

(C2) *(Stability of the PDE)*
There exist $\gamma > 0$ and $k, t \in \mathbb{N}$ satisfying $d/2 < k + \gamma$ and $(d-1)/2 < t + \gamma$, and $s, \ell \in \mathbb{R}$, so that for any $r > 0$ it holds that

$$\begin{aligned} \|u_1 - u_2\|_{H^\ell(\Omega)} \leq & C\big(\|\mathcal{P}(u_1) - \mathcal{P}(u_2)\|_{H^k(\Omega)} \\ & + \|\mathcal{B}(u_1) - \mathcal{B}(u_2)\|_{H^t(\partial\Omega)}\big) \quad \forall u_1, u_2 \in B_r(H^\ell(\Omega)), \end{aligned} \tag{3.14}$$

[2] We only assume the exponents are integers for simplicity but our arguments can be generalized to the case of non-integer indices.

$$\|\mathcal{P}(u_1)-\mathcal{P}(u_2)\|_{H^{k+\gamma}(\Omega)}+\|\mathcal{B}(u_1)-\mathcal{B}(u_2)\|_{H^{t+\gamma}(\partial\Omega)} \leq C\|u_1-u_2\|_{H^s(\Omega)}, \qquad \forall u_1,u_2\in B_r(H^s(\Omega)), \tag{3.15}$$

where $C=C(r)>0$ is a constant independent of the u_i.

(C3) $\mathcal{U}$ is continuously embedded in $H^s(\Omega)$. ◊

Observe that the above assumptions are analogous to Assumption 3.2 with the exception that we no longer work with the restricted Sobolev spaces H_0^k since we do not need to impose the boundary conditions. However, we need to state our stability results for both $\mathcal{P}$ and $\mathcal{B}$. We emphasize that the verification of condition (C2) remains a question in the analysis of PDEs. We are now ready to extend Theorem 3.3 to the case of non-trivial boundary conditions.

Theorem 3.8. *Suppose Assumption 3.7 is satisfied and let $u^\star\in\mathcal{U}$ denote the unique strong solution of (3.7). Let $u^\dagger$ be a minimizer of (3.8) with a set of collocation points $X\subset\overline{\Omega}$ where $X_\Omega\subset X$ denotes the collocation points in the interior of Ω and $X_{\partial\Omega}$ denotes the collocation points on the boundary $\partial\Omega$. Define the fill-distances*

$$h_\Omega := \sup_{\mathbf{x}'\in\Omega}\inf_{\mathbf{x}\in X_\Omega}|\mathbf{x}'-\mathbf{x}|, \qquad h_{\partial\Omega} := \sup_{\mathbf{x}'\in\partial\Omega}\inf_{\mathbf{x}\in X_{\partial\Omega}}\rho_{\partial\Omega}(\mathbf{x}',\mathbf{x}),$$

where $\rho_{\partial\Omega}:\partial\Omega\times\partial\Omega\to\mathbb{R}_+$ is the geodesic distance defined on $\partial\Omega$ (see Appendix A), and set $\bar{h}:=\max\{h_\Omega,h_{\partial\Omega}\}$. Then there exists a constant $h_0>0$ so that if $\bar{h}<h_0$ then

$$\|u^\dagger-u^\star\|_{H^s(\Omega)}\leq C\bar{h}^\gamma\|u^\star\|_{\mathcal{U}},$$

where $C>0$ is independent of $u^\dagger$ and $\bar{h}$.

Proof. The proof follows an identical approach to Theorem 3.3 and applies Theorem 3.1 with the appropriate setup. We take the operator $\mathcal{T}:u\mapsto\big(\mathcal{P}(u),\mathcal{B}(u)\big)$. We then choose the spaces $V_1\equiv H^\ell(\Omega)$, $V_2\equiv H^k(\Omega)\times H^t(\partial\Omega)$, $V_3\equiv H^{k+\gamma}(\Omega)\times H^{t+\gamma}(\Omega)$, and $V_4\equiv H^s(\Omega)$ where we equip V_2 with the norm $\|(f,g)\|_2:=\|f\|_{H^k(\Omega)}+\|g\|_{H^t(\partial\Omega)}$ and similarly for V_3 with the $H^k(\Omega)$ and $H^t(\partial\Omega)$ norms replaced by $H^{k+\gamma}(\Omega)$ and $H^{t+\gamma}(\partial\Omega)$ norms.

Analogously to the proof of Theorem 3.3, we can verify Conditions (A1), (A2), and (A3) by the hypothesis of the theorem. Condition (A6) is also satisfied since $u^\dagger$ is a minimizer of (3.13) and so $\|u^\dagger\|_{\mathcal{U}}\leq\|u^\star\|_{\mathcal{U}}$ as $u^\dagger$ satisfies more relaxed constraints.

It remains for us to verify (A4). Repeating the same argument as in the proof of Theorem 3.3, in the interior of Ω, yields the bound

$$\|\mathcal{P}(u^\dagger)-\mathcal{P}(u^\star)\|_{H^k(\Omega)}\leq Ch_\Omega^\gamma\|\mathcal{P}(u^\dagger)-\mathcal{P}(u^\star)\|_{H^{k+\gamma}(\Omega)}, \tag{3.16}$$

whenever $h_\Omega<h_1$ and h_1 is a sufficiently small constant that is independent of $u^\dagger$ and $u^\star$.

Let $\bar{g}=\mathcal{B}(u^\dagger)-\mathcal{B}(u^\star)$ which satisfies $\bar{g}(\mathbf{x})=0$ for all $\mathbf{x}\in X_{\partial\Omega}$ and so $\bar{g}\in H^{t+\gamma}(\partial\Omega)$ is zero on the set $X_{\partial\Omega}$. Then Proposition A.1 implies the existence of a constant $h_2>0$ so that whenever $h_{\partial\Omega}<h_2$ we have

$$\|\bar{g}\|_{H^t(\partial\Omega)}\leq Ch_{\partial\Omega}^\gamma\|\bar{g}\|_{H^{t+\gamma}(\partial\Omega)}.$$

Now take $h_0=\min\{h_1,h_2\}$ and combine the above bound with (3.16), and substitute the definition of $\bar{g}$ to get

$$\|\mathcal{P}(u^\dagger)-\mathcal{P}(u^\star)\|_{H^k(\Omega)}+\|\mathcal{B}(u^\dagger)-\mathcal{B}(u^\star)\|_{H^t(\partial\Omega)} \leq C\bar{h}^\gamma\big(\|\mathcal{P}(u^\dagger)-\mathcal{P}(u^\star)\|_{H^{k+\gamma}(\Omega)}+\|\mathcal{B}(u^\dagger)-\mathcal{B}(u^\star)\|_{H^{t+\gamma}(\partial\Omega)}\big),$$

whenever $\bar{h}<h_0$. This verifies (A4) with $\varepsilon\equiv C\bar{h}^\gamma$. □

Remark 3.9. We highlight that our statement of Theorem 3.8 can easily be extended to PDEs with mixed boundary conditions simply by modifying the norm that is chosen on the boundary, i.e., the spaces V_2 and V_4, so long as we can prove the requisite stability estimates in condition (C2). In particular, this idea will allow us to obtain errors for time-dependent PDEs, cast as a static PDE in an space-time domain Ω with the initial and boundary conditions imposed as mixed conditions on $\partial\Omega$. In fact, in the case of time-dependent PDEs we do not need to impose the boundary conditions on all of $\partial\Omega$ but only on a subset. ◊

We now return to our running example to verify Assumption 3.7 for the Darcy flow PDE.

Example 3.10 *(Nonlinear Darcy flow continued).* Consider the PDE (3.11) but this time with the boundary condition $u=g$ on $\partial\Omega$ for a function $g\in H^{t+\gamma}(\partial\Omega)$ with $t>\min\{3/2,(d-1)/2\}$ and $\gamma>0$. Now fix a function $\varphi\in H^{t+\gamma+1/2}(\Omega)$ so that its trace coincides with g and define $v=u-\varphi$ and observe that u solves the above PDE if v solves

$$\begin{cases} -\mathrm{div}\big(\exp(a)\nabla v\big)+\tau'(v)=f', & \mathbf{x}\in\Omega,\\ v=0, & \mathbf{x}\in\partial\Omega, \end{cases}$$

where we defined $\tau'(v) := \tau(v+\varphi)$ and $f' := f + \operatorname{div}\big(\exp(a)\nabla\varphi\big)$. Now observe that the functions τ' and f' still satisfy the same conditions as τ, f in Example 3.6 and so we obtain existence and uniqueness of the solutions v and in turn u.

Now consider two solutions u_1, u_2 arising from source terms f_1, f_2 and boundary data g_1, g_2. Then the error $w = u_1 - u_2$ solves the PDE

$$\begin{cases} -\operatorname{div}\big(\exp(a)\nabla w\big) = f + \tau(u_2) - \tau(u_1), & \mathbf{x} \in \Omega, \\ w = g_1 - g_2, & \mathbf{x} \in \partial\Omega. \end{cases}$$

By standard stability results for linear elliptic PDEs [55, Thm. 4.18] we have

$$\|w\|_{H^2(\Omega)} \le B(\|f_1 - f_2\|_{L^2(\Omega)} + \|\tau(u_1) - \tau(u_2)\|_{L^2(\Omega)} + \|g_1 - g_2\|_{H^{3/2}(\Omega)})$$

We can now repeat the same argument as in the final steps of Example 3.6 to get the bound

$$\|w\|_{H^2(\Omega)} \le \frac{B}{1 - B\beta}\left(\|f_1 - f_2\|_{H^2(\Omega)} + \|g_1 - g_2\|_{H^{3/2}(\Omega)}\right),$$

which verifies Assumption 3.7(ii) with $s = 2$ provided that $\beta B < 1$. ♢

3.4. The case of parametric PDEs

We now consider the setting of the parametric PDE (2.1). Our error estimates can be viewed as further extending Theorem 3.8 with additional assumptions due to the fact that we will need to approximate the solutions on the set $\Upsilon = \Omega \times \Theta$ as well as its relevant boundary which needs to be sufficiently regular for us to apply Proposition A.1. Beyond this technical point, the statement and proof of the result for parametric PDEs is identically to PDEs with boundary conditions and so we state our results succinctly, starting with the requisite assumptions on the parametric PDE.

Assumption 3.11. The following conditions hold:

(D1) $\Omega \subset \mathbb{R}^d$ and $\Theta \subset \mathbb{R}^p$ are compact sets such that $\partial\Omega$ and $\partial\Theta$ are smooth Riemannian manifolds of dimensions $d-1$ and $p-1$ respectively.

(D2) *(Stability of the parametric PDE)* There exist $\gamma > 0$ and $k, t \in \mathbb{N}$ satisfying $(d+p)/2 < k + \gamma$ and $(d+p-1)/2 < t + \gamma$, and Banach spaces V_1 and V_4 so that for any $r > 0$ it holds that

$$\begin{aligned} &\|u_1 - u_2\|_1 \\ &\quad \le C\left(\|\mathcal{P}(u_1) - \mathcal{P}(u_2)\|_{H^k(\Upsilon)} + \|\mathcal{B}(u_1) - \mathcal{B}(u_2)\|_{H^t(\partial\Upsilon)}\right) \quad \forall u_1, u_2 \in B_r(V_1), \end{aligned} \tag{3.17}$$

$$\begin{aligned} &\|\mathcal{P}(u_1) - \mathcal{P}(u_2)\|_{H^{k+\gamma}(\Upsilon)} + \|\mathcal{B}(u_1) - \mathcal{B}(u_2)\|_{H^{t+\gamma}(\partial\Upsilon)} \\ &\qquad \le C\|u_1 - u_2\|_4, \quad \forall u_1, u_2 \in B_r(V_4), \end{aligned} \tag{3.18}$$

where $C = C(r) > 0$ is a constant independent of the u_i.

(D3) $\mathcal{U}$ is continuously embedded in V_4. ♢

Unlike Assumptions 3.2 and 3.7 here we left the function spaces V_1 and V_4 as generic Banach spaces of functions $u : \Upsilon \mapsto \mathbb{R}$ since, for parametric PDEs, we can often obtain the desired stability results in non-standard norms, such as the mixed norm in Example 3.13 below, as opposed to the Sobolev norms used for the non-parametric PDE setting. More generally, one may also impose V_2, V_3 to be generic Banach spaces rather than the standard Sobolev spaces. The Sobolev space setting suffices for applications in this paper.

With the above assumptions we can now present our main result for the parametric PDE setting. The proof is omitted since it is identical to that of Theorem 3.8 except that (1) the argument on $\partial\Omega$ is now repeated for $\partial\Upsilon = \partial\Omega \times \Theta$ which is in general a smooth manifold with boundary but this modification does not affect any of the steps in the proof, and (2) the results are stated in terms of the norm on the space V_1.

Theorem 3.12. *Suppose Assumption 3.11 is satisfied and let $u^\star \in \mathcal{U}$ denote the unique strong solution of (2.1). Let $u^\dagger$ be a minimizer of (2.2) with a set of collocation points $S \subset \Upsilon \cup \partial\Upsilon$ where $S_\Upsilon \subset S$ denotes the collocation points in the interior of Υ and $S_{\partial\Upsilon}$ denotes the collocation points on the boundary $\partial\Upsilon$. Define the fill-distances*

$$h_\Upsilon := \sup_{\mathbf{s}' \in \Upsilon} \inf_{\mathbf{s} \in S_\Upsilon} |\mathbf{s}' - \mathbf{s}|, \qquad h_{\partial\Upsilon} := \sup_{\mathbf{s}' \in \partial\Upsilon} \inf_{\mathbf{s} \in S_{\partial\Upsilon}} \rho_{\partial\Upsilon}(\mathbf{s}', \mathbf{s}),$$

where $\rho_{\partial\Upsilon} : \partial\Upsilon \times \partial\Upsilon \to \mathbb{R}_+$ is the geodesic distance defined on $\partial\Upsilon$ (see Appendix A), and set $\bar{h} := \max\{h_\Upsilon, h_{\partial\Upsilon}\}$. Then there exists a constant $h_0 > 0$ so that if $\bar{h} < h_0$ then

$$\|u^\dagger - u^\star\|_1 \le C\bar{h}^\gamma \|u^\star\|_{\mathcal{U}},$$

where $C > 0$ is independent of $u^\dagger$ and $\bar{h}$.

We end this section by returning to our example of the Darcy flow PDE but this time in the setting where the coefficient a and the source f are dependent on a finite dimensional parameter $\boldsymbol{\theta}$. We will show that Assumption 3.11 can be verified in this case, with V_1 and V_4 taken as Banach spaces with mixed Sobolev and L^2 norms, and so Theorem 3.12 is applicable.

Example 3.13 *(1D Parametric Darcy flow PDE).* Consider the parametric elliptic PDE

$$\begin{cases} -\operatorname{div}\big(A(\mathbf{x},\boldsymbol{\theta})\nabla u\big) = f(\mathbf{x},\boldsymbol{\theta}) & \mathbf{x}\in\Omega, \\ u = 0, & \mathbf{x}\in\partial\Omega, \end{cases}$$

over a compact domain Ω and $\boldsymbol{\theta}\in\Theta$ where both Ω and Θ are assumed to satisfy condition (D1); e.g., take Ω and Θ to be unit balls. For simplicity we are ignoring the boundary operator in this case and imposing homogeneous Dirichlet boundary conditions. In this example we assume a is smooth in both $\mathbf{x}$ and $\boldsymbol{\theta}$, and there exists $m, M > 0$ such that $m \le A(\mathbf{x},\boldsymbol{\theta}) \le M$. As a concrete example we may take $A(\mathbf{x},\boldsymbol{\theta}) = \sum_{j=1}^{p} \theta_j \psi_j(\mathbf{x})$ where the ψ_j are a set of smooth functions on $\overline{\Omega}$ that are uniformly bounded from below.

First, the boundness of the operator $\mathcal{P}$ is straightforward to obtain, since a is smooth and its derivatives will be bounded in the bounded domain $\Omega\times\Theta$. More precisely, for any $\gamma > 0$, since $f = -\operatorname{div}\big(A(\mathbf{x},\boldsymbol{\theta})\nabla u\big) = -\nabla_{\mathbf{x}} A\cdot\nabla_{\mathbf{x}} u - A\Delta_{\mathbf{x}} u$, there exists some constant C independent of u and f such that

$$\|f\|_{H^\gamma(\Omega\times\Theta)} \le C\|u\|_{H^{\gamma+2}(\Omega\times\Theta)}.$$

Due to the linearity of the equation, by replacing u by $u_1 - u_2$ and noting that $f = \mathcal{P}u = \mathcal{P}u_1 - \mathcal{P}u_2$, we obtain the forward stability

$$\|\mathcal{P}u_1 - \mathcal{P}u_2\|_{H^\gamma(\Omega\times\Theta)} \le C\|u_1 - u_2\|_{H^{\gamma+2}(\Omega\times\Theta)}. \tag{3.19}$$

For the backward stability estimate, via integration by parts, we have

$$\begin{aligned} \int_\Upsilon A(\mathbf{x},\boldsymbol{\theta})|\nabla_{\mathbf{x}} u(\mathbf{x},\boldsymbol{\theta})|^2 \,\mathrm{d}\mathbf{x}\mathrm{d}\boldsymbol{\theta} &= \int_\Upsilon u(\mathbf{x},\boldsymbol{\theta}) f(\mathbf{x},\boldsymbol{\theta})\,\mathrm{d}\mathbf{x}\mathrm{d}\boldsymbol{\theta} \\ &\le \int_\Theta \|u(\cdot,\boldsymbol{\theta})\|_{L^2(\Omega)}\|f(\cdot,\boldsymbol{\theta})\|_{L^2(\Omega)}\,\mathrm{d}\boldsymbol{\theta} \\ &\le C_0 \int_\Theta \|\nabla_{\mathbf{x}} u(\cdot,\boldsymbol{\theta})\|_{L^2(\Omega)}\|f(\cdot,\boldsymbol{\theta})\|_{L^2(\Omega)}\,\mathrm{d}\boldsymbol{\theta} \\ &\le C_1 \|u\|_{L^2(\Theta,H_0^1(\Omega))}\|f\|_{L^2(\Theta,L^2(\Omega))}, \end{aligned}$$

where in the first and third inequalities, we used the Cauchy-Schwarz inequality; in the second inequality, we used the Poincaré inequality as $u(\cdot,\boldsymbol{\theta})$ is zero on $\partial\Omega$. Here we used the notation:

$$\|u\|^2_{L^2(\Theta,H_0^1(\Omega))} := \int_\Theta \|u(\cdot,\boldsymbol{\theta})\|^2_{H_0^1(\Omega)}\,\mathrm{d}\boldsymbol{\theta} \quad \text{and} \quad \|f\|^2_{L^2(\Theta,L^2(\Omega))} := \int_\Theta \|f(\cdot,\boldsymbol{\theta})\|^2_{L^2(\Omega)}\mathrm{d}\boldsymbol{\theta}.$$

Note that the $L^2(\Theta,L^2(\Omega))$ norm is also equivalent to the $L^2(\Omega\times\Theta)$ norm. Now, using the bound on A, we obtain that there exists a constant C such that

$$\|u\|_{L^2(\Theta,H^1(\Omega))} \le C\|f\|_{L^2(\Omega\times\Theta)}.$$

Similar to the proof for the forward stability, the backward stability follows by the linearity of the equation. We have

$$\|u_1 - u_2\|_{L^2(\Theta,H^1(\Omega))} \le C\|\mathcal{P}u_1 - \mathcal{P}u_2\|_{L^2(\Omega\times\Theta)}. \tag{3.20}$$

Thus it follows that we can verify condition (3.17) with the norm $\|\cdot\|_1 \equiv \|\cdot\|_{L^2(\Theta,H_0^1(\Omega))}$, $\|\cdot\|_4 = \|\cdot\|_{H^{\gamma+2}(\Omega\times\Theta)}$, and $k = 0$. ♢

3.5. Bounding fill-distances

Our bounds in Theorems 3.3 and 3.12 are given in terms of the fill distances h of our collocation points. In this section, we provide an upper bound of these fill distances in terms of the number of collocation points, under the assumptions that the points are randomly drawn according to uniform distributions both in the interior of the domains and their pertinent boundaries. Throughout this section we only consider the case of non-parametric PDEs, hence we work with Ω, assumed to be a compact subset of $\mathbb{R}^d$ with boundary $\partial\Omega$ which is a compact smooth manifold of dimension $d-1$. We focus on the non-parametric setting for simplicity and our results can easily be extended to the parametric PDE setting by simply replacing Ω with Υ as a compact subset of $\mathbb{R}^{d+p}$.

Proposition 3.14. *Suppose we sample M_Ω points in Ω and $M_{\partial\Omega}$ points on $\partial\Omega$, uniformly with respect to the canonical volume and surface measures. Let $\delta > 0$. Then, with probability at least $1-\delta$, the fill-in distances h_Ω and $h_{\partial\Omega}$ satisfy*

$$h_\Omega \le C\left(\frac{\log(M_\Omega/\delta)}{M_\Omega}\right)^{1/d}, \quad h_{\partial\Omega} \le C\left(\frac{\log(M_{\partial\Omega}/\delta)}{M_{\partial\Omega}}\right)^{1/(d-1)},$$

where C is a constant independent of M_Ω, $M_{\partial\Omega}$ and δ.

The proof of Proposition 3.14 can be found in Appendix B.

Let's combine Proposition 3.14 and previous error estimates to get error bounds regarding the number of collocation points. In the case of Theorem 3.3 where there is no boundary, we get

$$\|u^\dagger - u^\star\|_{H^s(\Omega)} \le C\left(\frac{\log(M_\Omega/\delta)}{M_\Omega}\right)^{\gamma/d} \|u^\star\|_{\mathcal{V}},$$

while in the case of Theorem 3.8 where boundary is considered, we have

$$\|u^\dagger - u^\star\|_{H^s(\Omega)} \le C\left(\left(\frac{\log(M_\Omega/\delta)}{M_\Omega}\right)^{\gamma/d} + \left(\frac{\log(M_{\partial\Omega}/\delta)}{M_{\partial\Omega}}\right)^{\gamma/(d-1)}\right)\|u^\star\|_{\mathcal{V}}.$$

More generally, we note that the bounds in Proposition 3.14 can be applied to the abstract setting in Theorem 3.1 when ϵ depends on the fill-in distance.

If s and γ are appropriately chosen such that the required assumptions hold, and $\gamma \ge d/2$, then the convergence rate is at least as fast as the Monte Carlo rate, for uniformly sampled collocation points. There is no curse of dimensionality in this case.

4. Numerical experiments

In this section, we study several numerical examples to demonstrate the interplay between the dimensionality of the problem and the regularity of the solution. Our theory demonstrates that this interplay is central to determining the convergence rate, and hence accuracy, of the methodology studied in this paper.

In Subsection 4.1, we consider a high dimensional elliptic PDE with smooth solutions. By varying the dimension of the problem and the frequency of the solution, we demonstrate dimension-benign convergence rates, and in particular the accuracy is better when the frequency of the solution is lower. In Subsection 4.2, we consider a high dimensional parametric PDE problem to illustrate the importance of choosing kernels that adapted to the regularity of the solution. In Subsection 4.3, we present a high dimensional Hamilton-Jacobi-Bellman (HJB) equation, which goes beyond our theory and demonstrates the interplay between dimensionality and regularity.

4.1. High dimensional PDEs

Consider the variable coefficient nonlinear elliptic PDEs

$$\begin{cases} -\nabla\cdot(A\nabla u) + u^3 = f, & \text{in } \Omega, \\ u = g, & \text{on } \partial\Omega. \end{cases} \tag{4.1}$$

We set $A(\mathbf{x}) = \exp\left(\sin\left(\sum_{j=1}^d \cos(x_j)\right)\right)$, and the ground truth solution

$$u^\star(\mathbf{x};\beta) = \exp(\sin(\beta\sum_{j=1}^d \cos(x_j))),$$

where we have a parameter β to control the frequency of u. The right hand side and boundary data are obtained using A and $u^\star$.

In the experiment, we choose the domain Ω to be the unit ball in $\mathbb{R}^d$ for $d = 2, 3, \dots, 6$. We sample $M_\Omega = 1000, 2000, 4000, 8000$ points uniformly in the interior, and respectively $M_{\partial\Omega} = 200, 400, 800, 1600$ points uniformly on the boundary.

After selecting the kernel function, the number of iteration steps in our algorithm is set to be 3 with initial solution 0. We sample another set of M_Ω test points and evaluate the L^2 error of the solution on these points. The results are averaged over 10 independent draws of the uniform collocation points.

In the first experiment, we choose the Matérn kernel with $\nu = 7/2$ and with lengthscale $\sigma = 0.25\sqrt{d}$. We choose $\beta = 1, 4$, to compare the convergence given ground truth with different frequencies. The results are shown in Fig. 2. It is clear that when β is small, the accuracy is better. The slopes of convergence curves also have a tendency to improve for $d \ge 3$ if we increase β.

In the second experiment, we fix $\beta = 4$, and choose the Matérn kernel with $\nu = 5/2, 9/2$ and with lengthscale $\sigma = 0.25\sqrt{d}$. Results are shown in Fig. 3. Comparing $\nu = 5/2, 9/2$ and $\nu = 7/2$ in the last example, we observe that increasing ν leads to faster convergence. This is due to the fact that the true solution is smooth. In dimension $d = 2$, we can identify the exact convergence rate as $\nu - 1$. In all dimensions, the rate is faster than the Monte Carlo rate. We observe that the regularity of the solution softens the effect of the curse of dimensionality, i.e., convergence rates are better in higher dimensions when β is smaller.

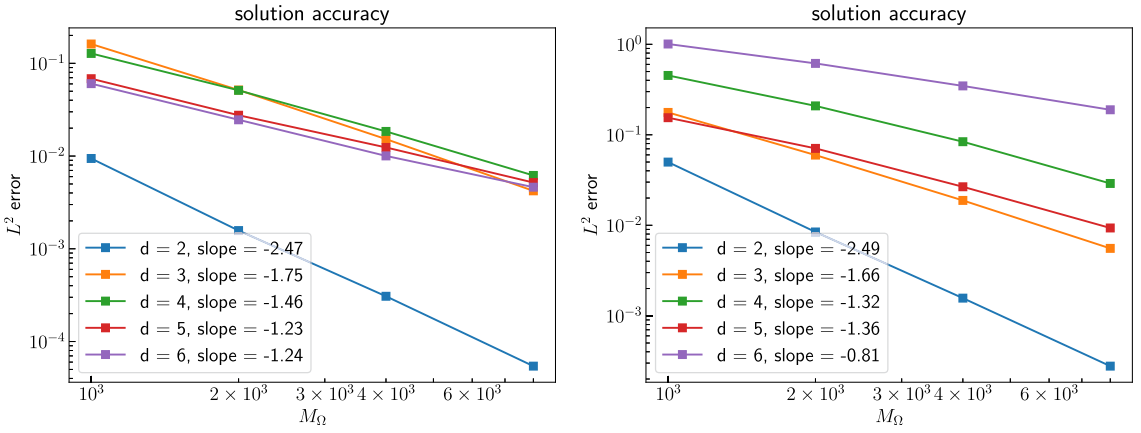

Fig. 2. L^2 test errors of solutions to Problem (4.1) as a function of the number of collocation points. Left: $\beta = 1$; right: $\beta = 4$. In both cases, we choose Matérn kernel with $\nu = 7/2$. Reported slopes in the legend denote empirical convergence rates.

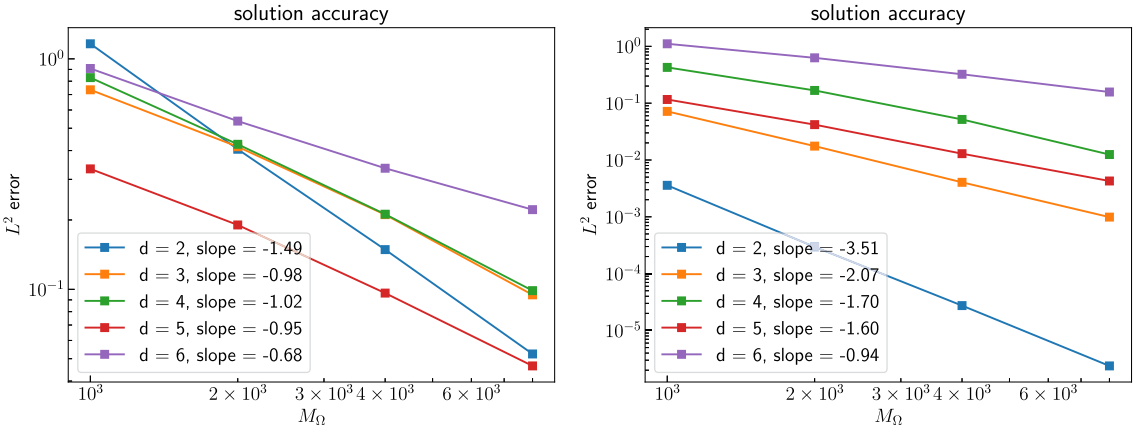

Fig. 3. L^2 test errors of solutions to Problem (4.1) as a function of the number of collocation points with $\beta = 4$. Left: Matérn kernel with $\nu = 5/2$; right: Matérn kernel with $\nu = 9/2$. Reported slopes in the legend denote empirical convergence rates.

4.2. Parametric PDEs

We consider a parametric version of the linear ($\tau = 0$) Darcy flow problem in Example 2.2:

$$\begin{cases} -\mathrm{div}\big(\exp(a(\mathbf{x},\boldsymbol{\theta}))\nabla u\big)(\mathbf{x}) = f(\mathbf{x}), & \mathbf{x}\in\Omega, \\ u(\mathbf{x}) = g(\mathbf{x}), & \mathbf{x}\in\partial\Omega. \end{cases} \tag{4.2}$$

Following the general form (2.1), we aim to obtain the solution as a function taking values in the product space Υ. (4.2) can be rewritten in terms of $\mathbf{s} = (\mathbf{x},\boldsymbol{\theta})$ with new forcing terms $\hat{f}$ and $\hat{g}$ depending only on the first coordinate of $\mathbf{s}$

$$\begin{cases} -\mathrm{div}_{\mathbf{x}}\big(A(\mathbf{s})\nabla_{\mathbf{x}} u(\mathbf{s})\big)(\mathbf{s}) = \hat{f}(\mathbf{s}) = f(\mathbf{x}), & \mathbf{s}\in\Upsilon, \\ u(\mathbf{s}) = \hat{g}(\mathbf{s}) = g(\mathbf{x}), & \mathbf{s}\in\partial\Upsilon. \end{cases} \tag{4.3}$$

Recall that we defined $\partial\Upsilon = \partial\Omega\times\Theta$. For our numerical example, we let $d = 1$ and vary p. We set $A(x,\theta) = 2 + \theta_0 + \sum_{j=1}^{p} \frac{\theta_j}{j^k}\sin(\pi x + j)$, $f(x) = x$ and $g(x) = 0$, a similar setting as in [16]. We choose $\Omega = [0,1]$, and $\Theta = [0,1]^p$, for $p = 2,3,\ldots,6$. Note $1 \le A(\mathbf{s}) \le 4$ since the sum is in $[-1,1]$ for all p and $\theta\in\Theta$, matching the setting of Example 3.13.

We sample different M_Ω points uniformly in the interior, and $M_{\partial\Omega} = M_\Omega/10$ points uniformly on the boundary of x. We do two experiments with different choices of kernel, in the first (Fig. 4, left), a vanilla Gaussian kernel with different length scales for the x and θ dimension, and with a scaling of the length scale in θ proportional to $\sqrt{p}$. In the second one (Fig. 4, right), we adapt the Gaussian kernel to the decay in $A(x,\theta)$, by including the decay of $1/j^k$ in the norm in θ space used by the kernel. We see significant improvement in test error using this adaptation in high dimensions, which suggests future research directions of kernel adaptation to the specific form of the PDE. In all cases, we use a cross-validation procedure for hyperparameter tuning and we observe the average

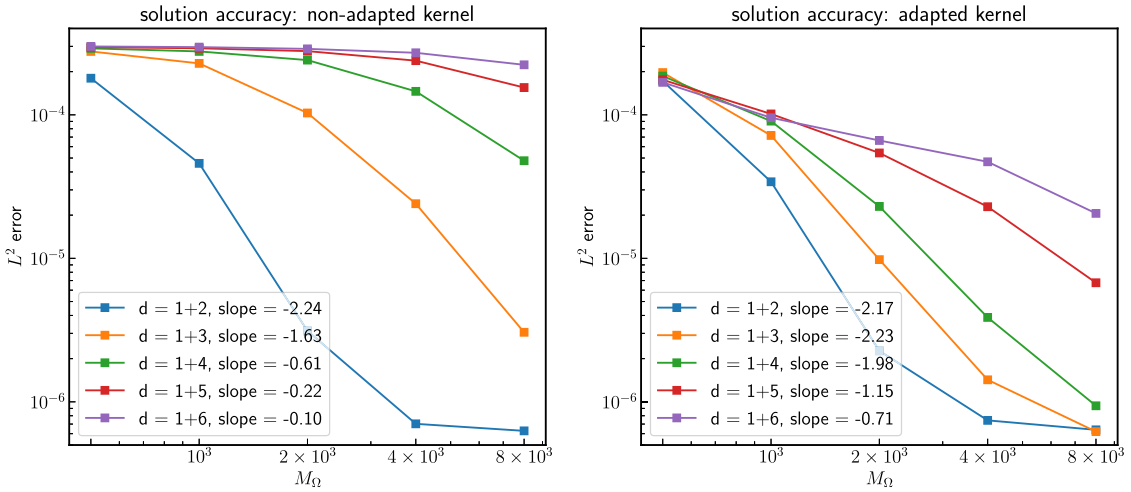

Fig. 4. L^2 test error of solutions to Problem (4.3) as a function of the number of collocation points. Left: vanilla gaussian kernel; Right: Gaussian kernel adapted to the regularity of A. Reported slopes in the legend denote empirical convergence rates.

L^2 test error on an independent set of test points for different values of p and M_Ω. Since $d = 1$ we computed our ground truth solution by numerically integrating Equation (4.3) using quadrature.

As mentioned, this problem was also explored by [16], in which sparse multivariate polynomials are used to estimate the solution with a rate independent of the number of parameters, provided the decay of the coefficient functions is large enough (in ℓ^p for some $0 < p < 1$). While this assumption is satisfied in this example, our method's convergence rate greatly depends on the dimension of θ when the kernel is not adapted to the particular equations and coefficients $A(x,\theta)$. Our results indicate improvement in the dependence of convergence rates on dimension when the kernel is adapted to the regularity of A. It remains open whether our kernel based approach (which is not specific to parametric equations) can achieve the same dimension independent convergence rates as the ones in [16] (which apply even in the countably infinite dimensional case and which they refer to as breaking the curse of dimensionality) for parametric elliptic PDEs with rapidly decreasing parametric dependence as specified above (this assumption implies a finite number of effective parameters).

4.3. High dimensional HJB equation

Consider a prototypical HJB equation:

$$\begin{aligned}(\partial_t + \Delta)V(\mathbf{x},t) - |\nabla V(\mathbf{x},t)|^2 &= 0\\ V(\mathbf{x},T) &= g(\mathbf{x}),\end{aligned} \tag{4.4}$$

where, $g(\mathbf{x}) = \log(\frac{1}{2} + \frac{1}{2}|\mathbf{x}|^2), \mathbf{x} \in \mathbb{R}^d, t \in [0,T]$. We are interested in solving $V(\mathbf{x}_0,0)$ for some $\mathbf{x}_0 \in \mathbb{R}^d$. We adopt the stochastic differential equation (SDE) formula for representing the solution of the PDEs, following [86,72]. More specifically, consider the SDE

$$dX_s = \sqrt{2}dW_s,\ X_0 = \mathbf{x}_0\,. \tag{4.5}$$

We define $Y_s = V(X_s,s), Z_s = \sqrt{2}\nabla V(X_s,s)$. By Ito's formula, one obtains

$$dY_s = \frac{1}{2}|Z_s|^2 ds + Z_s \cdot dW_s\,. \tag{4.6}$$

The strategy is to integrate the above SDE backward to Y_0. An implicit[3] Euler discretization from time t_{n+1} to t_n ($\Delta t = t_{n+1} - t_n$) leads to the following equation:

$$V(X_{t_{n+1}}, t_{n+1}) = V(X_{t_n}, t_n) + |\nabla V(X_{t_n}, t_n)|^2 \Delta t + \sqrt{2}\nabla V(X_{t_n}, t_n)\cdot \xi_{n+1}\sqrt{\Delta t}\,. \tag{4.7}$$

Algorithmically, we sample J different paths of the forward SDE in (4.5), namely $X^{(j)}_{t_n}, 1 \le j \le J$, using the Euler–Maruyama scheme. Then, backward in time, we apply our kernel method, namely to solve the following optimization problem

[3] Implicit because are integrating backwards in time.

Table 2
Numerical results for the HJB equation (4.4), computing the quantity $V(\mathbf{x}_0, 0)$.

σ	10	25	50	100	200
Computed solution $V(\mathbf{x}_0, 0)$	5.6042	4.6366	4.6039	4.6021	4.6021
Relative accuracy	22.10%	1.0154%	0.303%	0.2638%	0.2638%

$$\begin{cases} \underset{u\in\mathcal{U}}{\text{minimize}} & \|u\|_{\mathcal{U}} \\ \text{s.t.} & u(X_{t_n}^{(j)}, t_n) + |\nabla u(X_{t_n}^{(j)}, t_n)|^2 \Delta t + \sqrt{2}\nabla u(X_{t_n}^{(j)}, t_n)\cdot \xi_{n+1}\sqrt{\Delta t} = V(X_{t_{n+1}}^{(j)}, t_{n+1}) \end{cases} \tag{4.8}$$

to get the solution $V(\cdot, t_n)$, assuming $V(\cdot, t_{n+1})$ has been solved. Iterating this process, we end up with the solution $V(\mathbf{x}_0, 0)$. We can understand the algorithm as applying our kernel method iteratively with the sample path as the collocation points.

Experimentally, we consider $d = 100$ as in [86,72]. We aim to solve $V(\mathbf{x}_0, 0)$ for $\mathbf{x}_0 = 0$. The ground truth is $V(\mathbf{x}_0, 0) = 4.589992$ provided in [86]. We sample $J = 2000$ paths from $\mathbf{x}_0$ and choose the inverse quadratic kernel $k(\mathbf{x}, \mathbf{y}; \sigma) = \left(\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2d\sigma^2} + 1\right)^{-1}$. We use the "linearize-then-optimize" approach to compute an approximate solution to (4.8). The nugget term is set to be $\eta = 10^{-3}$. The result is shown in Table 2. We observe that a suitable choice of the lengthscale of the kernel is crucial to obtain an accurate solution. Compared to the relative accuracy of 0.171% (reported in [72]) using neural networks (DenseNet like architecture with 4 hidden layers) to solve (4.7), the accuracy of using kernel methods with a simple quadratic kernel is comparable. Moreover, the lengthscale of the kernel is very large, indicating that the solution behavior of this HJB equation is very smooth; similar "blessings of dimensionality" have been reported and discussed in [72], where they used a constant function (and the terminal function g) as ansatz to solve (4.7) and obtained very high accuracy.[4] Thus, this HJB example in dimension 100 demonstrates again the trade-off between the smoothness of the solution and the curse of dimensionality.

5. Conclusions

In this paper, we conducted an error analysis of GP and kernel based methods for solving PDEs. We provided convergence rates under the assumptions that (1) the solution belongs to the RKHS which is embedding to some Sobolev space of sufficient regularity, and (2) the underlying forward and inverse PDE operator is stable in corresponding Sobolev spaces.

Our analysis relies on the crucial minimizing norm property of the numerical solution in the kernel/GP methodology. The analysis could be seamlessly generalized to the function class of NNs and other norms such as non-quadratic norms if we can formulate the training process as a minimization problem over the related norm.

We emphasize that our convergence rates hold for the exact minimizer of the minimization problem. In practice, finding such a minimizer algorithmically can be a separate and challenging problem. Our numerical experience suggests that Gauss-Newton iterations usually perform well, and typically, 2-5 iterations are sufficient for convergence. Therefore, we can combine the error analysis in this paper and the fast implementation of the algorithm in [13] to obtain a near-linear complexity solver for nonlinear PDEs with rigorous accuracy guarantee.

It is worth mentioning that this paper focuses only on analyzing the MAP estimator within the GP interpretation. Exploring the posterior distribution of the GP can provide a means for quantifying uncertainty in the solution. In particular, analyzing the posterior contraction is an interesting direction for future research.

CRediT authorship contribution statement

Pau Batlle: Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis. **Yifan Chen:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Bamdad Hosseini:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Houman Owhadi:** Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Andrew M. Stuart:** Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[4] We anticipate that using the feature map perspective of kernel methods with constants and g as features will achieve a similar accuracy as in [72]. We did not pursue this here to avoid using strong prior information on the solution beyond regularity.

Acknowledgements

The authors gratefully acknowledge support by the Air Force Office of Scientific Research under MURI award number FA9550-20-1-0358 (Machine Learning and Physics-Based Modeling and Simulation). BH acknowledges support by the National Science Foundation grant number NSF-DMS-2208535 (Machine Learning for Bayesian Inverse Problems). HO also acknowledges support by the Department of Energy under award number DE-SC0023163 (SEA-CROGS: Scalable, Efficient and Accelerated Causal Reasoning Operators, Graphs and Spikes for Earth and Embedded Systems) and a Department of Defense Vannevar Bush Faculty Fellowship.

Appendix A. Sobolev sampling inequalities on manifolds

Below we collect useful sampling inequalities for Sobolev functions defined on smooth manifolds with corners. Following [46, Chs. 1, 16] we consider a smooth, compact Riemannian manifold $\mathcal{M} \subset \mathbb{R}^d$ of dimension $k \leq d$ with corners, i.e., a Riemannian manifold with a smooth structure with corners; see [46, Ch. 16]. On such a manifold we define the natural geodesic distance

$$\rho_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}, \qquad \rho_{\mathcal{M}}(x, y) := \inf \int_0^1 \|\dot{\ell}(t)\| dt,$$

where the infimum is taken over all piecewise smooth paths $\ell : [0, 1] \mapsto \mathcal{M}$ satisfying the boundary conditions $\ell(0) = x$ and $\ell(1) = y$, and $\|\dot{\ell}(t)\|$ is the length of the tangent vector $\dot{\ell}(T)$ under the Riemannian metric.

Following [31] (see also [83, Sec. 4.3]) we further consider the Sobolev spaces $H^k(\mathcal{M})$ of functions defined on $\mathcal{M}$ as follows: Let $\mathcal{A} = \{M_j, \Psi_j\}_{j=1}^N$ be an atlas for $\mathcal{M}$ and let $\{\kappa_j\}$ be a partition of unity of $\mathcal{M}$, subordinate to M_j. Then given functions $u : \mathcal{M} \to \mathbb{R}$ we define the Sobolev norms and the associated Sobolev spaces $H^s(\mathcal{M})$ as

$$H^s(\mathcal{M}) := \{u : \mathcal{M} \to \mathbb{R} \mid \|u\|_{H^s(\mathcal{M})} < +\infty\}, \qquad \|u\|_{H^s(\mathcal{M})} := \left(\sum_{j=1}^N \|\pi_j(u)\|^2_{H^s(\Xi_j)} \right)^{1/2},$$

where the maps π_j are defined as

$$\pi_j(f) := \begin{cases} \kappa_j(f(\Psi_j^{-1}(y))), & \text{if } y \in \Psi_j(M_j), \\ 0 & \text{otherwise,} \end{cases}$$

and the sets Ξ_j are given by

$$\Xi_j := \begin{cases} \mathbb{R}^k & \text{if } \Psi_j \text{ is an interior chart,} \\ \{(x_1, \dots, x_k) \in \mathbb{R}^k | x_1 \geq 0\} & \text{if } \Psi_j \text{ is a boundary chart,} \\ \{(x_1, \dots, x_k) \in \mathbb{R}^k | x_1 \geq 0, \dots, x_k \geq 0\} & \text{if } \Psi_j \text{ is a corner chart.} \end{cases}$$

Put simply, the Sobolev spaces $H^s(\mathcal{M})$ are functions on $\mathcal{M}$ that, locally after the flattening of the manifold belong to the standard Sobolev spaces H^s. With these notions at hand we then recall the following result of [31], which was proven by those authors for smooth embedded manifolds without boundary or corners. However, a brief investigation of the proof of that result reveals that it can immediately be generalized to our setting with manifolds with corners. In fact, the idea of the proof is to use the atlas to locally flatten the manifold and apply classic sampling theorems such as [2, Thm. 4.1] on each patch. The only difference in the case of manifolds with corners is that the patches do not only map to $\mathbb{R}^k$ but rather to the subspaces Ξ_j depending on whether the corresponding chart is an interior, boundary, or corner chart.

Proposition A.1 *([31, Lem. 10]). Suppose $\mathcal{M} \subset \mathbb{R}^d$ is a smooth, compact, Riemannian manifold with corners, of dimension k and let $s > k/2$ and $r \in \mathbb{N}$ satisfy $0 \leq r \leq \lceil s \rceil - 1$. Let $X \subset \mathcal{M}$ be a discrete set with mesh norm $h_{\mathcal{M}}$ defined as*

$$h_{\mathcal{M}} := \sup_{x' \in \mathcal{M}} \inf_{x \in X} \rho_{\mathcal{M}}(x, x').$$

Then there is a constant $h_0 > 0$ depending only on $\mathcal{M}$ such that if $h_{\mathcal{M}} < h_0$ and if $u \in H^s(\mathcal{M})$ satisfies $u|_X = 0$ then

$$\|u\|_{H^r(\mathcal{M})} \leq C h_{\mathcal{M}}^{s-r} \|u\|_{H^s(\mathcal{M})}.$$

Here $C > 0$ is a constant independent of $h_{\mathcal{M}}$ and u.

Appendix B. Bounds on fill distances

This section collects a result from [71] for bounding the fill-in distance for randomly distributed points on a manifold.

Assume $(\mathcal{M}, \rho)$ is a metric space, and μ is a finite positive Borel measure supported on $\mathcal{M}$. Let $X = \{x_1, ..., x_N\}$ be a set of N points, independently and randomly drawn from μ. Define the fill-in distance

$$h_{\mathcal{M}} = \sup_{x' \in \mathcal{M}} \inf_{x \in X} \rho(x, x'). \tag{B.1}$$

Then, [71, Thm. 2.1] implies the following:

Proposition B.1. *Suppose Φ is a continuous non-negative strictly increasing function on $(0, \infty)$ satisfying $\Phi(r) \to 0$ as $r \to 0^+$. If there exists a positive number r_0 such that $\mu(B(x, r)) \geq \Phi(r)$ holds for all $x \in \mathcal{M}$ and every $r < r_0$, then there exist positive constants c_1, c_2, c_3 and α_0 such that for any $\alpha > \alpha_0$, we have*

$$\mathbb{P}\left[h_{\mathcal{M}} \geq c_1 \Phi^{-1}\left(\frac{\alpha \log N}{N}\right)\right] \leq c_2 N^{1-c_3\alpha}. \tag{B.2}$$

We use this proposition to prove Proposition 3.14.

Proof of Proposition 3.14. We apply Proposition B.1. For the bounded domain $\Omega \subset \mathbb{R}^d$, we know that there exists a constant C such that $\Phi(r) = Cr^d$ will satisfy the assumption in Proposition B.1. Moreover, we choose α such that $c_2 M_\Omega^{1-c_3\alpha} \leq \delta$. This implies that $\alpha \geq \frac{1}{c_3 \log M_\Omega} \log(c_2 M_\Omega/\delta)$. Pick $\alpha = \frac{C'}{c_3 \log M_\Omega} \log(c_2 M_\Omega/\delta)$ for some $C' \geq 1$ such that $\alpha \geq \alpha_0$. Then Proposition B.1 shows that with probability at least $1 - \delta$,

$$h_\Omega \leq c_1 \Phi^{-1}\left(\frac{\alpha \log M_\Omega}{M_\Omega}\right) \leq C''\left(\frac{\log(M_\Omega/\delta)}{M_\Omega}\right)^{1/d},$$

where C'' is a constant independent of M_Ω and δ. The bound on $h_{\partial\Omega}$ can be proved similarly by choosing $\Phi(r) = Cr^{d-1}$. □

Appendix C. The choice of nugget terms

For numerical stability, we add a diagonal adaptive nugget term to the kernel matrix in our computation such that

$$u^{\ell+1}(x) = K(x, \boldsymbol{\phi}^l)[K(\boldsymbol{\phi}^l, \boldsymbol{\phi}^l) + \eta \mathrm{diag}(K(\boldsymbol{\phi}^l, \boldsymbol{\phi}^l))]^{-1} \begin{pmatrix} \left(f - u^\ell + \mathcal{P}'(u^\ell)u^\ell\right)|_{\mathbf{s}_\Omega} \\ \left(g - u^\ell + \mathcal{B}'(u^\ell)u^\ell\right)|_{\mathbf{s}_{\partial\Omega}} \end{pmatrix}$$

Typically $\eta = 10^{-10}$. This nugget term is similar to the adaptive nugget term proposed in [12]. It is much more effective than the naive choice of $K(\boldsymbol{\phi}^l, \boldsymbol{\phi}^l) + \eta I$, since the conditioning of the interior block and the boundary block in the kernel matrix differs dramatically.

Data availability

Data will be made available on request.

References

[1] B.O. Almroth, P. Stern, F.A. Brogan, Automatic choice of global shape functions in structural analysis, AIAA J. 16 (1978) 525–528.
[2] R. Arcangéli, M.C. López de Silanes, J.J. Torrens, An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing, Numer. Math. 107 (2007) 181–211.
[3] I. Babuška, J. Osborn, Can a finite element method perform arbitrarily badly?, Math. Comput. 69 (2000) 443–462.
[4] P. Batlle, M. Darcy, B. Hosseini, H. Owhadi, Kernel methods are competitive for operator learning, arXiv preprint, arXiv:2304.13202, 2023.
[5] J. Beck, R. Tempone, F. Nobile, L. Tamellini, On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods, Math. Models Methods Appl. Sci. 22 (2012) 1250023.
[6] R. Becker, M. Brunner, M. Innerberger, J.M. Melenk, D. Praetorius, Cost-optimal adaptive iterative linearized fem for semilinear elliptic pdes, ESAIM: Math. Model. Numer. Anal. 57 (2023) 2193–2225.
[7] A. Berlinet, C. Thomas-Agnan, Reproducing Kernel Hilbert Spaces in Probability and Statistics, Springer Science & Business Media, 2011.
[8] V.I. Bogachev, Gaussian Measures, American Mathematical Society, 1998.
[9] K. Böhmer, R. Schaback, A nonlinear discretization theory, J. Comput. Appl. Math. 254 (2013) 204–219.
[10] K. Böhmer, R. Schaback, A nonlinear discretization theory for meshfree collocation methods applied to quasilinear elliptic equations, Z. Angew. Math. Mech. (J. Appl. Math. Mech.) 100 (2020) e201800170.
[11] J. Chen, F. Schäfer, J. Huang, M. Desbrun, Multiscale Cholesky preconditioning for ill-conditioned problems, ACM Trans. Graph. 40 (2021) 1–13.
[12] Y. Chen, B. Hosseini, H. Owhadi, A.M. Stuart, Solving and learning nonlinear PDEs with Gaussian processes, J. Comput. Phys. 447 (2021) 110668.
[13] Y. Chen, H. Owhadi, F. Schäfer, Sparse Cholesky factorization for solving nonlinear PDEs via Gaussian processes, arXiv preprint, arXiv:2304.01294, 2023.
[14] Y. Chen, H. Owhadi, A. Stuart, Consistency of empirical Bayes and kernel flow for hierarchical parameter estimation, Math. Comput. (2021).
[15] K.C. Cheung, L. Ling, R. Schaback, h^2-convergence of least-squares kernel collocation methods, SIAM J. Numer. Anal. 56 (2018) 614–633.
[16] A. Chkifa, A. Cohen, R. DeVore, C. Schwab, Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs, ESAIM: Math. Model. Numer. Anal. 47 (2012) 253–280.
[17] A. Chkifa, A. Cohen, C. Schwab, High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs, Found. Comput. Math. 14 (2014) 601–633.
[18] O.A. Chkrebtii, D.A. Campbell, B. Calderhead, M.A. Girolami, Bayesian solution uncertainty quantification for differential equations, Bayesian Anal. 11 (2016) 1239–1267.
[19] I. Cialenco, G.E. Fasshauer, Q. Ye, Approximation of stochastic partial differential equations by a kernel-based collocation method, Int. J. Comput. Math. 89 (2012) 2543–2561.

[20] J. Cockayne, C. Oates, T. Sullivan, M. Girolami, Probabilistic numerical methods for PDE-constrained Bayesian inverse problems, AIP Conf. Proc. (2017) 060001.
[21] J. Cockayne, C.J. Oates, T.J. Sullivan, M. Girolami, Bayesian probabilistic numerical methods, SIAM Rev. 61 (2019) 756–789.
[22] A. Cohen, R. DeVore, Approximation of high-dimensional parametric PDEs, Acta Numer. 24 (2015) 1–159.
[23] A. Cohen, R. DeVore, C. Schwab, Convergence rates of best n-term Galerkin approximations for a class of elliptic sPDEs, Found. Comput. Math. 10 (2010) 615–646.
[24] M. Darcy, B. Hamzi, G. Livieri, H. Owhadi, P. Tavallali, One-shot learning of stochastic differential equations with data adapted kernels, Phys. D: Nonlinear Phenom. 444 (2023) 133583.
[25] T. De Ryck, S. Mishra, Error analysis for physics-informed neural networks (pinns) approximating Kolmogorov PDEs, Adv. Comput. Math. 48 (2022) 1–40.
[26] G.E. Fasshauer, Solving differential equations with radial basis functions: multilevel methods and smoothing, Adv. Comput. Math. 11 (1999) 139–159.
[27] R.P. Feynman, Cargo cult science, in: The Art and Science of Analog Circuit Design, Elsevier, 1998, pp. 55–61.
[28] B. Fornberg, N. Flyer, Solving PDEs with radial basis functions, Acta Numer. 24 (2015) 215–258.
[29] C. Franke, R. Schaback, Convergence order estimates of meshless collocation methods using radial basis functions, Adv. Comput. Math. 8 (1998) 381–399.
[30] C. Franke, R. Schaback, Solving partial differential equations by collocation using radial basis functions, Appl. Math. Comput. 93 (1998) 73–82.
[31] E. Fuselier, G.B. Wright, Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates, SIAM J. Numer. Anal. 50 (2012) 1753–1776.
[32] R.G. Ghanem, P.D. Spanos, Stochastic Finite Elements: a Spectral Approach, Dover Publications, 2003.
[33] P. Giesl, H. Wendland, Meshless collocation: error estimates with application to dynamical systems, SIAM J. Numer. Anal. 45 (2007) 1723–1741.
[34] D. Gilbarg, N.S. Trudinger, Elliptic Partial Differential Equations of Second Order, Springer, 1977.
[35] T.G. Grossmann, U.J. Komorowska, J. Latz, C.-B. Schönlieb, Can physics-informed neural networks beat the finite element method?, arXiv preprint, arXiv: 2302.04107, 2023.
[36] M.D. Gunzburger, C.G. Webster, G. Zhang, Stochastic finite element methods for partial differential equations with random input data, Acta Numer. 23 (2014) 521–650.
[37] J.S. Hesthaven, G. Rozza, B. Stamm, et al., Certified Reduced Basis Methods for Parametrized Partial Differential Equations, vol. 590, Springer, 2016.
[38] Y. Hon, R. Schaback, Solvability of partial differential equations by meshless kernel methods, Adv. Comput. Math. 28 (2008) 283–299.
[39] E.J. Kansa, Multiquadrics—a scattered data approximation scheme with applications to computational fluid-dynamics—i surface approximations and partial derivative estimates, Comput. Math. Appl. 19 (1990) 127–145.
[40] E.J. Kansa, Multiquadrics—a scattered data approximation scheme with applications to computational fluid-dynamics—ii solutions to parabolic, hyperbolic and elliptic partial differential equations, Comput. Math. Appl. 19 (1990) 147–161.
[41] R. Kempf, H. Wendland, C. Rieger, Kernel-based reconstructions for parametric PDEs, in: IWMMPDE 2017: Meshfree Methods for Partial Differential Equations IX 9, Springer, 2019, pp. 53–71.
[42] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. R. Stat. Soc., Ser. B, Stat. Methodol. 63 (2001) 425–464.
[43] F.M. Larkin, Gaussian measure in Hilbert space and applications in numerical analysis, J. Math. 2 (1972).
[44] O. Le Maître, O.M. Knio, Spectral Methods for Uncertainty Quantification: with Applications to Computational Fluid Dynamics, Springer Science & Business Media, 2010.
[45] J. Lee, Y. Bahri, R. Novak, S.S. Schoenholz, J. Pennington, J. Sohl-Dickstein, Deep neural networks as Gaussian processes, arXiv preprint, arXiv:1711.00165, 2017.
[46] J.M. Lee, Introduction to Smooth Manifolds, Springer, 2012.
[47] Z. Li, N. Kovachki, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, A. Anandkumar, Fourier neural operator for parametric partial differential equations, in: International Conference on Learning Representations, 2020.
[48] L. Ling, R. Opfer, R. Schaback, Results on meshless collocation techniques, Eng. Anal. Bound. Elem. 30 (2006) 247–253.
[49] L. Ling, R. Schaback, Stable and convergent unsymmetric meshless collocation methods, SIAM J. Numer. Anal. 46 (2008) 1097–1115.
[50] D. Long, N. Mrvaljevic, S. Zhe, B. Hosseini, A kernel approach for pde discovery and operator learning, arXiv preprint, arXiv:2210.08140, 2022.
[51] D. Long, Z. Wang, A. Krishnapriyan, R. Kirby, S. Zhe, M. Mahoney, Autoip: a united framework to integrate physics into Gaussian processes, in: International Conference on Machine Learning, in: PMLR, 2022, pp. 14210–14222.
[52] L. Lu, P. Jin, G. Pang, Z. Zhang, G.E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, Nat. Mach. Intell. 3 (2021) 218–229.
[53] Y. Lu, H. Chen, J. Lu, L. Ying, J. Blanchet, Machine learning for elliptic PDEs: fast rate generalization bound, neural scaling law and minimax optimality, arXiv preprint, arXiv:2110.06897, 2021.
[54] D.J. Lucia, P.S. Beran, W.A. Silva, Reduced-order modeling: new approaches for computational physics, Prog. Aerosp. Sci. 40 (2004) 51–117.
[55] W. McLean, Strongly Elliptic Systems and Boundary Integral Equations, Cambridge University Press, 2000.
[56] J.M. Melenk, On n-widths for elliptic problems, J. Math. Anal. Appl. 247 (2000) 272–289.
[57] R. Meng, X. Yang, Sparse Gaussian processes for solving nonlinear PDEs, arXiv preprint, arXiv:2205.03760, 2022.
[58] C. Mou, X. Yang, C. Zhou, Numerical methods for mean field games based on Gaussian processes and Fourier features, J. Comput. Phys. 460 (2022) 111188.
[59] K. Muandet, K. Fukumizu, B. Sriperumbudur, B. Schölkopf, et al., Kernel mean embedding of distributions: a review and beyond, Found. Trends Mach. Learn. 10 (2017) 1–141.
[60] R.M. Neal, Priors for infinite networks, Bayesian Lear. Neural Netw. (1996) 29–53.
[61] F. Nobile, R. Tempone, C.G. Webster, An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (2008) 2411–2442.
[62] F. Nobile, R. Tempone, C.G. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM J. Numer. Anal. 46 (2008) 2309–2345.
[63] A.K. Noor, J.M. Peters, Reduced basis technique for nonlinear analysis of structures, AIAA J. 18 (1980) 455–462.
[64] H. Owhadi, Bayesian numerical homogenization, Multiscale Model. Simul. 13 (2015) 812–828.
[65] H. Owhadi, Do ideas have shape? Idea registration as the continuous limit of artificial neural networks, Phys. D: Nonlinear Phenom. 444 (2023) 133592.
[66] H. Owhadi, C. Scovel, Operator-Adapted Wavelets, Fast Solvers, and Numerical Homogenization: From a Game Theoretic Approach to Numerical Approximation and Algorithm Design, Cambridge University Press, 2019.
[67] H. Owhadi, G.R. Yoo, Kernel flows: from learning kernels from data into the abyss, J. Comput. Phys. 389 (2019) 22–47.
[68] A. Pinkus, N-Widths in Approximation Theory, vol. 7, Springer Science & Business Media, 2012.
[69] M. Raissi, P. Perdikaris, G.E. Karniadakis, Numerical Gaussian processes for time-dependent and nonlinear partial differential equations, SIAM J. Sci. Comput. 40 (2018) A172–A198.
[70] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.
[71] A. Reznikov, E.B. Saff, The covering radius of randomly distributed points on a manifold, Int. Math. Res. Not. 2016 (2016) 6065–6094.
[72] L. Richter, L. Sallandt, N. Nüsken, Solving high-dimensional parabolic PDEs using the tensor train format, arXiv preprint, arXiv:2102.11830, 2021.
[73] S. Särkkä, Linear operators and stochastic partial differential equations in Gaussian process regression, in: International Conference on Artificial Neural Networks, Springer, 2011, pp. 151–158.

[74] R. Schaback, Convergence of unsymmetric kernel-based meshless collocation methods, SIAM J. Numer. Anal. 45 (2007) 333–351.
[75] R. Schaback, A computational tool for comparing all linear pde solvers: error-optimal methods are meshless, Adv. Comput. Math. 41 (2015) 333–355.
[76] R. Schaback, All well-posed problems have uniformly stable and convergent discretizations, Numer. Math. 132 (2016) 597–630.
[77] R. Schaback, H. Wendland, Kernel techniques: from machine learning to meshless methods, Acta Numer. 15 (2006) 543.
[78] F. Schäfer, M. Katzfuss, H. Owhadi, Sparse Cholesky factorization by Kullback–Leibler minimization, SIAM J. Sci. Comput. 43 (2021) A2019–A2046.
[79] F. Schäfer, T.J. Sullivan, H. Owhadi, Compression, inversion, and approximate pca of dense kernel matrices at near-linear computational complexity, Multiscale Model. Simul. 19 (2021) 688–730.
[80] B. Scholkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2018.
[81] Y. Shin, J. Darbon, G.E. Karniadakis, On the convergence of physics informed neural networks for linear second-order elliptic and parabolic type PDEs, arXiv preprint, arXiv:2004.01806, 2020.
[82] L.P. Swiler, M. Gulian, A.L. Frankel, C. Safta, J.D. Jakeman, A survey of constrained Gaussian process regression: approaches and implementation challenges, J. Mach. Learn. Model. Comput. 1 (2020).
[83] M. Taylor, Partial Differential Equations I: Basic Theory, Springer Science & Business Media, 2013.
[84] A.W. van der Vaart, J.H. van Zanten, Reproducing kernel Hilbert spaces of Gaussian priors, in: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh, Institute of Mathematical Statistics, 2008, pp. 200–222.
[85] J. Wang, J. Cockayne, O. Chkrebtii, T.J. Sullivan, C. Oates, et al., Bayesian numerical methods for nonlinear partial differential equations, Stat. Comput. 31 (2021) 1–20.
[86] E. Weinan, J. Han, A. Jentzen, Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations, Commun. Math. Stat. 5 (2017) 349–380.
[87] E. Weinan, B. Yu, The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems, Commun. Math. Stat. 6 (2018) 1–12.
[88] H. Wendland, Scattered Data Approximation, Cambridge University Press, 2004.
[89] C.K.I. Williams, C.E. Rasmussen, Gaussian Processes for Machine Learning, The MIT Press, 2006.
[90] A.G. Wilson, Z. Hu, R. Salakhutdinov, E.P. Xing, Deep kernel learning, in: Artificial Intelligence and Statistics, in: PMLR, 2016, pp. 370–378.
[91] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.
[92] Q. Ye, Kernel-based methods for stochastic partial differential equations, arXiv preprint, arXiv:1303.5381, 2013.