Bayesian Probabilistic Numerical Methods

Tim Sullivan[1,2], Jon Cockayne[3], Chris Oates[4,5] & Mark Girolami[5,6]

Uncertainty Quantification for Stochastic Systems and Applications

IPAM, California, US

14 November 2017

[1]Free University of Berlin, DE
[2]Zuse Institute Berlin, DE
[3]University of Warwick, UK
[4]Newcastle University, UK
[5]Alan Turing Institute, London, UK
[6]Imperial College London, UK

A Probabilistic Treatment of Numerics?

- The last 5 years have seen a renewed interest in probabilistic perspectives on numerical tasks such as quadrature, ODE and PDE solution, optimisation, etc., building upon a long history of such ideas (Poincaré, 1896; Larkin, 1970; Diaconis, 1988; Skilling, 1992).
- Many ways to motivate this modelling choice:
    - To a statistician's eye, numerical tasks look like inverse problems.
    - Worst-case errors are often too pessimistic — perhaps we should adopt an average-case viewpoint (Traub et al., 1988; Ritter, 2000)?
    - If discretisation error is not properly accounted for, then biased and over-confident inferences result. But the necessary numerical analysis in nonlinear and evolutionary contexts is **hard**!
    - Accounting for the impact of discretisation error in a statistical way allows forward and Bayesian inverse problems to speak a common statistical language.
- To make these ideas precise and to relate them to one another, some concrete definitions are needed!

Outline

An Inference Perspective on Numerics

An Abstract View of Numerical Methods

An abstract setting for numerical tasks consists of three spaces and two functions:

- $\mathcal{X}$, where an unknown/variable object x or u lives; $\dim \mathcal{X} = \infty$
- $\mathcal{A}$, where we observe information $A(x)$, via a function $A \colon \mathcal{X} \to \mathcal{A}$; $\dim \mathcal{A} < \infty$
- $\mathcal{Q}$, with a quantity of interest $Q \colon \mathcal{X} \to \mathcal{Q}$.

An abstract setting for numerical tasks consists of three spaces and two functions:

- $\mathcal{X}$, where an unknown/variable object x or u lives; $\dim \mathcal{X} = \infty$
- $\mathcal{A}$, where we observe information $A(x)$, via a function $A \colon \mathcal{X} \to \mathcal{A}$; $\dim \mathcal{A} < \infty$
- $\mathcal{Q}$, with a quantity of interest $Q \colon \mathcal{X} \to \mathcal{Q}$.

Example (Quadrature)

$$\mathcal{X} = C^0([0,1];\mathbb{R}) \qquad \mathcal{A} = ([0,1] \times \mathbb{R})^m \qquad \mathcal{Q} = \mathbb{R}$$

$$A(u) = (t_i, u(t_i))_{i=1}^m \qquad Q(u) = \int_0^1 u(t)\,\mathrm{d}t$$

An abstract setting for numerical tasks consists of three spaces and two functions:

- $\mathcal{X}$, where an unknown/variable object x or u lives; $\dim \mathcal{X} = \infty$
- $\mathcal{A}$, where we observe information $A(x)$, via a function $A \colon \mathcal{X} \to \mathcal{A}$; $\dim \mathcal{A} < \infty$
- $\mathcal{Q}$, with a quantity of interest $Q \colon \mathcal{X} \to \mathcal{Q}$.

Example (Solving the PDE $-\Delta u = f$ in strong form)

$$\mathcal{X} = H_0^1(\Omega; \mathbb{R}) \qquad \mathcal{A} = (\Omega \times \mathbb{R})^m \qquad \mathcal{Q} = \mathcal{X}$$

$$A(u) = (t_i, -\Delta u(t_i))_{i=1}^m \qquad Q(u) = u$$

An Abstract View of Numerical Methods

An abstract setting for numerical tasks consists of three spaces and two functions:

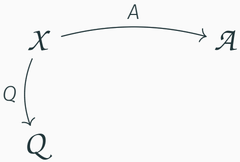
- $\mathcal{X}$, where an unknown/variable object x or u lives; $\dim \mathcal{X} = \infty$
- $\mathcal{A}$, where we observe information $A(x)$, via a function $A \colon \mathcal{X} \to \mathcal{A}$; $\dim \mathcal{A} < \infty$
- $\mathcal{Q}$, with a quantity of interest $Q \colon \mathcal{X} \to \mathcal{Q}$.

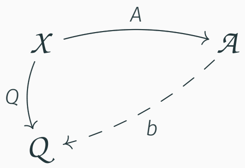
Example (Quadrature)

$$\mathcal{X} = C^0([0,1];\mathbb{R}) \qquad \mathcal{A} = ([0,1] \times \mathbb{R})^m \qquad \mathcal{Q} = \mathbb{R}$$

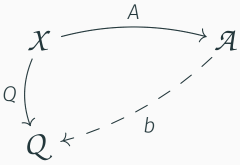
$$A(u) = (t_i, u(t_i))_{i=1}^m \qquad Q(u) = \int_0^1 u(t)\,\mathrm{d}t$$

- Conventional numerical methods are cleverly-designed functions $b \colon \mathcal{A} \to \mathcal{Q}$: they estimate $Q(x)$ by $b(A(x))$.
- Gauss/Sard/Larkin (1970): Does $b \circ A \approx Q$?
- N.B. *Some* methods try to "invert" A, form an estimate of x, then apply Q.


A
X
𝒜
Q
Q



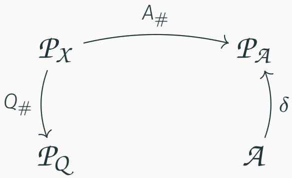
$$b\colon \mathcal{A} \to \mathcal{Q}$$

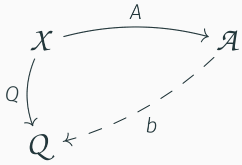


$b\colon \mathcal{A} \to \mathcal{Q}$

Go
Probabilistic!

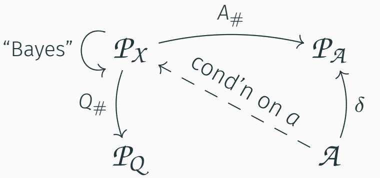
$\bullet \dashrightarrow \mathcal{P}_\bullet$



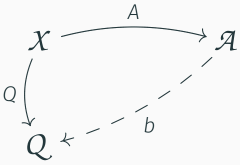


$b\colon \mathcal{A} \to \mathcal{Q}$

Go
Probabilistic!
$\bullet \dashrightarrow \mathcal{P}_\bullet$



$B\colon \mathcal{P}_\mathcal{X} \times \mathcal{A} \to \mathcal{P}_\mathcal{Q}$

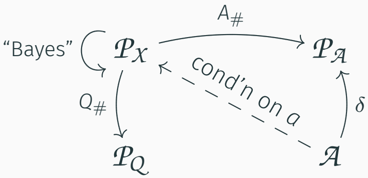


$b \colon \mathcal{A} \to \mathcal{Q}$

Go

Probabilistic!

$\bullet \dashrightarrow \mathcal{P}_\bullet$



$B \colon \mathcal{P}_\mathcal{X} \times \mathcal{A} \to \mathcal{P}_\mathcal{Q}$

Example (Quadrature)

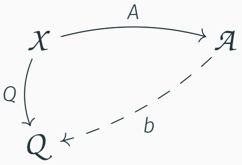
$$\mathcal{X} = C^0([0,1];\mathbb{R}) \qquad \mathcal{A} = ([0,1] \times \mathbb{R})^m \qquad \mathcal{Q} = \mathbb{R}$$

$$A(u) = (t_i, u(t_i))_{i=1}^m \qquad Q(u) = \int_0^1 u(t)\,\mathrm{d}t$$

A deterministic numerical method uses only the spaces and data to produce a point estimate of the integral.

A probabilistic numerical method converts an additional belief about the integrand into a belief about the integral.

Rev. Bayes Does Some Numerics I

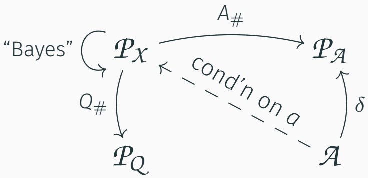


$b\colon \mathcal{A} \to \mathcal{Q}$

Go

Probabilistic!

$\bullet \dashrightarrow \mathcal{P}_\bullet$



$B\colon \mathcal{P}_{\mathcal{X}} \times \mathcal{A} \to \mathcal{P}_{\mathcal{Q}}$

Definition (Bayesian PNM)

A PNM B, with prior $\mu \in \mathcal{P}_{\mathcal{X}}$, is **Bayesian** for a quantity of interest Q if its output is the push-forward of the conditional distribution $\mu^a := \mu(\cdot|a)$ through Q:

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

Definition (Bayesian PNM)

A PNM B, with prior $\mu \in \mathcal{P}_{\mathcal{X}}$, is Bayesian for a quantity of interest Q if its output is the push-forward of the conditional distribution $\mu^a := \mu(\cdot|a)$ through Q:

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

Definition (Bayesian PNM)

A PNM B, with prior $\mu \in \mathcal{P}_{\mathcal{X}}$, is **Bayesian** for a quantity of interest Q if its output is the push-forward of the conditional distribution $\mu^a := \mu(\cdot|a)$ through Q:

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

Example (Sul′din, 1959, 1960)

- Under the Gaussian Brownian motion prior on $\mathcal{X} = C^0([0, 1]; \mathbb{R})$, the posterior mean / MAP estimator for the definite integral is the trapezoidal rule, i.e. integration using linear interpolation.
- The integrated Brownian motion prior corresponds to integration using cubic spline interpolation.

Method	QoI $Q(x)$	Information $A(x)$	Non-Bayesian PNMs	Bayesian PNMs[1]
Integrator	$\int x(t)\nu(\mathrm{d}t)$	$\{x(t_i)\}_{i=1}^n$	Approximate Bayesian Quadrature Methods [Osborne et al., 2012b,a, Gunter et al., 2014]	Bayesian Quadrature [Diaconis, 1988, O'Hagan, 1991, Ghahramani and Rasmussen, 2002, Briol et al., 2016]
	$\int f(t)x(\mathrm{d}t)$	$\{t_i\}_{i=1}^n$ s.t. $t_i \sim x$	Kong et al. [2003], Tan [2004], Kong et al. [2007]	
	$\int x_1(t)x_2(\mathrm{d}t)$	$\{(t_i, x_1(t_i))\}_{i=1}^n$ s.t. $t_i \sim x_2$		Oates et al. [2016]
Optimiser	$\arg\min x(t)$	$\{x(t_i)\}_{i=1}^n$		Bayesian Optimisation [Mockus, 1989][6]
		$\{\nabla x(t_i)\}_{i=1}^n$		Hennig and Kiefel [2013]
		$\{(x(t_i), \nabla x(t_i))\}_{i=1}^n$		Probabilistic Line Search [Mahsereci and Hennig, 2015]
		$\{\mathbb{I}[t_{\min} < t_i]\}_{i=1}^n$		Probabilistic Bisection Algorithm [Horstein, 1963][5]
		$\{\mathbb{I}[t_{\min} < t_i] + \text{error}\}_{i=1}^n$	Waeber et al. [2013]	
Linear Solver	$x^{-1}b$	$\{xt_i\}_{i=1}^n$		Probabilistic Linear Solvers [Hennig, 2015, Bartels and Hennig, 2016]
ODE Solver	x	$\{\nabla x(t_i)\}_{i=1}^n$	Filtering Methods for IVPs [Schober et al., 2014, Chkrebtii et al., 2016, Kersting and Hennig, 2016, Teymur et al., 2016, Schober et al., 2016][4] Finite Difference Methods [John and Wu, 2017][7]	Skilling [1992]
		∇x + rounding error	Hull and Swenson [1966], Mosbach and Turner [2009][2]	
	$x(t_{\text{end}})$	$\{\nabla x(t_i)\}_{i=1}^n$	Stochastic Euler [Krebs, 2016]	
PDE Solver	x	$\{Dx(t_i)\}_{i=1}^n$	Chkrebtii et al. [2016]	Probabilistic Meshless Methods [Owhadi, 2015a,b, Cockayne et al., 2016, Raissi et al., 2016]
		Dx + discretisation error	Conrad et al. [2016][3]	

Generalising Bayes' Theorem

- Thus, we are expressing PNMs in terms of Bayesian inverse problems (Stuart, 2010).
- But a naïve interpretation of Bayes' rule makes no sense here, because

$$\operatorname{supp}(\mu^a) \subseteq \mathcal{X}^a := \{x \in \mathcal{X} \mid A(x) = a\},$$

  typically $\mu(\mathcal{X}^a) = 0$, and — in contrast to typical statistical inverse problems — we think of the observation process as noiseless.
- E.g. quadrature example from earlier, with $A(u) = (t_i, u(t_i))_{i=1}^m$.
- Thus, we cannot take the usual approach of defining μ^a via its prior density as

$$\frac{\mathrm{d}\mu^a}{\mathrm{d}\mu}(x) \propto \text{likelihood}(x|a)$$

  because this density 'wants' to be the indicator function $\mathbb{1}[x \in \mathcal{X}^a]$.
- While linear-algebraic tricks work for linear conditioning of Gaussians, in general we condition on events of measure zero using disintegration.

Write

$$\mu(f) \equiv \mathbb{E}_{\mu}[f] \equiv \int_{\mathcal{X}} f(x)\, \mu(\mathrm{d}x)$$

Definition (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 a measurable map $A\colon \mathcal{X} \to \mathcal{A}$ if:

- $\mu^{a}(\mathcal{X} \setminus \mathcal{X}^{a}) = 0$ for $A_{\#}\mu$-almost all $a \in \mathcal{A}$; (support)

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

- $a \mapsto \mu^{a}(f)$ is measurable; (measurability)
- $\mu(f) = A_{\#}\mu\big(\mu^{a}(f)\big)$. (conditioning/reconstruction)

$$\text{i.e.} \quad \int_{\mathcal{X}} f(x)\, \mu(\mathrm{d}x) = \int_{\mathcal{A}} \left[\int_{\mathcal{X}^{a}} f(x)\, \mu^{a}(\mathrm{d}x) \right] (A_{\#}\mu)(\mathrm{d}a).$$

Theorem (Disintegration theorem (Chang and Pollard, 1997, Thm. 1))

Let $\mathcal{X}$ be a metric space and let $\mu \in \mathcal{P}_{\mathcal{X}}$ be inner regular. If the Borel σ-algebra on $\mathcal{X}$ is countably generated and contains all singletons $\{a\}$ for $a \in \mathcal{A}$, then there is an essentially unique disintegration $\{\mu^a\}_{a \in \mathcal{A}}$ of μ with respect to A. (If $\{\nu^a\}_{a \in \mathcal{A}}$ is another such disintegration, then $\{a \in \mathcal{A} : \mu^a \neq \nu^a\}$ is an $A_{\#}\mu$-null set.)

Example (Trivial example)

A (essentially, the!) disintegration of uniform measure on $\{(x_1, x_2) \mid x_1, x_2 \in \mathcal{X}\}$ with respect to 'vertical projection' $A((x_1, x_2)) := x_1$ is that μ^a is uniform measure on the vertical fibre $\mathcal{X}^a = \{(a, x_2) \mid x_2 \in [0, 1]\}$.

In general, disintegrations cannot be computed exactly — we have to work approximately.

Numerical Disintegration

- The exact disintegration "$\mu^a(\mathrm{d}x) \propto \mathbb{1}[A(x) = a]\, \mu(\mathrm{d}x)$" can be accessed numerically via relaxation, with approximation guarantees provided $a \mapsto \mu^a$ is 'nice', e.g. $A_{\#}\mu \in \mathcal{P}_{\mathcal{A}}$ has a smooth Lebesgue density.
- Consider relaxed posterior $\mu^a_\delta(\mathrm{d}x) \propto \phi(\|A(x) - a\|_{\mathcal{A}}/\delta)\, \mu(\mathrm{d}x)$.
  - Essentially any $\phi\colon [0, \infty) \to [0, 1]$ tending continuously to 1 at 0 and decaying quickly enough to 0 at ∞ will do.
  - E.g. $\phi(r) := \mathbb{1}[r < 1]$ or $\phi(r) := \exp(-r^2)$.
- Integral probability metric with respect to a normed space $\mathcal{F}$ of test functions:

$$d_{\mathcal{F}}(\mu, \nu) := \sup\bigl\{ |\mu(f) - \nu(f)| \bigm| \|f\|_{\mathcal{F}} \leq 1 \bigr\}.$$

- $\mathcal{F}$ = bounded continuous functions with uniform norm $\leftrightarrow$ total variation.
- $\mathcal{F}$ = bounded Lipschitz continuous functions with Lipschitz norm $\leftrightarrow$ Wasserstein.

$$\text{“}\mu^{a}(\mathrm{d}x) \propto \mathbb{1}[A(x) = a]\,\mu(\mathrm{d}x)\text{”}$$
$$\mu^{a}_{\delta}(\mathrm{d}x) \propto \phi(\|A(x) - a\|_{\mathcal{A}}/\delta)\,\mu(\mathrm{d}x)$$
$$d_{\mathcal{F}}(\mu, \nu) := \sup\{|\mu(f) - \nu(f)| \,|\, \|f\|_{\mathcal{F}} \leq 1\}$$

Theorem (Cockayne et al., 2017b, Theorem 4.3)

If $a \mapsto \mu^{a}$ is γ-Hölder from $(\mathcal{A}, \|\cdot\|_{\mathcal{A}})$ into $(\mathcal{P}_{\mathcal{X}}, d_{\mathcal{F}})$, i.e.

$$d_{\mathcal{F}}(\mu^{a}, \mu^{a'}) \leq C\|a - a'\|^{\gamma} \quad \text{for } a, a' \in \mathcal{A},$$

then so too is the approximation $\mu^{a}_{\delta} \approx \mu^{a}$ as a function of δ:

$$d_{\mathcal{F}}(\mu^{a}, \mu^{a}_{\delta}) \leq C\delta^{\gamma} \quad \text{for } A_{\#}\mu\text{-almost all } a \in \mathcal{A}.$$

(The change of Hölder constants depends only on the rate of decay of ϕ.)

To sample μ_δ^a we take inspiration from rare event simulation and use tempering schemes to sample the posterior.

Set $\delta_0 > \delta_1 > \ldots > \delta_N$ and consider

$$\mu_{\delta_0}^a,\ \mu_{\delta_1}^a,\ \ldots,\ \mu_{\delta_N}^a$$

- $\mu_{\delta_0}^a$ is easy to sample — often $\mu_{\delta_0}^a = \mu$.
- $\mu_{\delta_N}^a$ has δ_N close to zero and is hard to sample.
- Intermediate distributions define a “ladder” which takes us from prior to posterior.
- Even within this framework, there is considerable choice of sampling scheme, e.g. brute-force MCMC, SMC, QMC, pCN, ...

A multivalent boundary value problem:

$$
\begin{aligned}
u''(t) - u(t)^2 &= -t && \text{for } t \geq 0 \\
u(0) &= 0 \\
u(t)/\sqrt{t} &\to 1 && \text{as } t \to +\infty
\end{aligned}
$$

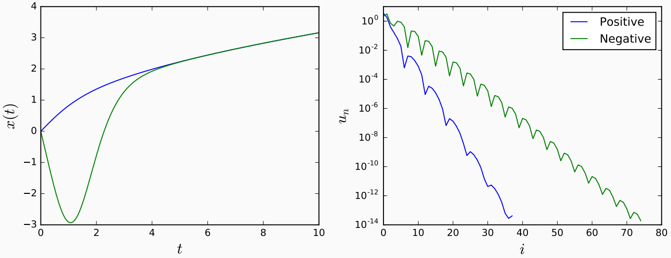


Figure 1: The two solutions of Painlevé's first transcendental and their spectra in the orthonormal Chebyshev polynomial basis over $[0, 10]$.

A multivalent boundary value problem:

$$u''(t) - u(t)^2 = -t \qquad \text{for } t \geq 0$$
$$u(0) = 0$$
$$u(10) = \sqrt{10}$$

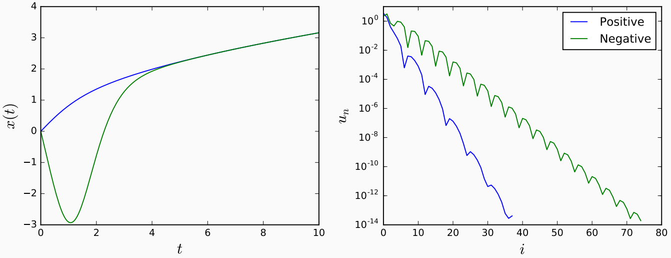
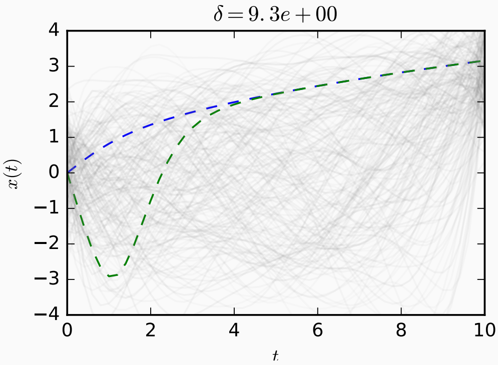


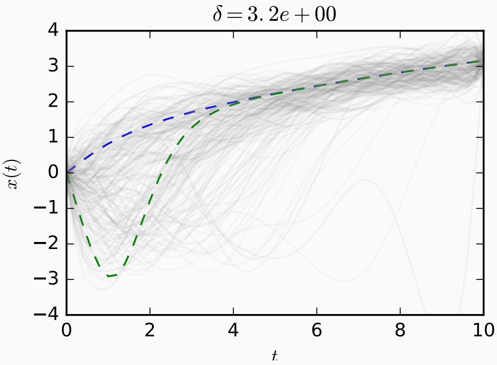
Figure 1: The two solutions of Painlevé's first transcendental and their spectra in the orthonormal Chebyshev polynomial basis over $[0, 10]$.

- We use SMC-based numerical disintegration with $\phi(r) := \exp(-r^2)$, 1600 δ-values log-spaced from $\delta = 10$ to $\delta = 10^{-4}$, appluing/observing the PDE at 15 equi-spaced points in $[0, 10]$.
- A centred Gauss or Cauchy prior on Chebyshev coefficients recovers the positive solution — can bias to get the negative. ✓/✗

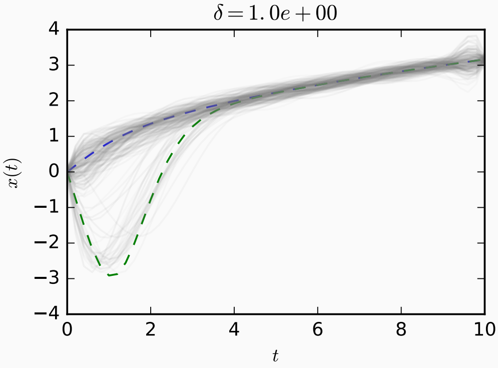


Example: Painlevé's First Transcendental II

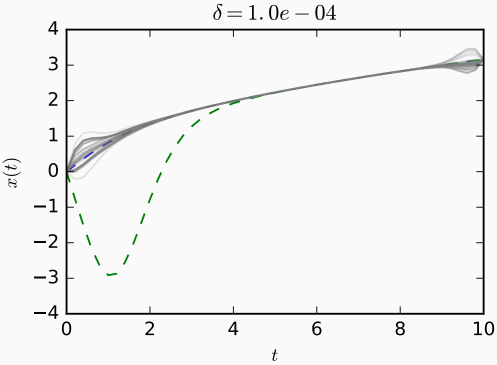
- We use SMC-based numerical disintegration with $\phi(r) := \exp(-r^2)$, 1600 δ-values log-spaced from $\delta = 10$ to $\delta = 10^{-4}$, appluing/observing the PDE at 15 equi-spaced points in $[0, 10]$.
- A centred Gauss or Cauchy prior on Chebyshev coefficients recovers the positive solution — can bias to get the negative. ✓/✗



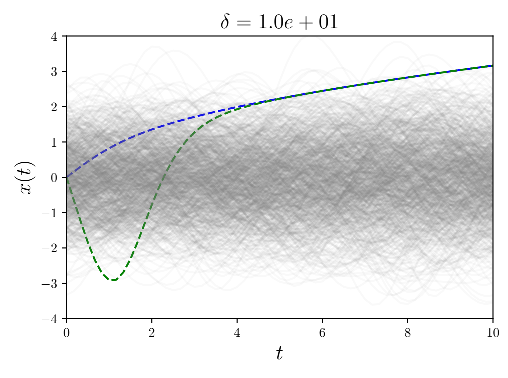
- We use SMC-based numerical disintegration with $\phi(r) := \exp(-r^2)$, 1600 δ-values log-spaced from $\delta = 10$ to $\delta = 10^{-4}$, appluing/observing the PDE at 15 equi-spaced points in $[0, 10]$.
- A centred Gauss or Cauchy prior on Chebyshev coefficients recovers the positive solution — can bias to get the negative. ✓/✗



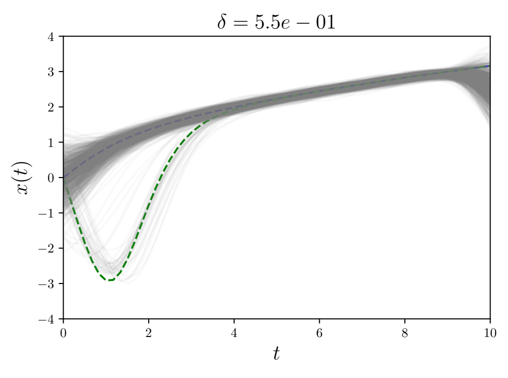
- We use SMC-based numerical disintegration with $\phi(r) := \exp(-r^2)$, 1600 δ-values log-spaced from $\delta = 10$ to $\delta = 10^{-4}$, appluing/observing the PDE at 15 equi-spaced points in $[0, 10]$.
- A centred Gauss or Cauchy prior on Chebyshev coefficients recovers the positive solution — can bias to get the negative. ✓/✗



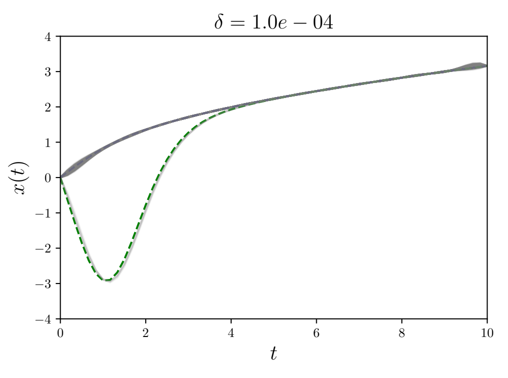
- The choice of sampler does matter: replacing SMC with parallel tempered pCN with 100 δ-values log-spaced from $\delta = 10$ to $\delta = 10^{-4}$ and 10^8 iterations relieves the positive bias. ✓
- Both solutions survive to small δ, approximately the same proportions as the posterior densities at the two exact solutions. ⁉



- The choice of sampler does matter: replacing SMC with parallel tempered pCN with 100 δ-values log-spaced from $\delta = 10$ to $\delta = 10^{-4}$ and 10^8 iterations relieves the positive bias. ✓
- Both solutions survive to small δ, approximately the same proportions as the posterior densities at the two exact solutions. ⁉



- The choice of sampler does matter: replacing SMC with parallel tempered pCN with 100 δ-values log-spaced from $\delta = 10$ to $\delta = 10^{-4}$ and 10^8 iterations relieves the positive bias. ✓
- Both solutions survive to small δ, approximately the same proportions as the posterior densities at the two exact solutions. ⁉



Optimal Information

Suppose we have a loss function $L\colon \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$.

- The worst-case error for a classical numerical method $b\colon \mathcal{A} \to \mathcal{Q}$ is

$$e_{\mathrm{WC}}(A, b) := \sup_{x \in \mathcal{X}} L\big(b(A(x)), Q(x)\big).$$

- The average-case error under a probability measure $\mu \in \mathcal{P}_{\mathcal{X}}$ is

$$e_{\mathrm{AC}}(A, b) := \int_{\mathcal{X}} L\big(b(A(x)), Q(x)\big)\, \mu(\mathrm{d}x).$$

  Kadane and Wasilkowski (1985) show that the minimiser b is a non-random Bayes decision rule, and the minimiser A is "optimal information" for this task.

- A BPNM B has "no choice" but to be $Q_{\sharp}\mu^{a}$ once $A(x) = a$ is given; optimality of A means minimising

$$e_{\mathrm{BPN}}(A) := \int_{\mathcal{X}} \left[\int_{\mathcal{Q}} L(q, Q(x))\, (Q_{\sharp}\mu^{A(x)})(\mathrm{d}q) \right] \mu(\mathrm{d}x).$$

Theorem (Cockayne et al., 2017b)

For a quadratic loss $L(q, q') := \|q - q'\|_{\mathcal{Q}}^2$ on a Hilbert space $\mathcal{Q}$, optimal information for BPNM and ACE coincide (though the minimal values may differ).

Theorem (Cockayne et al., 2017b)

For a quadratic loss $L(q, q') := \|q - q'\|_{\mathcal{Q}}^2$ on a Hilbert space $\mathcal{Q}$, optimal information for BPNM and ACE coincide (though the minimal values may differ).

Example

Decide whether or not a card drawn fairly at random is ♦, incurring unit loss if you guess wrongly; can choose to be told whether the card is red (A_1) or is non-♣ (A_2).

$$\mathcal{X} = \{♣, ♦, ♥, ♠\} \qquad \mathcal{A} = \{0, 1\} \subset \mathbb{R} \qquad \mathcal{Q} = \{0, 1\} \subset \mathbb{R}$$

$$A_1(x) = \mathbb{1}[x \in \{♦, ♥\}] \qquad A_2(x) = \mathbb{1}[x \in \{♦, ♥, ♠\}] \qquad Q(x) = \mathbb{1}[x = ♦]$$

$$\mu = \mathrm{Unif}_{\mathcal{X}} \qquad L(q, q') = \mathbb{1}[q \neq q']$$

Which information operator, A_1 or A_2, is better? (Note that $e_{\mathrm{WC}}(A_i, b) = 1$ for all deterministic b!)

Optimal Information: AC $\neq$ BPN!

$$\mathcal{X} = \{♣, ♦, ♥, ♠\} \qquad \mathcal{A} = \{0, 1\} \subset \mathbb{R} \qquad \mathcal{Q} = \{0, 1\} \subset \mathbb{R}$$

$$A_1(x) = \mathbb{1}[x \in \{♦, ♥\}] \qquad A_2(x) = \mathbb{1}[x \in \{♦, ♥, ♠\}] \qquad Q(x) = \mathbb{1}[x = ♦]$$

$$\mu = \mathrm{Unif}_{\mathcal{X}} \qquad L(q, q') = \mathbb{1}[q \neq q']$$

reality $\rightarrow$		♣		♦		♥		♠		
$e_{AC}(A_1, b)$	$= \frac{1}{4}($	$\mathbb{1}[b(0) \neq 0]$	$+$	$\mathbb{1}[b(1) \neq 1]$	$+$	$\mathbb{1}[b(1) \neq 0]$	$+$	$\mathbb{1}[b(0) \neq 0]$	$)$	
$e_{AC}(A_1, b \equiv 0)$	$= \frac{1}{4}($	0	$+$	1	$+$	0	$+$	0	$)$	$= \frac{1}{4}$
$e_{AC}(A_1, b = \mathrm{id})$	$= \frac{1}{4}($	0	$+$	0	$+$	1	$+$	0	$)$	$= \frac{1}{4}$

Optimal Information: AC $\neq$ BPN!

$$\mathcal{X} = \{♣, ♦, ♥, ♠\} \qquad \mathcal{A} = \{0,1\} \subset \mathbb{R} \qquad \mathcal{Q} = \{0,1\} \subset \mathbb{R}$$

$$A_1(x) = \mathbb{1}[x \in \{♦, ♥\}] \qquad A_2(x) = \mathbb{1}[x \in \{♦, ♥, ♠\}] \qquad Q(x) = \mathbb{1}[x = ♦]$$

$$\mu = \mathrm{Unif}_{\mathcal{X}} \qquad L(q, q') = \mathbb{1}[q \neq q']$$

reality →		♣		♦		♥		♠		
$e_{AC}(A_1, b)$	$= \frac{1}{4}($	$\mathbb{1}[b(0) \neq 0]$	$+$	$\mathbb{1}[b(1) \neq 1]$	$+$	$\mathbb{1}[b(1) \neq 0]$	$+$	$\mathbb{1}[b(0) \neq 0]$	$)$	
$e_{AC}(A_1, b \equiv 0)$	$= \frac{1}{4}($	0	$+$	1	$+$	0	$+$	0	$)$	$= \frac{1}{4}$
$e_{AC}(A_1, b = \mathrm{id})$	$= \frac{1}{4}($	0	$+$	0	$+$	1	$+$	0	$)$	$= \frac{1}{4}$
$e_{AC}(A_2, b)$	$= \frac{1}{4}($	$\mathbb{1}[b(0) \neq 0]$	$+$	$\mathbb{1}[b(1) \neq 1]$	$+$	$\mathbb{1}[b(1) \neq 0]$	$+$	$\mathbb{1}[b(1) \neq 0]$	$)$	
$e_{AC}(A_2, b \equiv 0)$	$= \frac{1}{4}($	0	$+$	1	$+$	0	$+$	0	$)$	$= \frac{1}{4}$

Optimal Information: AC $\neq$ BPN!

$$\mathcal{X} = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\} \qquad \mathcal{A} = \{0,1\} \subset \mathbb{R} \qquad \mathcal{Q} = \{0,1\} \subset \mathbb{R}$$

$$A_1(x) = \mathbb{1}[x \in \{\diamondsuit, \heartsuit\}] \qquad A_2(x) = \mathbb{1}[x \in \{\diamondsuit, \heartsuit, \spadesuit\}] \qquad Q(x) = \mathbb{1}[x = \diamondsuit]$$

$$\mu = \mathrm{Unif}_{\mathcal{X}} \qquad L(q, q') = \mathbb{1}[q \neq q']$$

reality $\rightarrow$		$\clubsuit$		$\diamondsuit$		$\heartsuit$		$\spadesuit$	
$e_{\mathrm{AC}}(A_1, b)$	$= \frac{1}{4}($	$\mathbb{1}[b(0) \neq 0]$	$+$	$\mathbb{1}[b(1) \neq 1]$	$+$	$\mathbb{1}[b(1) \neq 0]$	$+$	$\mathbb{1}[b(0) \neq 0]$	$)$
$e_{\mathrm{AC}}(A_1, b \equiv 0)$	$= \frac{1}{4}($	0	$+$	1	$+$	0	$+$	0	$) = \frac{1}{4}$
$e_{\mathrm{AC}}(A_1, b = \mathrm{id})$	$= \frac{1}{4}($	0	$+$	0	$+$	1	$+$	0	$) = \frac{1}{4}$
$e_{\mathrm{AC}}(A_2, b)$	$= \frac{1}{4}($	$\mathbb{1}[b(0) \neq 0]$	$+$	$\mathbb{1}[b(1) \neq 1]$	$+$	$\mathbb{1}[b(1) \neq 0]$	$+$	$\mathbb{1}[b(1) \neq 0]$	$)$
$e_{\mathrm{AC}}(A_2, b \equiv 0)$	$= \frac{1}{4}($	0	$+$	1	$+$	0	$+$	0	$) = \frac{1}{4}$
$e_{\mathrm{BPN}}(A_1)$	$= \frac{1}{4}($	$\mathbb{E}_{Q_\sharp \mu^0} L(\cdot, 0)$	$+$	$\mathbb{E}_{Q_\sharp \mu^1} L(\cdot, 1)$	$+$	$\mathbb{E}_{Q_\sharp \mu^1} L(\cdot, 0)$	$+$	$\mathbb{E}_{Q_\sharp \mu^0} L(\cdot, 0)$	$)$
	$= \frac{1}{4}($	$(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0)$	$+$	$(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1)$	$+$	$(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0)$	$+$	$(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0)$	$) = \frac{1}{4}$

Optimal Information: AC $\neq$ BPN!

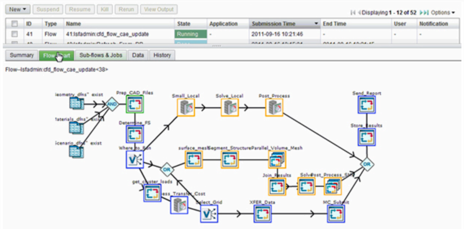
$$\mathcal{X} = \{\clubsuit, \diamondsuit, \heartsuit, \spadesuit\} \qquad \mathcal{A} = \{0,1\} \subset \mathbb{R} \qquad \mathcal{Q} = \{0,1\} \subset \mathbb{R}$$

$$A_1(x) = \mathbb{1}[x \in \{\diamondsuit, \heartsuit\}] \qquad A_2(x) = \mathbb{1}[x \in \{\diamondsuit, \heartsuit, \spadesuit\}] \qquad Q(x) = \mathbb{1}[x = \diamondsuit]$$

$$\mu = \mathrm{Unif}_{\mathcal{X}} \qquad L(q,q') = \mathbb{1}[q \neq q']$$

reality $\rightarrow$	$\clubsuit$	$\diamondsuit$	$\heartsuit$	$\spadesuit$	
$e_{AC}(A_1, b) = \frac{1}{4}($	$\mathbb{1}[b(0) \neq 0] +$	$\mathbb{1}[b(1) \neq 1] +$	$\mathbb{1}[b(1) \neq 0] +$	$\mathbb{1}[b(0) \neq 0]$	$)$
$e_{AC}(A_1, b \equiv 0) = \frac{1}{4}($	$0 +$	$1 +$	$0 +$	0	$) = \frac{1}{4}$
$e_{AC}(A_1, b = \mathrm{id}) = \frac{1}{4}($	$0 +$	$0 +$	$1 +$	0	$) = \frac{1}{4}$
$e_{AC}(A_2, b) = \frac{1}{4}($	$\mathbb{1}[b(0) \neq 0] +$	$\mathbb{1}[b(1) \neq 1] +$	$\mathbb{1}[b(1) \neq 0] +$	$\mathbb{1}[b(1) \neq 0]$	$)$
$e_{AC}(A_2, b \equiv 0) = \frac{1}{4}($	$0 +$	$1 +$	$0 +$	0	$) = \frac{1}{4}$
$e_{BPN}(A_1) = \frac{1}{4}($	$\mathbb{E}_{Q_\sharp \mu^0} L(\cdot, 0) +$	$\mathbb{E}_{Q_\sharp \mu^1} L(\cdot, 1) +$	$\mathbb{E}_{Q_\sharp \mu^1} L(\cdot, 0) +$	$\mathbb{E}_{Q_\sharp \mu^0} L(\cdot, 0)$	$)$
$= \frac{1}{4}($	$(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0) +$	$(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1) +$	$(\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0) +$	$(\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0)$	$) = \frac{1}{4}$
$e_{BPN}(A_2) = \frac{1}{4}($	$\mathbb{E}_{Q_\sharp \mu^0} L(\cdot, 0) +$	$\mathbb{E}_{Q_\sharp \mu^1} L(\cdot, 1) +$	$\mathbb{E}_{Q_\sharp \mu^1} L(\cdot, 0) +$	$\mathbb{E}_{Q_\sharp \mu^1} L(\cdot, 0)$	$)$
$= \frac{1}{4}($	$(1 \cdot 0)$	$+ (\frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 1)$	$+ (\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 0)$	$+ (\frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot 0)$	$) = \frac{1}{3}$

Coherent Pipelines of BPNMs

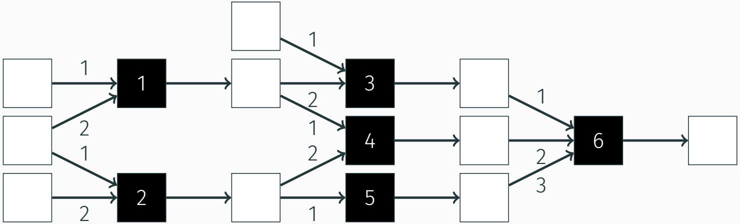


- Numerical methods usually form part of pipelines.
- Prime example: a PDE solve is a *forward model* in an *inverse problem*.
- Motivation for PNMs in the context of Bayesian inverse problems:

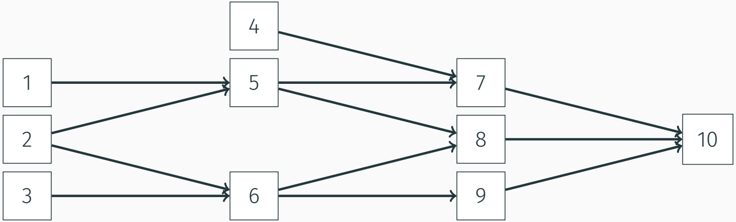
  Make the forward and inverse problem
  speak the same statistical language!

- We can compose PNMs in series, e.g. $B_2(B_1(\mu, a_1), a_2)$ is formally $B(\mu, (a_1, a_2))$…
  although figuring out what the spaces $\mathcal{X}_i$, $\mathcal{A}_i$ and operators A_i etc. are is a headache!

- More generally, we compose PNMs in a graphical way by allowing input information nodes (□) to feed into method nodes (■), which in turn output new information.
- (Pictures are easier than formal definitions!)

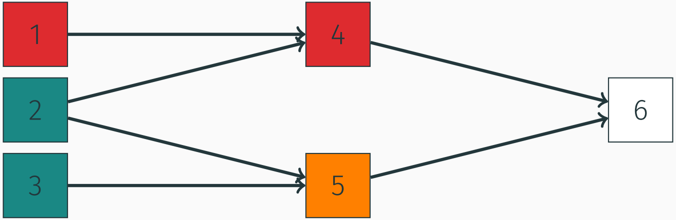


- More generally, we compose PNMs in a graphical way by allowing input information nodes (□) to feed into method nodes (■), which in turn output new information.
- (Pictures are easier than formal definitions!)



We define the corresponding **dependency graph** by replacing each □→■→□ by □→□, and we number the vertices in an increasing fashion, so that $\boxed{i} \to \boxed{i'}$ implies $i < i'$.

Coherence II



Definition

A prior is coherent for the dependency graph if every node Y_k is conditionally independent of all older non-parent nodes Y_i given its direct parent nodes Y_j.

$$Y_k \perp\!\!\!\perp Y_{\{1,\ldots,k-1\}\setminus \mathrm{parents}(k)} \mid Y_{\mathrm{parents}(k)}$$

This is weaker than the Markov condition for directed acyclic graphs (Lauritzen, 1991): we do not insist that the variables at the source nodes are independent.

Theorem (Cockayne et al., 2017b, Theorem 5.9)

If a pipeline of PNMs is such that

- *the prior is coherent for the dependence graph, and*
- *the component PNMs are all Bayesian*

then the pipeline is the Bayesian pipeline $\boxed{\text{sources}} \rightarrow \blacksquare \rightarrow \square$.

Theorem (Cockayne et al., 2017b, Theorem 5.9)

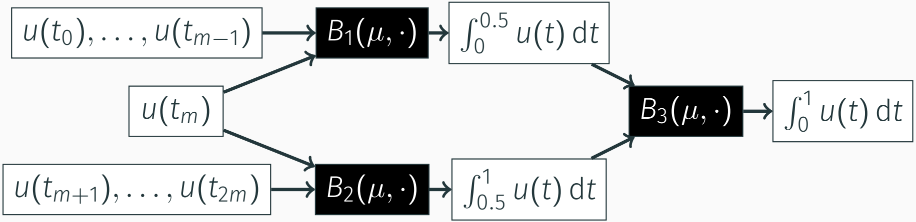
If a pipeline of PNMs is such that

- *the prior is coherent for the dependence graph, and*
- *the component PNMs are all Bayesian*

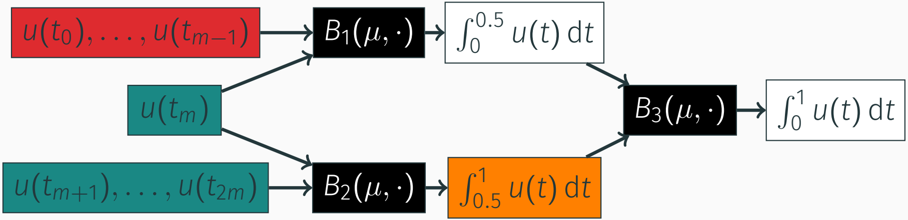
then the pipeline is the Bayesian pipeline $\boxed{\text{sources}} \to \blacksquare \to \square$.

- Redundant structure in the pipeline (recycled information) will break coherence, and hence Bayesianity of the pipeline.
- In principle, coherence and hence being Bayesian depend upon the prior.
- This should not be surprising — as a loose analogy, one doesn't expect the trapezoidal rule to be a good way to integrate very smooth functions.

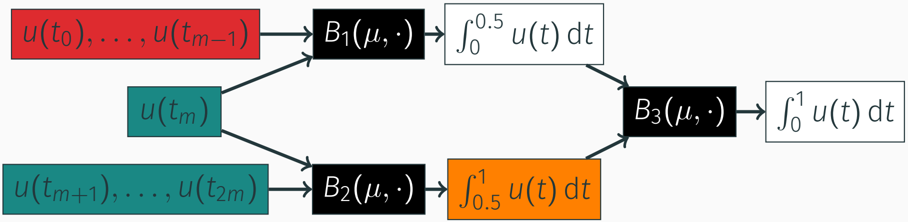
Split Integration: Coherence



- Integrate a function over $[0, 1]$ in two steps using nodes $0 \leq t_0 < \cdots < t_{m-1} < 0.5$, $t_m = 0.5$, and $t_{m+1} < \cdots < t_{2m} \leq 1$.



- Integrate a function over $[0, 1]$ in two steps using nodes $0 \leq t_0 < \cdots < t_{m-1} < 0.5$, $t_m = 0.5$, and $t_{m+1} < \cdots < t_{2m} \leq 1$.
- Is $\int_{0.5}^{1} u(t)\,\mathrm{d}t$ independent of $u(t_0), \ldots, u(t_{m-1})$ given $u(t_m), \ldots, u(t_{2m})$?

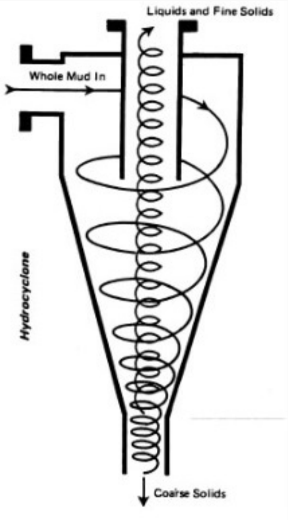


- Integrate a function over $[0, 1]$ in two steps using nodes $0 \leq t_0 < \cdots < t_{m-1} < 0.5$, $t_m = 0.5$, and $t_{m+1} < \cdots < t_{2m} \leq 1$.
- Is $\int_{0.5}^{1} u(t)\,\mathrm{d}t$ independent of $u(t_0), \ldots, u(t_{m-1})$ given $u(t_m), \ldots, u(t_{2m})$?
- For a Brownian motion prior on the integrand, yes. ✓
- For a Brownian motion prior on the derivative of the integrand, no. ✗
- This leads to the complicated issue of eliciting an appropriate prior that respects the problem's structure. ⁉

Application to Industrial Process Monitoring

Hydrocyclones (Oates, Cockayne, and Ackroyd, 2017)

- Hydrocyclones are used in industry as an alternative to centrifuges or filtration systems to separate fluids of different densities or particulate matter from a fluid.
- Monitoring is an essential control component, but usually cannot be achieved visually: Gutierrez et al. (2000) propose electrical impedance tomography as an alternative.
- EIT is an indirect imaging technique in which the conductivity field in the interior — which correlates with many material properties of interest — is inferred from boundary conditions.
- In its Bayesian formulation, this is a well-posed inverse problem (Dunlop and Stuart, 2016a,b) closely related to Calderón's problem (Uhlmann, 2009).



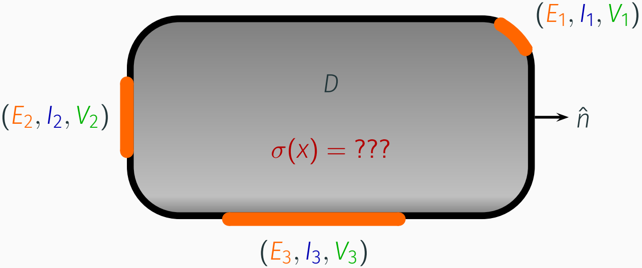
Complete Electrode Model (Cheng et al., 1989; Somersalo et al., 1992)

The interior conductivity field σ and electrical potential field v and the applied boundary currents I_i, measured voltages V_i, and known contact impedances ζ_i are related by

$$-\nabla \cdot \sigma(x)\nabla v(x) = 0 \qquad x \in D; \qquad \int_{E_i} \sigma(x)\frac{\partial v(x)}{\partial \hat{n}}\,\mathrm{d}x = I_i \qquad x \in E_i, i = 1, \dots, m;$$

$$v(x) + \zeta_i \sigma(x)\frac{\partial v(x)}{\partial \hat{n}} = V_i \qquad x \in E_i; \qquad \sigma(x)\frac{\partial v(x)}{\partial \hat{n}} = 0 \qquad x \in \partial D \setminus \bigcup_{i=1}^{m} E_i.$$

Furthermore, we consider a vector of such models, with multiple current stimulation patterns, at multiple points in time, for a time-dependent field $\sigma(t, x)$.



- Sampling from the posterior(s) requires repeatedly solving the forward PDE.
- We use the probabilistic meshless method of Cockayne et al. (2016, 2017a):
  - a Gaussian process extension of symmetric collocation;
  - a BPNM for a Gaussian prior and linear elliptic PDEs of this type.
- PMM allows us to:
  - account for uncertainty arising from the PDE having no explicit solution;
  - use coarser discretisations of the PDE to solve the problem faster while still providing meaningful UQ.

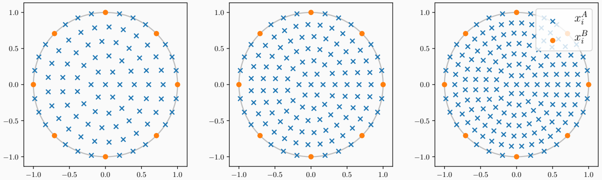


Figure 2: Like collocation, PMM imposes the PDE relation at $n_{\mathcal{A}}$ interior nodes and boundary conditions at $n_{\mathcal{B}}$ boundary nodes.

- For the inverse problem we use a Karhunen–Loève series prior:

$$\log \sigma(t, x; \omega) = \sum_{k=1}^{\infty} k^{-\alpha} \psi_k(t; \omega) \phi_k(x),$$

  with the ψ_k being a-priori independent Brownian motions in t.
- Like Dunlop and Stuart (2016a), we assume additive Gaussian observational noise with variance $\gamma^2 > 0$, independently on each E_i.
- We adopt a filtering formulation, inferring $\sigma(t_i, \cdot; \cdot)$ sequentially.
- Within each data assimilation step, the Bayesian update is performed by SMC with $P \in \mathbb{N}$ weighted particles and a pCN transition kernel (which uses point evaluations of σ directly and avoids truncation of the KL expansion).
- Real-world data obtained at 49 regular time intervals: rapid injection between frames 10 and 11, followed by diffusion and rotation of the liquids.

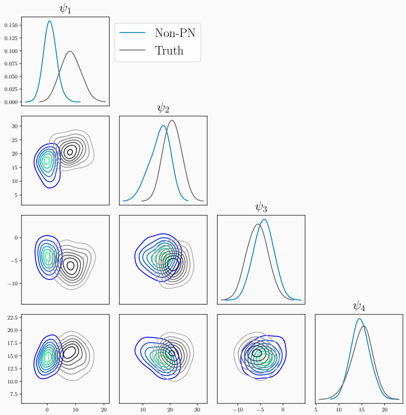


Figure 3: A small number $n_{\mathcal{A}} + n_{\mathcal{B}} = 71$ of collocation points was used to discretise the PDE, but the uncertainty due to discretisation was not modelled. The reference posterior distribution over the coefficients ψ_k is plotted (grey) and compared to the approximation to the posterior obtained when the PDE is discretised and the discretisation error is not modelled (blue, "Non-PN"). The approximate posterior is highly biased.

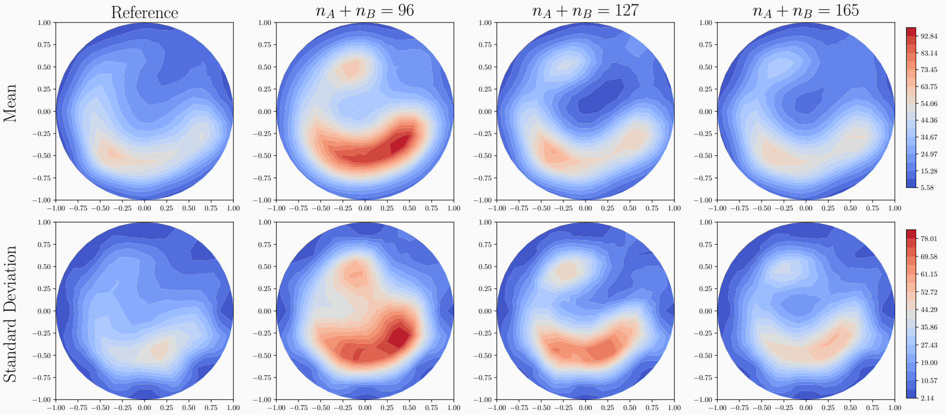


Figure 4: Posterior means and standard-deviations for the recovered conductivity field at $t = 14$. The first column shows the reference solution, obtained using symmetric collocation with a large number of collocation points. The remaining columns show the recovered field when PMM is used with $n_{\mathcal{A}} + n_{\mathcal{B}}$ collocation points.

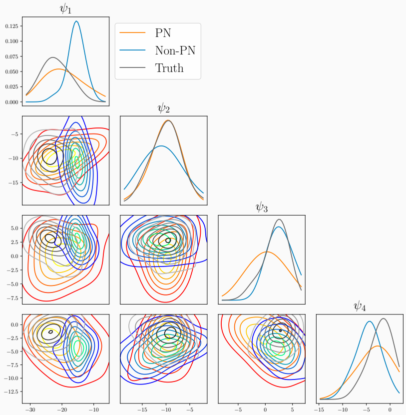


Figure 5: Posterior distribution over the coefficients ψ_k at the final time. A small number $n_{\mathcal{A}} + n_{\mathcal{B}} = 71$ of collocation points was used to discretise the PDE. The reference posterior distribution over the coefficients ψ_k is plotted (grey) and compared to the approximation to the posterior obtained when discretisation of the PDE is not modelled (blue, "Non-PN") and modelled (orange, "PN").

EIT Comments

- Typically PDE discretisation error in BIPs is ignored, or its contribution is bounded through detailed numerical analysis (Schwab and Stuart, 2012). Theoretical bounds are difficult in the temporal setting due to propagation and accumulation of errors
- As a modelling choice, the PN approach eases these difficulties. As with the Painlevé example, this is a statistically correct implementation of the assumptions, but it is (at present) costly. ✓/✗
- Furthermore, Markov temporal evolution of the conductivity field was assumed; this is likely incorrect, since time derivatives of this field will vary continuously. Even a-priori knowledge about the spin direction is neglected at present. ✗
- Again, we see a need for priors that are "physically reasonable" and statistically/computationally appropriate. ⁉

Closing Remarks

Closing Remarks

- Numerical methods can be characterised in a Bayesian fashion. ✓
- This does not coincide with average-case analysis and IBC. ✓
- BPNMs can be composed into pipelines, e.g. for inverse problems. ✓
- Bayes' rule as disintegration $\to$ (expensive!) numerical implementation. ✓/✗
  - Lots of room to improve computational cost and bias. ⁉
  - Departures from the 'Bayesian gold standard' can be assessed in terms of cost-accuracy tradeoff. ⁉
- How to choose/design an appropriate prior? ⁉
- Full details and further applications in

Cockayne, Oates, Sullivan, and Girolami (2017b)
"Bayesian probabilistic numerical methods"
arXiv:1702.03673.

Closing Remarks

- Numerical methods can be characterised in a Bayesian fashion. ✓
- This does not coincide with average-case analysis and IBC. ✓
- BPNMs can be composed into pipelines, e.g. for inverse problems. ✓
- Bayes' rule as disintegration $\to$ (expensive!) numerical implementation. ✓/✗
  - Lots of room to improve computational cost and bias. ⁇
  - Departures from the 'Bayesian gold standard' can be assessed in terms of cost-accuracy tradeoff. ⁇
- How to choose/design an appropriate prior? ⁇
- Full details and further applications in

Cockayne, Oates, Sullivan, and Girolami (2017b)
"Bayesian probabilistic numerical methods"
arXiv:1702.03673.

Thank You

J. T. Chang and D. Pollard. Conditioning as disintegration. *Statist. Neerlandica*, 51(3):287–317, 1997. doi:10.1111/1467-9574.00056.

K.-S. Cheng, D. Isaacson, J. C. Newell, and D. G. Gisser. Electrode models for electric current computed tomography. *IEEE Trans. Biomed. Eng.*, 36(9), 1989. doi:10.1109/10.35300.

J. Cockayne, C. Oates, T. J. Sullivan, and M. Girolami. Probabilistic meshless methods for partial differential equations and Bayesian inverse problems, 2016. arXiv:1605.07811.

J. Cockayne, C. Oates, T. J. Sullivan, and M. Girolami. Probabilistic numerical methods for PDE-constrained Bayesian inverse problems. In G. Verdoolaege, editor, *Proceedings of the 36th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume 1853 of *AIP Conference Proceedings*, pages 060001–1–060001–8, 2017a. doi:10.1063/1.4985359.

J. Cockayne, C. Oates, T. J. Sullivan, and M. Girolami. Bayesian probabilistic numerical methods, 2017b. arXiv:1702.03673.

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

M. M. Dunlop and A. M. Stuart. The Bayesian formulation of EIT: analysis and algorithms. *Inv. Probl. Imaging*, 10(4): 1007–1036, 2016a. doi:10.3934/ipi.2016030.

M. M. Dunlop and A. M. Stuart. MAP estimators for piecewise continuous inversion. *Inv. Probl.*, 32(10):105003, 50, 2016b. doi:10.1088/0266-5611/32/10/105003.

M. Giry. A categorical approach to probability theory. In *Categorical aspects of topology and analysis (Ottawa, Ont., 1980)*, volume 915 of *Lecture Notes in Math.*, pages 68–85. Springer, Berlin-New York, 1982.

J. Gutierrez, T. Dyakowski, M. Beck, and R. Williams. Using electrical impedance tomography for controlling hydrocyclone underflow discharge. 108(2):180–184, 2000.

J. B. Kadane and G. W. Wasilkowski. Average case ϵ-complexity in computer science. A Bayesian view. In *Bayesian Statistics, 2 (Valencia, 1983)*, pages 361–374. North-Holland, Amsterdam, 1985.

F. M. Larkin. Optimal approximation in Hilbert spaces with reproducing kernel functions. *Math. Comp.*, 24:911–921, 1970. doi:10.2307/2004625.

S. Lauritzen. *Graphical Models*. Oxford University Press, 1991.

C. J. Oates, J. Cockayne, and R. G. Ackroyd. Bayesian probabilistic numerical methods for industrial process monitoring, 2017. arXiv:1707.06107.

H. Poincaré. *Calcul des Probabilites*. Georges Carré, Paris, 1896.

K. Ritter. *Average-Case Analysis of Numerical Problems*, volume 1733 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2000. doi:10.1007/BFb0103934.

References III

C. Schwab and A. M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inv. Probl.*, 28(4):045003, 32, 2012. doi:10.1088/0266-5611/28/4/045003.

J. Skilling. Bayesian solution of ordinary differential equations. In C. R. Smith, G. J. Erickson, and P. O. Neudorfer, editors, *Maximum Entropy and Bayesian Methods*, volume 50 of *Fundamental Theories of Physics*, pages 23–37. Springer, 1992. doi:10.1007/978-94-017-2219-3.

E. Somersalo, M. Cheney, and D. Isaacson. Existence and uniqueness for electrode models for electric current computed tomography. *SIAM J. Appl. Math.*, 52(4):1023–1040, 1992. doi:10.1137/0152060.

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

A. V. Sul′din. Wiener measure and its applications to approximation methods. I. *Izv. Vysš. Učebn. Zaved. Matematika*, 6(13): 145–158, 1959.

A. V. Sul′din. Wiener measure and its applications to approximation methods. II. *Izv. Vysš. Učebn. Zaved. Matematika*, 5 (18):165–179, 1960.

J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski. *Information-Based Complexity*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1988. With contributions by A. G. Werschulz and T. Boult.

G. Uhlmann. Electrical impedance tomography and Calderón's problem. *Inv. Probl.*, 25(12):123011, 39, 2009. doi:10.1088/0266-5611/25/12/123011.