Probabilistic Meshless Methods for Modelling Numerical Error in Inverse Problems

Jon Cockayne[1], Chris J. Oates[2], Tim Sullivan[1], Mark Girolami[1]

1: University of Warwick, 2: University of Technology Sydney

Introduction

Probabilistic Numerics (PN) reinterprets classical numerical tasks as statistical inference tasks. PDEs are an ideal candidate for such approaches, as for complex problems in high dimensions error bounds are crude and eliminating error is computationally unfeasible. Such problems would benefit from a rigorous probabilistic quantification of error.

Recent work (1) has developed a probabilistic scheme for solution of ODEs, and PDEs in the case when an orthonormal basis can be constructed. We aim to widen this work to general bases by considering probabilistic approaches to meshless methods.

Approximation of PDEs

We wish to solve a system of operator equations for $u(\mathbf{x})$ defined on Ω (with boundary $\partial\Omega$):

$$\begin{aligned} \mathcal{A}u(\mathbf{x}) &= f(\mathbf{x}) && (\text{in } \Omega) \\ \mathcal{B}u(\mathbf{x}) &= g(\mathbf{x}) && (\text{on } \partial\Omega) \end{aligned}$$

A general method for determining u numerically is to choose a basis for the solution space $\{\phi_j\}$ and construct an approximation $\hat{u}$:

$$\hat{u}(\mathbf{x}) = \sum_j c_j \phi_j(\mathbf{x})$$

Different choices of basis yield different numerical methods; the finite element (FE) method uses orthonormal ϕ_j, which is computationally tractable, but requires manual construction of a 'mesh' on which the basis is defined. This mesh means that the method often breaks down for complex problems, such as impact problems or deformations.

Meshless methods use arbitrary basis functions which makes them more resilient for such complex problems, but the solution is more computationally intensive to obtain.

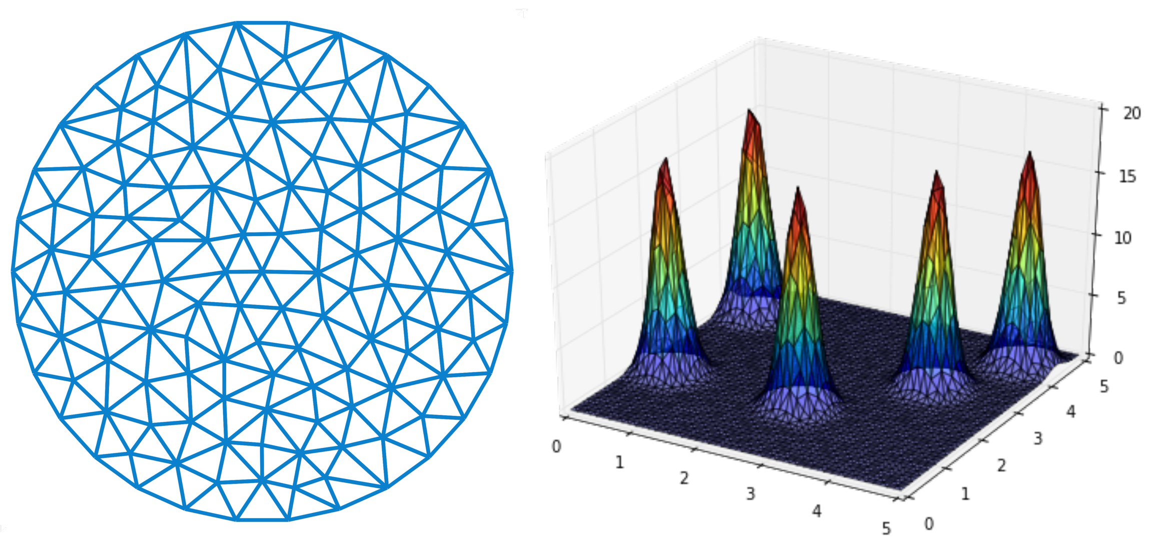

Fig 1: Illustration of basis for finite elements (left) vs. meshless methods (right)

Forward Problem

When operators $\mathcal{A}$ and $\mathcal{B}$ are linear the problem is more tractable. We proceed to choose a basis $\{\phi_j\}$ and place a prior over weights $\{c_j\}$. This is a Bayesian approach to collocation (4).

Optimal Basis Functions

Owhadi (2) showed that by exploiting *Green's functions* for the system we can construct an optimal kernel when using a GP prior for u:

$$k(x,x') = \int_\Omega G(x,z)G(x',z')\Lambda(z,z')dzdz'$$

We can show that this is the kernel for an RKHS:

$$\{u : \mathcal{A}u \in H, \mathcal{B}u = 0\}$$

where H is the RKHS with kernel Λ.

This allows us to approximate $\hat{u}$ as a GP in the linear case, and suggests a choice of basis for collocation – however k and G are problem-dependent and may be difficult to obtain for some operators.

Bayesian Collocation

This is a closer analogue to the finite element method or traditional meshless methods – we choose a basis and place a prior over the weights in the basis expansion.

In the linear setting a multivariate Gaussian prior gives a Gaussian posterior for $\hat{u}$, from which we can easily sample.

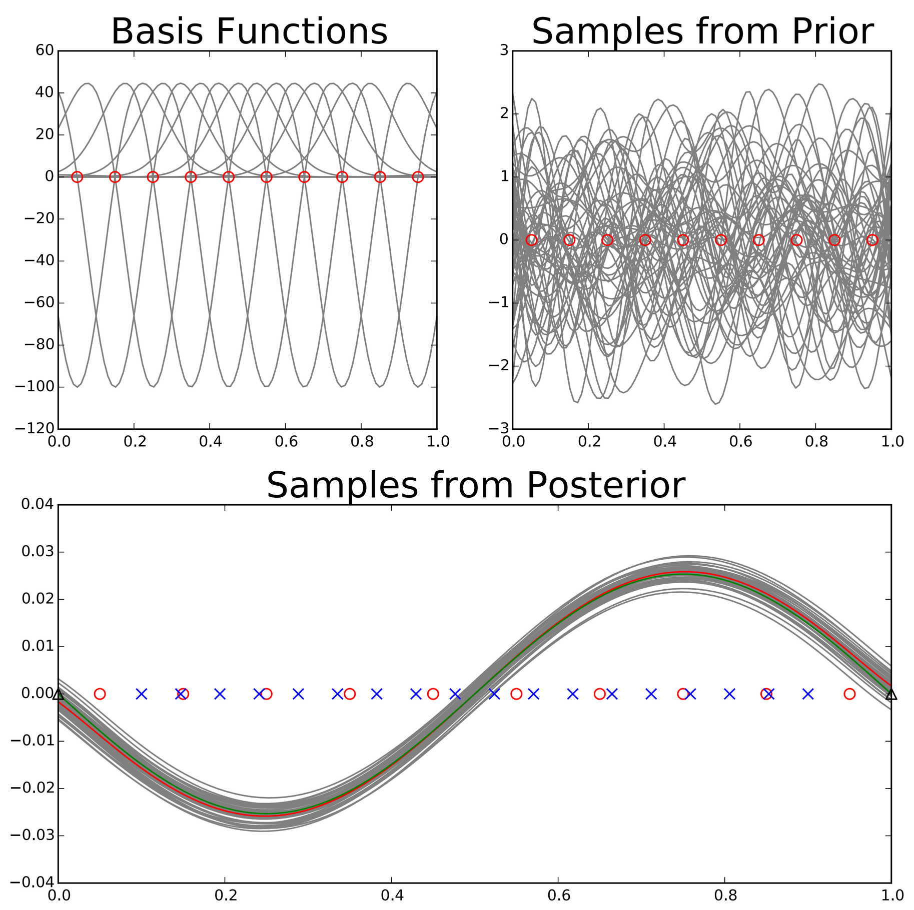

Fig 2: Poisson's equation in 1D on domain [0,1], with Dirichlet boundary conditions. Samples drawn from the Bayesian Collocation approach.

Inverse Problem

In this case we are interested in performing inference on parameters θ of a PDE model, based on observations y_i of some physical system at locations x_i. Examples might be determining coefficents of the model, or the initial conditions. We obtain a posterior:

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

The likelihood $p(y|\theta)$ can be found as:

$$p(y|\theta) := \prod_i p(y_i | u(x_i;\theta))$$

which is dependent on an exact value for u at each x_i. Simply using our approximation $\hat{u}$ assumes that error in the approximation is negligible. Approximations with negligible error may be very expensive to generate, and if the approximation is not sufficiently robust this can lead to posterior contraction around incorrect parameter values.

Instead we use an approximate likelihood incorporating the solver error, similar to (1):

$$p_{\mathrm{PN}}(\theta|y) = \int \prod_i p(y_i|u)p(u|\theta)\mathrm{d}u$$

For linear operators this can be evaluated explicitly, and we can draw samples from the posterior $p(\theta|y)$ with MCMC.

Example

We used this method to infer θ in:

$$\Delta u = \sin(\theta x)$$

with $\Omega = (0,1)$ and $u(0) = u(1) = 0$. Data was generated with $\theta = 2\pi$, and observations of the function were corrupted with white noise. We compare the shape of posterior for θ in the PN approach vs. assuming no error in an FE approximation of $\hat{u}$.

In Fig 3 we see that the posteriors from FE approximation are skewed to one side of the true value, approaching it as the number of basis functions increases. Further the distributions are narrow, suggesting high confidence in incorrect values.

The PN approach shows less pronounced skew in the posteriors, with lower confidence in the skewed estimate as reflected by the wider distributions.

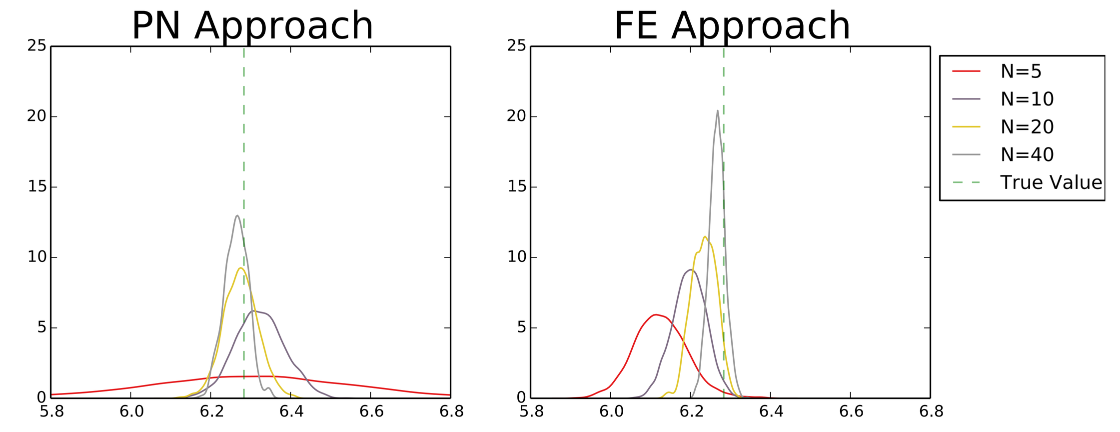

Fig 3: Inferring θ for $\Delta u = \sin(\theta x)$.

Nonlinear Problems

We can no longer use a GP – the operators are no longer linear so Gaussianity is not preserved. Instead we apply Bayesian Collocation.

Neither the posterior for $\hat{u}$ nor our approximate likelihood for the inverse problem will have a closed form, so we need to sample from each of these with MCMC. For the inverse problem a pseudo-marginal approach will be required to evaluate the integral.

Furthermore with a large number of basis functions the problem is very high-dimensional. This suggests to use infinite-dimensional MCMC schemes like in (3).

Further Work

Further work focusses on application of the method to more practical linear problems. To be able to compete with meshless methods and with the finite element method we need to make the required matrix inversions easier by using compactly supported basis functions. Work also needs to be done to understand how we should optimally place our basis functions in the domain.

In addition we aim to develop a methodology to apply these techniques to general nonlinear problems. In these cases we believe the PN approach will have a clear advantage, since understanding the resulting error from finite computational budget is of critical importance.

We are also examining other inference problems related to PDEs which might benefit from a PN approach. Examples include model selection and multiscale methods.

References

(1) Conrad, P.R., Girolami, M., Särkkä, S., Stuart, A., and Zygalakis, K. (2015). Probability measures for numerical solutions of differential equations. arXiv preprint arXiv:1506.04592.
(2) Owhadi, H. (2015). Bayesian numerical homogenization. arXiv preprint arXiv:1406.6668.
(3) Beskos, A. (2014). A stable manifold MCMC method for high dimensions. Statistics and Probability Letters, 90(1), 46–52.
(4) Fasshauer, G. E. (1999). Solving Partial Differential Equations by Collocation with Radial Basis Functions: Multilevel Methods and Smoothing. Advances in Computational Mathematics, 11(2-3):139–159