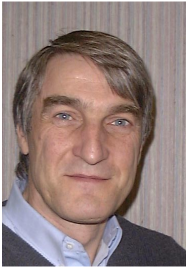
Part IV

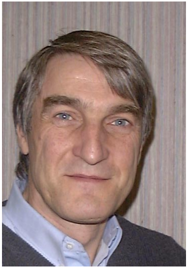


Some Bayesian Numerical Analysis (with discussion)

A. O'HAGAN, University of Nottingham

In: Bayesian Statistics (Eds. Bernardo, Berger, Dawid and Smith), 4, 345-363, 1992.

> *Bayesian approaches to interpolation, quadrature and optimisation are discussed, based on representing prior information about the function in question in terms of a Gaussian process. Emphasis is placed on how different methods are appropriate when the function is cheap or expensive to evaluate. A particular case of expensive functions is a regression function, where 'evaluation' consists of gaining observations (with the small added complication of measurement error).*



Some Bayesian Numerical Analysis (with discussion)

A. O'HAGAN, University of Nottingham

In: Bayesian Statistics (Eds. Bernardo, Berger, Dawid and Smith), 4, 345-363, 1992.

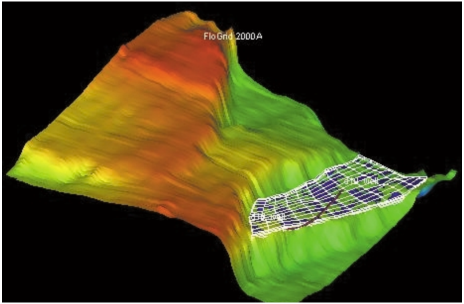
> *Bayesian approaches to interpolation, quadrature and optimisation are discussed, based on representing prior information about the function in question in terms of a Gaussian process. Emphasis is placed on how different methods are appropriate when the function is cheap or expensive to evaluate. A particular case of expensive functions is a regression function, where 'evaluation' consists of gaining observations (with the small added complication of measurement error).*

Eighth Job: Solution of PDEs

Darcy's Law

Consider a dynamical system with unknown parameters, e.g. Darcy's law:

$$-\nabla \cdot (\theta(t)\nabla x(t)) = g(t) \quad t \in D$$
$$x(t) = b(t) \quad t \in \partial D$$



$$-\nabla \cdot (\theta(t)\nabla x(t)) = g(t) \quad t \in D$$
$$x(t) = b(t) \quad t \in \partial D$$

Problem 1

Generally $x(t)$ does not have a closed-form. This is usually known as a forward problem.

Solution

We will construct a Bayesian Probabilistic Numerical Method for PDEs.

Darcy's Law

$$
\begin{aligned}
-\nabla \cdot (\theta(t) \nabla x(t)) &= g(t) \quad t \in D \\
x(t) &= b(t) \quad t \in \partial D
\end{aligned}
$$

Problem 1
Generally $x(t)$ does not have a closed-form. This is usually known as a forward problem.

Solution
We will construct a Bayesian Probabilistic Numerical Method for PDEs.

$$
\begin{aligned}
-\nabla \cdot (\theta(t) \nabla x(t)) &= g(t) \quad t \in D \\
x(t) &= b(t) \quad t \in \partial D
\end{aligned}
$$

Problem 2
To make predictions with the PDE, coefficients $\theta(t)$ must be estimated. This is usually known as an inverse problem.

Solution
We will show how to propagate discretisation uncertainty from the forward problem into a (Bayesian) inverse problem.

$$-\nabla \cdot (\theta(t)\nabla x(t)) = g(t) \quad t \in D$$
$$x(t) = b(t) \quad t \in \partial D$$

Problem 2

To make predictions with the PDE, coefficients $\theta(t)$ must be estimated. This is usually known as an inverse problem.

Solution

We will show how to propagate discretisation uncertainty from the forward problem into a (Bayesian) inverse problem.

Using an inaccurate forward solver in an inverse problem can produce biased and overconfident posteriors.

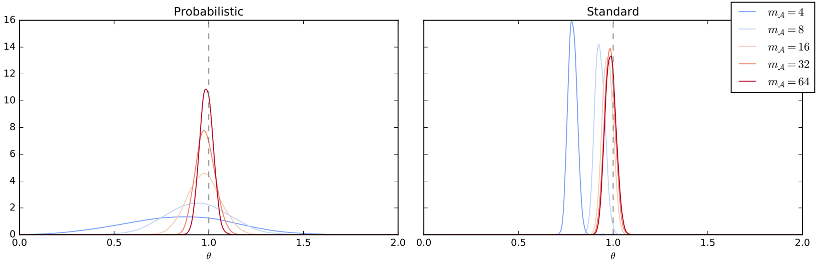


Figure: Comparison of inverse problem posteriors produced using a PN forward solver (left) vs. no PN (right).

Forward Problem

Replace the PDE operators with the abstract operators $\mathcal{A}$ and $\mathcal{B}$

$$-\nabla \cdot (\theta(t)\nabla x(t)) = g(t) \quad t \in D$$
$$x(t) = b(t) \quad t \in \partial D$$

Abstract Formulation

Replace the PDE operators with the abstract operators $\mathcal{A}$ and $\mathcal{B}$

$$\begin{aligned} \mathcal{A}x(t) &= g(t) \quad t \in D \\ \mathcal{B}x(t) &= b(t) \quad t \in \partial D \end{aligned}$$

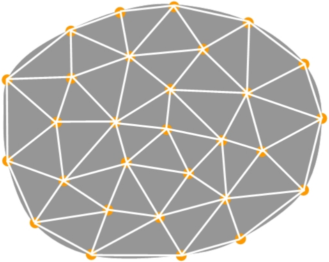
Abstract Formulation

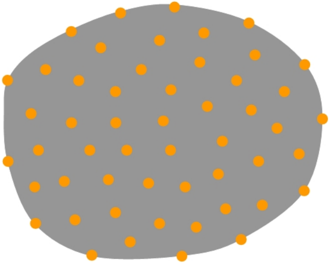
Replace the PDE operators with the abstract operators $\mathcal{A}$ and $\mathcal{B}$

$$\begin{aligned} \mathcal{A}x(t) &= g(t) \quad t \in D \\ \mathcal{B}x(t) &= b(t) \quad t \in \partial D \end{aligned}$$

Generally a solution $x(t)$ is not available in closed-form. Solvers are based on discretising the problem:

- Finite Differences
- Finite Volumes
- Symmetric Collocation





An example of a meshless method is symmetric collocation:

Let $k(t,t')$ to be a positive definite function, let $T = \{t_i\}_{i=1}^n$ and let

$$\begin{aligned} \hat{x}(t) &= \sum_{i=1}^{N} w_i \bar{\mathcal{A}} k(t, t_i) \\ &= \boldsymbol{w}^\top \bar{\mathcal{A}} K(t, T) \end{aligned}$$

where $\bar{\mathcal{A}}$ denotes the adjoint of $\mathcal{A}$ and

$$\bar{\mathcal{A}} K(t,T) := \begin{bmatrix} \bar{\mathcal{A}} k(t,t_1) \\ \vdots \\ \bar{\mathcal{A}} k(t,t_n) \end{bmatrix}.$$

An example of a meshless method is symmetric collocation:

Let $k(t,t')$ to be a positive definite function, let $T = \{t_i\}_{i=1}^n$ and let

$$\begin{aligned} \hat{x}(t) &= \sum_{i=1}^{N} w_i \bar{\mathcal{A}} k(t, t_i) \\ &= \boldsymbol{w}^\top \bar{\mathcal{A}} K(t, T) \end{aligned}$$

where $\bar{\mathcal{A}}$ denotes the adjoint of $\mathcal{A}$ and

$$\bar{\mathcal{A}} K(t, T) := \begin{bmatrix} \bar{\mathcal{A}} k(t, t_1) \\ \vdots \\ \bar{\mathcal{A}} k(t, t_n) \end{bmatrix}.$$

For linear $\mathcal{A}$, the weights $\boldsymbol{w}$ are uniquely determined by enforcing that $\mathcal{A}\hat{x}(t_i) = g_i := g(t_i)$ at each $i = 1, \dots, n$:

$$\boldsymbol{w} := [\mathcal{A}\bar{\mathcal{A}}K(T,T)]^{-1}\boldsymbol{g}$$

so that (and we ignore boundary conditions to reduce notation)

$$\hat{x}(t) = \bar{\mathcal{A}}K(t,T)[\mathcal{A}\bar{\mathcal{A}}K(T,T)]^{-1}\boldsymbol{g}.$$

If k is positive definite then it defines a Reproducing Kernel Hilbert Space and standard methods can be used to analyse the symmetric collocation method; e.g. Chapter 16 of Wendland [2004].

What about a Bayesian Probabilistic Numerical Method?

Symmetric Collocation

For linear $\mathcal{A}$, the weights $\boldsymbol{w}$ are uniquely determined by enforcing that $\mathcal{A}\hat{x}(t_i) = g_i := g(t_i)$ at each $i = 1, \ldots, n$:

$$\boldsymbol{w} := [\mathcal{A}\bar{\mathcal{A}}K(T,T)]^{-1}\boldsymbol{g}$$

so that (and we ignore boundary conditions to reduce notation)

$$\hat{x}(t) = \bar{\mathcal{A}}K(t,T)[\mathcal{A}\bar{\mathcal{A}}K(T,T)]^{-1}\boldsymbol{g}.$$

If k is positive definite then it defines a Reproducing Kernel Hilbert Space and standard methods can be used to analyse the symmetric collocation method; e.g. Chapter 16 of Wendland [2004].

What about a Bayesian Probabilistic Numerical Method?

For linear $\mathcal{A}$, the weights $\boldsymbol{w}$ are uniquely determined by enforcing that $\mathcal{A}\hat{x}(t_i) = g_i := g(t_i)$ at each $i = 1, \ldots, n$:

$$\boldsymbol{w} := [\mathcal{A}\bar{\mathcal{A}}K(T,T)]^{-1}\boldsymbol{g}$$

so that (and we ignore boundary conditions to reduce notation)

$$\hat{x}(t) = \bar{\mathcal{A}}K(t,T)[\mathcal{A}\bar{\mathcal{A}}K(T,T)]^{-1}\boldsymbol{g}.$$

If k is positive definite then it defines a Reproducing Kernel Hilbert Space and standard methods can be used to analyse the symmetric collocation method; e.g. Chapter 16 of Wendland [2004].

What about a Bayesian Probabilistic Numerical Method?

Let $P_x : x \sim \mathcal{GP}(0, k)$ be a Gaussian prior and consider the information operator

$$A(x) = \begin{bmatrix} \mathcal{A}x(t_1) \\ \vdots \\ \mathcal{A}x(t_n) \end{bmatrix}.$$

The Quantity of Interest here is just $Q(x) = x$.

Then the posterior $P_{x|a}$ is also Gaussian:

$$\begin{aligned} P_{x|a} &: x \sim \mathcal{GP}(m_1, \Sigma_1) \\ m_1(t) &= \bar{\mathcal{A}}K(t, T) \left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1} \boldsymbol{g} \\ \Sigma_1(t, t') &= k(t, t') - \bar{\mathcal{A}}K(t, T) \left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1} \mathcal{A}K(T, t') \end{aligned}$$

See e.g. Cockayne et al. [2016], Särkkä [2011], Cialenco et al. [2012], Owhadi [2015].

Observation: The mean function is the same as in symmetric collocation!

Let $P_x : x \sim \mathcal{GP}(0, k)$ be a Gaussian prior and consider the information operator

$$A(x) = \begin{bmatrix} \mathcal{A}x(t_1) \\ \vdots \\ \mathcal{A}x(t_n) \end{bmatrix}.$$

The Quantity of Interest here is just $Q(x) = x$.

Then the posterior $P_{x|a}$ is also Gaussian:

$$\begin{aligned} P_{x|a} : x &\sim \mathcal{GP}(m_1, \Sigma_1) \\ m_1(t) &= \bar{\mathcal{A}}K(t, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1} \boldsymbol{g} \\ \Sigma_1(t, t') &= k(t, t') - \bar{\mathcal{A}}K(t, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1} \mathcal{A}K(T, t') \end{aligned}$$

See e.g. Cockayne et al. [2016], Särkkä [2011], Cialenco et al. [2012], Owhadi [2015].

Observation: The mean function is the same as in symmetric collocation!

Let $P_x : x \sim \mathcal{GP}(0, k)$ be a Gaussian prior and consider the information operator

$$A(x) = \begin{bmatrix} \mathcal{A}x(t_1) \\ \vdots \\ \mathcal{A}x(t_n) \end{bmatrix}.$$

The Quantity of Interest here is just $Q(x) = x$.

Then the posterior $P_{x|a}$ is also Gaussian:

$$\begin{aligned} P_{x|a} : x &\sim \mathcal{GP}(m_1, \Sigma_1) \\ m_1(t) &= \bar{\mathcal{A}}K(t, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1}\boldsymbol{g} \\ \Sigma_1(t, t') &= k(t, t') - \bar{\mathcal{A}}K(t, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1}\mathcal{A}K(T, t') \end{aligned}$$

See e.g. Cockayne et al. [2016], Särkkä [2011], Cialenco et al. [2012], Owhadi [2015].

Observation: The mean function is the same as in symmetric collocation!

A Probabilistic Numerical Method

For the probabilistic numerical method, RKHS results reveal that:

$$P_{x|a}\{x' : \|x' - x\|_2 < \epsilon\} = 1 - O\left(\frac{h^{2\beta-2\rho-d}}{\epsilon}\right)$$

- h the fill distance of $T = \{t_i\}_{i=1}^n$
- β is related to the kernel k (e.g. order of the Sobolev native space, in the case of a Matérn kernel)
- $\rho < \beta - d/2$ the order of the differential operator $\mathcal{A}$
- d the dimension of D

Full details can be found in Cockayne et al. [2016].

Inverse Problem

We have solved the forward problem...

$$\begin{aligned} -\nabla \cdot (\theta(t) \nabla x(t)) &= g(t) \quad t \in D \\ x(t) &= b(t) \quad t \in \partial D \end{aligned}$$

Now we need to incorporate the forward posterior measure $P_{x|a}$ into the posterior measure for the inverse problem, θ

We have solved the forward problem...

$$-\nabla \cdot (\theta(t)\nabla x(t)) = g(t) \quad t \in D$$
$$x(t) = b(t) \quad t \in \partial D$$

Now we need to incorporate the forward posterior measure $P_{x|a}$ into the posterior measure for the inverse problem, θ

Inverse Problem: Given noisy data e.g.

$$y_i = x(t_i^{\text{obs}}; \theta) + \xi_i$$

$i = 1, \ldots, M$, estimate θ.

Could define a misfit

$$\|\boldsymbol{x}(\cdot;\theta) - \boldsymbol{y}\|_2$$

and seek to minimise it?

- If $\theta \in \mathbb{R}^N$ and $M < N$ then there will be many minimizers.
- If θ is a function then the problem will always be underdetermined.
- Noise ξ may be such that $\boldsymbol{y}$ is not attainable for any θ

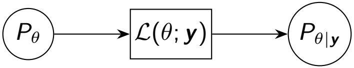
Could define a misfit

$$\|\boldsymbol{x}(\cdot;\theta) - \boldsymbol{y}\|_2$$

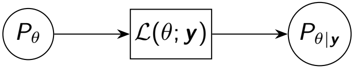
and seek to minimise it?

- If $\theta \in \mathbb{R}^N$ and $M < N$ then there will be many minimizers.
- If θ is a function then the problem will always be underdetermined.
- Noise ξ may be such that $\boldsymbol{y}$ is not attainable for any θ

Bayesian Inverse Problem [Stuart, 2010]:



Bayesian Inverse Problem [Stuart, 2010]:



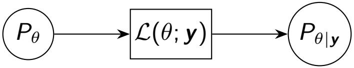
- Prior: P_θ, belief about θ before observing information.
- Likelihood $\mathcal{L}$: a model for "how likely" particular θ are, e.g.:

$$\mathcal{L}(\theta; \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x}(\cdot;\theta) - \boldsymbol{y}\|_2^2}{2\sigma^2}\right)$$

- Posterior: $P_{\theta|\boldsymbol{y}}$, belief about θ after observing $\boldsymbol{y}$.

The Inverse Problem

Bayesian Inverse Problem [Stuart, 2010]:



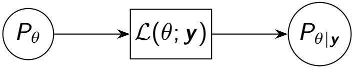
- Prior: P_θ, belief about θ before observing information.
- Likelihood $\mathcal{L}$: a model for "how likely" particular θ are, e.g.:

$$\mathcal{L}(\theta; \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x}(\cdot;\theta) - \boldsymbol{y}\|_2^2}{2\sigma^2}\right)$$

- Posterior: $P_{\theta|\boldsymbol{y}}$, belief about θ after observing $\boldsymbol{y}$.

The Inverse Problem

Bayesian Inverse Problem [Stuart, 2010]:



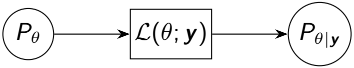
- Prior: P_θ, belief about θ before observing information.
- Likelihood $\mathcal{L}$: a model for "how likely" particular θ are, e.g.:

$$\mathcal{L}(\theta; \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x}(\cdot;\theta) - \boldsymbol{y}\|_2^2}{2\sigma^2}\right)$$

- Posterior: $P_{\theta|\boldsymbol{y}}$, belief about θ after observing $\boldsymbol{y}$.

The Inverse Problem

Bayesian Inverse Problem [Stuart, 2010]:



The posterior can be found by Bayes Theorem:

$$\frac{\mathrm{d}P_{\theta|\boldsymbol{y}}}{\mathrm{d}P_\theta} \quad \propto \quad \mathcal{L}(\theta; \boldsymbol{y})$$

Discretisation Error

In PDE inverse problems the likelihood $\mathcal{L}(\theta; \boldsymbol{y})$ depends on the unknown solution $x(\cdot; \theta)$ of the PDE.

Assuming the data in the inverse problem is:

$$y_i = x(t_i^{\text{obs}}) + \xi_i \qquad i = 1, \dots, n$$
$$\boldsymbol{\xi} \sim N(\boldsymbol{0}, \Gamma)$$

implies the standard likelihood:

$$\mathcal{L}(\theta; \boldsymbol{y}) \sim N(\boldsymbol{y}; \boldsymbol{x}(\cdot; \theta), \Gamma)$$

This is intractable because $x(\cdot; \theta)$ is unknown.

In PDE inverse problems the likelihood $\mathcal{L}(\theta; \boldsymbol{y})$ depends on the unknown solution $x(\cdot;\theta)$ of the PDE.

Assuming the data in the inverse problem is:

$$y_i = x(t_i^{\text{obs}}) + \xi_i \qquad\qquad i = 1, \dots, n$$
$$\boldsymbol{\xi} \sim N(\mathbf{0}, \Gamma)$$

implies the standard likelihood:

$$\mathcal{L}(\theta; \boldsymbol{y}) \sim N(\boldsymbol{y}; \boldsymbol{x}(\cdot;\theta), \Gamma)$$

This is intractable because $x(\cdot;\theta)$ is unknown.

Discretisation Error

Common approach: replace x with $\hat{x}_N$ given by some numerical solver, and "hope for the best":

$$\hat{\mathcal{L}}_N(\theta; \mathbf{y}) = \exp\left(-\frac{\|\hat{\mathbf{x}}_N(\cdot;\theta) - \mathbf{y}\|_2^2}{2\sigma^2} \right)$$

...which we have already seen can go wrong!

Seminal results in Stuart [2010] shows that under certain assumptions, the convergence of $\hat{x}^N \to x$ transfers to a rate in the approximate posterior $P^N_{\theta|\mathbf{y}} \to P_{\theta|\mathbf{y}}$:

$$\left| \log \hat{\mathcal{L}}_N(\theta; \mathbf{y}) - \log \mathcal{L}(\theta; \mathbf{y}) \right| \leq C\varphi(N)$$

for some constant C.

But this says nothing about the error in the non-asymptotic limit!

Common approach: replace x with $\hat{x}_N$ given by some numerical solver, and "hope for the best":

$$\hat{\mathcal{L}}_N(\theta; \boldsymbol{y}) = \exp\left(-\frac{\|\hat{\boldsymbol{x}}_N(\cdot;\theta) - \boldsymbol{y}\|_2^2}{2\sigma^2}\right)$$

...which we have already seen can go wrong!

Seminal results in Stuart [2010] shows that under certain assumptions, the convergence of $\hat{x}^N \to x$ transfers to a rate in the approximate posterior $P^N_{\theta|\boldsymbol{y}} \to P_{\theta|\boldsymbol{y}}$:

$$\left|\log \hat{\mathcal{L}}_N(\theta; \boldsymbol{y}) - \log \mathcal{L}(\theta; \boldsymbol{y})\right| \leq C\varphi(N)$$

for some constant C.

But this says nothing about the error in the non-asymptotic limit!

Common approach: replace x with $\hat{x}_N$ given by some numerical solver, and "hope for the best":

$$\hat{\mathcal{L}}_N(\theta; \boldsymbol{y}) = \exp\left(-\frac{\|\hat{\boldsymbol{x}}_N(\cdot;\theta) - \boldsymbol{y}\|_2^2}{2\sigma^2}\right)$$

...which we have already seen can go wrong!

Seminal results in Stuart [2010] shows that under certain assumptions, the convergence of $\hat{x}^N \to x$ transfers to a rate in the approximate posterior $P^N_{\theta|\boldsymbol{y}} \to P_{\theta|\boldsymbol{y}}$:

$$\left|\log \hat{\mathcal{L}}_N(\theta; \boldsymbol{y}) - \log \mathcal{L}(\theta; \boldsymbol{y})\right| \leq C\varphi(N)$$

for some constant C.

But this says nothing about the error in the non-asymptotic limit!

An elegant solution based on the Bayesian Probabilistic Numerical Method: Marginalise the unknown solution x according to the output $P_{x|a}$ of the Probabilistic Numerical Method, to obtain a "PN" likelihood:

$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x) \mathrm{d}P_{x|a}$$
$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1)$$

where m_1 and Σ_1 arise from the Probabilistic Numerical Method. e.g.

$$\Sigma_1 = K(T^{\text{obs}}, T^{\text{obs}}) - \bar{\mathcal{A}}K(T^{\text{obs}}, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1}\mathcal{A}K(T, T^{\text{obs}})$$

This carries similar convergence results to the "standard" method as the number n of points in $T = \{t_i\}_{i=1}^{n}$ is increased (strictly, as the fill distance h is decreased).

However, unlike the standard method, it provides full uncertainty quantification.

Let's see a couple of applications...

Forward $\mapsto$ Inverse Problem

An elegant solution based on the Bayesian Probabilistic Numerical Method: Marginalise the unknown solution x according to the output $P_{x|a}$ of the Probabilistic Numerical Method, to obtain a "PN" likelihood:

$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x)\mathrm{d}P_{x|a}$$
$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1)$$

where m_1 and Σ_1 arise from the Probabilistic Numerical Method. e.g.

$$\Sigma_1 = K(T^{\text{obs}}, T^{\text{obs}}) - \bar{\mathcal{A}}K(T^{\text{obs}}, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1}\mathcal{A}K(T, T^{\text{obs}})$$

This carries similar convergence results to the "standard" method as the number n of points in $T = \{t_i\}_{i=1}^n$ is increased (strictly, as the fill distance h is decreased).

However, unlike the standard method, it provides full uncertainty quantification.

Let's see a couple of applications...

An elegant solution based on the Bayesian Probabilistic Numerical Method: Marginalise the unknown solution x according to the output $P_{x|a}$ of the Probabilistic Numerical Method, to obtain a "PN" likelihood:

$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x)\mathrm{d}P_{x|a}$$
$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1)$$

where m_1 and Σ_1 arise from the Probabilistic Numerical Method. e.g.

$$\Sigma_1 = K(T^{\text{obs}}, T^{\text{obs}}) - \bar{\mathcal{A}}K(T^{\text{obs}}, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1}\mathcal{A}K(T, T^{\text{obs}})$$

This carries similar convergence results to the "standard" method as the number n of points in $T = \{t_i\}_{i=1}^n$ is increased (strictly, as the fill distance h is decreased).

However, unlike the standard method, it provides full uncertainty quantification.

Let's see a couple of applications...

An elegant solution based on the Bayesian Probabilistic Numerical Method: Marginalise the unknown solution x according to the output $P_{x|a}$ of the Probabilistic Numerical Method, to obtain a "PN" likelihood:

$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x)\mathrm{d}P_{x|a}$$

$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1)$$

where m_1 and Σ_1 arise from the Probabilistic Numerical Method. e.g.

$$\Sigma_1 = K(T^{\text{obs}}, T^{\text{obs}}) - \bar{\mathcal{A}}K(T^{\text{obs}}, T)\left[\mathcal{A}\bar{\mathcal{A}}K(T, T)\right]^{-1}\mathcal{A}K(T, T^{\text{obs}})$$

This carries similar convergence results to the "standard" method as the number n of points in $T = \{t_i\}_{i=1}^n$ is increased (strictly, as the fill distance h is decreased).

However, unlike the standard method, it provides full uncertainty quantification.

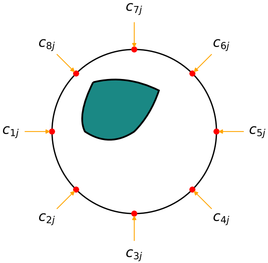
Let's see a couple of applications...

Electrical Impedance Tomography

A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.

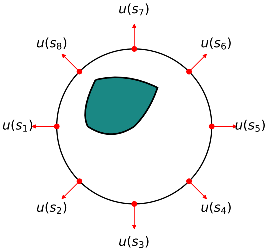
Electrical Impedance Tomography

A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.



Many patterns of current c_{ij}, $j = 1, \ldots, N_c$ injected through boundary electrodes t_i^{obs}, $i = 1, \ldots, N_s$

A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.



Resulting voltage measured: $y_i = x(t_i^{\text{obs}}) - x(t_{\text{ref}}) + \epsilon_i$

Electrical Impedance Tomography

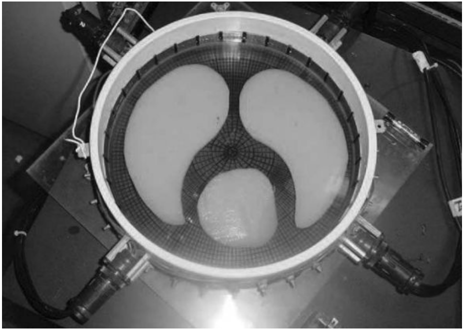
A medical imaging technique. Goal: reconstruct interior conductivity field of a patient, to detect tumors.

Governing equations are essentially Darcy's law:

$$-\nabla \cdot (\theta(t)\nabla x(t)) = 0 \qquad\qquad t \in D$$

$$\theta(t_i^{\text{obs}})\frac{\partial x}{\partial n}(t_i^{\text{obs}}) = c_{ij} \qquad\qquad i = 1, \dots, N_S$$

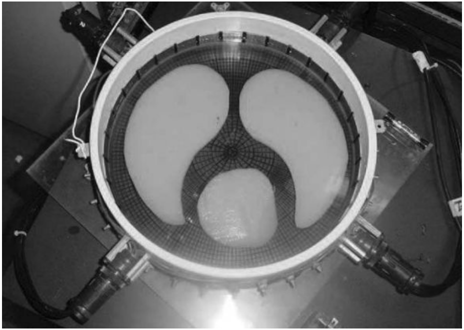
Experiments due to Isaacson et al. [2004].



- Tank filled with saline.
- Three targets:
  - "Heart shaped": higher conductivity.
  - "Lung shaped": lower conductivity.
- 32 equally spaced electrodes.
- Simultaneously stimulated for 31 different stimulation patterns.

Experimental Set-Up

Experiments due to Isaacson et al. [2004].



- Tank filled with saline.
- Three targets:
  - "Heart shaped": higher conductivity.
  - "Lung shaped": lower conductivity.
- 32 equally spaced electrodes.
- Simultaneously stimulated for 31 different stimulation patterns.

A Hard Problem...

- High dimensional (992) observations.
- Observations are only of the boundary - weak information.
- Target $\theta(\cdot)$ is infinite-dimensional.
- The "ideal" likelihood $\mathcal{L}(\theta; \boldsymbol{y})$ requires exact solution of the PDE.

Posteriors obtained using the PN likelihood

$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x) \mathrm{d}P_{x|a}$$

$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1).$$

Focus on varying the number n of points in $T = \{t_i\}_{i=1}^{n}$ that are used.

Computation facilitated with Markov chain Monte Carlo, based on the preconditioned Crank-Nicholson proposal.

- High dimensional (992) observations.
- Observations are only of the boundary - weak information.
- Target $\theta(\cdot)$ is infinite-dimensional.
- The "ideal" likelihood $\mathcal{L}(\theta; \boldsymbol{y})$ requires exact solution of the PDE.

Posteriors obtained using the PN likelihood

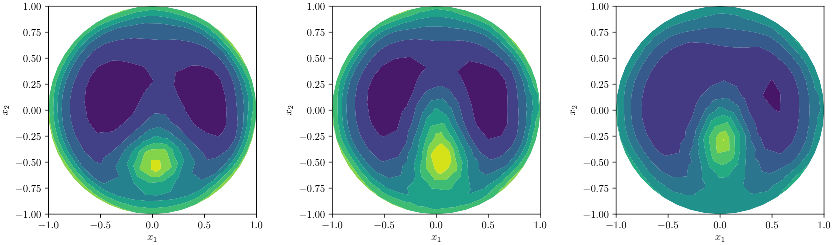
$$\mathcal{L}_n(\theta; \boldsymbol{y}) \propto \int p(\boldsymbol{y}|\theta, x) \mathrm{d}P_{x|a}$$
$$\implies \boldsymbol{y}|\theta \sim N(\boldsymbol{m}_1, \Gamma + \Sigma_1).$$

Focus on varying the number n of points in $T = \{t_i\}_{i=1}^n$ that are used.

Computation facilitated with Markov chain Monte Carlo, based on the preconditioned Crank-Nicholson proposal.

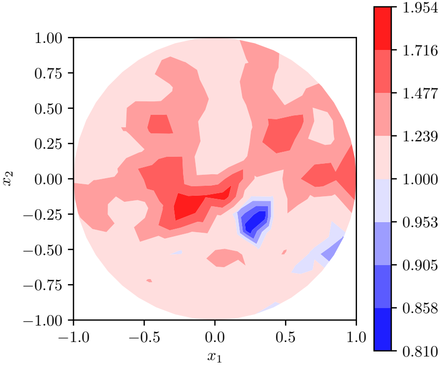
Recovered Fields

Posterior means $m(t) = \mathbb{E}_{\boldsymbol{y}}[\theta(t)]$:



(a) $n = 96$ (b) $n = 127$ (c) $n = 165$

Ratio of (pointwise) posterior variance $v(t) = \mathbb{V}_{\boldsymbol{y}}[\theta(t)]$ computed from the PN posterior based on $\mathcal{L}_n$ and the "standard" posterior based on $\hat{\mathcal{L}}_N$ with $n = N = 96$:



Allen–Cahn

A prototypical non-linear PDE:

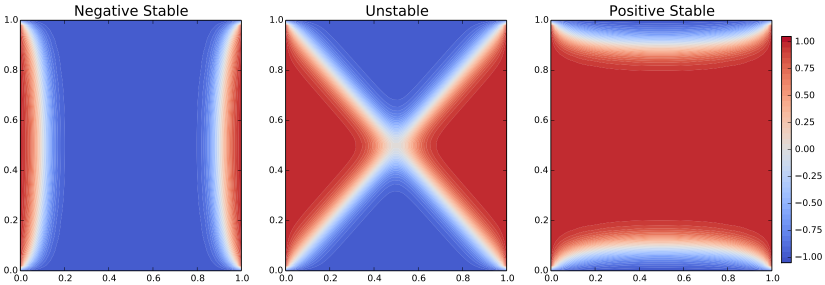
$$\begin{aligned} -\theta \nabla^2 x(t) + \theta^{-1}(x(t)^3 - x(t)) &= 0 && t \in (0,1)^2 \\ x(t) &= 1 && t_1 \in \{0,1\}\,; 0 < t_2 < 1 \\ x(t) &= -1 && t_2 \in \{0,1\}\,; 0 < t_1 < 1 \end{aligned}$$

Goal: Infer θ from (16) noisy observations $y_i = x(t_i^{\text{obs}}) + \epsilon_i$ (over a regular grid $\{t_i^{\text{obs}}\}$ in the interior).

A prototypical non-linear PDE:

$$
\begin{aligned}
-\theta \nabla^2 x(t) + \theta^{-1}(x(t)^3 - x(t)) &= 0 \qquad t \in (0,1)^2 \\
x(t) &= 1 \qquad t_1 \in \{0,1\}\,;\, 0 < t_2 < 1 \\
x(t) &= -1 \quad\; t_2 \in \{0,1\}\,;\, 0 < t_1 < 1
\end{aligned}
$$

True data-generating parameter was $\theta = 0.04$. Leads to multiple solutions:



Nonlinear PDE $\implies$ the conjugate Gaussian structure is broken!

Numerical disintegration?

A simpler "trick" for semi-linear PDEs:

Nonlinear PDE $\implies$ the conjugate Gaussian structure is broken!

Numerical disintegration?

A simpler "trick" for semi-linear PDEs:

Nonlinear PDE $\implies$ the conjugate Gaussian structure is broken!

Numerical disintegration?

A simpler "trick" for semi-linear PDEs:

$$-\theta \nabla^2 x(t) + \theta^{-1}(x(t)^3 - x(t)) = 0 \quad (1)$$

split the operator...

$$-\theta \nabla^2 x(t) - \theta^{-1} x(t) = z \quad (2)$$

$$\theta^{-1} x(t)^3 = -z \quad (3)$$

(1) = (2) + (3)

Nonlinear PDE $\implies$ the conjugate Gaussian structure is broken!

Numerical disintegration?

A simpler "trick" for semi-linear PDEs:

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

...and invert

$$\begin{aligned} -\theta \nabla^2 x(t) - \theta^{-1} x(t) &= z \\ x(t) &= \sqrt[3]{-\theta z} \end{aligned}$$

Nonlinear PDE $\implies$ the conjugate Gaussian structure is broken!

Numerical disintegration?

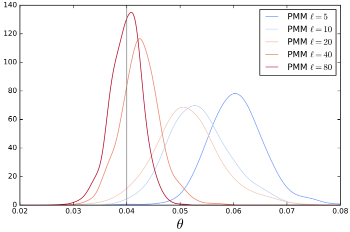
A simpler "trick" for semi-linear PDEs: $\implies$ Solve the new system

$$\begin{aligned} \mathcal{A}_1 x(t) &:= -\theta \nabla^2 x(t) - \theta^{-1} x(t) &&= z \\ \mathcal{A}_2 x(t) &:= x(t) &&= \sqrt[3]{-\theta z} \end{aligned}$$

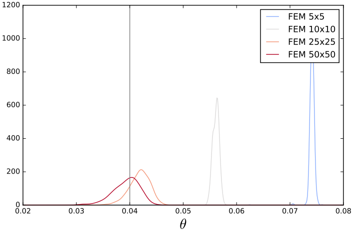
... and z can be marginalised by importance sampling[2].

[2] Details in Cockayne et al. [2016]

Allen–Cahn: Inverse Problem



(a) Probabilistic Numerical Method



(b) Standard Method (FEA)

Comparison of posteriors for θ obtained with (a) the probabilistic PDE solver and (b) a standard PDE solver based on Finite Element Analysis (FEA).

Ninth Job: Characterise Optimal Information

Original example from Sul'din (1959):

Consider

$$\mathcal{X} = \{x : [0,1] \to \mathbb{R} \text{ such that } x(0) = 0\}$$

and numerical integration:

$$\begin{aligned} A(x) &= \begin{bmatrix} x(t_1) \\ \vdots \\ x(t_n) \end{bmatrix} \\ Q(x) &= \int_0^1 x(t) \mathrm{d}t \end{aligned}$$

Here the prior distribution P_x will be the Weiner measure on $\mathcal{X}$.

Our goal is to determine the average case optimal method (w.r.t. P_x) of the form

$$b(a) = \sum_{i=1}^{n} w_i a_i \qquad \left(= \sum_{i=1}^{n} w_i x(t_i) \right)$$

i.e. choose optimal weights $w_1, \ldots, w_n$ and knots $t_1, \ldots, t_n$ to minimise the average error.

Optimality here is measured with the loss function $L(q, q') = (q - q')^2$.

Step #1: An explicit expression for the average error

$$\int [b(A(x)) - Q(x)]^2 P_x(\mathrm{d}x)$$

$$= \int_{\mathcal{X}} \left(\sum_{i=1}^{n} w_i x(t_i) - \int_0^1 x(t)\mathrm{d}t \right)^2 P_x(\mathrm{d}x)$$

$$= \int_{\mathcal{X}} \left(\int_0^1 x(t)\mathrm{d}t \right)^2 P_x(\mathrm{d}x) - 2\sum_{i=1}^{n} w_i \int_{\mathcal{X}} \left(\int_0^1 x(t)\mathrm{d}t \right) \cdot x(t_i) P_x(\mathrm{d}x)$$

$$+ \sum_{i,j=1}^{n} w_i w_j \operatorname{cov}(x(t_i), x(t_j)) \qquad \text{(Fubini)}$$

$$= \frac{1}{3} - 2\sum_{i=1}^{n} w_i \cdot \left(t_i - \frac{t_i^2}{2} \right) + \sum_{i,j=1}^{n} w_i w_j \min(t_i, t_j) \quad \text{(Def'n of } P_x\text{)}$$

Step #1: An explicit expression for the average error

$$\int [b(A(x)) - Q(x)]^2 P_x(\mathrm{d}x)$$

$$= \int_{\mathcal{X}} \left(\sum_{i=1}^{n} w_i x(t_i) - \int_0^1 x(t)\mathrm{d}t \right)^2 P_x(\mathrm{d}x)$$

$$= \int_{\mathcal{X}} \left(\int_0^1 x(t)\mathrm{d}t \right)^2 P_x(\mathrm{d}x) - 2 \sum_{i=1}^{n} w_i \int_{\mathcal{X}} \left(\int_0^1 x(t)\mathrm{d}t \right) \cdot x(t_i) P_x(\mathrm{d}x)$$

$$+ \sum_{i,j=1}^{n} w_i w_j \operatorname{cov}(x(t_i), x(t_j)) \qquad \text{(Fubini)}$$

$$= \frac{1}{3} - 2 \sum_{i=1}^{n} w_i \cdot \left(t_i - \frac{t_i^2}{2} \right) + \sum_{i,j=1}^{n} w_i w_j \min(t_i, t_j) \quad (\text{Def'n of } P_x)$$

Step #1: An explicit expression for the average error

$$\int [b(A(x)) - Q(x)]^2 P_x(\mathrm{d}x)$$

$$= \int_{\mathcal{X}} \left(\sum_{i=1}^{n} w_i x(t_i) - \int_0^1 x(t)\mathrm{d}t \right)^2 P_x(\mathrm{d}x)$$

$$= \int_{\mathcal{X}} \left(\int_0^1 x(t)\mathrm{d}t \right)^2 P_x(\mathrm{d}x) - 2\sum_{i=1}^{n} w_i \int_{\mathcal{X}} \left(\int_0^1 x(t)\mathrm{d}t \right) \cdot x(t_i) P_x(\mathrm{d}x)$$

$$+ \sum_{i,j=1}^{n} w_i w_j \operatorname{cov}(x(t_i), x(t_j)) \qquad \text{(Fubini)}$$

$$= \frac{1}{3} - 2\sum_{i=1}^{n} w_i \cdot \left(t_i - \frac{t_i^2}{2} \right) + \sum_{i,j=1}^{n} w_i w_j \min(t_i, t_j) \quad \text{(Def'n of } P_x\text{)}$$

Step #2: Optimise weights given locations

$$\text{objective} = \frac{1}{3} - 2\sum_{i=1}^{n} w_i \cdot \left(t_i - \frac{t_i^2}{2}\right) + \sum_{i,j=1}^{n} w_i w_j \min(t_i, t_j)$$

$$= \frac{1}{3} - 2\boldsymbol{w} \cdot \boldsymbol{c} + \boldsymbol{w}' \cdot \Sigma \cdot \boldsymbol{w}$$

This is a quadratic problem with solution

$$\boldsymbol{w} = \Sigma^{-1}\boldsymbol{c}.$$

Step #2: Optimise weights given locations

$$\begin{aligned}\text{objective} &= \frac{1}{3} - 2\sum_{i=1}^{n} w_i \cdot \left(t_i - \frac{t_i^2}{2}\right) + \sum_{i,j=1}^{n} w_i w_j \min(t_i, t_j) \\ &= \frac{1}{3} - 2\boldsymbol{w} \cdot \boldsymbol{c} + \boldsymbol{w}' \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{w}\end{aligned}$$

This is a quadratic problem with solution

$$\boldsymbol{w} = \Sigma^{-1}\boldsymbol{c}.$$

Example: Optimal Information for an Integral

Step #2: Optimise weights given locations

$$\text{objective} = \frac{1}{3} - 2\sum_{i=1}^{n} w_i \cdot \left(t_i - \frac{t_i^2}{2}\right) + \sum_{i,j=1}^{n} w_i w_j \min(t_i, t_j)$$

$$= \frac{1}{3} - 2\boldsymbol{w} \cdot \boldsymbol{c} + \boldsymbol{w}' \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{w}$$

This is a quadratic problem with solution

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}\boldsymbol{c}.$$

Example: Optimal Information for an Integral

Step #2: Optimise weights given locations

The solution corresponds to the method:

$$b(a) = x(t_1) \cdot \frac{t_2}{2} + \sum_{i=2}^{n-1} x(t_i) \cdot \frac{t_{i+1} - t_{i-1}}{2} + x(t_n) \cdot \left(1 - \frac{t_n + t_{n-1}}{2}\right)$$

This is a trapezoidal rule, based on the data $x(t_i)$, the fact $x(0) = 0$, and the assumption $x(1) = x(t_n)$.

Example: Optimal Information for an Integral

Step #3: Optimise locations

Average case error of the trapezoidal rule:

$$\text{objective} \quad = \quad \frac{1}{3}(1 - t_n)^3 + \frac{1}{12}\sum_{i=1}^{n}(t_i - t_{i-1})^3$$

This can be minimised with elementary calculus.

The solution corresponds to the method:

$$b(a) = \frac{2}{2n+1}\sum_{i=1}^{n} a_i, \quad a_i = x(t_i), \quad x_i = \frac{2i}{2n+1}$$

so that the average case optimal method has evenly spaced knots.

But what about optimal information for Bayesian Probabilistic Numerical Methods?

Step #3: Optimise locations

Average case error of the trapezoidal rule:

$$\text{objective} \quad = \quad \frac{1}{3}(1-t_n)^3 + \frac{1}{12}\sum_{i=1}^{n}(t_i - t_{i-1})^3$$

This can be minimised with elementary calculus.

The solution corresponds to the method:

$$b(a) = \frac{2}{2n+1}\sum_{i=1}^{n} a_i, \quad a_i = x(t_i), \quad x_i = \frac{2i}{2n+1}$$

so that the average case optimal method has evenly spaced knots.

But what about optimal information for Bayesian Probabilistic Numerical Methods?

Step #3: Optimise locations

Average case error of the trapezoidal rule:

$$\text{objective} \quad = \quad \frac{1}{3}(1 - t_n)^3 + \frac{1}{12}\sum_{i=1}^{n}(t_i - t_{i-1})^3$$

This can be minimised with elementary calculus.

The solution corresponds to the method:

$$b(a) = \frac{2}{2n+1}\sum_{i=1}^{n} a_i, \quad a_i = x(t_i), \quad x_i = \frac{2i}{2n+1}$$

so that the average case optimal method has evenly spaced knots.

But what about optimal information for Bayesian Probabilistic Numerical Methods?

The contribution of Kadane and Wasilkowski [1985]:

Consider a classical numerical method (A, b) with information operator $A : \mathcal{X} \to \mathcal{A}$, such that $A \in \Lambda$ for some set Λ, and estimator $b : \mathcal{A} \to \mathcal{Q}$. Let $L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ be a loss function that is pre-specified. Then consider the minimal average case error

$$\inf_{A \in \Lambda, b} \int L(b(A(x)), Q(x)) \mathrm{d}P_x.$$

The minimiser $b(\cdot)$ is a non-randomised Bayes rule and the minimiser A is "optimal information" over Λ, or optimal experimental design for this numerical task.

Generalisation of optimal information to probabilistic numerical methods?

The contribution of Kadane and Wasilkowski [1985]:

Consider a classical numerical method (A, b) with information operator $A : \mathcal{X} \to \mathcal{A}$, such that $A \in \Lambda$ for some set Λ, and estimator $b : \mathcal{A} \to \mathcal{Q}$. Let $L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ be a loss function that is pre-specified. Then consider the minimal average case error

$$\inf_{A \in \Lambda, b} \int L(b(A(x)), Q(x)) \mathrm{d}P_x.$$

The minimiser $b(\cdot)$ is a non-randomised Bayes rule and the minimiser A is "optimal information" over Λ, or optimal experimental design for this numerical task.

Generalisation of optimal information to probabilistic numerical methods?

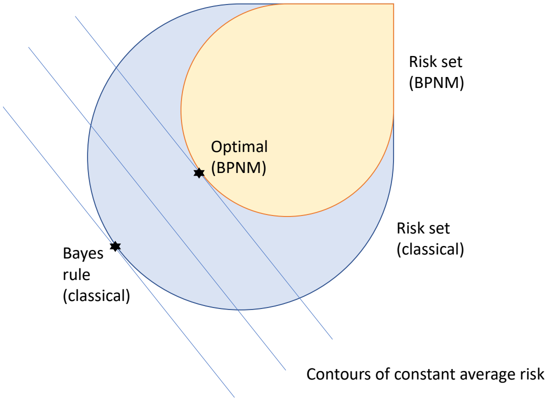
For Bayesian probabilistic numerical methods $B(P_x, a) = Q_{\#} P_{x|a}$, optimal information is defined as

$$\arg\inf_{A \in \Lambda} \int \int L(Q_{\#} P_{x|A(x)}(\omega), Q(x)) \mathrm{d}P_x \, \mathrm{d}\omega.$$

Important point: The Bayesian probabilistic numerical method output $Q_{\#} P_{x|a}$ will not in general be supported on the set of Bayes acts. This presents a non-trivial constraint on the risk set...

Optimal Information

Average Case Analysis $\overset{1985}{\leftrightarrow}$ Bayesian Decision Theory $\overset{?}{\leftrightarrow}$ Bayesian Probabilistic Numerical Methods



Optimal Information

In Cockayne et al. [2017] we established the following (new) result:

Let $(\mathcal{Q}, \langle \cdot, \cdot \rangle_{\mathcal{Q}})$ be an inner-product space with associated norm $\| \cdot \|_{\mathcal{Q}}$ and consider the canonical loss $L(q, q') = \|q - q'\|_{\mathcal{Q}}^2$. Then optimal information for Bayesian probabilistic numerical methods coincides with average-case optimal information.

The assumption is non-trivial:

Consider the following counter-example:

- $\mathcal{X} = \{b, c, d, e\}$,
- $Q(x) = 1[x = b]$,
- P_x uniform,
- $A(x) = 1[x \in S]$, where we are allowed either $S = \{b, c\}$ or $\{b, c, d\}$,
- $L(q, q') = 1[q \neq q']$.

Then average-case optimal information can be either $S = \{b, c\}$ or $\{b, c, d\}$. On the other hand, optimal information in the Bayesian probabilistic numerical context is just $S = \{b, c\}$.

Optimal Information

In Cockayne et al. [2017] we established the following (new) result:

Let $(\mathcal{Q}, \langle \cdot, \cdot \rangle_{\mathcal{Q}})$ be an inner-product space with associated norm $\| \cdot \|_{\mathcal{Q}}$ and consider the canonical loss $L(q, q') = \|q - q'\|_{\mathcal{Q}}^2$. Then optimal information for Bayesian probabilistic numerical methods coincides with average-case optimal information.

The assumption is non-trivial:

Consider the following counter-example:

- $\mathcal{X} = \{b, c, d, e\}$,
- $Q(x) = 1[x = b]$,
- P_x uniform,
- $A(x) = 1[x \in S]$, where we are allowed either $S = \{b, c\}$ or $\{b, c, d\}$,
- $L(q, q') = 1[q \neq q']$.

Then average-case optimal information can be either $S = \{b, c\}$ or $\{b, c, d\}$. On the other hand, optimal information in the Bayesian probabilistic numerical context is just $S = \{b, c\}$.

Optimal Information

In Cockayne et al. [2017] we established the following (new) result:

Let $(\mathcal{Q}, \langle \cdot, \cdot \rangle_{\mathcal{Q}})$ be an inner-product space with associated norm $\| \cdot \|_{\mathcal{Q}}$ and consider the canonical loss $L(q, q') = \|q - q'\|_{\mathcal{Q}}^2$. Then optimal information for Bayesian probabilistic numerical methods coincides with average-case optimal information.

The assumption is non-trivial:

Consider the following counter-example:

- $\mathcal{X} = \{b, c, d, e\}$,
- $Q(x) = 1[x = b]$,
- P_x uniform,
- $A(x) = 1[x \in S]$, where we are allowed either $S = \{b, c\}$ or $\{b, c, d\}$,
- $L(q, q') = 1[q \neq q']$.

Then average-case optimal information can be either $S = \{b, c\}$ or $\{b, c, d\}$. On the other hand, optimal information in the Bayesian probabilistic numerical context is just $S = \{b, c\}$.

Return to the original example of Sul'din (1959):

From the previous result, since $\mathcal{Q} = \mathbb{R}$ is an inner-product space equipped with the loss function $L(q, q') = (q - q')^2$, it follows that the optimal information for Bayesian probabilistic numerical method coincides with average case optimal information.

Thus the optimal Bayesian Probabilistic Numerical Method (w.r.t. P_x) is:

$$B(P_x, a) = \mathrm{N}\left(\frac{2}{2n+1} \sum_{i=1}^{n} a_i, \; \frac{1}{3(2n+1)^2} \right)$$

N.B. The variance $\frac{1}{3(2n+1)^2}$ is twice the optimal average error.

Example: Optimal Information for an Integral

Return to the original example of Sul'din (1959):

From the previous result, since $\mathcal{Q} = \mathbb{R}$ is an inner-product space equipped with the loss function $L(q, q') = (q - q')^2$, it follows that the optimal information for Bayesian probabilistic numerical method coincides with average case optimal information.

Thus the optimal Bayesian Probabilistic Numerical Method (w.r.t. P_x) is:

$$B(P_x, a) = \mathrm{N}\left(\frac{2}{2n+1}\sum_{i=1}^{n} a_i, \ \frac{1}{3(2n+1)^2}\right)$$

N.B. The variance $\frac{1}{3(2n+1)^2}$ is twice the optimal average error.

Example: Optimal Information for an Integral

Return to the original example of Sul'din (1959):

From the previous result, since $\mathcal{Q} = \mathbb{R}$ is an inner-product space equipped with the loss function $L(q, q') = (q - q')^2$, it follows that the optimal information for Bayesian probabilistic numerical method coincides with average case optimal information.

Thus the optimal Bayesian Probabilistic Numerical Method (w.r.t. P_x) is:

$$B(P_x, a) = \mathrm{N}\left(\frac{2}{2n+1} \sum_{i=1}^{n} a_i, \ \frac{1}{3(2n+1)^2} \right)$$

N.B. The variance $\frac{1}{3(2n+1)^2}$ is twice the optimal average error.

Conclusion

In Part IV it has been argued that:

- Over-confident inferences for unknown parameters in PDEs, due to ignoring discretisation error, can be mitigated with Bayesian Probabilistic Numerical Methods (BPNM).
- However, for general (non-linear) PDEs the "offline" computations can be difficult.
- Optimal information for BPNM is not always equivalent to average-case optimal information - but for "nice" problems the two are identical.

END OF PART IV

In Part IV it has been argued that:

- Over-confident inferences for unknown parameters in PDEs, due to ignoring discretisation error, can be mitigated with Bayesian Probabilistic Numerical Methods (BPNM).
- However, for general (non-linear) PDEs the "offline" computations can be difficult.
- Optimal information for BPNM is not always equivalent to average-case optimal information - but for "nice" problems the two are identical.

END OF PART IV

Conclusion

In Part IV it has been argued that:

- Over-confident inferences for unknown parameters in PDEs, due to ignoring discretisation error, can be mitigated with Bayesian Probabilistic Numerical Methods (BPNM).
- However, for general (non-linear) PDEs the "offline" computations can be difficult.
- Optimal information for BPNM is not always equivalent to average-case optimal information - but for "nice" problems the two are identical.

END OF PART IV

Conclusion

In Part IV it has been argued that:

- Over-confident inferences for unknown parameters in PDEs, due to ignoring discretisation error, can be mitigated with Bayesian Probabilistic Numerical Methods (BPNM).
- However, for general (non-linear) PDEs the "offline" computations can be difficult.
- Optimal information for BPNM is not always equivalent to average-case optimal information - but for "nice" problems the two are identical.

END OF PART IV

Conclusion

In Part IV it has been argued that:

- Over-confident inferences for unknown parameters in PDEs, due to ignoring discretisation error, can be mitigated with Bayesian Probabilistic Numerical Methods (BPNM).
- However, for general (non-linear) PDEs the "offline" computations can be difficult.
- Optimal information for BPNM is not always equivalent to average-case optimal information - but for "nice" problems the two are identical.

END OF PART IV