SIAM Review
Vol. 61, No. 4, pp. 756–789

Bayesian Probabilistic Numerical Methods*

Jon Cockayne†
Chris J. Oates‡
T. J. Sullivan§
Mark Girolami¶

Abstract. Over forty years ago average-case error was proposed in the applied mathematics literature as an alternative criterion with which to assess numerical methods. In contrast to worst-case error, this criterion relies on the construction of a probability measure over candidate numerical tasks, and numerical methods are assessed based on their average performance over those tasks with respect to the measure. This paper goes further and establishes *Bayesian probabilistic numerical methods* as solutions to certain inverse problems based upon the numerical task within the Bayesian framework. This allows us to establish general conditions under which Bayesian probabilistic numerical methods are well defined, encompassing both the nonlinear and non-Gaussian contexts. For general computation, a numerical approximation scheme is proposed and its asymptotic convergence established. The theoretical development is extended to pipelines of computation, wherein probabilistic numerical methods are composed to solve more challenging numerical tasks. The contribution highlights an important research frontier at the interface of numerical analysis and uncertainty quantification, and a challenging industrial application is presented.

Key words. probabilistic numerics, Bayesian methods, numerical analysis, information-based complexity

AMS subject classifications. 65N21, 65N75, 62-02, 62C10, 62G08, 62M40

DOI. 10.1137/17M1139357

*Received by the editors July 18, 2017; accepted for publication (in revised form) February 14, 2019; published electronically November 6, 2019.
https://doi.org/10.1137/17M1139357
Funding: The work of the third author was supported by the Freie Universität Berlin within the German Research Foundation's Excellence Initiative and Collaborative Research Centre 1114. The work of the fourth author was supported by EPSRC grants (EP/R034710/1, EP/R018413/1, EP/R004889/1, and EP/P020720/1), an EPSRC Established Career Fellowship (EP/J016934/3), and a Royal Academy of Engineering Research Chair. The work of the second and fourth authors was supported by the Lloyd's Register Foundation Programme on Data-Centric Engineering at the Alan Turing Institute. This material was based upon work partially supported by the National Science Foundation under grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the above-named funding agencies and institutions.

†Department of Statistics, University of Warwick, Coventry CV4 7AL, UK (j.cockayne@warwick.ac.uk).
‡School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK, and Alan Turing Institute, British Library, London NW1 2DB, UK (chris.oates@ncl.ac.uk).
§Freie Universität Berlin, Berlin 14195, Germany, and Zuse Institute Berlin, Berlin 14195, Germany (sullivan@zib.de).
¶Department of Mathematics, Imperial College London, London SW7 2AZ, UK, and Alan Turing Institute, British Library, London NW1 2DB, UK (m.girolami@imperial.ac.uk).

Contents

1 Introduction **757**

2 Probabilistic Numerical Methods **759**
2.1 Notation . . . 759
2.2 Definition of a PNM . . . 760
2.2.1 Classical Numerical Methods . . . 763
2.3 Bayesian PNMs . . . 763
2.4 Model Evidence . . . 764
2.5 The Disintegration Theorem . . . 764
2.6 Prior Construction . . . 765
2.7 Relation to Emulation . . . 766

3 Decision-Theoretic Treatment **766**
3.1 Loss and Risk . . . 766
3.2 Bayes Decision Rules . . . 767
3.2.1 Optimal Information . . . 767
3.3 Conditional Connection to Average-Case Analysis . . . 768

4 Numerical Disintegration **768**
4.1 Sequential Approximation of a Disintegration . . . 769
4.2 Monte Carlo Methods . . . 769

5 Computational Pipelines and PNMs **770**
5.1 Computational Pipelines . . . 770
5.2 Bayesian Computational Pipelines . . . 773

6 Numerical Experiments **774**
6.1 Poisson Equation . . . 775
6.2 The Painlevé ODE . . . 777
6.3 Application to Industrial Process Monitoring . . . 779

7 Discussion **783**

Acknowledgments **785**

References **785**

1. Introduction. Numerical computation underpins almost all of modern scientific and industrial research and development. The impact of a finite computational budget is that problems whose solutions are high- or infinite-dimensional, such as differential equations, must be discretized in order to be solved. The result is an approximation to the object of interest. The declining rate of processor improvement as physical limits are reached stands in contrast to the surge in complexity of modern inference problems, and as a result the error incurred by discretization is attracting increased interest (e.g., [11]).

The use of single-precision arithmetic to permit finer temporal resolution in modern climate models exemplifies this situation. However, when computing in single precision, a detailed time discretization can *increase* total error, due to the increased

number of single-precision computations, and so in practice some form of trade-off is sought [37]. It has been argued that statistical considerations can permit principled error control strategies for such models [39].

Well-designed numerical methods aim to mitigate discretization errors of all forms [78]. Nonetheless, the introduction of error is unavoidable and it is the role of the numerical analyst to provide control of this error [72]. The central theoretical results of numerical analysis have not in general been obtained through statistical considerations. However, the connection of discretization error to statistics was noted as far back as [40], in which it was argued that round-off error can be modeled using a series of independent random perturbations to a standard numerical method. Other sources of numerical error, such as time discretization in the iterative numerical solution of a differential equation, can be highly structured in a manner that is not easily captured by a probability model; see [44] and [41, section 2.8]. To address these issues, the field of *probabilistic numerics* has emerged with the aim of performing formal uncertainty quantification for the mathematical object, such as the solution of a differential equation, being approximated.

The foundations of probabilistic numerics were laid in the 1970s and 1980s, when an important shift in emphasis occurred from the descriptive statistical models of the 1960s to the use of formal inference modalities that generalize across classes of numerical tasks. In a remarkable series of papers, [54, 55, 56, 57, 58, 59, 60], Larkin presented now-classical though then little-known results in probabilistic numerics, in particular establishing the correspondence between a Gaussian measure on a Hilbert space and its associated average-case optimal numerical method. Rediscovered and reemphasized on a number of occasions, the role for statisticians in this new outlook was clearly captured in [43]. The 1980s culminated in development of Bayesian optimization methods [67, 91], as well as the relation of smoothing splines to Bayesian estimation [50, 23].

The modern notion of a probabilistic numerical method (henceforth PNM) was described in [39]; these are algorithms whose output is a distribution over possible values of a deterministic quantity of interest (QoI), such as the value of an integral. Recent research in this field includes PNMs for numerical linear algebra [38, 15], numerical solution of ordinary differential equations (ODEs; [83, 17, 14]), numerical solution of partial differential equations (PDEs; [74, 16]), and numerical integration [73, 9]. For a more comprehensive list see supplement section SM3.

Open Problems. Despite numerous recent successes and achievements, there is currently no general statistical foundation for PNMs. For instance, at present it is not clear under what conditions a PNM is well defined, except for in the standard conjugate Gaussian framework considered in [58]. This limits the extent to which domain-specific knowledge, such as boundedness of an integrand or monotonicity of a solution to a differential equation, can be leveraged in PNMs. In contrast, classical numerical methods often exploit such information to achieve substantial reduction in discretization error. For instance, finite element methods for the solution of a PDE proceed based on a mesh that is designed to be more refined in areas of the domain in which greater variation of the solution is anticipated [87].

Furthermore, although PNMs have been proposed for many standard numerical tasks, the lack of common theoretical foundations makes methodological comparisons difficult. Again taking PDEs as an example, [16] placed a probability distribution on the solution of the PDE, whereas [17] placed a probability distribution on the discretization error of a numerical method. The uncertainty modeled in each case is

fundamentally different, yet at present there is no framework in which to articulate the relationship between the two approaches. Furthermore, though PNMs are often reported as being "Bayesian" there is no clear definition of what this ought to entail.

A more profound consequence of the lack of common foundation occurs when we seek to compose multiple PNMs. For example, multiphysics cardiac models involve coupled ODEs and PDEs which must each be discretized and approximately solved to estimate a clinical QoI [69]. The composition of successive discretizations leads to nontrivial error propagation and accumulation that could be quantified, in a statistical sense, with PNMs. However, proper composition of multiple PNMs requires that these PNMs share common statistical foundations to ensure coherence of the overall statistical output. These foundations remain to be established.

Contributions. This paper aims to establish rigorous foundations for PNMs.

First, we argue for an explicit definition of a Bayesian PNM. Our framework generalizes the seminal work of [58] and builds on the modern and popular mathematical framework of [88]. This illuminates subtle distinctions among existing methods and clarifies the sense in which a non-Bayesian PNM can be considered to approximate a Bayesian PNM.

Second, we establish when PNMs are well defined outside of the conjugate Gaussian context. For exploration of nonlinear, non-Gaussian models, a numerical approximation scheme is developed and shown to asymptotically approach the posterior distribution of interest. Our aim here is not to develop new or more computationally efficient PNMs, but to understand when such a development can be well defined.

Third, we establish that Bayesian PNMs can be meaningfully composed. This is a central motivation for probabilistic numerics; in isolation, the error of a numerical method can often be studied and understood but, when composed into a pipeline, analysis of the resulting error structure becomes more difficult. The real power of probabilistic numerics lies in its application to pipelines of numerical methods, where the probabilistic formulation permits analysis of variance (ANOVA) to understand the contribution of each discretization to the overall uncertainty in the output. This paper introduces conditions under which a composition of PNMs can be considered to provide meaningful output, so that ANOVA can be justified.

Structure of the Paper. In section 2 we argue for an explicit definition of Bayesian PNMs and establish when such methods are well defined. Section 3 establishes connections to other related fields, in particular in relation to evaluating the performance of PNMs. In section 4 we develop useful numerical approximations to the output of Bayesian PNMs. Section 5 develops the theory of composition for Bayesian PNMs. (This is somewhat technical and readers unfamiliar with graphical models may wish to omit this section on first reading.) Finally, section 6 presents applications of the techniques discussed in this paper. All proofs can be found in the supplementary material.

2. Probabilistic Numerical Methods.

The aim of this section is to provide rigorous mathematical and statistical foundations for PNMs.

2.1. Notation.

For a measurable space $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $\mathcal{P}_{\mathcal{X}}$ denotes the set of all probability distributions on $(\mathcal{X}, \Sigma_{\mathcal{X}})$. For $\mu, \mu' \in \mathcal{P}_{\mathcal{X}}$ we write $\mu \ll \mu'$ when μ is absolutely continuous with respect to μ'. The unit Dirac measure at $x \in \mathcal{X}$ will be denoted $\delta(x) \in \mathcal{P}_{\mathcal{X}}$. Let $1[S]$ denote the indicator function of an event $S \in \Sigma_{\mathcal{X}}$. For a measurable function $f\colon \mathcal{X} \to \mathbb{R}$ and $\mu \in \mathcal{P}_{\mathcal{X}}$, we will write $\mu(f) \coloneqq \int f(x)\,\mu(\mathrm{d}x)$ and $\|f\|_\infty \coloneqq \sup_{x\in\mathcal{X}} |f(x)|$. The pointwise product of two functions f and g is denoted

$f \cdot g$. For a function T, $T_\#$ denotes the associated push-forward operator[1] that acts on measures on the domain of T. Let $\perp\!\!\!\perp$ denote conditional independence. The subset $\ell^p \subset \mathbb{R}^\infty$ consists of those sequences (u_i) for which $\sum_{i=1}^\infty |u_i|^p < \infty$. The set of continuous functions on $(0,1)$ will be denoted $C(0,1)$.

2.2. Definition of a PNM. To build intuition, first consider numerical approximation of the Lebesgue integral $\int x(t)\,\nu(\mathrm{d}t)$ for some integrable function $x\colon D \to \mathbb{R}$, with respect to a measure ν on D. Here we may directly interrogate the integrand $x(t)$ at any $t \in D$, but unless D is finite we cannot evaluate x at all $t \in D$ with a finite computational budget. The role of a numerical method here is to approximate such integrals based on finite information $\{x(t_i)\}_{i=1}^n$ at some collection of locations $\{t_i\}_{i=1}^n$ that can be specified.

To see the abstract structure of this problem, assume that x exists in a measurable space $(\mathcal{X}, \Sigma_\mathcal{X})$. Information about x is provided through an *information operator* $A\colon \mathcal{X} \to \mathcal{A}$ whose codomain is a measurable space $(\mathcal{A}, \Sigma_\mathcal{A})$. Thus, for the Lebesgue integration problem, the information operator is

$$A(x) = [x(t_1), \ldots, x(t_n)]^\top = a \in \mathcal{A}. \tag{2.1}$$

The space $\mathcal{X}$, in this case a space of functions, can be high- or infinite-dimensional, but the space $\mathcal{A}$ of information is assumed to be finite-dimensional in accordance with our finite computational budget. In this paper we make explicit a QoI $Q(x)$, defined by a map $Q\colon \mathcal{X} \to \mathcal{Q}$ into a measurable space $(\mathcal{Q}, \Sigma_\mathcal{Q})$. This captures the idea that x itself may not be the object of interest for the numerical problem; for the Lebesgue integration illustration, the QoI is $Q(x) = \int x(t)\,\nu(\mathrm{d}t)$.

The standard approach to such computational problems is to construct an algorithm which, when applied, produces some approximation $\hat{q}(a)$ of $Q(x)$, based on the information $a \in \mathcal{A}$, whose theoretical convergence can be studied. A successful algorithm will often tailor the information operator A to the QoI Q. For example, classical Gaussian cubature specifies *sigma points* $\{t_i\}_{i=1}^n$ at which the integrand is to be evaluated, based on exact integration of polynomial test functions of degree $2n-1$. The probabilistic numerical approach instead begins with the introduction of a random variable $X\colon \Omega \to \mathcal{X}$, formally a $\Sigma_\mathcal{X}$-measurable function on a probability space Ω that will henceforth be left implicit. The true state $x \in \mathcal{X}$ is fixed but unknown; the random variable X is an abstract device used to represent epistemic uncertainty about x before the information operator has been evaluated [39]. This is now formalized.

Definition 2.1 (belief distribution). *An element $\mu \in \mathcal{P}_\mathcal{X}$ is a* belief distribution[2] *for x if it carries the formal semantics of belief about the true, unknown state variable x.*

Thus we may consider μ to be the law of X. The construction of an appropriate belief distribution μ for a specific numerical task is not the focus of this research and has been considered in detail in previous work; see section SM2 in the Supplement for an overview of this material. Rather we consider the problem of how one *updates* the

[1] Recall that, for measurable $T\colon \mathcal{X} \to \mathcal{A}$, the push-forward $T_\#\mu$ of a distribution $\mu \in \mathcal{P}_\mathcal{X}$ is defined as $T_\#\mu(A) = \mu(T^{-1}(A))$ for all $A \in \Sigma_\mathcal{A}$.

[2] Two remarks are in order: First, we have avoided the use of "prior" as this abstract framework encompasses both Bayesian and non-Bayesian PNMs (to be defined). Second, the use of "belief" differs from the set-valued *belief functions* in Dempster–Shafer theory, which do not require that $\mu(E) + \mu(E^\mathrm{c}) = 1$ [85].

belief distribution μ in response to the information $A(x) = a$. Generic approaches to updating belief distributions, which generalize Bayesian inference beyond the unique update demanded by Bayes' theorem, have been formalized in [6, 19].

DEFINITION 2.2 (probabilistic numerical method). *Let $(\mathcal{X}, \Sigma_{\mathcal{X}})$, $(\mathcal{A}, \Sigma_{\mathcal{A}})$, and $(\mathcal{Q}, \Sigma_{\mathcal{Q}})$ be measurable spaces and let $A\colon \mathcal{X} \to \mathcal{A}$, $Q\colon \mathcal{X} \to \mathcal{Q}$, and $B\colon \mathcal{P}_{\mathcal{X}} \times \mathcal{A} \to \mathcal{P}_{\mathcal{Q}}$, where A and Q are measurable functions. The pair $M = (A, B)$ is called a* probabilistic numerical method *for estimation of a QoI Q. The map A is called an* information operator*, and the map B is called a* belief update operator*.*

The output of a PNM is a distribution $B(\mu, a) \in \mathcal{P}_{\mathcal{Q}}$. This is assigned the formal status of a belief distribution for the QoI $Q(x)$, based on both the initial belief μ about the state x and the information a that are input to the PNM. An objection sometimes raised to this construction is that x itself is not random. This work does not propose that x should be considered as such; the random variable X is distinguished in the notation as a formal statistical device used to represent epistemic uncertainty [42, 65]. Thus, there is no distinction from traditional statistics, in which x represents a fixed but unknown parameter and X encodes epistemic uncertainty about this parameter.

To strengthen intuition we now give specific examples of established PNMs.

EXAMPLE 2.3 (probabilistic integration). *Consider the Lebesgue integration problem discussed earlier. Take $D \subseteq \mathbb{R}^d$, $\mathcal{X}$ a separable Banach space of real-valued functions on D, and $\Sigma_{\mathcal{X}}$ the Borel σ-algebra for $\mathcal{X}$. The space $(\mathcal{X}, \Sigma_{\mathcal{X}})$ is endowed with a Gaussian belief distribution $\mu \in \mathcal{P}_{\mathcal{X}}$. Given information $A(x) = a$, define μ^a to be the restriction of μ to those functions that interpolate x at the points $\{t_i\}_{i=1}^n$; that μ^a is again Gaussian follows from linearity of the information operator (see* [7] *for details). The QoI remains $Q(x) = \int x(t)\, \nu(\mathrm{d}t)$.*

This problem was first considered in [58]*. The belief update operator proposed therein, and later considered in* [22, 73] *and others, was $B(\mu, a) = Q_{\#}\mu^a$. Since Gaussians are closed under linear projection, the PNM output $B(\mu, a)$ is a univariate Gaussian. Specifically, if μ has mean function $m\colon \mathcal{X} \to \mathbb{R}$ and covariance function $k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, then*

$$B(\mu, a) = N(z^\top \mathrm{K}^{-1}(a - \bar{m}), z_0 - z^\top \mathrm{K}^{-1} z),$$

where $\bar{m}, z \in \mathbb{R}^n$ are defined as $\bar{m}_i = m(t_i)$, $z_i = \int k(t, t_i)\, \nu(\mathrm{d}t)$, $\mathrm{K} \in \mathbb{R}^{n \times n}$ is defined as $\mathrm{K}_{i,j} = k(t_i, t_j)$, and $z_0 = \int k(t, t')\, (\nu \times \nu)(\mathrm{d}(t, t')) \in \mathbb{R}$. This method was extensively studied in [9]*, with a table of ν and k for which z and z_0 possess a closed form provided therein.*

Note that the information operator proposed above conceptualizes the integrand as a black box. This may seem unnatural in situations where an explicit formula for the integrand is provided. However, the integrand is treated in the same manner in both the design and the analysis of classical cubature methods, so this part of our framework is relatively standard [22]*. Furthermore, the mean of $B(\mu, a)$ can be made to coincide with classical cubature rules for different choices of μ and A* [82, 47, 46]*. The crucial distinction between PNMs and classical numerical methods is the* distributional *nature of $B(\mu, a)$, which carries the formal semantics of belief about the QoI. The full distribution $B(\mu, a)$ was examined in* [9]*, which established contraction to the exact value of the integral under smoothness conditions on the covariance function and on the integrand. See also* [45, 48]*.*

To limit scope we have focused on the situation where the information operator A is prespecified. However, Monte Carlo methods are constructed by sampling the points

$\{t_i\}_{i=1}^n$ from the underlying distribution ν, rather than specifying a fixed point set. In the present framework this corresponds to randomly selecting an information operator from a set of such operators, where the randomness arises from the randomly sampled evaluation locations.

EXAMPLE 2.4 (probabilistic meshless method). *As a canonical example of a PDE, take the following elliptic problem with Dirichlet boundary conditions:*

$$\begin{aligned} -\nabla \cdot (\kappa \nabla x) &= f && \text{in } D, \\ x &= g && \text{on } \partial D, \end{aligned} \tag{2.2}$$

where we assume $D \subset \mathbb{R}^d$ and $\kappa \colon D \to \mathbb{R}^{d \times d}$ is a known coefficient. Let $\mathcal{X}$ be a separable Banach space of appropriately differentiable real-valued functions and take $\Sigma_{\mathcal{X}}$ to be the Borel σ-algebra for $\mathcal{X}$. In contrast to the first example, the QoI here is $Q(x) = x$, as the goal is to make inferences about the solution of the PDE itself.

Such problems were considered in [16]*, wherein μ was restricted to be a Gaussian distribution on $(\mathcal{X}, \Sigma_{\mathcal{X}})$ with mean function $m = 0$ and covariance function $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. The information operator was constructed by choosing finite sets of locations $T_1 = \{t_{1,1}, \ldots, t_{1,n_1}\} \subset D$ and $T_2 = \{t_{2,1}, \ldots, t_{2,n_2}\} \subset \partial D$ at which the system* (2.2) *was evaluated, so that*

$$A(x) = \begin{bmatrix} -\nabla \cdot (\kappa(t_{1,1}) \nabla x(t_{1,1})) \\ \vdots \\ -\nabla \cdot (\kappa(t_{1,n_1}) \nabla x(t_{1,n_1})) \\ x(t_{2,1}) \\ \vdots \\ x(t_{2,n_2}) \end{bmatrix}, \qquad a = \begin{bmatrix} f(t_{1,1}) \\ \vdots \\ f(t_{1,n_1}) \\ g(t_{2,1}) \\ \vdots \\ g(t_{2,n_2}) \end{bmatrix}.$$

The belief update operator was chosen to be $B(\mu, a) = \mu^a$, where μ^a is the restriction of μ to those functions for which $A(x) = a$ is satisfied. In the setting of a linear system of PDEs such as that in (2.2)*, the distribution $B(\mu, a)$ is again Gaussian* [7]*. Full details are provided in* [16].

*As in the previous example, the mean of $B(\mu, a)$ coincides with the numerical solution to the PDE provided by a classical method (*symmetric collocation*;* [30]*). The distribution $B(\mu, a)$ provides uncertainty quantification for the exact solution and can again be shown to contract to the exact solution under smoothness conditions* [16]*. Indeed, the standard deviation of $B(\mu, a)$ is closely related to a standard error bound which can be derived for that method, as shown in Proposition* 4.1 *of* [16]*. Specifically it was shown that $|x(t) - \hat{x}(t)| \leq \sigma(t) \|x\|_k$, where $\hat{x}(t)$ is the mean and $\sigma(t)$ the standard deviation of $X(t) \sim B(\mu, a)$, while $\|\cdot\|_k$ is the norm of the reproducing kernel Hilbert space whose kernel is k. This method was analyzed in detail for a specific choice of belief distribution μ, in an impressive contribution from* [75].

In this example the standard method for approximately solving the PDE is to first discretize the system using, for example, a finite-element method. For a sufficiently fine discretization the error in the discrete approximation is negligible, and the dominant error is associated with solving the linear system that arises from the discretization. Recently PNMs have been developed for solving linear systems (e.g., [15] *or* [38]*), and when applied in this setting such methods could provide a viable approach to quantifying uncertainty in the solution of the discretized PDE.*

2.2.1. Classical Numerical Methods. Many standard numerical methods fit into the above framework, as can be seen by taking

$$B(\mu, a) = \delta \circ b(a) \tag{2.3}$$

independent of the distribution μ. Here the function $b\colon \mathcal{A} \to \mathcal{Q}$ is a classical numerical method for solving the problem of interest, such as the symmetric collocation method in Example 2.4. The function $\delta\colon \mathcal{Q} \to \mathcal{P}_{\mathcal{Q}}$ maps $b(a) \in \mathcal{Q}$ to a Dirac measure centered on $b(a)$, converting the output of this numerical method to a probability measure in our framework.

Note that not *all* numerical methods are particular instances of this framework. In particular, adaptive numerical methods would correspond to a sequence of information operators based on a filtration; this setting is more general than that we consider in the present paper.

2.3. Bayesian PNMs. Having defined a PNM, we now state the central definition of this paper, that is, of a *Bayesian* PNM (BPNM). Define μ^a to be the conditional distribution of the random variable X, given the event $A(X) = a$. For now we assume that this can be defined without ambiguity and reserve a more technical treatment of conditional probabilities for section 2.5.

This work follows [58] and casts the problem of determining x in (2.1) as a problem of Bayesian inversion, a framework now popular in applied mathematics and uncertainty quantification [88]. In a standard Bayesian inverse problem the observed quantity a is typically corrupted with measurement error, which is described by a likelihood. Inferences are obtained, under mild assumptions, through a suitably general version of Bayes' theorem [88, section 2.2]. For a PNM, however, the information is *not* corrupted with measurement error. As a result, the support of the likelihood is a null set under the prior, making even rather general statements of Bayes' theorem ill-defined when the random variables involved are infinite-dimensional. This necessitates a new definition.

Definition 2.5 (Bayesian probabilistic numerical method). *A probabilistic numerical method $M = (A, B)$ is said to be* Bayesian[3] *for a QoI Q if, for all $\mu \in \mathcal{P}_{\mathcal{X}}$, the output*

$$B(\mu, a) = Q_{\#}\mu^a \quad \text{for } A_{\#}\mu\text{-almost all } a \in \mathcal{A}.$$

That is, a PNM is Bayesian if the output of the PNM is the push-forward of the conditional distribution μ^a through Q. This definition is familiar from the examples in section 2.2, which are both examples of BPNMs.

For BPNMs we adopt the traditional terminology in which μ is the *prior* for x and the output $Q_{\#}\mu^a$ is the *posterior* for $Q(x)$. Since the information a is noiseless, the distribution μ^a is the restriction of the prior to the set of all $x \in \mathcal{X}$ such that $A(x) = a$. The prior thus determines the extent of posterior uncertainty, so it is crucial that the prior is properly calibrated. The prior is necessarily specific to the problem being addressed and should, minimally, encode basic mathematical properties of the state x that are known from numerical analysis, such as smoothness and any symmetries that may be present. It should also be calibrated so that the truth is expected to lie

[3] The use of "Bayesian" contrasts with [6], in which all belief update operators represent Bayesian learning algorithms to some greater or lesser extent. An alternative term could be "lossless," since all the information in a is conditioned upon to produce the posterior μ^a.

in a high probability region, a calibration that might in part be based upon heuristics and expert judgment. Selection of the prior is discussed further in section SM2 of the supplement.

Note that, for fixed A and μ, the Bayesian choice of belief update operator B (if it exists) is uniquely defined. It is emphasized that the class of BPNMs is a subclass of all PNMs. A literature survey of both Bayesian and non-Bayesian PNMs is provided in section SM3 of the supplement. Our analysis is focused on BPNMs due to their appealing philosophical interpretation and ease of generalization to pipelines of computation in section 5. For non-Bayesian PNMs, careful definition and analysis of the belief update operator is necessary to enable proper interpretation of the uncertainty quantification being provided. In particular, the analysis of non-Bayesian PNMs may present considerable challenges in the context of computational pipelines, whereas for BPNMs this is shown in section 5 to be straightforward.

Notably, in both Examples 2.3 and 2.4, certain choices of prior and information yield posteriors whose modes coincide with a classical numerical method. This is due to a deep connection between BPNMs and Bayesian decision rules which will be discussed in more detail in section 3.3. However, correspondence with classical methods should not be considered a fundamental design criterion for BPNMs and there exist BPNMs for which this is not the case (e.g., [15, 93]). Equally, BPNMs have been proposed to tackle problems that have not been well studied in numerical analysis (e.g., [96]).

2.4. Model Evidence. A cornerstone of the Bayesian framework is the model evidence, or marginal likelihood [66]. Let $\mathcal{A} \subseteq \mathbb{R}^n$ be equipped with the Lebesgue reference measure λ, such that $A_{\#}\mu$ admits a density $p_A = \mathrm{d}A_{\#}\mu/\mathrm{d}\lambda$. Then the *model evidence* $p_A(a)$, based on the information that $A(x) = a$, can be used as the basis for Bayesian model comparison. Two prior distributions μ and $\tilde{\mu}$ can be compared through the Bayes factor

$$\mathrm{BF} \coloneqq \frac{\tilde{p}_A(a)}{p_A(a)} = \frac{\mathrm{d}A_{\#}\tilde{\mu}}{\mathrm{d}A_{\#}\mu}(a), \tag{2.4}$$

where $\tilde{p}_A = \mathrm{d}A_{\#}\tilde{\mu}/\mathrm{d}\lambda$. A large Bayes factor means that the prior $\tilde{\mu}$ assigns significantly more probability to the observed data a than μ, and thus should be preferred. The second expression is independent of the reference measure λ and so is valid for more general $\mathcal{A}$.

The model evidence has been explored in connection with the design of BPNMs. For the integration and linear PDE examples 2.3 and 2.4, the model evidence has a closed form and was investigated in [9, 16]. Section 6 considers the model evidence in the context of nonlinear ODEs and PDEs.

2.5. The Disintegration Theorem. The purpose of this section is to formalize μ^a and to determine conditions under which it exists and is well defined. From Definition 2.5, the output of a BPNM is $B(\mu, a) = Q_{\#}\mu^a$. If μ^a exists, the push-forward $Q_{\#}\mu^a$ exists as Q is assumed to be measurable; thus, in this section, we focus on the rigorous definition of μ^a.

Unlike many problems of Bayesian inversion, proceeding by an analogue of Bayes' theorem is not possible. Let $\mathcal{X}^a = \{x \in \mathcal{X} : A(x) = a\}$. Then we observe that, if it is measurable, $\mathcal{X}^a$ is often a set of zero measure under μ. Standard techniques for infinite-dimensional generalization of Bayes' theorem rely on constructing a posterior distribution based on its Radon–Nikodym derivative with respect to the prior [88].

However, when $\mu^a \not\ll \mu$ no Radon–Nikodym derivative exists and we must turn to other approaches to ensure a BPNM is well-defined.

Conditioning on null sets was first formalized in the celebrated construction of measure-theoretic probability by [53]. The central challenge is to establish uniqueness of conditional probabilities. For this work we exploit the *disintegration theorem* to ensure our constructions are well-defined. The definition below is due to [21, p.78], and a statistical introduction to disintegration can be found in [13].

DEFINITION 2.6 (disintegration). *For $\mu \in \mathcal{P}_{\mathcal{X}}$, a collection $\{\mu^a\}_{a\in\mathcal{A}} \subset \mathcal{P}_{\mathcal{X}}$ is a* disintegration *of μ with respect to the (measurable) map $A\colon \mathcal{X} \to \mathcal{A}$ if:*

1. *(Concentration:) $\mu^a(\mathcal{X} \setminus \mathcal{X}^a) = 0$ for $A_{\#}\mu$-almost all $a \in \mathcal{A}$;*

and, for each measurable $f\colon \mathcal{X} \to [0,\infty)$,

2. *(Measurability:) $a \mapsto \mu^a(f)$ is measurable;*
3. *(Conditioning:) $\mu(f) = \int \mu^a(f)\, A_{\#}\mu(\mathrm{d}a)$.*

The concept of disintegration extends the usual concept of conditioning of random variables to the case where $\mathcal{X}^a$ is a null set, in a way closely related to regular conditional distributions [53]. Existence is guaranteed under quite weak conditions:

THEOREM 2.7 (disintegration theorem [13, Theorem 1]). *Let $\mathcal{X}$ be a metric space, let $\Sigma_{\mathcal{X}}$ be its Borel σ-algebra, and consider a Radon distribution $\mu \in \mathcal{P}_{\mathcal{X}}$. Let $\Sigma_{\mathcal{A}}$ be countably generated and contain the singletons $\{a\}$ for each $a \in \mathcal{A}$. Then there exists a disintegration $\{\mu^a\}_{a\in\mathcal{A}}$ of μ with respect to A. Moreover, if $\{\nu^a\}_{a\in\mathcal{A}}$ is another such disintegration, then $\{a \in \mathcal{A} : \mu^a \neq \nu^a\}$ is a $A_{\#}\mu$-null set.*

The requirement that μ is Radon always holds if $\mathcal{X}$ is a separable complete metric space or, more generally, a *Radon space*. The requirement that $\Sigma_{\mathcal{A}}$ is countably generated includes the standard case where $\mathcal{A} = \mathbb{R}^n$ with the Borel σ-algebra. Theorem 2.7 implies that $\{\mu^a\}_{a\in\mathcal{A}}$ exists and is essentially unique for all of the examples considered in this paper. Thus, under mild conditions, we have established that BPNMs are well-defined, in that an essentially unique disintegration $\{\mu^a\}_{a\in\mathcal{A}}$ exists. Variational definition of μ^a has been offered as an alternative approach for when the existence of a disintegration is difficult to establish [33, p. 3].

2.6. Prior Construction. The Gaussian distribution is popular as a prior in the PNM literature for its tractability, in the senses that its finite-dimensional distributions take a closed form and that an explicit conditioning formula exists. More general priors, such as Besov priors [18, 61] and Cauchy priors [89], are less easily accessed. In this section we summarize a common construction for these prior distributions, designed to ensure that a disintegration will exist.

Let $\{\phi_i\}_{i=0}^{\infty}$ denote a normalized Schauder basis for $\mathcal{X}$, assumed to be a separable Banach space in this section. Any $x \in \mathcal{X}$ can be represented through an expansion

$$x = x_0 + \sum_{i=0}^{\infty} u_i \phi_i \tag{2.5}$$

for some fixed element $x_0 \in \mathcal{X}$ and a sequence $u \in \mathbb{R}^{\infty}$. Construction of measures $\mu \in \mathcal{P}_{\mathcal{X}}$ is then reduced to construction of almost-surely convergent measures on $\mathbb{R}^{\infty}$ and studying the push-forward of such measures into $\mathcal{X}$. In particular, this will ensure that $\mu \in \mathcal{P}_{\mathcal{X}}$ is Radon as $\mathcal{X}$ is a separable complete metric space, a key requirement for the existence of a disintegration $\{\mu^a\}_{a\in\mathcal{A}}$.

To this end it is common to split the coefficients u_i into stochastic and deterministic components; let $\xi \in \mathbb{R}^{\infty}$ represent an i.i.d. sequence of random variables, and

$\gamma \in \ell^p$ for some $p \in [1, \infty)$. Then with $u_i = \gamma_i \xi_i$, we require that $u \in \ell^1$ almost-surely for the prior distribution to be well-posed. Different choices of (ξ, γ) give rise to different distributions on $\mathcal{X}$. For instance, $\xi_i \sim \text{Uniform}(-1, 1)$, $\gamma \in \ell^1$ is termed a *uniform* prior, and $\xi_i \sim \mathcal{N}(0, 1)$, $\gamma \in \ell^2$ gives a *Gaussian* prior, where γ determines the regularity of the covariance operator [7]. The choice of $\xi_i \sim \text{Cauchy}(0, 1)$ gives a *Cauchy* prior; here we require $\gamma \in \ell^1 \cap \ell \log \ell$ for $\mathcal{X}$ a separable Banach space, or $\gamma \in \ell^2$ for when $\mathcal{X}$ is a Hilbert space; see [89].

A range of prior specifications will be explored in section 6, including non-Gaussian prior distributions for the numerical solution of nonlinear ODEs.

2.7. Relation to Emulation. The characterization of a numerical task as a Bayesian inverse problem was mentioned in section 2.3. In this section emulation [81], another method bearing some resemblance to PNMs, will be discussed.

An *emulator* is a regression model that is used to predict black box output, based on a limited input-output training dataset, at input locations other than those at which it was trained. In this specific sense, several BPNMs can be regarded as being based on emulators:

- The PNM proposed in Example 2.3 is mathematically equivalent to a Gaussian process emulator of the integrand, with the uncertainty being pushed forward onto the value of the integral.
- The PNM proposed in Example 2.4 is mathematically equivalent to a Gaussian process emulator of the gradient field f and boundary field g of the PDE, with the uncertainty in these fields being pushed forward onto the solution space of the PDE.

However, the output of a BPNM differs from the traditional sense in which emulation is used and understood. Indeed, due to disintegration along the information operator, the output of a BPNM inherits structure and characteristics of the mathematical problem being solved. In a PDE-constrained inverse problem, for example, emulators are traditionally used to interpolate or extrapolate in parameter space, whereas BPNMs can be used to quantify the error in the forward problem that is due to discretization of the PDE.

3. Decision-Theoretic Treatment. Next we assess the performance of PNMs from a decision-theoretic perspective [5] and explore connections to average-case analysis of classical numerical methods [79]. Note that the treatment here is agnostic to whether or not the PNM in question is Bayesian, and also encompasses classical numerical methods. Throughout, the existence of a disintegration $\{\mu^a\}_{a \in \mathcal{A}}$ will be assumed.

3.1. Loss and Risk. Consider a generic loss function $L \colon \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ where $L(q^\dagger, q)$ describes the loss incurred when the true QoI $q^\dagger = Q(x)$ is estimated with $q \in \mathcal{Q}$. Integrability of L is assumed. The belief update operator B returns a distribution over $\mathcal{Q}$ which can be cast as a randomized decision rule for estimation of $q^\dagger$. For randomized decision rules, the *risk function* $r \colon \mathcal{Q} \times \mathcal{P}_\mathcal{Q} \to \mathbb{R}$ is defined as

$$r(q^\dagger, \nu) = \int L(q^\dagger, q)\, \nu(\mathrm{d}q) \,.$$

The *average risk* of the PNM $M = (A, B)$ with respect to $\mu \in \mathcal{P}_\mathcal{X}$ is defined as

$$R(\mu, M) = \int r(Q(x), B(\mu, A(x)))\, \mu(\mathrm{d}x). \tag{3.1}$$

Here a state $x \sim \mu$ is drawn at random and the risk of the PNM output $B(\mu, A(x))$ is computed; the procedure is then averaged. We follow the convention of terming $R(\mu, M)$ the *Bayes risk* of the PNM, though the usual objection that a frequentist expectation enters into the definition of the Bayes risk could be raised.

Next, we consider a sequence $A^{(n)}$ of information operators indexed such that $A^{(n)}(x)$ is n-dimensional (i.e., n pieces of information are provided about x).

DEFINITION 3.1 (contraction). *A sequence $M^{(n)} = (A^{(n)}, B^{(n)})$ of PNMs is said to* contract *at a* rate r_n *under a belief distribution μ if $R(\mu, M^{(n)}) = O(r_n)$.*

This definition allows for the comparison of both classical and probabilistic numerical methods [43, 22]. In each case an important goal is to determine methods $M^{(n)}$ that contract as quickly as possible for a given distribution μ that defines the Bayes risk. This is the approach taken in average-case analysis [79] and will be discussed in section 3.3. For Examples 2.3 and 2.4 of BPNMs, [9] and [16] established rates of contraction for particular prior distributions μ; we refer the reader to those papers for more detail.

3.2. Bayes Decision Rules. A (possibly randomized) decision rule is said to be a *Bayes rule* if it achieves the minimum Bayes risk among all decision rules. In the context of PNMs, let

$$\mathfrak{B}(A) = \arg\inf_B R(\mu, (A, B)).$$

That is, for fixed A, $\mathfrak{B}(A)$ is the set of all belief update operators that achieve minimum Bayes risk.

This raises the natural question of which belief update operators yield Bayes rules. Although the definition of a Bayes rule applies generically to both probabilistic and deterministic numerical methods, it can be shown[4] that, if $\mathfrak{B}(A)$ is nonempty, there exists a $B \in \mathfrak{B}(A)$ which takes the form of a classical numerical method, as expressed in (2.3). Thus, in general, *BPNMs do not constitute Bayes rules*, as the extra uncertainty inflates the Bayes risk. Nonetheless, there is a natural connection between BPNMs and Bayes rules, as exposed in [43].

THEOREM 3.2. *Let $M = (A, B)$ be a BPNM for the QoI Q. Let $(\mathcal{Q}, \langle\cdot,\cdot\rangle_{\mathcal{Q}})$ be an inner product space and let the loss function L have the form $L(q^\dagger, q) = \|q^\dagger - q\|_{\mathcal{Q}}^2$, where $\|\cdot\|_{\mathcal{Q}}$ is the norm induced by the inner product. Then the mean of the distribution $B(\mu, a)$ is a Bayes rule for estimation of $q^\dagger$.*

This well-known fact from Bayesian decision theory[5] is interesting in light of recent research in constructing PNMs whose mean functions correspond to classical numerical methods [83, 38, 82, 90, 84]. Theorem 3.2 explains the results in Examples 2.3 and 2.4, in which both instances of BPNMs were demonstrated to be centered on an established classical method.

3.2.1. Optimal Information. The previous section considered selection of the belief update operator B, but not the information operator A. The choice of A determines the Bayes risk for a PNM, which leads to a problem of experimental design to minimize that risk. The theoretical study of optimal information for classical methods is the focus of the information-based complexity literature [70, 92, 94]. Here we characterize optimal information for BPNMs.

[4]The proof is included in supplementary section SM4.1.

[5]This is the fact that "the Bayes act is the posterior mean under squared-error loss" [5].

Consider the choice of A from a fixed set Λ of candidate information operators. For example, for the task of numerical integration, Λ could represent all possible choices of locations $\{t_i\}_{i=1}^n$ where the integrand is to be evaluated. For BPNMs, one can ask for optimal information,

$$A_\mu \in \operatorname*{arg\,inf}_{A\in\Lambda} \{R(\mu, M) \ : \ M = (A, B),\ B(\mu, a) = Q_{\#}\mu^a\},$$

where we have made explicit the fact that the optimal information depends on the choice of prior μ.

3.3. Conditional Connection to Average-Case Analysis. The decision theoretic framework in section 3.1 is somewhat related to the average-case analysis (ACA) of classical numerical methods [79]. In ACA the performance of a classical numerical method $b\colon \mathcal{A} \to \mathcal{Q}$ is studied in terms of the Bayes risk $R(\mu, M)$ given in (3.1), for the PNM $M = (A, B)$ with belief operator $B(\mu, a) = \delta \circ b(a)$ as in (2.3). ACA is concerned with the study of optimal information:

$$A^*_\mu \in \operatorname*{arg\,inf}_{A\in\Lambda} \left\{ \inf_b R(\mu, M) \ : \ M = (A, B),\ B(\mu, a) = \delta \circ b(a) \right\}.$$

In general there is no reason to expect A_μ and A^*_μ to coincide, since BPNMs are not Bayes rules.[6] Indeed, an explicit example where $A_\mu \neq A^*_\mu$ is presented in section SM4.2 of the supplement. However, we can establish sufficient conditions under which optimal information for a BPNM is the same as optimal information according to ACA.

THEOREM 3.3. *Let $(\mathcal{Q}, \langle\cdot,\cdot\rangle_{\mathcal{Q}})$ be an inner product space and the loss function L have the form $L(q^\dagger, q) = \|q^\dagger - q\|^2_{\mathcal{Q}}$, where $\|\cdot\|_{\mathcal{Q}}$ is the norm induced by the inner product. Then the optimal information A_μ for a BPNM and A^*_μ for ACA are identical.*

It is emphasized that this result is *not* a trivial consequence of the correspondence between Bayes rules and worst-case optimal methods exposed in [43], as evidenced by the counterexample in section SM4.2. To the best of our knowledge, information-based complexity research has studied A^*_μ but not A_μ. Theorem 3.3 establishes that, for the squared-norm loss, we can use results on optimal information from the ACA literature to construct optimal BPNMs. An example where optimal information for integration is derived based on Theorem 3.3 is included in section SM4.3.

This completes our performance assessment for PNMs; next we turn to computational matters.

4. Numerical Disintegration. In this section we describe a numerical method to access the output from a BPNM. The approach that we describe is based on constructing an approximation to μ^a that can be sampled using potentially sophisticated Monte Carlo methods.

Constructive approximation of μ^a is not trivial. The approach considered in this work is based on sampling from an approximate distribution μ^a_δ which converges in an appropriate sense to μ^a as $\delta \to 0$; this is in a similar spirit to the approach of [1]. We note, however, that because Monte Carlo sampling incurs a high computational cost, the method introduced in this section should not be considered a practical approach in general. Rather, it is presented to demonstrate that numerical methods for approximating the output of BPNMs *can* be constructed, and to provide a benchmark for future, more expedient numerical methods that may be developed.

[6] The distribution $Q_{\#}\mu^a$ will in general not be supported on the set of Bayes acts.

4.1. Sequential Approximation of a Disintegration. Suppose that $\mathcal{A} = \mathbb{R}^n$ and endow $\mathcal{A}$ with the structure of a Hilbert space. Let $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ denote a decreasing function that is continuous at 0, with $\phi(0) = 1$ and $\lim_{r\to\infty} \phi(r) = 0$. Consider, for $\delta > 0$,

$$\mu_\delta^a(\mathrm{d}x) \coloneqq \frac{1}{Z_\delta^a}\phi\left(\frac{\|A(x) - a\|_{\mathcal{A}}}{\delta}\right)\mu(\mathrm{d}x), \qquad Z_\delta^a \coloneqq \int_{\mathcal{A}} \phi\left(\frac{\|\tilde{a} - a\|_{\mathcal{A}}}{\delta}\right) p_A(\tilde{a})\,\mathrm{d}\tilde{a},$$

where the normalization constant Z_δ^a is nonzero as p_A is bounded away from 0 on a neighborhood of $a \in \mathcal{A}$, and ϕ is bounded away from 0 on a sufficiently small interval $[0, \gamma]$. Our aim is to approximate μ^a with μ_δ^a for small bandwidth parameter δ. The construction, which generalizes approximate Bayesian computation [20] and is similar in spirit to [27], ensures that $\mu_\delta^a \ll \mu$. The role of ϕ is to admit states $x \in \mathcal{X}$ for which $A(x)$ is close to, but not necessarily equal to, a, and it is assumed to be sufficiently regular.

ASSUMPTION 4.1. *For each $\alpha > 0$ we have that $C_\phi^\alpha \coloneqq \int r^{\alpha+n-1}\phi(r)\,\mathrm{d}r < \infty$.*

We also require a regularity assumption on $A_{\#}\mu$.

ASSUMPTION 4.2. *The distribution $A_{\#}\mu \in \mathcal{P}_{\mathcal{A}}$ admits a Lipschitz density p_A with respect to Lebesgue measure on $\mathcal{A}$, with the property that $p_A(a) > 0$ for all $a \in \mathcal{A}$.*

To discuss the convergence of μ_δ^a to μ^a we must first specify a mode of convergence in $\mathcal{P}_{\mathcal{X}}$. Let $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ be a normed space of (measurable) functions $f\colon \mathcal{X} \to \mathbb{R}$. Let

$$d_{\mathcal{F}}(\nu, \nu') \coloneqq \sup_{\|f\|_{\mathcal{F}} \le 1} |\nu(f) - \nu'(f)| \quad \text{for } \nu, \nu' \in \mathcal{P}_{\mathcal{X}}.$$

Many common probability metrics such as the total variation and Wasserstein distances arise in this way [68]. However, not all spaces of functions $\mathcal{F}$ lead to useful theory. In particular, the total variation distance between μ^a and $\mu^{a'}$ for $a \neq a'$ will be 1 in general. Furthermore, depending on the choice of $\mathcal{F}$, $d_{\mathcal{F}}$ may not be a metric. Sufficient conditions for weak convergence with respect to $\mathcal{F}$ are now established.

ASSUMPTION 4.3. *The map $a \mapsto \mu^a$ is almost everywhere α-Hölder continuous in $d_{\mathcal{F}}$ for some $\alpha > 0$, i.e., $d_{\mathcal{F}}(\mu^a, \mu^{a'}) \le C_\mu^\alpha \|a - a'\|_{\mathcal{A}}^\alpha$ for some constant $C_\mu^\alpha > 0$ and for $A_{\#}\mu$-almost all $a, a' \in \mathcal{A}$.*

THEOREM 4.4. *Let $\bar{C}_\phi^\alpha \coloneqq C_\phi^\alpha / C_\phi^0$. Then, for each fixed a we have that, for $\delta > 0$ sufficiently small,*

$$d_{\mathcal{F}}(\mu_\delta^a, \mu^a) \le C_\mu^\alpha (1 + \bar{C}_\phi^\alpha)\delta^\alpha.$$

This result justifies the approximation of μ^a by μ_δ^a whenever the QoI can be well approximated using integrals with respect to $\mathcal{F}$. This result is stronger than that of earlier work, such as [77], in that it holds for infinite-dimensional $\mathcal{X}$, though it also relies upon the stronger Hölder continuity assumption. The work of [1] establishes convergence of similar approximating distributions to the disintegration, but the Hölder continuity assumption allows the distributional error to be bounded. The impact of ϕ on convergence rates is discussed in section SM5 of the supplementary material.

4.2. Monte Carlo Methods. The previous section established a sequence of well-defined distributions μ_δ^a that converge, in a weak sense, to the exact disintegration

μ^a. By construction, $\mu^a_\delta \ll \mu^a$, which is sufficient to allow standard Monte Carlo methods to be used. The construction of Monte Carlo methods is decoupled from the core material in the main text and the main methodological considerations are well documented [35].

For the experiments reported in subsequent sections two approaches were explored; a sequential Monte Carlo (SMC) method [26] and a parallel tempering method [34]. This provided a transparent sampling scheme, whose nonasymptotic approximation error can be theoretically understood. In particular, such tempering schemes provide robust estimators of model evidence that can be used for Bayesian model comparison. Full details of the Monte Carlo methods used, along with associated theoretical analysis for the SMC method, are contained in section SM5.1 of the supplementary material.

5. Computational Pipelines and PNMs. The last theoretical development in this paper concerns how multiple PNMs can be combined. These results represent an important contribution, as composition has been repeatedly cited as an important motivation for PNMs [39, 17, 14, 16], but how this ought to be accomplished has not yet been addressed. Nevertheless, this material is rather technical and the reader may wish to omit this section on first reading.

Contemporary applications typically rely on the composition of several numerical methods, but the question of when PNMs can be composed, and when the output of such a composition is meaningful, has not yet been addressed. It has been shown that accumulated discretization error can have a highly nontrivial impact on computational output [80, 3, 4], and that the richer, structured output provided by PNMs may help to better describe this impact. The authors of [39] also highlighted using composed PNMs for an analysis of variance, to determine the dominant sources of discretization error. If these applications are to be pursued, then it is vital that the output of composed PNMs is meaningful.

Composed PNMs have appeared before in the literature. In recent work, [14], [17], and [16] used PNMs within a broader statistical procedure to estimate parameters in systems of ODEs and PDEs. The probabilistic description of discretization error was incorporated into the data-likelihood, resulting in posterior distributions for parameters with inflated uncertainty to account for the inferential impact of discretization error. However, beyond these limited works, no examination of the composition of PNMs has been performed.

This section defines a *pipeline* as an abstract graphical object that may be combined with a collection of *compatible* PNMs. It is proved that when compatible BPNMs are employed in the pipeline, the distributional output of the pipeline carries a Bayesian interpretation under an explicit conditional independence condition on the prior μ.

To build intuition, for the simple case where two BPNMs are composed in series, our results provide conditions for when, informally, the output $B_2(B_1(\mu, a_1), a_2)$ corresponds to a single Bayesian procedure $B(\mu, (a_1, a_2))$. To reduce the notational and technical burden, in this section we will not provide rigorous measure theoretic details, but these details broadly follow the same pattern as in section 2.5.

5.1. Computational Pipelines. To analyze pipelines of PNMs, we consider n such methods $M_1, \ldots, M_n$, where each method $M_i = (A_i, B_i)$ is defined on a common[7]

[7]This is without loss of generality, since $\mathcal{X}$ can be taken as the union of all state spaces required by the individual methods.

state space $\mathcal{X}$ and targets a QoI $Q_i \in \mathcal{Q}_i$. A pipeline will be represented as a directed graphical model, wherein the QoIs Q_i from parent methods constitute information operators for child methods. It may be the case that a method will take quantities from multiple parents as input. To allow for this, we suppose that the information operator $A_i \colon \mathcal{X} \to \mathcal{A}_i$ can be decomposed into components $A_{i,j} \colon \mathcal{X} \to \mathcal{A}_{i,j}$ such that $A_i = (A_{i,1}, \ldots, A_{i,m(i)})$ and $\mathcal{A}_i = \mathcal{A}_{i,1} \times \cdots \times \mathcal{A}_{i,m(i)}$. Thus, each component $A_{i,j}$ can be thought of as the QoI output by one of the parents of the method M_i.

Without loss of generality we designate the nth QoI Q_n to be the *principal* QoI. That is, the purpose of the computational pipeline is to estimate $Q_n(x)$. The case of multiple principal QoI is a simple extension and will not be described. Nodes with no children are called *terminal* nodes, while nodes with no parents are called *source nodes*. We denote by A the set of all source nodes.

DEFINITION 5.1 (pipeline). *A* pipeline P *is a directed acyclic graph such that:*

- *Nodes are of two kinds:* Information *nodes are depicted by* $\square$*, and* method *nodes are depicted by* $\blacksquare$*.*
- *The graph is bipartite, so that edges connect a method node to an information node or vice versa; that is, edges are of the form* $\square \to \blacksquare$ *or* $\blacksquare \to \square$*.*
- *There are* n *method nodes, each with a unique label in* $\{1, \ldots, n\}$*.*
- *The method node labeled* i *has* $m(i)$ *parents and one child. Its in-edges are assigned a unique label in* $\{1, \ldots, m(i)\}$*.*
- *There is a unique terminal node and it is the child of method node* n*. This represents the principal QoI* Q_n*.*

EXAMPLE 5.2 (distributed integration). *As a thought experiment, consider partitioning the domain of integration for a simple one-dimensional integral:*

$$\underbrace{\int_0^1 x(t)\,\mathrm{d}t}_{\text{(c)}} = \underbrace{\int_0^{0.5} x(t)\,\mathrm{d}t}_{\text{(a)}} + \underbrace{\int_{0.5}^1 x(t)\,\mathrm{d}t}_{\text{(b)}}. \tag{5.1}$$

To keep the presentation simple we consider an integral over $[0,1]$ *with information obtained at* $2m+1$ *equidistant knots* $t_i = i/2m$*. Let* M_1 *be a BPNM for estimating* $Q_1(x) =$ *(a) and* M_2 *be a BPNM for estimating* $Q_2(x) =$ *(b).*

In terms of our notational convention, we divide the information operator into four components: $A_{i,j}$ *for* $i, j \in \{1,2\}$*.* $A_{1,1}$ *and* $A_{2,2}$ *contain the information unique to* M_1 *and* M_2*. Specifically,*

$$A_{1,1}(x) = \begin{bmatrix} x(t_1) \\ \vdots \\ x(t_{m-1}) \end{bmatrix}, \qquad A_{2,2}(x) = \begin{bmatrix} x(t_{m+1}) \\ \vdots \\ x(t_{2m}) \end{bmatrix}.$$

$A_{1,2}$ *and* $A_{2,1}$ *contain the information that is shared between the two methods; that is,* $A_{1,2} = A_{2,1} = \{x(t_m)\}$*. To complete the specification we need a third PNM for estimation of* $Q_3(x) =$ *(c), which we denote* M_3 *and which combines the outputs of* M_1 *and* M_2 *by simply adding them together. Formally this has information operator* $A_3(x) = (A_{3,1}(x), A_{3,2}(x))$*, where* $A_{3,1}(x) =$ *(a) and* $A_{3,2}(x) =$ *(b). Its belief update operator is given by* $B_3(\mu, (a_{3,1}, a_{3,2})) = \delta(a_{3,1} + a_{3,2})$*. Figure* 1 *shows an intuitive graphical representation of this set-up. The pipeline* P *itself, which is identical to Figure* 1 *but with specific node and edge labels, is shown in Figure* 2(a)*.*

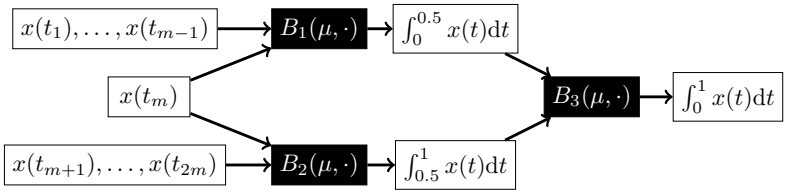
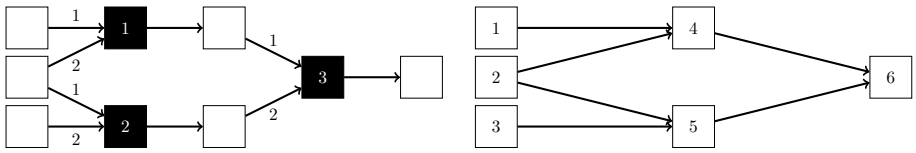


Fig. 1 *An intuitive representation of Example* 5.2.



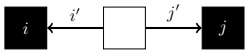
(a) Pipeline (b) Dependence Graph

Fig. 2 *Computation as a graphical model.* (a) *The pipeline P corresponding to Figure* 1. (b) *Dependence graph $G(P)$ corresponding to the pipeline P. The nodes are indexed with a topological ordering.*

In general, the method node labeled i is taken to represent the method M_i. The in-edge to this node labeled j is taken to represent the information provided by the relationship $A_{i,j}(x) = a_{i,j}$. Here $a_{i,j}$ can be either deterministic information provided to the pipeline, or statistical information derived from the output of another PNM. To make this formal and to "match the input-output spaces" we next define what it means for the collection of methods M_i to be compatible with the pipeline P. Informally, this describes the conditions that must be satisfied for method nodes in a pipeline to be able to connect to each other.

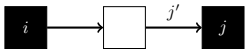
DEFINITION 5.3 (compatible). *The collection $(M_1, \ldots, M_n)$ of PNMs is* compatible *with the pipeline P if the following two requirements are satisfied:*

(i) *For a motif*



we have that $A_{i,i'} = A_{j,j'}$ and $\mathcal{A}_{i,i'} = \mathcal{A}_{j,j'}$.

(ii) *For a motif*



we have that $\mathcal{Q}_i = \mathcal{A}_{j,j'}$.

Note that we do not require the converse of (i) at this stage; that is, the same information can be represented by more than one node in the pipeline. This permits the existence of redundant nodes in the pipeline, meaning that information is not recycled. It will transpire that pipelines with such redundancy do not admit a Bayesian treatment.

The role of the pipeline P is to specify the order in which information, either deterministic or statistical, is propagated through the collection of PNMs. This is illustrated next.

EXAMPLE 5.4 (propagation of information). *For the pipeline in Figure* 2(a)*, the propagation of information proceeds as follows:*

1. *The source nodes, representing* $A(x) = \{A_{1,1}(x), A_{1,2}(x) = A_{2,1}(x), A_{2,2}(x)\}$*, are evaluated as* $\{a_{1,1}, a_{1,2} = a_{2,1}, a_{2,2}\}$*. This represents all the information on* x *that is provided.*
2. *The distributions* $\mu^{(1)} \coloneqq B_1(\mu, (a_{1,1}, a_{1,2}))$ *and* $\mu^{(2)} \coloneqq B_2(\mu, (a_{2,1}, a_{2,2}))$ *are computed.*
3. *The push-forward distribution* $\mu^{(3)} \coloneqq (B_3)_{\#}(\mu, \mu^{(1)} \times \mu^{(2)})$ *is computed.*

Here $\mu^{(1)} \times \mu^{(2)}$ *is defined on the Cartesian product* $\Sigma_{\mathcal{A}_{3,1}} \times \Sigma_{\mathcal{A}_{3,2}}$ *with independent components* $\mu^{(1)}$ *and* $\mu^{(2)}$*. The notation* $(B_3)_{\#}$ *refers to the push-forward of the function* $B_3(\mu, \cdot)$ *through its second argument. The distribution* $\mu^{(3)}$ *is the output of the pipeline and represents belief about the principal QoI* $Q_3(x)$*.*

The procedure in Example 5.4 can be formalized, but to keep the presentation and notation succinct, we leave this implicit.

DEFINITION 5.5 (computation). *For a collection* $(M_1, \ldots, M_n)$ *of PNMs that are compatible with a pipeline* P*, the* computation $P(M_1, \ldots, M_n)$ *is defined as the PNM with information operator* A *and belief update operator* B *that takes* μ *and* $A(x) = a$ *as input and returns the distribution* $\mu^{(n)}$ *as its output* $B(\mu, a)$*, obtained through the procedure outlined in Example* 5.4*.*

That is, the *computation* $P(M_1, \ldots, M_n)$ is itself a PNM for the principal QoI Q_n. Note that this definition includes a classical numerical work-flow just as a PNM encompasses a standard numerical method.

5.2. Bayesian Computational Pipelines. Since $P(M_1, \ldots, M_n)$ is itself a PNM, there is a natural definition for when such a computation can be called Bayesian.

DEFINITION 5.6 (Bayesian computation). *Denote by* (A, B) *the information and belief update operators associated with the computation* $P(M_1, \ldots, M_n)$ *and consider a disintegration* $\{\mu^a\}_{a \in \mathcal{A}}$ *of* μ *with respect to the information operator* A*. The computation* $P(M_1, \ldots, M_n)$ *is said to be* Bayesian *for the QoI* Q_n *if* $B(\mu, a) = (Q_n)_{\#}\mu^a$ *for* $A_{\#}\mu$*-almost all* $a \in \mathcal{A}$*.*

This is clearly an appealing property; the output of a Bayesian computation can be interpreted as a posterior distribution over the QoI $Q_n(x)$ given the prior μ and the information $A(x)$. Or, more informally, the "pipeline is lossless with information." However, at face value it seems difficult to verify whether or not a given computation $P(M_1, \ldots, M_n)$ is Bayesian, since it depends on both the individual PNMs M_i and the pipeline P through which they are combined. Our next task is to establish verifiable sufficient conditions, for which we require another definition.

DEFINITION 5.7 (dependence graph). *The* dependence graph *of a pipeline* P *is the directed acyclic graph* $G(P)$ *obtained by taking the pipeline* P*, removing the method nodes, and replacing all* $\square \to \blacksquare \to \square$ *motifs with direct edges* $\square \to \square$*.*

The dependency graph for Example 5.2 is shown in Figure 2(b).

For a computation $P(M_1, \ldots, M_n)$, each of the J distinct nodes in $G(P)$ can be associated with a random variable Y_j where either $Y_j = A_{k,l}(X)$ for some k, l, when the node is a source, or otherwise $Y_j = Q_k(X)$ for some k. Randomness here is understood to be due to $X \sim \mu$, so that the distribution of the $\{Y_j\}_{j=1}^J$ is a consequence of μ. The convention used here is that the Y_j are indexed according to a topological ordering on $G(P)$, which has the properties that (i) the source nodes correspond to indices $1, \ldots, I$, and (ii) the final random variable is $Y_J = Q_n(X)$.

DEFINITION 5.8 (coherence). *Consider a computation $P(M_1, \ldots, M_n)$. Denote by $\pi(j) \subseteq \{1, \ldots, j-1\}$ the parent set of node j in the dependence graph $G(P)$. Then we say that $\mu \in \mathcal{P}_{\mathcal{X}}$ is* coherent *for the computation $P(M_1, \ldots, M_n)$ if the implied joint distribution of the random variables $Y_1, \ldots, Y_J$ satisfies, for each $j = I+1, \ldots, J$, the conditional independence relation*

$$Y_j \perp\!\!\!\perp Y_{\{1,\ldots,j-1\}\setminus\pi(j)} \mid Y_{\pi(j)}.$$

Note that this is weaker than the Markov condition for directed acyclic graphs (see [62]), since we do not insist that the variables represented by the source nodes are independent. It is emphasized that, for a given $\mu \in \mathcal{P}_{\mathcal{X}}$, the coherence condition can in general be verified.

The following result provides verifiable sufficient conditions which ensure that a computation composed of individual BPNMs is a Bayesian computation.

THEOREM 5.9. *Let $M_1, \ldots, M_n$ be BPNMs and let $\mu \in \mathcal{P}_{\mathcal{X}}$ be coherent for the computation $P(M_1, \ldots, M_n)$. Then the computation $P(M_1, \ldots, M_n)$ is Bayesian for the QoI Q_n.*

Conversely, if non-Bayesian PNMs are combined, then the computation $P(M_1, \ldots, M_n)$ need not be Bayesian. An ancestral sampling method for sampling from the output of a pipeline of PNMs is described in section SM5.8 of the supplement.

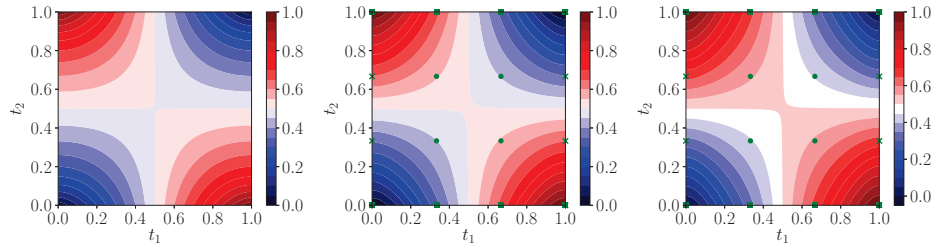
EXAMPLE 5.10 (Example 5.2, continued). *The random variables Y_i in this example are $Y_1 = \{X(t_i)\}_{i=1}^{m-1}$, $Y_2 = X(t_m)$, $Y_3 = \{X(t_i)\}_{i=m+1}^{2m}$, $Y_4 = \int_0^{0.5} X(t)\,\mathrm{d}t$, $Y_5 = \int_{0.5}^{1} X(t)\,\mathrm{d}t$. From $G(P)$ in Figure* 2(b)*, the coherence condition in Definition* 5.8 *requires that the nontrivial conditional independences $Y_4 \perp\!\!\!\perp Y_3 \mid \{Y_1, Y_2\}$ and $Y_5 \perp\!\!\!\perp Y_1 \mid \{Y_2, Y_3\}$ hold. Thus, the distribution μ is coherent for the computation $P(M_1, M_2, M_3)$ if and only if, for $X \sim \mu$, the associated information variables satisfy*

$$\int_0^{0.5} X(t)\,\mathrm{d}t \perp\!\!\!\perp \{X(t_i)\}_{i=m+1}^{2m} \,\Big|\, \{X(t_i)\}_{i=1}^{m},$$
$$\int_{0.5}^{1} X(t)\,\mathrm{d}t \perp\!\!\!\perp \{X(t_i)\}_{i=1}^{m-1} \,\Big|\, \{X(t_i)\}_{i=m}^{2m}.$$

The distribution μ induced by the Weiner process on x satisfies these conditions, since under μ the stochastic process $\{x(t) : t > t_m\}$ is conditionally independent of its history $\{x(t) : t < t_m\}$ given the current state $x(t_m)$. Thus, for this choice of μ, Theorem 5.9 *affirms that $P(M_1, M_2, M_3)$ is a Bayesian computation and parallel computation of (a) and (b) in* (5.1) *can be justified from a Bayesian standpoint. However, for the alternative belief distribution induced by the Weiner process on $\partial^s x$, $s \geq 1$, this condition is not satisfied and the computation $P(M_1, M_2, M_3)$ is not Bayesian. To turn this into a Bayesian procedure for these alternative belief distributions it would be required that $A_{1,2}(x)$ provides information about all the derivatives $\partial^k x(t_m)$, $k \leq s$.*

Thus, we have established locally verifiable conditions that guarantee that the propagation of uncertainty through a computational pipeline is meaningful. This concludes our theoretical development.

6. Numerical Experiments. In this final section of the paper we present three numerical experiments. The first is a linear PDE, the second is a nonlinear ODE, and the third is an application to a problem in industrial process monitoring, described by a pipeline of PNMs. In each case we experiment with non-Gaussian prior distributions and, in doing so, go beyond previous work.



(a) Model solution (b) Gaussian prior (c) Cauchy prior

Fig. 3 (a) *Model solution $x(t)$, $t = (t_1, t_2)$, generated by application of a finite element method based on a triangular mesh of 50×50 elements.* (b), (c) *Posterior means for the solution x of the Poisson equation, with $n = 16$ and different choices of prior distribution. Design points for the interior, Dirichlet, and Neumann boundary conditions are indicated by green dots, green squares, and green crosses, respectively.*

6.1. Poisson Equation. Our first illustration is an instance of the Poisson equation, a linear PDE with mixed Dirichlet–Neumann boundary conditions:

$$-\nabla^2 x(t) = 0, \qquad t \in (0,1)^2, \tag{6.1}$$

$$x(t) = t_1, \qquad t_1 \in [0,1], \qquad t_2 = 0, \tag{6.2}$$

$$x(t) = 1 - t_1, \qquad t_1 \in [0,1], \qquad t_2 = 1, \tag{6.3}$$

$$\partial x / \partial t_2 = 0, \qquad t_2 \in (0,1), \qquad t_1 \in \{0,1\}. \tag{6.4}$$

Figure 3(a) shows a model solution to this system, generated with a finite element Galerkin method on a fine mesh. While this is a relatively simple linear PDE, it is a useful illustrative example. A detailed theoretical treatment of such problems under a Gaussian prior can be found in [16].

Tensor products of orthogonal polynomials were used for prior specification:

$$\phi_i(t) = C_j(2t_1 - 1)C_k(2t_2 - 1),$$

$i + j \leq N_c$. The polynomials C_i were chosen to be normalized Chebyshev polynomials of the first kind. Prior specification then follows the formulation given in section 2.6, where the remaining parameters were chosen to be $x_0 \equiv 1$ and $\gamma_i = (i+1)^{-2}$. The random variables ξ_i were taken to be either Gaussian or Cauchy and the polynomial basis was truncated to $N = 45$ terms, corresponding to a maximum polynomial degree of $N_C = 8$. Note that closed-form expressions are available for analysis under the Gaussian prior [16] but, to demonstrate the numerical disintegration approach from section 4, these were not exploited.

The information operator was defined by a set of locations $t_i \in [0,1]^2$, $i = 0, \ldots, N_t$, where either the interior condition or one of the boundary conditions was enforced. Denote by $\{t^{I,i}\}$ the set of interior points, by $\{t^{D,j}\}$ the set of Dirichlet boundary points, and by $\{t^{N,k}\}$ the set of Neumann boundary points, where $i = 1, \ldots, N_I$, $j = 1, \ldots, N_D$, and $k = 1, \ldots, N_N$, with $n = N_I + N_D + N_N$. Then the information operator is given by the concatenation of the conditions defined above:

$$A(x) = [A^I(x)^\top, A^D(x)^\top, A^N(x)^\top]^\top,$$

$$A^I(x) = \begin{bmatrix} -\nabla^2 x(t^{I,1}) \\ \vdots \\ -\nabla^2 x(t^{I,N_I}) \end{bmatrix}, \quad A^D(x) = \begin{bmatrix} x(t^{D,1}) \\ \vdots \\ x(t^{D,N_D}) \end{bmatrix}, \quad A^N(x) = \begin{bmatrix} \frac{\partial}{\partial t_1} x(t^{N,1}) \\ \vdots \\ \frac{\partial}{\partial t_1} x(t^{N,N_N}) \end{bmatrix}.$$

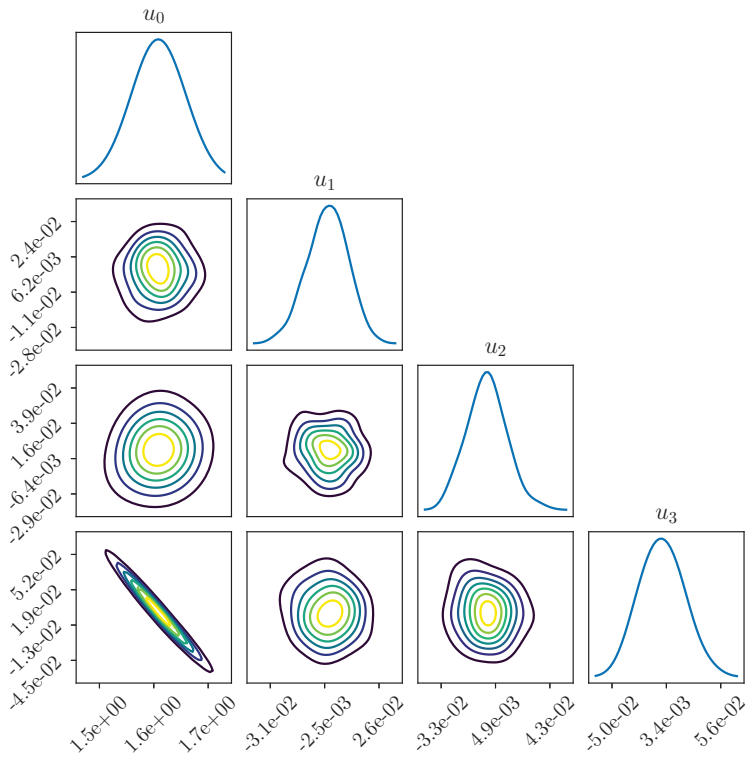
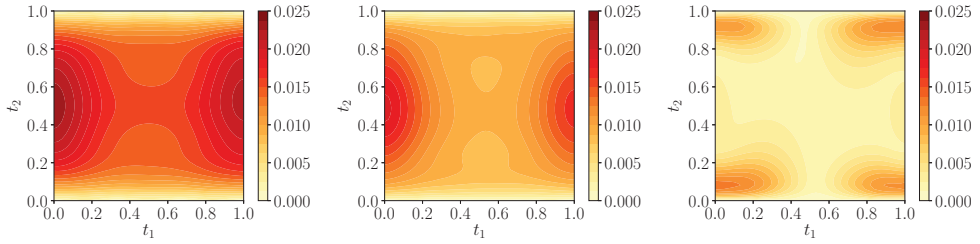


Fig. 4 *Posterior distributions for the first four coefficients of the spectrum for the solution x of the Poisson equation, obtained with Monte Carlo methods and numerical disintegration, based on $\delta = 0.0008$, $n = 16$. The Gaussian prior was used.*

The BPNM output was approximated by numerical disintegration and sampled with a Monte Carlo method whose implementation is reported in section SM5.6 of the supplementary material. Figures 3(b) and 3(c) show the mean of the posterior distribution for Gaussian and Cauchy priors with $n = 16$, with little qualitative difference observed. The mean functions match closely the model solution as given in Figure 3(a).

The posterior distribution of the spectrum $\{u_i\}$ was also investigated. Figure 4 shows the posterior distribution of these coefficients and it is seen that the correlation structure between coefficients (e.g., u_0 and u_3) is nontrivial.

Figure 5(a) shows that the posterior variance is lowest near the Dirichlet boundaries where the solution is known, and it peaks where the Neumann condition is imposed. This is to be expected, as evaluations of the Neumann boundary condition provide less information about the solution itself. The remainder of Figure 5 shows convergence of the posterior distribution as the number of design points is varied, for $n = 16, 25, 36$. In each case a Gaussian prior was used. As expected, the standard deviation in the posterior distribution is seen to decrease as the number of design points is increased. At $n = 36$, the shape of the region of highest uncertainty changes markedly, with the most uncertain region lying between the Dirichlet boundary and the first evaluation points on the Neumann boundary. This is likely due to the fact that the number of evaluation points is approaching the size of the polynomial basis;



(a) $n = 16$ (b) $n = 25$ (c) $n = 36$

Fig. 5 *Heat map of the pointwise standard deviation for the solution x to the Poisson equation as the number n of design points is varied. In each case a Gaussian prior has been used.*

when the number of points equals the size of the basis, the system is completely determined. Thus, we need $N \gg n$ in order for the discretization error to be properly quantified.

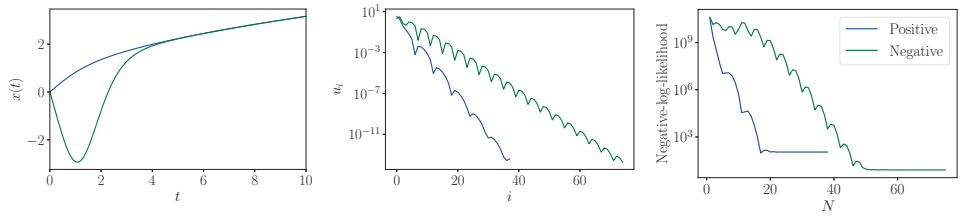
6.2. The Painlevé ODE. In this section a BPNM is developed to solve a nonlinear ODE based on Painlevé's first transcendental,

$$\begin{aligned} x'' &= x^2 - t, & t &\in [0, \infty), \\ x(0) &= 0, & & \\ t^{-1/2} x(t) &\to 1 & &\text{as } t \to \infty\,. \end{aligned}$$

Two distinct solutions are known, illustrated in Figure 6(a). These model solutions were obtained using the deflation technique described in [29]. To permit computation, the right-boundary condition was relaxed by truncating the domain to $[0, 10]$ and using the modified condition $x(10) = \sqrt{10}$. Such systems for which multiple solutions exist have been studied before in the context of PNMs, in both [14] and [16]. It was noted in both papers that existence of multiple solutions can present a substantial challenge to classical numerical methods. Existing techniques for this include the deflation method of [29]. While these still have a substantially lower cost than the numerical disintegration method applied here, the manner in which PNMs can place mass on multiple possible solutions is philosophically appealing.

The spectrum plot in Figure 6(b) shows the coefficients $\{u_i\}$ obtained when each solution is represented in a basis of L_2-orthonormal Chebyshev polynomials. By L_2-orthonormality and Parseval's identity, the slower decay for the negative solution compared to the positive solution is equivalent to the negative solution having a larger L_2 norm. This explains the general preference that optimization-based numerical solvers—and the results now presented—have for the positive solution.

To build a BPNM, a prior μ was defined using a series expansion as in (2.5). The basis functions were $\phi_i(t) = C_i(\frac{1}{2}(t-5))$, where the C_i were normalized Chebyshev polynomials of the first kind. Both Gaussian and Cauchy priors were considered by taking $u_i := \gamma_i \xi_i$, where the ξ_i were taken to be either standard Gaussian or standard Cauchy and the polynomial basis truncated to $N = 40$ terms. In each case, $x_0(t) \equiv 0$. In accordance with the exponential convergence rate for spectral methods when the solution to the system is a smooth function, the sequence of scale parameters was set to $\gamma_i = \alpha\beta^{-i}$, where $\alpha = 8$ and $\beta = 1.5$. These values were chosen by inspection of the true spectra to ensure that both solutions were in the support of the prior.



(a) Solutions (b) Spectra (c) Likelihoods

Fig. 6 (a) *Two distinct solutions for the Painlevé ODE.* (b) *The true coefficients* $\{u_i\}$*, as determined by a model solver (the MATLAB package* `chebfun`*).* (c) *Negative log-likelihoods for the point estimates of coefficients for the positive and negative solutions given by* `chebfun`*, as the truncation level* N *is varied.*

The information operator A is defined by the choice of locations $\{t_j\}_{j=1}^m$ at which the posterior will be constrained. Experiments for several values of m were performed. In each case $t_1 = 0$, $t_m = 10$, and the remaining t_j were equispaced on $[0, 10]$. To be explicit, the information operator was

$$A(x) = \begin{bmatrix} x''(t_1) - (x(t_1))^2 \\ \vdots \\ x''(t_m) - (x(t_m))^2 \\ x(0) \\ x(10) \end{bmatrix}$$

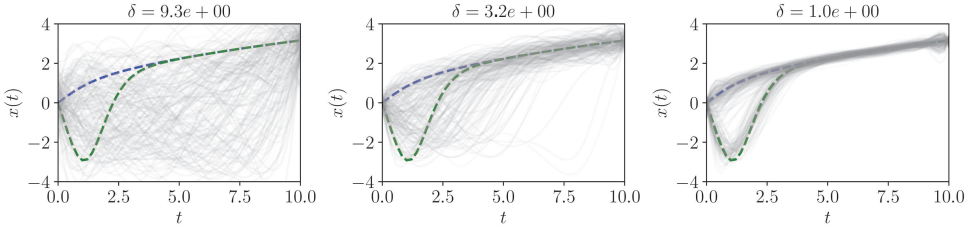
with the last two elements enforcing the boundary conditions. Thus, the information was $a = [-t_1, \ldots, -t_m, 0, \sqrt{10}]$, which is $n = m + 2$ dimensional. The BPNM output $B(\mu, a)$ was approximated via numerical disintegration and sampled with Monte Carlo methods, the details of which are provided in section SM5.7 of the supplementary material.

Results for a selection of bandwidths δ, with $n = 17$, are shown in Figure 7. Note that a strong preference for the positive solution is expressed at the smallest δ, with mass around both solutions at larger δ. For the Gaussian prior, some mass remained around the negative solution at the smallest δ, while this was not so for the Cauchy prior. This reflects the fact that, for a collection of independent univariate Cauchy random variables, one element is likely to be significantly larger in magnitude than the others, which favors faster decay for the remaining elements.

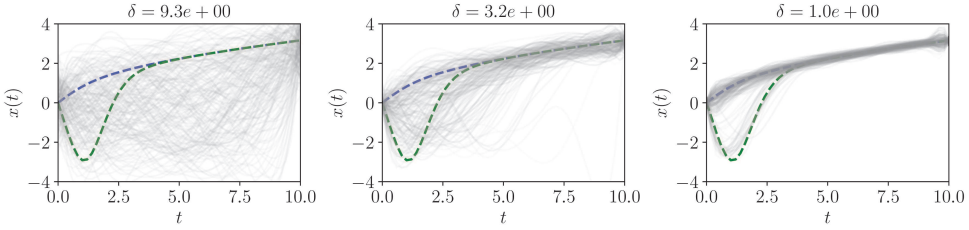
Using the method of thermodynamic integration described in section SM5.4, model evidence was estimated with Monte Carlo for both the Gaussian and the Cauchy priors at $n = 15$. The Bayes factor for the Cauchy, compared to the Gaussian, prior was approximately 20.3, which constitutes strong evidence in favor of a Cauchy prior for this problem at the given level of discretization.

Figure 8 shows the posterior distributions for first six coefficients u_i at $n = 17$ and $\delta = 1$. Strong multimodality is clear, as well as skewed correlation structure between the coefficients. Illustration of such posteriors for smaller δ is difficult as the posteriors become extremely peaked.

Figure 9 shows convergence of the posterior distribution as n is increased. Of particular interest is that for $n = 12$, the posterior based on a Gaussian prior becomes trimodal. For each prior, the posterior mass settles on the positive solution to the system at $n = 22$, in accordance with this solution having smaller L_2 norm. This



(a) Gaussian Prior



(b) Cauchy Prior.

Fig. 7 *Posterior samples for the Painlevé system for $n = 17$. Blue and green dashed lines show the positive and negative solutions determined with* `chebfun`*. Gray lines are samples from an approximation to the posterior provided by numerical disintegration (bandwidth parameter δ).*

perhaps reflects the fact that, while in the limiting case both solutions should have an equal likelihood, the curvature of the likelihood at each mode may differ. Prior truncation may also be influential; in Figure 6(c) the log-likelihood of the negative solution increases at a slower rate than that of the positive solution. Thus, while in the setting of an infinite prior series neither solution should be preferred, in practice truncation might bias one solution over the other. Last, it is clear that the parameters α and β may also have a significant effect on which solution is preferred. Further theoretical work will be required to understand many of these phenomena.

Of particular interest is how a preference for the negative solution could be encoded into a PNM. Owing to the flexible specification of the information operator, there is considerable choice in this matter. An elegant approach is the introduction of additional, inequality-based information $x'(0) \leq 0$. Such information can be difficult to incorporate in standard numerical algorithms, but is of interest in many physical problems [51]. For BPNMs we can extend the information operator to include $1[x'(0) \leq 0]$. Posterior distributions obtained in this way for the Gaussian prior at $n = 17$ are shown in Figure 10. Note that all posterior mass has now settled close to the negative solution. This highlights the simplicity with which BPNMs can encode a preference for a particular solution when multiple solutions exist.

6.3. Application to Industrial Process Monitoring. This final experiment illustrates how statistical models for discretization error can be propagated through a pipeline of computation to model how these errors are accumulated.

Hydrocyclones are machines used to separate solid particles from a liquid in which they are suspended, or two liquids of different densities, using centrifugal forces. High

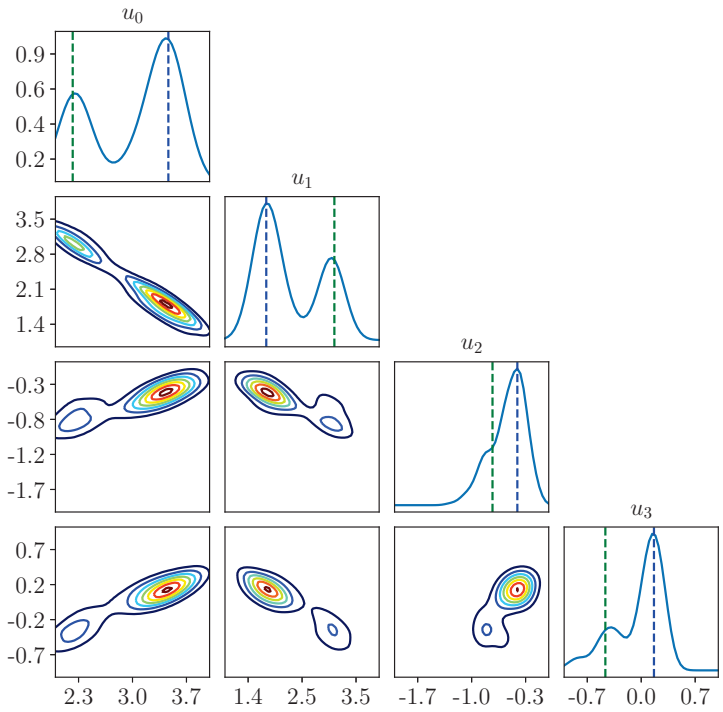


Fig. 8 *Posterior distributions for the first four coefficients obtained with numerical disintegration (bandwidth $\delta = 1$) at $n = 17$. Dashed lines show the coefficient values for the positive (blue) and negative (green) solutions determined with* `chebfun`*. The Gaussian prior was used.*

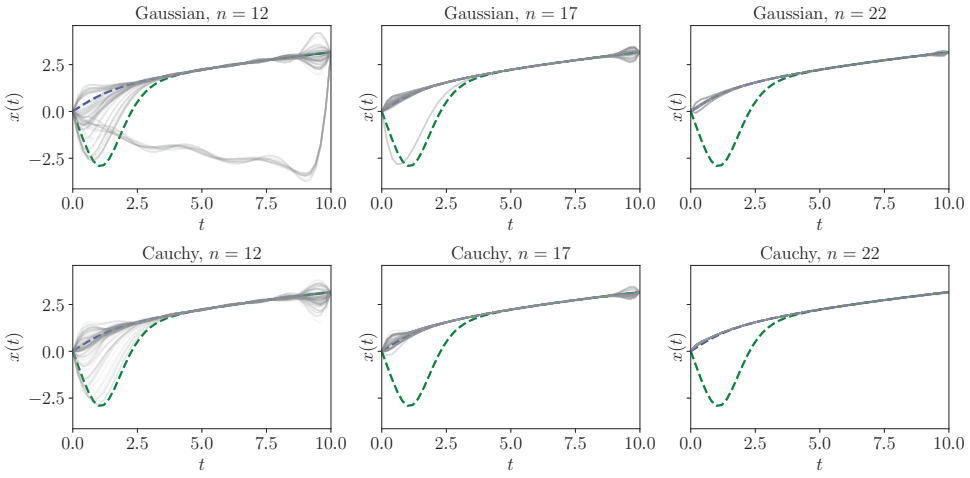


Fig. 9 *Convergence for the numerical disintegration scheme as n is increased. Top: Gaussian prior. Bottom: Cauchy prior. In all cases $\delta = 10^{-4}$.*

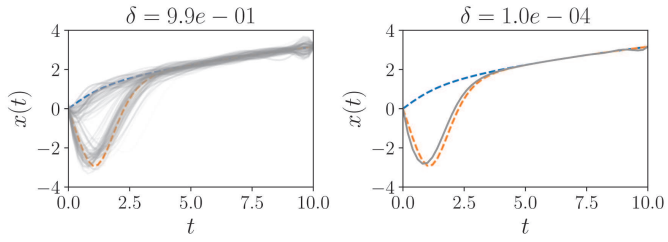


Fig. 10 *Posterior distribution at $n = 17$, based on a Gaussian prior, with the negative boundary condition enforced. Left: $\delta = 0.99$. Right: $\delta = 0.0001$.*

pressure fluid is injected into the top of a tank to create a vortex. The induced centrifugal force causes denser material to move to the wall of the tank, while lighter material concentrates in the center, from where it can be extracted. They have widespread applications, including in areas such as environmental engineering and the petrochemical industry [86]. For safe operation and to ensure that the materials are well separated, the hydrocyclone must be monitored to allow adjustment of the input flow rate [8]. However, direct monitoring is impossible owing to the opaque walls of the equipment and high interior pressure. Electrical impedance tomography (EIT) has been proposed to allow monitoring of the contents [36].

EIT is a technique that allows recovery of an interior conductivity field based upon measurements of voltage obtained from applying a stimulating current on the boundary. It is suited to this problem, since the two materials in the hydrocyclone will generally be of different conductivities. In its simplified form [10], EIT is described by a linear PDE with boundary conditions that incorporate the stimulating currents and measured voltages:

$$
\begin{aligned}
-\nabla \cdot (a(t)\nabla x(t)) &= 0, \qquad && t \in D, \\
a(t)\frac{\partial x}{\partial n}(t) &= \begin{cases} c_e, & t = t^e, \\ 0, & t \in \partial D \setminus \{t^e\}_{e=1}^{N_e}, \end{cases}
\end{aligned} \tag{6.5}
$$

where D denotes the domain, modeling the hydrocyclone tank, e indexes the stimulating electrodes, $t_e \in \partial D$ are the corresponding locations of the electrodes on ∂D, a is the unknown conductivity field to be determined, and $\frac{\partial}{\partial n}$ denotes the derivative with respect to the outward pointing normal vector. The electrode t^1 is referred to as the *reference* electrode. The vector $c = (c_1, \ldots, c_{N_e})$ denotes the stimulation current pattern. Several stimulation patterns were considered, denoted c^j, $j = 1, \ldots, N_j$.

The experimental data described in [95] were considered. In the experiment, a cylindrical perspex tank was used with a single ring of eight electrodes. Translation invariance in the vertical direction means that the contents are effectively a single two-dimensional region and electrical conductivity can be modeled as a two-dimensional field. At the start of the experiment, a mixing impeller created a rotational flow, was then removed, and, after a few seconds, concentrated potassium chloride solution was carefully injected into the water initially filling the tank. Data, denoted y_τ, were collected at regular time intervals by application of several stimulation patterns $c^1, \ldots, c^{N_j}$.

To formulate the statistical problem, consider parameterizing the conductivity field as $a(\tau, t)$, where $\tau \in [0, T]$ is a temporal index while $t \in D$ is the spatial coordinate

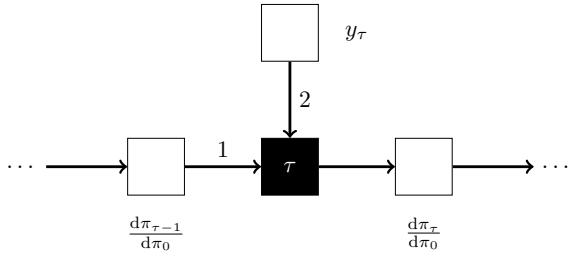


Fig. 11 *Pipeline for hydrocyclone application: The method node (black) represents the use of a probabilistic numerical PDE solver, which is incorporated into the likelihood for evolving the particles according to a Markov transition kernel.*

and D is the circular domain representing the perspex tank in the experiment. A log-Gaussian prior was placed over the conductivity field so that $\log a$ is a Gaussian process with separable covariance function $k_a((\tau,t),(\tau',t')) := \lambda \min(\tau,\tau') \exp(-\frac{1}{2}\ell^{-2}\|t - t'\|^2)$, where ℓ is a length-scale parameter representing the anticipated spatial variation of the conductivity field and λ is a parameter controlling the amplitude of the field. Here ℓ was fixed to $\ell = 0.3$, while $\lambda = 10^{-3}$. The problem of estimating a based on data can be well-posed in the Bayesian framework [28]. Full details of this experiment can be found in the accompanying report [71].

Our aim is to use a BPNM to account for the effect of discretization of the PDE on inferences that are made on the conductivity field. For fixed τ, a Gaussian prior was posited for x, with covariance $k_x(t,t') := \exp(-\frac{1}{2}\ell_x^{-2}\|t - t'\|^2)$, where ℓ_x was fixed to $\ell_x = 0.3$. The associated BPNM, a probabilistic meshless method (PMM), was described in Example 2.4. The statistical inference procedure is formulated as a pipeline in Figure 11. It is assumed that the desired outcome is to monitor the contents of the tank while the current contents are being mixed. This suggests a particle filter approach where a PMM M_τ is employed to handle the intractable likelihood $p(y_\tau|a_\tau)$ that involves the exact solution of a PDE. The distribution of a_τ given $y_1,\ldots,y_\tau$ is denoted π_τ and the computation $P(M_1,\ldots,M_\tau)$ is Bayesian provided that the particle approximation error is overlooked.

To briefly illustrate the method, Figure 12 presents posterior means for the field $a(\tau,\cdot)$ for each postinjection time point $\tau = 1,\ldots,8$. These are based on a particle approximation of size $P = 200$ with method nodes based upon a BPNM, as in Example 2.4 with $n = 119$ design points. The high conductivity region representing the potassium chloride solution can be seen rotating through the domain in the frames $\tau \geq 2$ after injection, with its conductivity reducing as it mixes with the water.

The full posterior distribution over the conductivity field is inflated as a result of explicitly modeling the discretization error. Figure 13 shows the integrated posterior standard deviation of the field for $\tau = 1,\ldots,8$ for both the "PN" pipeline, as described above, and a "non-PN" pipeline in which a symmetric collocation PDE solver[8] was used to approximately solve the PDE. The parameters of the symmetric collocation solver were identical to those used in the PMM. In the left panel we observe some structural periodicity, present in both the PN and non-PN pipelines. This may be due to the rotation of the medium causing the area of high conductivity to periodically

[8]Recall that the PMM has a corresponding symmetric collocation solution to the PDE as its mean function.

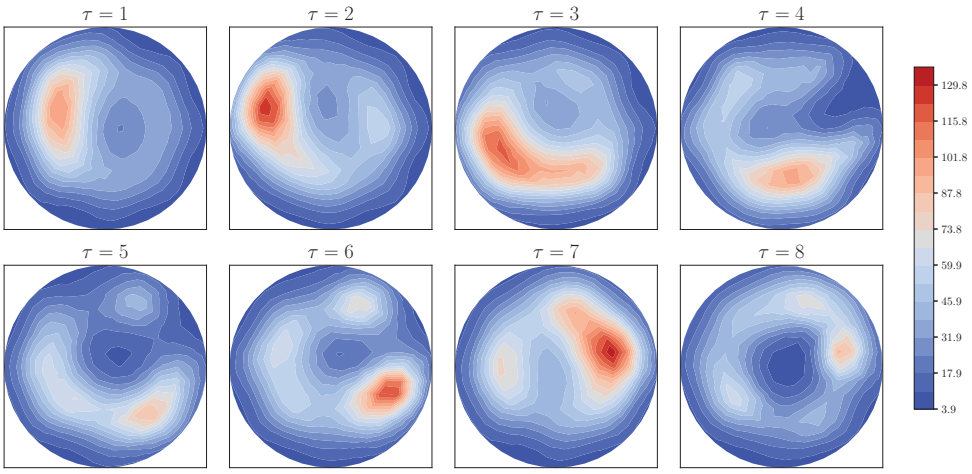


Fig. 12 *Posterior mean conductivity fields recovered in the hydrocyclone experiment, for the first eight frames postinjection of the fluid.*

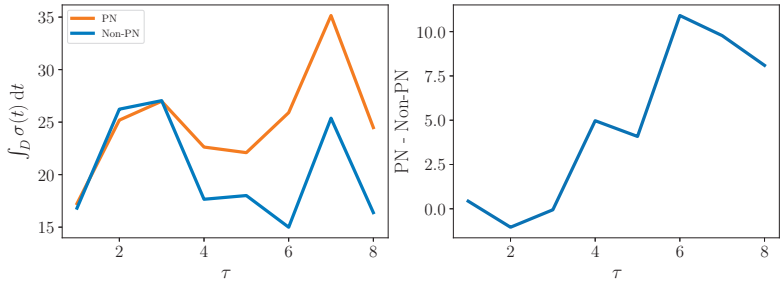


Fig. 13 *Left: Integrated standard deviation $\sigma(t)$ over the domain $t \in D$ for the first $\tau = 1, \ldots, 8$ frames postinjection, for both the probabilistic ("PN") and non-probabilistic ("non-PN") approaches described in the main text. Right: The difference between these two quantities.*

reach an area of the domain, relative to the eight sensors, in which it is particularly easy to recover. With this periodicity subtracted in the right panel, there was a clear increase in posterior uncertainty in the PN pipeline compared to the non-PN version. This suggests that we have quantified and propagated uncertainty due to successive discretization of the PDE at each time point. See [71] for further analysis of these results and of the quality of the uncertainty quantification being provided.

7. Discussion. This paper has established statistical foundations for PNMs in general and investigated the Bayesian case in detail. Through connection to Bayesian inverse problems [88], we have established when BPNMs can be well-defined and when the output can be considered meaningful. The presentation touched on several important issues and a brief discussion of the most salient points follows.

Bayesian vs. Non-Bayesian PNMs. The focus on BPNMs was motivated by the observation that the output of a pipeline of PNMs can be guaranteed to admit a valid Bayesian interpretation if the constituent PNMs are individually Bayesian and the prior distribution is coherent. By Theorem 5.9, prior coherence can be established

at a local level, essentially via a local Markov condition, so that BPNMs provide an extensible modeling framework as required to solve more challenging numerical tasks. These results support a research strategy that focuses on BPNMs, so that uncertainty can be meaningfully propagated.

On the other hand, there are pragmatic reasons why either approximations to BPNMs or indeed non-Bayesian PNMs might be useful. The predominant reason would be to circumvent the computational costs that can be associated with BPNMs, such as the use of numerical disintegration developed in this work. Recent research efforts, such as [83, 84] and [49] for the solution of ODEs, have aimed for computational costs that are competitive with classical methods, at the expense of fully Bayesian estimation for the solution of the ODE. Such methods are of interest as non-Bayesian PNMs, but their role in computational pipelines remains unclear. Our contribution serves to make this explicit.

Computational Cost. The numerical disintegration algorithm proposed in this paper should not be considered a practical means of implementing BPNMs in general. Indeed, the results reported required orders of magnitude more computational effort to generate than model solutions given by a single instance of a classical numerical method. Thus, there is clearly scope for further methodological work to produce more practical algorithms for BPNMs. This is the subject of current research.

Prior Elicitation. Throughout this work we assumed that a belief distribution μ was provided. The question of *whose* belief is represented in μ has been discussed by several authors and a chronology is included in section SM2 of the supplementary material. Of these perspectives we mention in particular [39], wherein μ is the belief of an agent that "we get to design." This offers a connection to frequentist statistics, in that an agent can be designed to ensure favorable frequentist properties hold. In the context of computational pipelines, the challenge of eliciting a coherent prior is closely connected to the challenge of eliciting a single unified prior based on the conflicting input of multiple experts [32, 2].

A robust statistics perspective is also relevant. One such approach would be to consider a generalized Bayes risk (3.1) wherein the state variable X is assumed to be drawn from a distribution $\tilde{\mu} \neq \mu$. This offers an opportunity to derive BPNMs that are robust to certain forms of prior misspecification. This direction was not considered in the present paper, but has been pursued in the ACA literature for classical numerical methods [79, Chapter IV, section 4]. However, in general, the specification of prior distributions for robust inference on an infinite-dimensional state space can be difficult. The consistency and robustness of Bayesian inference procedures—particularly with respect to perturbations of the prior such as those arising from numerical approximations—in such settings is a subtle topic, with both positive [12, 25, 52, 63] and negative [24, 31, 76] results depending upon fine topological and geometric detail.

Estimation. This paper has focused on foundations and further methodological work will be required to establish sufficient conditions under which $B(\mu, A_n(x^\dagger))$ converges, in an appropriate sense, to an atom on a single element $q^\dagger = Q(x^\dagger)$ representing the data-generating QoI in the limit as the amount of information, n, is increased. There are three questions here: (i) when is $q^\dagger$ identifiable from the given information, (ii) at what rate does $B(\mu, A_n(x^\dagger))$ concentrate on $q^\dagger$, and (iii) when can the uncertainty quantification provided be considered to be well calibrated?

Generalization and Extensions. Two more directions are highlighted as extensions of this work. First, note that in this paper the information operator $A\colon \mathcal{X} \to \mathcal{A}$ is treated as a deterministic object. However, in some applications there is auxiliary randomness in the acquisition of information. For our integration example, nodes t_i might arise as a random sample, or observations $x(t_i)$ themselves might occur with measurement error, for example, due to finite precision arithmetic. Then a more elaborate model $A\colon \mathcal{X} \times \Omega \to \mathcal{A}$ would be required, where Ω is a probability space that injects randomness into the information operator. This is the setting of, for instance, randomized quasi-Monte Carlo methods. Future work will extend the framework of PNMs to include randomized information operators of this kind.

As a second direction, recall that in an adaptive algorithm the choice of information is made iteratively, based on information previously observed. For some tasks, such as that described in section SM4.3 and its generalizations discussed there, it can be proven that adaptive algorithms do not outperform nonadaptive algorithms in average-case error [64]. However, outside this setting adaptation can be beneficial and should be investigated in the context of BPNMs.

Acknowledgments. The authors are grateful to Amazon for the provision of AWS credits and to the authors of the C++ library `Eigen` and the Python library `Eigency`.

REFERENCES

[1] N. L. Ackerman, C. E. Freer, and D. M. Roy, *On computability and disintegration*, Math. Structures Comput. Sci., 27 (2017), pp. 1287–1314, https://doi.org/10.1017/s0960129516000098. (Cited on pp. 768, 769)

[2] I. Albert, S. Donnet, C. Guihenneuc-Jouyaux, S. Low-Choy, K. Mengersen, and J. Rousseau, *Combining expert opinions in prior elicitation*, Bayesian Anal., 7 (2012), pp. 503–531, https://doi.org/10.1214/12-BA717. (Cited on p. 784)

[3] T. V. Anderson, *Efficient, Accurate, and non-Gaussian Error Propagation through Nonlinear, Closed-Form, Analytical System Models*, Master's thesis, Department of Mechanical Engineering, Brigham Young University, 2011. (Cited on p. 770)

[4] I. Babuška and G. Söderlind, *On roundoff error growth in elliptic problems*, ACM Trans. Math. Software, 44 (2018), pp. 1–22, https://doi.org/10.1145/3134444. (Cited on p. 770)

[5] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer Series in Statistics, Springer-Verlag, New York, 1985, https://doi.org/10.1007/978-1-4757-4286-2. (Cited on pp. 766, 767)

[6] P. G. Bissiri, C. C. Holmes, and S. G. Walker, *A general framework for updating belief distributions*, J. R. Stat. Soc. Ser. B. Stat. Methodol., 78 (2016), pp. 1103–1130, https://doi.org/10.1111/rssb.12158. (Cited on pp. 761, 763)

[7] V. I. Bogachev, *Gaussian Measures*, Math. Surveys Monogr. 62, AMS, Providence, RI, 1998, https://doi.org/10.1090/surv/062. (Cited on pp. 761, 762, 766)

[8] D. Bradley, *The Hydrocyclone*, Internat. Ser. Monogr. Chem. Engrg. 4, Elsevier, 2013. (Cited on p. 781)

[9] F.-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic, *Probabilistic integration: A role in statistical computation? (with discussion)*, Stat. Sci., 34 (2019), pp. 1–22, https://doi.org/10.1214/18-STS660. (Cited on pp. 758, 761, 764, 767)

[10] A.-P. Calderón, *On an inverse boundary value problem*, in Seminar on Numerical Analysis and Its Applications to Continuum Physics (Rio de Janeiro, 1980), Soc. Brasil. Mat., Rio de Janeiro, 1980, pp. 65–73, https://doi.org/10.1590/s0101-82052006000200002. (Cited on p. 781)

[11] M. A. Capistrán, J. A. Christen, and S. Donnet, *Bayesian analysis of ODEs: Solver optimal accuracy and Bayes factors*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 829–849, https://doi.org/10.1137/140976777. (Cited on p. 757)

[12] I. Castillo and R. Nickl, *On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures*, Ann. Statist., 42 (2014), pp. 1941–1969, https://doi.org/10.1214/14-AOS1246. (Cited on p. 784)

[13] J. T. Chang and D. Pollard, *Conditioning as disintegration*, Statist. Neerlandica, 51 (1997), pp. 287–317, https://doi.org/10.1111/1467-9574.00056. (Cited on p. 765)
[14] O. A. Chkrebtii, D. A. Campbell, B. Calderhead, and M. A. Girolami, *Bayesian solution uncertainty quantification for differential equations*, Bayesian Anal., 11 (2016), pp. 1239–1267, https://doi.org/10.1214/16-BA1017. (Cited on pp. 758, 770, 777)
[15] J. Cockayne, C. J. Oates, I. C. F. Ipsen, and M. Girolami, *A Bayesian conjugate gradient method*, Bayesian Anal., to appear. (Cited on pp. 758, 762, 764)
[16] J. Cockayne, C. Oates, T. J. Sullivan, and M. Girolami, *Probabilistic Meshless Methods for Partial Differential Equations and Bayesian Inverse Problems*, preprint, https://arxiv.org/abs/1605.07811, 2016. (Cited on pp. 758, 762, 764, 767, 770, 775, 777)
[17] P. R. Conrad, M. Girolami, S. Särkkä, A. M. Stuart, and K. C. Zygalakis, *Statistical analysis of differential equations: Introducing probability measures on numerical solutions*, Stat. Comput., 27 (2017), pp. 1065–1082, https://doi.org/10.1007/s11222-016-9671-0. (Cited on pp. 758, 770)
[18] M. Dashti, S. Harris, and A. Stuart, *Besov priors for Bayesian inverse problems*, Inverse Probl. Imaging, 6 (2012), pp. 183–200, https://doi.org/10.3934/ipi.2012.6.183. (Cited on p. 765)
[19] M. de Carvalho, G. L. Page, and B. J. Barney, *On the geometry of Bayesian inference*, Bayesian Anal., (2018), https://doi.org/10.1214/18-ba1112. (Cited on p. 761)
[20] P. Del Moral, A. Doucet, and A. Jasra, *An adaptive sequential Monte Carlo method for approximate Bayesian computation*, Stat. Comput., 22 (2012), pp. 1009–1020, https://doi.org/10.1007/s11222-011-9271-y. (Cited on p. 769)
[21] C. Dellacherie and P.-A. Meyer, *Probabilities and Potential*, North-Holland, Amsterdam, New York, 1978, https://doi.org/10.1016/s0304-0208(08)x7141-5. (Cited on p. 765)
[22] P. Diaconis, *Bayesian numerical analysis*, in Statistical Decision Theory and Related Topics IV, Vol. 1, Springer, New York, 1988, pp. 163–175, https://doi.org/10.1007/978-1-4613-8768-8_20. (Cited on pp. 761, 767)
[23] P. Diaconis and D. Freedman, *Frequency properties of Bayes rules*, in Scientific Inference, Data Analysis, and Robustness (Madison, Wis., 1981), Publ. Math. Res. Center Univ. Wisconsin 48, Academic Press, Orlando, FL, 1983, pp. 105–115, https://doi.org/10.1016/b978-0-12-121160-8.50011-9. (Cited on p. 758)
[24] P. Diaconis and D. A. Freedman, *On the consistency of Bayes estimates*, Ann. Statist., 14 (1986), pp. 1–67, https://doi.org/10.1214/aos/1176349830. (Cited on p. 784)
[25] J. L. Doob, *Application of the theory of martingales*, in Le Calcul des Probabilités et ses Applications, Colloques Internationaux du Centre National de la Recherche Scientifique 13, Centre National de la Recherche Scientifique, Paris, 1949, pp. 23–27. (Cited on p. 784)
[26] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, Springer-Verlag, New York, 2001, https://doi.org/10.1007/978-1-4757-3437-9. (Cited on p. 770)
[27] L. L. Duan, A. L. Young, A. Nishimura, and D. B. Dunson, *Bayesian Constraint Relaxation*, preprint, https://arxiv.org/abs/1801.01525, 2018. (Cited on p. 769)
[28] M. M. Dunlop and A. M. Stuart, *The Bayesian formulation of EIT: Analysis and algorithms*, Inverse Probl. Imaging, 10 (2016), pp. 1007–1036, https://doi.org/10.3934/ipi.2016030. (Cited on p. 782)
[29] P. E. Farrell, Á. Birkisson, and S. W. Funke, *Deflation techniques for finding distinct solutions of nonlinear partial differential equations*, SIAM J. Sci. Comput., 37 (2015), pp. A2026–A2045, https://doi.org/10.1137/140984798. (Cited on p. 777)
[30] G. E. Fasshauer, *Solving differential equations with radial basis functions: Multilevel methods and smoothing*, Adv. Comput. Math., 11 (1999), pp. 139–159, https://doi.org/10.1023/A:1018919824891. (Cited on p. 762)
[31] D. A. Freedman, *On the asymptotic behavior of Bayes' estimates in the discrete case*, Ann. Math. Statist., 34 (1963), pp. 1386–1403, https://doi.org/10.1214/aoms/1177703871. (Cited on p. 784)
[32] S. French, *Aggregating expert judgement*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Math. RACSAM, 105 (2011), pp. 181–206, https://doi.org/10.1007/s13398-011-0018-6. (Cited on p. 784)
[33] N. Garcia Trillos and D. Sanz-Alonso, *The Bayesian update: Variational formulations and gradient flows*, Bayesian Anal., (2018), https://doi.org/10.1214/18-BA1137. (Cited on p. 765)
[34] C. J. Geyer, *Markov chain Monte Carlo maximum likelihood*, in Proceedings of the 23rd Symposium on the Interface of Computing Science and Statistics, Interface Foundation of North America, 1991, pp. 156–163. (Cited on p. 770)

[35] M. Girolami and B. Calderhead, *Riemann manifold Langevin and Hamiltonian Monte Carlo methods*, J. R. Stat. Soc. Ser. B Stat. Methodol., 73 (2011), pp. 123–214, https://doi.org/10.1111/j.1467-9868.2010.00765.x. (Cited on p. 770)

[36] J. A. Gutierrez, T. Dyakowski, M. S. Beck, and R. A. Williams, *Using electrical impedance tomography for controlling hydrocyclone underflow discharge*, Powder Technol., 108 (2000), pp. 180–184, https://doi.org/10.1016/s0967-0661(97)00233-5. (Cited on p. 781)

[37] R. Harvey and D. Verseghy, *The reliability of single precision computations in the simulation of deep soil heat diffusion in a land surface model*, Clim. Dynam., 46 (2015), pp. 3865–3882, https://doi.org/10.1007/s00382-015-2809-5. (Cited on p. 758)

[38] P. Hennig, *Probabilistic interpretation of linear solvers*, SIAM J. Optim., 25 (2015), pp. 234–260, https://doi.org/10.1137/140955501. (Cited on pp. 758, 762, 767)

[39] P. Hennig, M. A. Osborne, and M. Girolami, *Probabilistic numerics and uncertainty in computations*, Proc. A, 471 (2015), art. 20150142, https://doi.org/10.1098/rspa.2015.0142. (Cited on pp. 758, 760, 770, 784)

[40] P. Henrici, *Error Propagation for Difference Method*, John Wiley and Sons, New York, London, 1963. (Cited on p. 758)

[41] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002, https://doi.org/10.1137/1.9780898718027. (Cited on p. 758)

[42] J. B. Kadane, *Principles of Uncertainty*, Texts in Statistical Science Series, CRC Press, Boca Raton, FL, 2011, https://doi.org/10.1201/b11322. (Cited on p. 761)

[43] J. B. Kadane and G. W. Wasilkowski, *Average case ϵ-complexity in computer science: A Bayesian view*, in Bayesian Statistics, Elsevier, North-Holland, Amsterdam, 1985, pp. 361–374. (Cited on pp. 758, 767, 768)

[44] W. Kahan, *The Improbability of Probabilistic Error Analyses for Numerical Computations*, in U. California, Berkeley Statistics Colloquium, 1996, www.cs.berkeley.edu/~wkahan/improber.pdf. (Cited on p. 758)

[45] M. Kanagawa, B. K. Sriperumbudur, and K. Fukumizu, *Convergence guarantees for kernel-based quadrature rules in misspecified settings*, in Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NIPS), 2016, pp. 3288–3296. (Cited on p. 761)

[46] T. Karvonen, C. J. Oates, and S. Särkkä, *A Bayes-Sard cubature method*, in Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NIPS), 2019, pp. 5882–5893. (Cited on p. 761)

[47] T. Karvonen and S. Särkkä, *Classical quadrature rules via Gaussian processes*, in 27th IEEE International Workshop on Machine Learning for Signal Processing, IEEE, 2017, pp. 1–6. (Cited on p. 761)

[48] T. Karvonen and S. Särkkä, *Fully symmetric kernel quadrature*, SIAM J. Sci. Comput., 40 (2018), pp. A697–A720, https://doi.org/10.1137/17m1121779. (Cited on p. 761)

[49] H. Kersting and P. Hennig, *Active uncertainty calibration in Bayesian ODE solvers*, in Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI), AUAI Press, Arlington, VA, 2016, pp. 309–318. (Cited on p. 784)

[50] G. S. Kimeldorf and G. Wahba, *Spline functions and stochastic processes*, Sankhyā Ser. A, 32 (1970), pp. 173–180. (Cited on p. 758)

[51] D. Kinderlehrer and G. StampacThia, *An Introduction to Variational Inequalities and Their Applications*, SIAM, 2000, https://doi.org/10.1137/1.9780898719451. Reprint of the Academic Press 1980 original. (Cited on p. 779)

[52] B. J. K. Kleijn and A. W. van der Vaart, *The Bernstein–Von-Mises theorem under misspecification*, Electron. J. Stat., 6 (2012), pp. 354–381, https://doi.org/10.1214/12-EJS675. (Cited on p. 784)

[53] A. N. Kolmogorov, *Foundations of Probability*, Ergebnisse Der Mathematik, 1933. (Cited on p. 765)

[54] J. Kuelbs, F. M. Larkin, and J. A. Williamson, *Weak probability distributions on reproducing kernel Hilbert spaces*, Rocky Mt. J. Math., 2 (1972), pp. 369–378, https://doi.org/10.1216/RMJ-1972-2-3-369. (Cited on p. 758)

[55] F. Larkin, *Probabilistic estimation of poles or zeros of functions*, J. Approx. Theory, 27 (1979), pp. 355–371, https://doi.org/10.1016/0021-9045(79)90124-2. (Cited on p. 758)

[56] F. M. Larkin, *Estimation of a non-negative function*, BIT, 9 (1969), pp. 30–52. (Cited on p. 758)

[57] F. M. Larkin, *Optimal approximation in Hilbert spaces with reproducing kernel functions*, Math. Comp., 24 (1970), pp. 911–921, https://doi.org/10.2307/2004625. (Cited on p. 758)

[58] F. M. Larkin, *Gaussian measure in Hilbert space and applications in numerical analysis*, Rocky Mountain J. Math., 2 (1972), pp. 379–421, https://doi.org/10.1216/RMJ-1972-2-3-379. (Cited on pp. 758, 759, 761, 763)

[59] F. M. Larkin, *Probabilistic error estimates in spline interpolation and quadrature*, in Information Processing 74 (Proc. IFIP Congress, Stockholm, 1974), North-Holland, Amsterdam, 1974, pp. 605–609. (Cited on p. 758)
[60] F. M. Larkin, *A modification of the secant rule derived from a maximum likelihood principle*, BIT, 19 (1979), pp. 214–222, https://doi.org/10.1007/BF01930851. (Cited on p. 758)
[61] M. Lassas, E. Saksman, and S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors*, Inverse Probl. Imaging, 3 (2009), pp. 87–122, https://doi.org/10.3934/ipi.2009.3.87. (Cited on p. 765)
[62] S. Lauritzen, *Graphical Models*, Oxford University Press, 1991. (Cited on p. 774)
[63] L. Le Cam, *On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates*, Univ. California Publ. Statist., 1 (1953), pp. 277–329. (Cited on p. 784)
[64] D. Lee and G. W. Wasilkowski, *Approximation of linear functionals on a Banach space with a Gaussian measure*, J. Complexity, 2 (1986), pp. 12–43, https://doi.org/10.1016/0885-064X(86)90021-X. (Cited on p. 785)
[65] D. V. Lindley, *Understanding Uncertainty*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ, revised ed., 2014, https://doi.org/10.1002/9781118650158.indsp2. (Cited on p. 761)
[66] D. J. C. MacKay, *Bayesian interpolation*, Neural Comput., 4 (1992), pp. 415–447. (Cited on p. 764)
[67] J. Mockus, *Bayesian Approach to Global Optimization: Theory and Applications*, Springer Science & Business Media, 1989. (Cited on p. 758)
[68] A. Müller, *Integral probability metrics and their generating classes of functions*, Adv. in Appl. Probab., 29 (1997), pp. 429–443, https://doi.org/10.2307/1428011. (Cited on p. 769)
[69] S. Niederer, L. Mitchell, N. Smith, and G. Plank, *Simulating human cardiac electrophysiology on clinical time-scales*, Front. in Physiol., 2 (2011), https://doi.org/10.3389/fphys.2011.00014. (Cited on p. 759)
[70] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*, European Mathematical Society, 2010. (Cited on p. 767)
[71] C. J. Oates, J. Cockayne, R. G. Aykroyd, and M. A. Girolami, *Bayesian probabilistic numerical methods in time-dependent state estimation for industrial hydrocyclone equipment*, J. Amer. Statist. Assoc., 2019, pp. 1–27. (Cited on pp. 782, 783)
[72] W. L. Oberkampf and C. J. Roy, *Verification and Validation in Scientific Computing*, Cambridge University Press, Cambridge, 2013. (Cited on p. 758)
[73] A. O'Hagan, *Bayes–Hermite quadrature*, J. Statist. Plann. Inference, 29 (1991), pp. 245–260, https://doi.org/10.1016/0378-3758(91)90002-V. (Cited on pp. 758, 761)
[74] H. Owhadi, *Bayesian numerical homogenization*, Multiscale Model. Simul., 13 (2015), pp. 812–828, https://doi.org/10.1137/140974596. (Cited on p. 758)
[75] H. Owhadi, *Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games*, SIAM Rev., 59 (2017), pp. 99–149, https://doi.org/10.1137/15M1013894. (Cited on p. 762)
[76] H. Owhadi, C. Scovel, and T. J. Sullivan, *On the brittleness of Bayesian inference*, SIAM Rev., 57 (2015), pp. 566–582, https://doi.org/10.1137/130938633. (Cited on p. 784)
[77] J. Pfanzagl, *Conditional distributions as derivatives*, Ann. Probab., 7 (1979), pp. 1046–1050. (Cited on p. 769)
[78] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., Cambridge University Press, Cambridge, UK, 2007. (Cited on p. 758)
[79] K. Ritter, *Average-Case Analysis of Numerical Problems*, Lecture Notes in Math. 1733, Springer-Verlag, Berlin, 2000, https://doi.org/10.1007/BFb0103934. (Cited on pp. 766, 767, 768, 784)
[80] C. Roy, *Review of discretization error estimators in scientific computing*, in Proceedings of 48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, AIAA, 2010. (Cited on p. 770)
[81] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, *Design and analysis of computer experiments*, Stat. Sci., 4 (1989), pp. 409–423, https://doi.org/10.1214/ss/1177012413. (Cited on p. 766)
[82] S. Särkkä, J. Hartikainen, L. Svensson, and F. Sandblom, *On the relation between Gaussian process quadratures and sigma-point methods*, J. Adv. Inf. Fusion, 11 (2016), pp. 31–46. (Cited on pp. 761, 767)
[83] M. Schober, D. K. Duvenaud, and P. Hennig, *Probabilistic ODE solvers with Runge–Kutta means*, in Proceedings of Advances in Neural Information Processing Systems (NIPS), 2014, pp. 739–767. (Cited on pp. 758, 767, 784)

[84] M. Schober, S. Särkkä, and P. Hennig, *A probabilistic model for the numerical solution of initial value problems*, Stat. Comput., 29 (2019), pp. 99–122, https://doi.org/10.1007/s11222-017-9798-7. (Cited on pp. 767, 784)

[85] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976. (Cited on p. 760)

[86] R. Sripriya, M. Kaulaskar, S. Chakraborty, and B. Meikap, *Studies on the performance of a hydrocyclone and modeling for flow characterization in presence and absence of air core*, Chem. Eng. Sci., 62 (2007), pp. 6391–6402, https://doi.org/10.1016/j.ces.2007.07.046. (Cited on p. 781)

[87] G. Strang and G. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, NJ, 1973. (Cited on p. 758)

[88] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061. (Cited on pp. 759, 763, 764, 783)

[89] T. J. Sullivan, *Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors*, Inverse Probl. Imaging, 11 (2017), pp. 857–874, https://doi.org/10.3934/ipi.2017040. (Cited on pp. 765, 766)

[90] O. Teymur, K. Zygalakis, and B. Calderhead, *Probabilistic linear multistep methods*, in Proceedings of Advances in Neural Information Processing Systems (NIPS), 2016, pp. 4314–4321. (Cited on p. 767)

[91] A. Törn and A. Žilinskas, *Global Optimization*, Lecture Notes in Comput. Sci. 350, Springer-Verlag, Berlin, 1989, https://doi.org/10.1007/3-540-50871-6. (Cited on p. 758)

[92] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Computer Science and Scientific Computing, Academic Press, Inc., Boston, MA, 1988. (Cited on p. 767)

[93] J. Wang, J. Cockayne, and C. Oates, *On the Bayesian solution of differential equations*, in Proceedings of the 38th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 2018. (Cited on p. 764)

[94] G. Wasilkowski, *Optimal algorithms for linear problems with Gaussian measures*, Rocky Mountain J. Math., 16 (1986), pp. 727–750, https://doi.org/10.1216/rmj-1986-16-4-727. (Cited on p. 767)

[95] R. M. West, S. Meng, R. G. Aykroyd, and R. A. Williams, *Spatial-temporal modeling for electrical impedance imaging of a mixing process*, Rev. Sci. Instrum., 76 (2005), art. 073703, https://doi.org/10.1063/1.1947882. (Cited on p. 781)

[96] X. Xi, F.-X. Briol, and M. Girolami, *Bayesian quadrature for multiple related integrals*, in Proceedings of the 35th International Conference on Machine Learning (ICML), 2018, pp. 5373–5382. (Cited on p. 764)