Probabilistic Numerics

Selected Highlights from Working Group II

C. Oates[1,2] and **T. J. Sullivan**[3,4]

SAMSI QMC Transition Workshop
SAMSI, Durham, NC, US, 7–9 May 2018

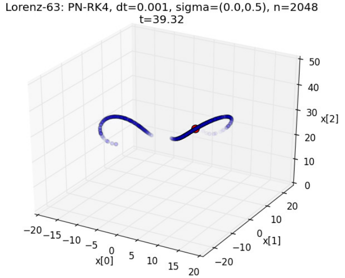
[1]Alan Turing Institute, London, UK
[2]Newcastle University, UK
[3]Free University of Berlin, DE
[4]Zuse Institute Berlin, DE

Introduction

Working Group II: Probabilistic Numerics

WG II Mission Statement

The accuracy and robustness of numerical predictions that are based on mathematical models depend critically upon the construction of accurate discrete approximations to key quantities of interest. The exact error due to approximation will be unknown to the analyst, but worst-case upper bounds can often be obtained. This working group aims, instead, to develop Probabilistic Numerical Methods, which provide the analyst with a richer, probabilistic quantification of the numerical error in their output, thus providing better tools for reliable statistical inference.



PN ensemble solution of the Lorenz-63 system; convergence rate of the (non-Gaussian!) solution distribution to the exact solution is given by Lie, Stuart, and Sullivan (2017a).

Why Probabilistic Numerics?

- The last 5 years have seen a renewed interest, at various levels of rigour, in probabilistic perspectives on numerical tasks — e.g. quadrature, ODE and PDE solution, optimisation.
- A long heritage: Poincaré (1896); Larkin (1970); Diaconis (1988); Skilling (1992).
- There are 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; Trefethen, 2008)?
  - "Big data" problems often require (random) subsampling.
  - If discretisation error is not properly accounted for, then biased and over-confident inferences result (Conrad et al., 2016).
  - Accounting for the impact of discretisation error in a statistical way allows forward and Bayesian inverse problems to speak a common statistical language.
- We think that some concrete definitions and theory-building are needed!
- We also think that concrete *applications* are needed!

Who has Taken Part?

Members* of the Working Group

	Name	Affiliation	WG Role
1.	Alessandro Barp	Imperial College London, UK	
2.	David Bortz	University of Colorado, US	
3.	François-Xavier Briol	University of Warwick, UK and Imperial College London, UK	RG Chair
4.	Ben Calderhead	Imperial College London, UK	
5.	Oksana Chkrebtii	Ohio State University, US	
6.	Jon Cockayne	University of Warwick, UK	
7.	Vanja Dukic	University of Colorado, US	
8.	Ruituo Fan	University of North Carolina, US	
9.	Mark Girolami	Imperial College London, UK and the Alan Turing Institute, UK	
10.	Jan Hannig	University of North Carolina, US	
11.	Philipp Hennig	Max Planck Institute, Tübingen, DE	
12.	Fred Hickernell	Illinois Institute of Technology, US	
13.	Toni Karvonen	Aalto University, FI	
14.	Han Cheng Lie	Freie Universität Berlin, DE	RG Chair
15.	Chris Oates	Newcastle University, UK and the Alan Turing Institute, UK	WG Leader
16.	Houman Owhadi	California Institute of Technology, US	
17.	Jagadeeswaran Rathinavel	Illinois Institute of Technology US	
18.	Florian Schäfer	California Institute of Technology, US	
19.	Andrew Stuart	California Institute of Technology, US	
20.	Tim Sullivan	Freie Universität Berlin, DE and Zuse Institute Berlin, DE	WG Leader
21.	Onur Teymur	Imperial College London, UK	
22.	Junyang Wang	Newcastle University, UK	

What Have We Been Up To?

Publications acknowledging SAMSI support

	Published	
1.	Briol, Cockayne, Teymur, Yoo, Schober, and Hennig (2016)	SAMSI Optimization 2016–2017
2.	Dukic and Bortz (2018)	
3.	Oates, Niederer, Lee, Briol, and Girolami (2017b)	
	Submitted and under review	
4.	Cockayne, Oates, Sullivan, and Girolami (2017)	Best Paper Award at JSM 2018
5.	Lie, Stuart, and Sullivan (2017a)	
6.	Lie and Sullivan (2017)	
7.	Lie, Sullivan, and Teckentrup (2017b)	
8.	Oates, Cockayne, and Ackroyd (2017a)	
9.	Schäfer, Sullivan, and Owhadi (2017)	
10.	Karvonen, Oates, and Särkkä (2018)	
11.	Cockayne, Oates, and Girolami (2018)	
12.	Xi, Briol, and Girolami (2018)	
	In preparation	
13.	Chkrebtii (In preparation)	
14.	Hennig, Kersting, and Sullivan (In preparation)	
15.	Rathinavel and Hickernell (In preparation)	

- Teleconference over Skype, later Webex.
- 1 hour session every 1 or 2 weeks.
- 22 presentations by WG members and guests on
  - PN history, e.g. the work of Mike Larkin in the 1970s;
  - Ongoing research on the WG's topics of interest: quadrature, random Bayesian inverse problems, probabilistic linear algebra, information dynamics, …
- Speaker schedule, technicalities etc. kindly coordinated by **François-Xavier Briol** (Warwick & Imperial College London) and **Han Cheng Lie** (Freie Universität Berlin).

SAMSI–Lloyd's–Turing Workshop on PN, 11.–13.04.2018

- Venue: the **Alan Turing Institute**, the UK's national institute for data science, housed in the **British Library**, Euston Road, London
- 34 participants from US, UK, FR, FI, DE, CH
- 17 talks, 4 research sessions, and 1 panel discussion

prob-num.github.io

Generously supported by:

The Alan Turing Institute

Lloyd's Register Foundation

samsi

NSF Duke NCSU UNC

PN Minisymposium @ SIAM UQ18, 16.–19.04.2018

“Probabilistic Numerical Methods for Quantification of Discretisation Error”

- $3 \times$ 2-hour minisymposium at SIAM UQ18
- Organisers: Mark Girolami (Imperial & Turing), Philipp Hennig (MPI Tübingen), Chris Oates (Newcastle & Turing), and Tim Sullivan (FU Berlin / ZIB)

MS4 Monday 09:30–11:30		MS17 Monday 14:00–16:00		MS32 Tuesday 09:10–11:10	
Sullivan		Kanagawa		Hickernell	
Campbell	$\int u(x)\,\mathrm{d}x$	Oates	$-\nabla \cdot (\kappa \nabla u) = f$	Briol	$\int u(x)\,\mathrm{d}x$
Cockayne	$Ax = b$	Teymur	$\frac{\mathrm{d}}{\mathrm{d}t} u = f(t, u)$	Gessner	$\int u(x)\,\mathrm{d}x$
Kersting	$\frac{\mathrm{d}}{\mathrm{d}t} u = f(t, u)$	Schäfer	$Ax = b$	Karvonen	$\int u(x)\,\mathrm{d}x$

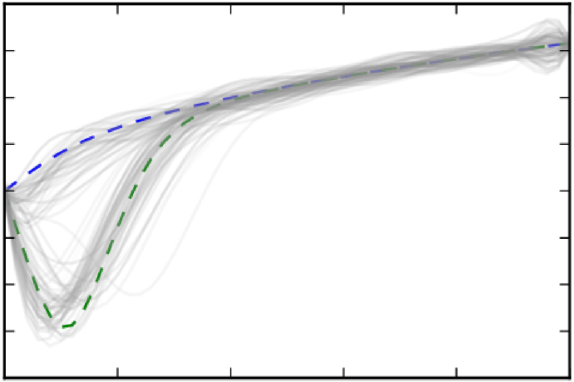
Research Tour I

Cockayne, Oates, Sullivan, and Girolami (2017) arXiv:1702.03673
[*Best Student Paper Award, ASA Section on Bayesian Statistical Science at JSM!]

Goal is to formulate a *Bayesian* approach to traditional "numerical tasks", such as the solution of a differential equation

$$\frac{\mathrm{d}u}{\mathrm{d}t} = f(t, u)$$

that enables uncertainty quantification due to space and time discretisation, necessitated by a finite computational budget.



- Prior: $u \sim \mathbb{P}$

- Prior: $u \sim \mathbb{P}$
- Posterior:

$$\mathbb{P}_n \ : \ \lim_{\epsilon \downarrow 0} \frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0}$$

where

$$\frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0} = \exp\left(-\frac{1}{\epsilon} \sum_{i=1}^{n} \left| \frac{\mathrm{d}u}{\mathrm{d}t}(t_i, x_i) - f(t_i, x_i) \right|^2 \right)$$

Bayesian Probabilistic Numerical Methods

- Prior: $u \sim \mathbb{P}$
- Posterior:

$$\mathbb{P}_n \;:\; \lim_{\epsilon \downarrow 0} \frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0}$$

  where

$$\frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0} = \exp\left(-\frac{1}{\epsilon} \sum_{i=1}^{n} \left| \frac{\mathrm{d}u}{\mathrm{d}t}(t_i, x_i) - f(t_i, x_i) \right|^2 \right)$$

- Relationship to average case analysis and optimal algorithms?

- Prior: $u \sim \mathbb{P}$
- Posterior:

$$\mathbb{P}_n \; : \; \lim_{\epsilon \downarrow 0} \frac{d\mathbb{P}_\epsilon}{d\mathbb{P}_0}$$

where

$$\frac{d\mathbb{P}_\epsilon}{d\mathbb{P}_0} = \exp\left(-\frac{1}{\epsilon}\|A(u) - a\|^2\right) \qquad \text{how to design } A?$$

- Relationship to average case analysis and optimal algorithms?

- Prior: $u \sim \mathbb{P}$
- Posterior:

$$\mathbb{P}_n \ : \ \lim_{\epsilon \downarrow 0} \frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0}$$

where

$$\frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0} = \exp\left(-\frac{1}{\epsilon}\|A(u) - a\|^2\right) \qquad \text{how to design } A?$$

- Relationship to average case analysis and optimal algorithms?
- Design criterion:

$$\arg\min_{A \in \mathcal{A}} \int \underbrace{d}_{\text{e.g. Wasserstein}} \left(\delta(u^\dagger), \mathbb{P}_n\right) \mathrm{d}\mathbb{P}(u^\dagger)$$

is in general distinct to average case analysis!

Bayesian Probabilistic Numerical Methods

- Prior: $u \sim \mathbb{P}$
- Posterior:

$$\mathbb{P}_n \ : \ \lim_{\epsilon \downarrow 0} \frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0}$$

  where

$$\frac{\mathrm{d}\mathbb{P}_\epsilon}{\mathrm{d}\mathbb{P}_0} = \exp\left(-\frac{1}{\epsilon}\|A(u) - a\|^2\right) \qquad \text{how to design } A?$$

- Relationship to average case analysis and optimal algorithms?
- Design criterion:

$$\operatorname*{arg\,min}_{A \in \mathcal{A}} \int \underbrace{d}_{\text{e.g. Wasserstein}} \left(\delta(u^\dagger), \mathbb{P}_n\right) \mathrm{d}\mathbb{P}(u^\dagger)$$

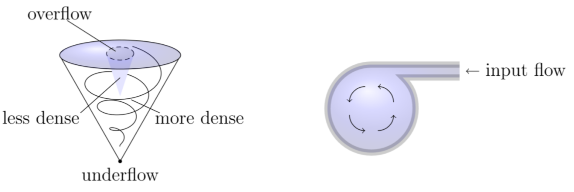
  is in general distinct to average case analysis!
- Closure under composition (see next project...)

Oates, Cockayne, and Ackroyd (2017a) arXiv:1707.06107

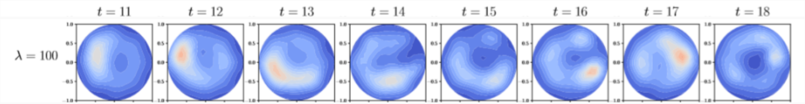
Goal is to perform Bayesian uncertainty quantification for the time-dependent conductivity field $a(x;t)$ s.t.

$$\nabla \cdot (a \nabla u) = 0 \quad \text{in } D$$
$$\int_{E_i} a \nabla u \cdot \mathrm{n} \, \mathrm{d}\sigma = I_i$$
$$u + \zeta_i a \nabla u \cdot \mathrm{n} = U_i \quad \text{on } E_i$$
$$a \nabla u \cdot \mathrm{n} = 0 \quad \text{on } \partial D \setminus \bigcup_{i=1}^{m} E_i$$

based on noisy observations of $U_i(t)$ whilst propagating uncertainty due to discretisation of the PDE.

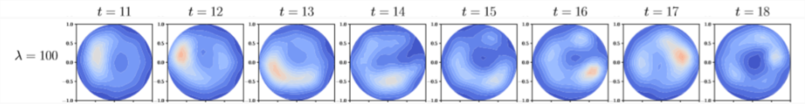


(a) Hydrocyclone tank schematic (b) Cross-section (top of tank)



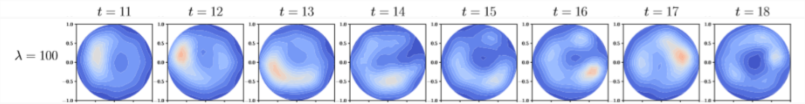
- The hydrocyclone operates at a high velocity, so this demands coarse discretisation of the PDE.

Bayesian PNMs for Industrial Process Monitoring



- The hydrocyclone operates at a high velocity, so this demands coarse discretisation of the PDE.
- Exploits closure of Bayesian PNM under composition!

Bayesian PNMs for Industrial Process Monitoring



- The hydrocyclone operates at a high velocity, so this demands coarse discretisation of the PDE.
- Exploits closure of Bayesian PNM under composition!
- Better reflection of uncertainty on the unknown field.

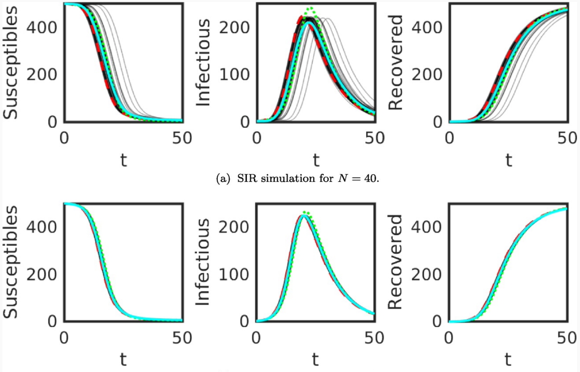
UQ using PN: Application to Models in Mathematical Epidemiology

Dukic and Bortz (2018) *Inv. Probl. Sci. Engrng.* 28(2):223–232

Goal is to perform Bayesian uncertainty quantification for the solution to an epidemiological model

$$S(t) = -\beta S(t)I(t)$$
$$I(t) = \beta S(t)I(t) - \gamma I(t)$$
$$R(t) = \gamma I(t)$$

that captures uncertainty due to finite computational budget.



(a) SIR simulation for $N = 40$.

(b) SIR simulation for $N = 80$.

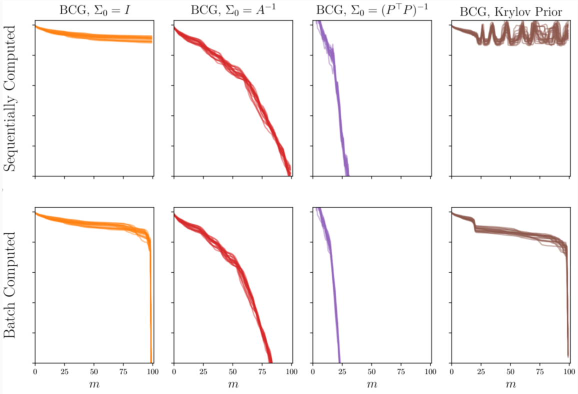
A Bayesian Conjugate Gradient Method

Cockayne, Oates, and Girolami (2018) arXiv:1801.05242

Goal is to perform Bayesian uncertainty quantification for the solution $x \in \mathbb{R}^N$ to a linear system

$$Ax = b$$

where only $n \ll N$ vector-matrix multiplications can be performed.



A Bayesian Conjugate Gradient Method

See **Jon Cockayne**'s talk!

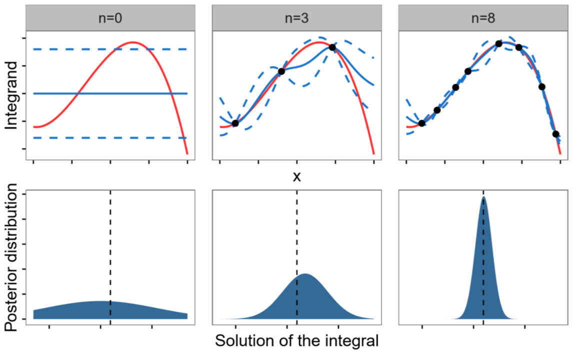
Rathinavel and Hickernell (In preparation)

Goal is to perform Bayesian uncertainty quantification for an integral

$$\int f(x)\,\mathrm{d}\mu(x)$$

where

- $f(x)$ is expensive to evaluate
- uncertainty estimates should be "well-calibrated"
- computational overhead should be $O(n \log n)$



Ask Fred!

Ask Fred!

- Fix a tolerance $0 < \tau \ll 1$.

Ask Fred!

- Fix a tolerance $0 < \tau \ll 1$.
- Prior: $f \sim \mathrm{GP}(m, k)$ where $m \equiv m_\theta$ and $k \equiv k_\theta$, $\theta \in \Theta$.

Ask Fred!

- Fix a tolerance $0 < \tau \ll 1$.
- Prior: $f \sim \mathrm{GP}(m, k)$ where $m \equiv m_\theta$ and $k \equiv k_\theta$, $\theta \in \Theta$.
- For each $n = 1, 2, \ldots$
  - Select
    $$\hat{\theta} = \operatorname*{arg\,max}_{\theta \in \Theta} \; \mathrm{prob}(f(x_1), \ldots, f(x_n) | \theta)$$
    (i.e. empirical Bayes)
  - If 99% of the mass of
    $$\mathrm{prob}\left(\int f(x) \, \mathrm{d}\mu(x) \;\middle|\; f(x_1), \ldots, f(x_n), \hat{\theta} \right)$$
    lies in an interval A of width τ, then break.
- Return the credible set A.

Ask Fred!

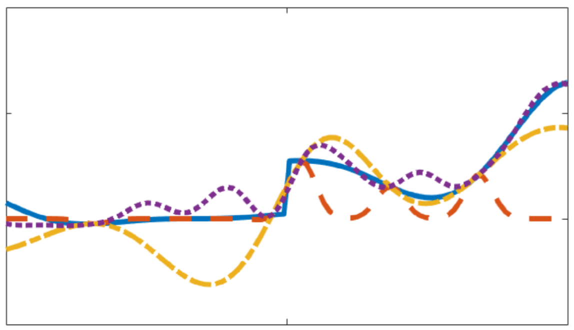
- Fix a tolerance $0 < \tau \ll 1$.
- Prior: $f \sim \mathrm{GP}(m, k)$ where $m \equiv m_\theta$ and $k \equiv k_\theta$, $\theta \in \Theta$.
- For each $n = 1, 2, \ldots$
  - Select
    $$\hat{\theta} = \operatorname*{arg\,max}_{\theta \in \Theta} \mathrm{prob}(f(x_1), \ldots, f(x_n) | \theta)$$
    (i.e. empirical Bayes)
  - If 99% of the mass of
    $$\mathrm{prob}\left(\int f(x) \, \mathrm{d}\mu(x) \,\middle|\, f(x_1), \ldots, f(x_n), \hat{\theta} \right)$$
    lies in an interval A of width τ, then break.
- Return the credible set A.
- Efficient computation of evidence at $O(n \log n)$ cost.

Xi, Briol, and Girolami (2018) arXiv:1801.04153

Goal is to perform Bayesian uncertainty quantification for the related integrals

$$\int f_1(x)\,\mathrm{d}x$$
$$\vdots$$
$$\int f_m(x)\,\mathrm{d}x$$

where each $f_i(x)$ is expensive to evaluate.



- Main tool: vector-valued GP/RKHS

Bayesian Quadrature for Multiple Related Integrals

- Main tool: vector-valued GP/RKHS
- Kernels:

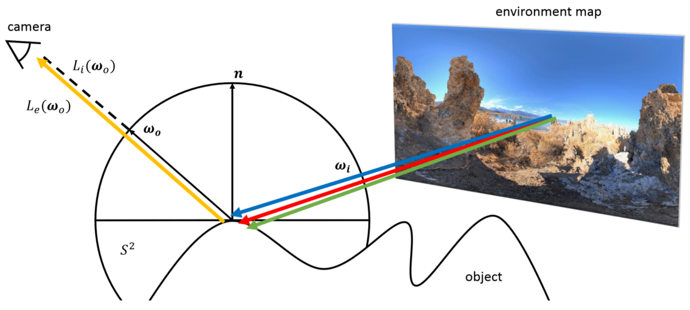
$$k((x,i),(y,j)) = k_1(i,j)k_2(x,y)$$

Bayesian Quadrature for Multiple Related Integrals

- Main tool: vector-valued GP/RKHS
- Kernels:

$$k((x,i),(y,j)) = k_1(i,j)k_2(x,y)$$

- Interesting applied work around elicitation of k_1:



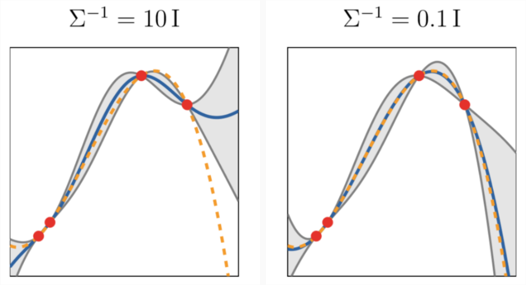
A Bayes–Sard Cubature Method

Karvonen et al. (2018) arXiv:1804.03016

Given a cubature rule

$$\hat{\mu}(f) = \sum_{i=1}^{n} w_i f(x_i)$$
$$\approx \int f(x)\,\mathrm{d}\mu(x)$$

can we cast $\hat{\mu}$ as a Bayes rule in a decision-theoretic framework?



A Bayes–Sard Cubature Method

- Pick $\phi_i(x)$ such that $\phi_i(x_j) = \delta_{i,j}$ and $\int \phi_i(x) \, \mathrm{d}\mu(x) = \frac{1}{n}$.

A Bayes–Sard Cubature Method

- Pick $\phi_i(x)$ such that $\phi_i(x_j) = \delta_{i,j}$ and $\int \phi_i(x) \, \mathrm{d}\mu(x) = \frac{1}{n}$.
- Form a regression model:

$$
\begin{aligned}
f(x) &= \sum_{i=1}^{n} \beta_i \phi_i(x) + g(x) \\
\beta_i &\sim \text{Uniform} \\
g &\sim \mathrm{GP}(0, k)
\end{aligned}
$$

A Bayes–Sard Cubature Method

- Pick $\phi_i(x)$ such that $\phi_i(x_j) = \delta_{i,j}$ and $\int \phi_i(x)\,\mathrm{d}\mu(x) = \frac{1}{n}$.
- Form a regression model:

$$\begin{aligned} f(x) &= \sum_{i=1}^{n} \beta_i \phi_i(x) + g(x) \\ \beta_i &\sim \mathrm{Uniform} \\ g &\sim \mathrm{GP}(0, k) \end{aligned}$$

- Then

$$\int f(x)\,\mathrm{d}\mu(x) \,\Big|\, f(x_1), \ldots, f(x_n) \sim \mathrm{N}\Big(\hat{\mu}(f), \sigma^2\Big)$$

where σ is the worst case error of the cubature rule $\hat{\mu}$ in the RKHS $H(k)$.

A Bayes–Sard Cubature Method

- Pick $\phi_i(x)$ such that $\phi_i(x_j) = \delta_{i,j}$ and $\int \phi_i(x) \, \mathrm{d}\mu(x) = \frac{1}{n}$.
- Form a regression model:

$$f(x) = \sum_{i=1}^{n} \beta_i \phi_i(x) + g(x)$$

$$\beta_i \sim \text{Uniform}$$

$$g \sim \text{GP}(0, k)$$

- Then

$$\int f(x) \, \mathrm{d}\mu(x) \Big| f(x_1), \ldots, f(x_n) \sim \mathrm{N}\left(\hat{\mu}(f), \sigma^2\right)$$

where σ is the worst case error of the cubature rule $\hat{\mu}$ in the RKHS $H(k)$.
- Enables Bayesian uncertainty quantification for QMC?

Probabilistic Models for Integration Error in Assessment of Functional Cardiac Models

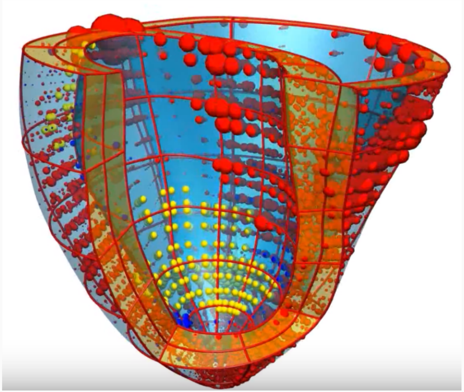
Oates, Niederer, Lee, Briol, and Girolami (2017b)
Advances in Neural Information Processing Systems (NIPS 2017)

Goal is to perform Bayesian uncertainty quantification for the integral

$$\int f(x)p(x)\,\mathrm{d}x$$

where

- $f(x)$ is expensive to evaluate
- $p(x)$ can only be sampled
- $p(x)$ is expensive to sample



Probabilistic Models for Integration Error in Assessment of Functional Cardiac Models

- Assessment of functional cardiac models is difficult.

Probabilistic Models for Integration Error in Assessment of Functional Cardiac Models

- Assessment of functional cardiac models is difficult.
- Prior: $f \sim \mathrm{GP}(m, k)$, $p \sim \mathrm{DirichletProcessMixtureModel}(\alpha, P_0, \phi(\cdot, \cdot))$ (ϕ a RBF)

Probabilistic Models for Integration Error in Assessment of Functional Cardiac Models

- Assessment of functional cardiac models is difficult.
- Prior: $f \sim \mathrm{GP}(m, k)$, $p \sim \mathrm{DirichletProcessMixtureModel}(\alpha, P_0, \phi(\cdot, \cdot))$ (ϕ a RBF)
- Data: $\{x_i\}_{i=1}^n \sim p(x)$ and $f(x_i) \in \mathbb{R}$

Probabilistic Models for Integration Error in Assessment of Functional Cardiac Models

- Assessment of functional cardiac models is difficult.
- Prior: $f \sim \mathrm{GP}(m, k)$, $p \sim \mathrm{DirichletProcessMixtureModel}(\alpha, P_0, \phi(\cdot, \cdot))$ (ϕ a RBF)
- Data: $\{x_i\}_{i=1}^n \sim p(x)$ and $f(x_i) \in \mathbb{R}$
- Posterior can be sampled:
  - $\tilde{p} \sim \mathrm{Posterior}(p | \{x_1, \ldots, x_n\})$
  - $\tilde{p}(\cdot) = \sum_{i=1}^{\infty} \tilde{w}_i \phi(\cdot, \tilde{x}_i)$
  - $\int k(\cdot, x)\, \mathrm{d}\tilde{p}(x) = \sum_{i=1}^{\infty} \tilde{w}_i \int k(\cdot, x) \phi(x, \tilde{x}_i)\, \mathrm{d}x$

Probabilistic Models for Integration Error in Assessment of Functional Cardiac Models

- Assessment of functional cardiac models is difficult.
- Prior: $f \sim \mathrm{GP}(m, k)$, $p \sim \mathrm{DirichletProcessMixtureModel}(\alpha, P_0, \phi(\cdot, \cdot))$ (ϕ a RBF)
- Data: $\{x_i\}_{i=1}^n \sim p(x)$ and $f(x_i) \in \mathbb{R}$
- Posterior can be sampled:
  - $\tilde{p} \sim \mathrm{Posterior}(p | \{x_1, \ldots, x_n\})$
  - $\tilde{p}(\cdot) = \sum_{i=1}^{\infty} \tilde{w}_i \phi(\cdot, \tilde{x}_i)$
  - $\int k(\cdot, x) \, \mathrm{d}\tilde{p}(x) = \sum_{i=1}^{\infty} \tilde{w}_i \int k(\cdot, x) \phi(x, \tilde{x}_i) \, \mathrm{d}x$
  - For conjugate (k, ϕ) have a closed-form kernel mean embedding!

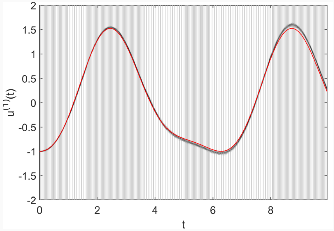
Research Tour II

Chkrebtii (In preparation)

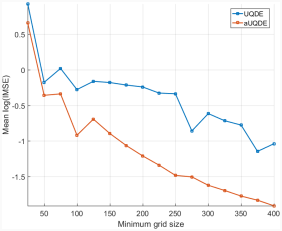
- For state-space probabilistic ODE solvers, the selection of the discretization grid is a problem of statistical design.

Chkrebtii (In preparation)

- For state-space probabilistic ODE solvers, the selection of the discretization grid is a problem of statistical design.
- A maximum entropy design yields a closed-form objective function that can be used to control the step size — step length decreases when the predicted and the actual model evaluations differ, i.e. when the state changes quickly.



Marginal sample paths (gray) over the unknown state, the exact solution shown in red. Gray lines in the background illustrate the adaptive time step.



Mean over 100 simulation runs of the logarithm of IMSE for the adaptive (aUQDES) and equally spaced grid probabilistic numerical solver (UQDES).

Lie and Sullivan (2017) arXiv:1708.02516

- A **mode** of a probability measure μ on a Banach space $\mathcal{U}$ is a point $u^\star \in \mathcal{U}$ of maximum probability — in the case of a Bayesian posterior, a **MAP estimator**.
- Without a reference uniform measure, modes must be defined intrinsically.

Lie and Sullivan (2017) arXiv:1708.02516

- A **mode** of a probability measure μ on a Banach space $\mathcal{U}$ is a point $u^\star \in \mathcal{U}$ of maximum probability — in the case of a Bayesian posterior, a **MAP estimator**.
- Without a reference uniform measure, modes must be defined intrinsically.
- A **strong mode** (Dashti et al., 2013):

$$u^\star \text{ is a strong mode of } \mu \iff \lim_{r \to 0} \frac{\sup_{u \in \mathcal{U}} \mu(B(u,r))}{\mu(B(u^\star,r))} \leq 1.$$

- A **weak mode** (Helin and Burger, 2015) with respect to a subspace $E \subset \mathcal{U}$:

$$u^\star \text{ is an } E\text{-weak mode of } \mu \iff \lim_{r \to 0} \frac{\sup_{v \in E} \mu(B(u^\star + v,r))}{\mu(B(u^\star,r))} \leq 1.$$

Lie and Sullivan (2017) arXiv:1708.02516

- A **mode** of a probability measure μ on a Banach space $\mathcal{U}$ is a point $u^\star \in \mathcal{U}$ of maximum probability — in the case of a Bayesian posterior, a **MAP estimator**.
- Without a reference uniform measure, modes must be defined intrinsically.
- A **strong mode** (Dashti et al., 2013):

$$u^\star \text{ is a strong mode of } \mu \iff \lim_{r \to 0} \frac{\sup_{u \in \mathcal{U}} \mu(B(u,r))}{\mu(B(u^\star,r))} \leq 1.$$

- A **weak mode** (Helin and Burger, 2015) with respect to a subspace $E \subset \mathcal{U}$:

$$u^\star \text{ is an } E\text{-weak mode of } \mu \iff \lim_{r \to 0} \frac{\sup_{v \in E} \mu(B(u^\star + v,r))}{\mu(B(u^\star,r))} \leq 1.$$

- All strong modes are weak modes, but are all weak modes strong modes?
- Lie and Sullivan (2017) replace the norm ball B by any open, bounded neighbourhood K of 0 in a topological vector space $\mathcal{U}$.

The set K is said to have *locally inwardly translatable boundary* ∂K if, for all $z \in \partial K$, there exists $v \in \mathcal{U} \setminus \{0\}$ and an open neighbourhood W of z such that, for all $0 < \lambda < 1$,

$$\lambda v + W \cap \partial K \subset K. \tag{LITB}$$

The *coincident limiting ratios* condition holds for $x \in \mathcal{U}$ and $E \subset \mathcal{U}$ if

$$\lim_{r \to 0} \frac{\sup_{v \in E} \mu(K(x+v, r))}{\mu(K(x, r))} = \lim_{r \to 0} \frac{\sup_{z \in \mathcal{U}} \mu(K(z, r))}{\mu(K(x, r))}. \tag{CLR}$$

The set K is said to have *locally inwardly translatable boundary* ∂K if, for all $z \in \partial K$, there exists $v \in \mathcal{U} \setminus \{0\}$ and an open neighbourhood W of z such that, for all $0 < \lambda < 1$,

$$\lambda v + W \cap \partial K \subset K. \tag{LITB}$$

The *coincident limiting ratios* condition holds for $x \in \mathcal{U}$ and $E \subset \mathcal{U}$ if

$$\lim_{r \to 0} \frac{\sup_{v \in E} \mu(K(x+v,r))}{\mu(K(x,r))} = \lim_{r \to 0} \frac{\sup_{z \in \mathcal{U}} \mu(K(z,r))}{\mu(K(x,r))}. \tag{CLR}$$

Theorem 1 (Lie and Sullivan, 2017)

If K satisfies (LITB)*, and if E is topologically dense in $\mathcal{U}$, then* (CLR) *holds for all $x \in \mathcal{U}$.*

Let E contain the origin or be topologically dense in a neighbourhood of the origin, and let $u^\star$ be an E-weak mode. Then $u^\star$ is a strong mode if and only if $u^\star$ and E satisfy (CLR).

Hence, if E is topologically dense in $\mathcal{U}$ and K satisfies (LITB)*, then $u^\star$ is a strong mode if and only if it is an E-weak mode.*

Random integrators for ODEs

Lie, Stuart, and Sullivan (2017a) arXiv:1703.03680

- Aim: provide a randomised numerical solution to an ODE, where the stochasticity in the solution represents the accumulated impact of truncation error along the flow.
- Numerical analysis objective: quantify the convergence rate of such methods.

Random integrators for ODEs

Lie, Stuart, and Sullivan (2017a) arXiv:1703.03680

- Aim: provide a randomised numerical solution to an ODE, where the stochasticity in the solution represents the accumulated impact of truncation error along the flow.
- Numerical analysis objective: quantify the convergence rate of such methods.
- For a randomised numerical solution $(U_k)_{k=0}^{T/h}$ with time step $h > 0$, we seek a result of the form

$$\mathbb{E}\left[\sup_{0 \leq k \leq T/h} |U_k - u(kh)|^2 \right] \leq Ch^r,$$

and relate r to the order of deterministic methods.

Random integrators for ODEs

Lie, Stuart, and Sullivan (2017a) arXiv:1703.03680

- Aim: provide a randomised numerical solution to an ODE, where the stochasticity in the solution represents the accumulated impact of truncation error along the flow.
- Numerical analysis objective: quantify the convergence rate of such methods.
- For a randomised numerical solution $(U_k)_{k=0}^{T/h}$ with time step $h > 0$, we seek a result of the form

$$\mathbb{E}\left[\sup_{0 \leq k \leq T/h} |U_k - u(kh)|^2\right] \leq Ch^r,$$

  and relate r to the order of deterministic methods.
- See the talk of **Han Cheng Lie** later for further discussion — what kinds of integrators, what kinds of vector fields / flows, and what technical conditions.

Schäfer, Sullivan, and Owhadi (2017) arXiv:1706.02205

- The (cubic) cost of inverting a Gram matrix $\Theta := (G(x_i, x_j))_{i,j\in\mathcal{I}}$ of a kernel G is a major computational bottleneck in Gaussian process techniques.

Fast operations on kernel matrices

Schäfer, Sullivan, and Owhadi (2017) arXiv:1706.02205

- The (cubic) cost of inverting a Gram matrix $\Theta := (G(x_i, x_j))_{i,j \in \mathcal{I}}$ of a kernel G is a major computational bottleneck in Gaussian process techniques.
- A probabilistic interpretation of Cholesky factorisation yields a simple novel algorithm for inversion of "nice" dense kernel matrices at near-linear cost.
- Simple idea: perform a zero-fill-in incomplete Cholesky factorisation `ichol(0)` with respect to some sparsity pattern $S_\rho \subset \mathcal{I} \times \mathcal{I}$, $\rho > 0$ an "interaction radius".

Fast operations on kernel matrices

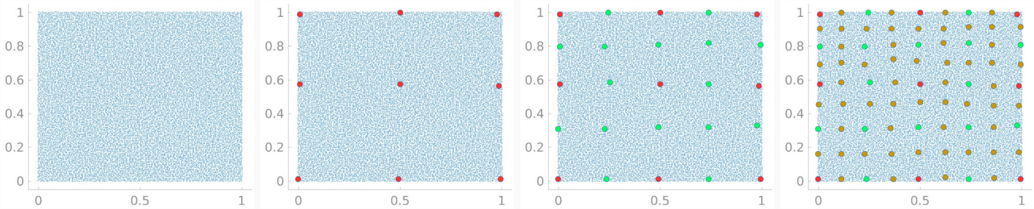
Schäfer, Sullivan, and Owhadi (2017) arXiv:1706.02205

- The (cubic) cost of inverting a Gram matrix $\Theta := (G(x_i, x_j))_{i,j \in \mathcal{I}}$ of a kernel G is a major computational bottleneck in Gaussian process techniques.
- A probabilistic interpretation of Cholesky factorisation yields a simple novel algorithm for inversion of "nice" dense kernel matrices at near-linear cost.
- Simple idea: perform a zero-fill-in incomplete Cholesky factorisation `ichol(0)` with respect to some sparsity pattern $S_\rho \subset \mathcal{I} \times \mathcal{I}$, $\rho > 0$ an "interaction radius".
- The clever part: a coarse-to-fine ordering of the x_i yielding near-linear sparsity pattern in near-linear complexity. This can be justified using quite easy heuristics, and proved using the theory of gamblets (Owhadi and Scovel, 2017).

Schäfer, Sullivan, and Owhadi (2017) arXiv:1706.02205

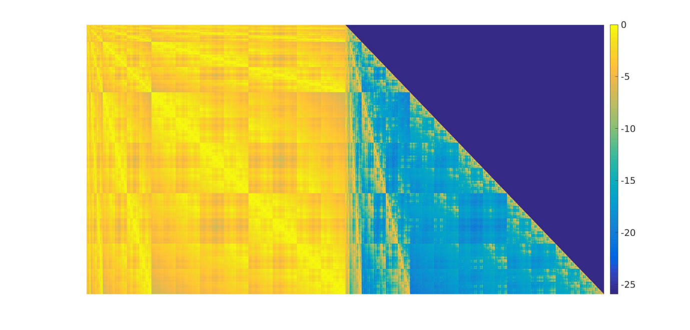
- The (cubic) cost of inverting a Gram matrix $\Theta := (G(x_i, x_j))_{i,j\in\mathcal{I}}$ of a kernel G is a major computational bottleneck in Gaussian process techniques.
- A probabilistic interpretation of Cholesky factorisation yields a simple novel algorithm for inversion of "nice" dense kernel matrices at near-linear cost.
- Simple idea: perform a zero-fill-in incomplete Cholesky factorisation `ichol(0)` with respect to some sparsity pattern $S_\rho \subset \mathcal{I} \times \mathcal{I}$, $\rho > 0$ an "interaction radius".
- The clever part: a coarse-to-fine ordering of the x_i yielding near-linear sparsity pattern in near-linear complexity. This can be justified using quite easy heuristics, and proved using the theory of gamblets (Owhadi and Scovel, 2017).
- Setting: $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain; $s > d/2$ is an integer, $\mathcal{L} \colon H_0^s(\Omega) \to H^{-s}(\Omega)$ is a linear, local, bounded, invertible, positive and self-adjoint operator and $G = \mathcal{L}^{-1}$ its Green's function; $\{x_i\}_{i\in\mathcal{I}} \subset \Omega$ is a "nice" point set.

Fast operations on kernel matrices



Above: ordering a near-uniform data set from coarse to fine.

Below: a dense kernel matrix in this ordering, and its `ichol(0)` factor.



Theorem 2 (Schäfer et al., 2017)

Under mild technical conditions, the `ichol(0)` *factorization of the sparsified kernel matrix* $\Theta_\rho := \Theta \mathbb{1}_{S_\rho}$ *has computational complexity* $\mathrm{O}(N \log(N) \rho^d)$ *in space and* $\mathrm{O}(N \log^2(N) \rho^{2d})$ *in time; finding* S_ρ *can also be achieved at this complexity.*

Theorem 2 (Schäfer et al., 2017)

Under mild technical conditions, the `ichol(0)` *factorization of the sparsified kernel matrix* $\Theta_\rho := \Theta \mathbb{1}_{S_\rho}$ *has computational complexity* $\mathrm{O}(N \log(N) \rho^d)$ *in space and* $\mathrm{O}(N \log^2(N) \rho^{2d})$ *in time; finding* S_ρ *can also be achieved at this complexity.*

Theorem 3 (Schäfer et al., 2017)

Let $\{x_i\}_{i \in \mathcal{I}} \subset \Omega$ *be such that*

$$\frac{\max_{x \in \Omega} \min_{i \in \mathcal{I}} \operatorname{dist}(x_i, x)}{\min_{i \neq j \in \mathcal{I}} \left(\operatorname{dist}\left(x_i, \{x_j\} \cup \partial\Omega\right)\right)} \leq \frac{1}{\delta}$$

and define $\Theta_{i,j} := G(x_i, x_j)$*. Then the* `ichol(0)` *factor* L_ρ *of* Θ_ρ *has approximation error*

$$\left\| \Theta - P L_\rho L_\rho^{\mathrm{T}} P^{\mathrm{T}} \right\| \leq C \operatorname{poly}(N) \exp(-\gamma \rho).$$

In particular, an ε*-approximation of* Θ *can be obtained in computational complexity* $\mathrm{O}(N \log(N) \log^d(N/\varepsilon))$ *in space and* $\mathrm{O}(N \log^2(N) \log^{2d}(N/\varepsilon))$ *in time.*

Lie, Sullivan, and Teckentrup (2017b) arXiv:1712.05717

Bayesian inverse problem à la Stuart (2010)

BIP with prior μ_0 on $\mathcal{U}$, data $y \in \mathcal{Y}$, and negative log-likelihood $\Phi \colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$: realise the posterior μ^y on $\mathcal{U}$

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u;y))}{Z(y)}. \tag{1}$$

Lie, Sullivan, and Teckentrup (2017b) arXiv:1712.05717

Bayesian inverse problem à la Stuart (2010)

BIP with prior μ_0 on $\mathcal{U}$, data $y \in \mathcal{Y}$, and negative log-likelihood $\Phi \colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$: realise the posterior μ^y on $\mathcal{U}$

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u;y))}{Z(y)}. \tag{1}$$

- How does replacing Φ by a randomised numerical approximation Φ_N impact μ^y?
  - Φ_N could be a kriging/GP surrogate for Φ (Stuart and Teckentrup, 2018);
  - $\mathcal{Y}$ could be high-dimensional and Φ_N could result from subsampling;
  - a deterministic forward model $G \colon \mathcal{U} \to \mathcal{Y}$ inside Φ could be replaced by a PN forward model G_N in Φ_N.
- Goal: transfer the (probabilistic) convergence rate $\Phi_N \to \Phi$ to a (probabilistic) convergence rate $\mu_N^y \to \mu^y$.

- Replacing Φ by Φ_N in (1), we obtain a random approximation μ_N^{samp} of μ:

$$\frac{\mathrm{d}\mu_N^{\mathrm{samp}}}{\mathrm{d}\mu_0}(u) := \frac{\exp(-\Phi_N(u))}{Z_N^{\mathrm{samp}}}, \qquad (2)$$

$$Z_N^{\mathrm{samp}} := \mathbb{E}_{\mu_0}\left[\exp(-\Phi_N(\cdot))\right].$$

Random and deterministic approximate posteriors

- Replacing Φ by Φ_N in (1), we obtain a random approximation μ_N^{samp} of μ:

$$\frac{\mathrm{d}\mu_N^{\mathrm{samp}}}{\mathrm{d}\mu_0}(u) := \frac{\exp(-\Phi_N(u))}{Z_N^{\mathrm{samp}}}, \qquad (2)$$

$$Z_N^{\mathrm{samp}} := \mathbb{E}_{\mu_0}\left[\exp(-\Phi_N(\cdot))\right].$$

- Taking the expectation of the random likelihood gives a deterministic approximation:

$$\frac{\mathrm{d}\mu_N^{\mathrm{marg}}}{\mathrm{d}\mu_0}(u) := \frac{\mathbb{E}_{\nu_N}\left[\exp(-\Phi_N(u))\right]}{\mathbb{E}_{\nu_N}\left[Z_N^{\mathrm{samp}}\right]}.$$

Random and deterministic approximate posteriors

- Replacing Φ by Φ_N in (1), we obtain a random approximation μ_N^{samp} of μ:

$$\frac{\mathrm{d}\mu_N^{\text{samp}}}{\mathrm{d}\mu_0}(u) := \frac{\exp(-\Phi_N(u))}{Z_N^{\text{samp}}}, \qquad (2)$$

$$Z_N^{\text{samp}} := \mathbb{E}_{\mu_0}\left[\exp(-\Phi_N(\cdot))\right].$$

- Taking the expectation of the random likelihood gives a deterministic approximation:

$$\frac{\mathrm{d}\mu_N^{\text{marg}}}{\mathrm{d}\mu_0}(u) := \frac{\mathbb{E}_{\nu_N}\left[\exp(-\Phi_N(u))\right]}{\mathbb{E}_{\nu_N}\left[Z_N^{\text{samp}}\right]}.$$

- An alternative deterministic approximation can be obtained by taking the expected value of the density $(Z_N^{\text{samp}})^{-1}e^{-\Phi_N(u)}$ in (2). However, μ_N^{marg} provides a clear interpretation as the posterior obtained by the approximation of the true data likelihood $e^{-\Phi(u)}$ by $\mathbb{E}_{\nu_N}\left[e^{-\Phi_N(u)}\right]$, and is more amenable to sampling methods such as pseudo-marginal MCMC (Beaumont, 2003; Andrieu and Roberts, 2009).

Theorem 4 (Lie, Sullivan, and Teckentrup, 2017b)

For suitable Hölder exponents $p_1, p'_1, p_2, \dots$ quantifying the integrability of Φ and Φ_N, we obtain deterministic convergence $\mu_N^{\mathrm{marg}} \to \mu$ and mean-square convergence $\mu_N^{\mathrm{samp}} \to \mu$ in the Hellinger metric:

$$d_{\mathrm{H}}(\mu, \mu_N^{\mathrm{marg}}) \leq C \left\| \mathbb{E}_{\nu_N} \left[|\Phi - \Phi_N|^{p'_2} \right]^{1/p'_2} \right\|_{L^{2p'_1 p'_3}_{\mu_0}(\mathcal{U})},$$

$$\mathbb{E}_{\nu_N} \left[d_{\mathrm{H}}\left(\mu, \mu_N^{\mathrm{samp}}\right)^2 \right]^{1/2} \leq D \left\| \mathbb{E}_{\nu_N} \left[|\Phi - \Phi_N|^{2q'_1} \right]^{1/2q'_1} \right\|_{L^{2q'_2}_{\mu_0}(\mathcal{U})}.$$

Theorem 4 (Lie, Sullivan, and Teckentrup, 2017b)

For suitable Hölder exponents $p_1, p_1', p_2, \dots$ quantifying the integrability of Φ and Φ_N, we obtain deterministic convergence $\mu_N^{\mathrm{marg}} \to \mu$ and mean-square convergence $\mu_N^{\mathrm{samp}} \to \mu$ in the Hellinger metric:

$$d_{\mathrm{H}}\big(\mu, \mu_N^{\mathrm{marg}}\big) \leq C \left\| \mathbb{E}_{\nu_N} \left[|\Phi - \Phi_N|^{p_2'} \right]^{1/p_2'} \right\|_{L_{\mu_0}^{2p_1' p_3'}(\mathcal{U})},$$

$$\mathbb{E}_{\nu_N}\left[d_{\mathrm{H}}\big(\mu, \mu_N^{\mathrm{samp}}\big)^2 \right]^{1/2} \leq D \left\| \mathbb{E}_{\nu_N} \left[|\Phi - \Phi_N|^{2q_1'} \right]^{1/2q_1'} \right\|_{L_{\mu_0}^{2q_2'}(\mathcal{U})}.$$

- There are similar results for approximation of G by G_N (in a fixed quadratic misfit potential).
- One application: random solution of ODEs as in **Han Cheng Lie**'s talk.
- Another application: random reduction of high-dimensional data…

Example: Monte Carlo approximation of high-dimensional misfits

We consider a Monte Carlo approximation Φ_N of a quadratic potential Φ (Nemirovski et al., 2008; Shapiro et al., 2009), further applied and analysed in the MAP estimator context by Le et al. (2017). This approximation is particularly useful for data $y \in \mathbb{R}^J$, $J \gg 1$.

$$\begin{aligned}
\Phi(u) &:= \frac{1}{2}\left\| \Gamma^{-\frac{1}{2}}(y - G(u)) \right\|^2 \\
&= \frac{1}{2}\big(\Gamma^{-\frac{1}{2}}(y - G(u))\big)^{\mathsf{T}} \mathbb{E}[\sigma\sigma^{\mathsf{T}}] \big(\Gamma^{-\frac{1}{2}}(y - G(u))\big) && \text{where } \mathbb{E}[\sigma] = 0 \in \mathbb{R}^J, \mathbb{E}[\sigma\sigma^{\mathsf{T}}] = I_{J\times J} \\
&= \frac{1}{2}\mathbb{E}\left[\left| \sigma^{\mathsf{T}} \big(\Gamma^{-\frac{1}{2}}(y - G(u))\big) \right|^2 \right] \\
&\approx \frac{1}{2N} \sum_{i=1}^{N} \left| {\sigma^{(i)}}^{\mathsf{T}} \big(\Gamma^{-\frac{1}{2}}(y - G(u))\big) \right|^2 && \text{for i.i.d. } \sigma^{(1)}, \dots, \sigma^{(N)} \overset{\mathrm{d}}{=} \sigma \\
&= \frac{1}{2}\left\| \Sigma_N^{\mathsf{T}} \big(\Gamma^{-\frac{1}{2}}(y - G(u)) \right\|^2 && \text{for } \Sigma_N := \frac{1}{\sqrt{N}} [\sigma^{(1)} \cdots \sigma^{(N)}] \in \mathbb{R}^{J\times N} \\
&=: \Phi_N(u).
\end{aligned}$$

Le et al. (2017) suggest that a good choice for the $\mathbb{R}^J$-valued random vector σ would be one with independent and identically distributed (i.i.d.) entries from a sub-Gaussian probability distribution, e.g.

- the Gaussian distribution: $\sigma_j \sim \mathcal{N}(0, 1)$, for $j = 1, \dots, J$; and
- the ℓ-sparse distribution: for $\ell \in [0, 1)$, let $s := \frac{1}{1-\ell} \geq 1$ and set, for $j = 1, \dots, J$,

$$\sigma_j := \sqrt{s} \begin{cases} 1, & \text{with probability } \frac{1}{2s}, \\ 0, & \text{with probability } \ell = 1 - \frac{1}{s}, \\ -1, & \text{with probability } \frac{1}{2s}. \end{cases}$$

Example: Monte Carlo approximation of high-dimensional misfits

- Le et al. (2017) observe that, for large J and moderate $N \approx 10$, the random potential Φ_N and the original potential Φ are very similar, in particular having approximately the same minimisers and minimum values.
- Statistically, these correspond to the maximum likelihood estimators under Φ and Φ_N being very similar; after weighting by a prior, this corresponds to similarity of maximum a posteriori (MAP) estimators.
- Here, we study the BIP instead of the MAP problem, and thus the corresponding conjecture is that the deterministic posterior $\mathrm{d}\mu(u) \propto \exp(-\Phi(u))\,\mathrm{d}\mu_0(u)$ is well approximated by the random posterior $\mathrm{d}\mu_N^{\mathrm{samp}}(u) \propto \exp(-\Phi_N(u))\,\mathrm{d}\mu_0(u)$.

Example: Well-posedness of BIPs with Monte Carlo misfits

Applying the general results to this setting gives the following transfer of the Monte Carlo convergence rate from the approximation of Φ to the approximation of μ:

Proposition 5

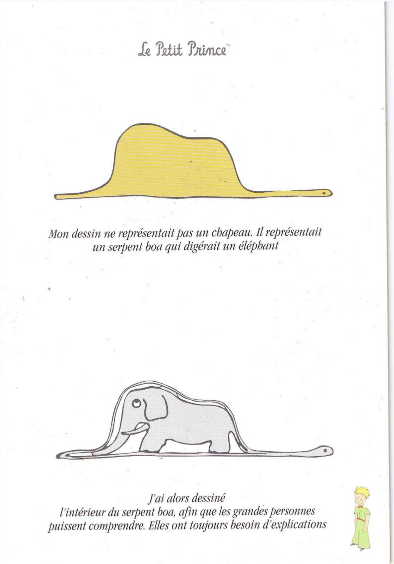
Suppose that the entries of σ are i.i.d. ℓ-sparse, for some $\ell \in [0, 1)$, and that $\Phi \in L^2_{\mu_0}(\mathcal{U})$. Then there exists a constant C, independent of N, such that

$$\left(\mathbb{E}_\sigma\left[d_{\mathrm{H}}\left(\mu, \mu_N^{\mathrm{samp}}\right)^2\right]\right)^{1/2} \leq \frac{C}{\sqrt{N}}.$$

For technical reasons to do with the non-compactness of the support and finiteness of MGFs of maxima, the current proof technique does not work for the Gaussian case.

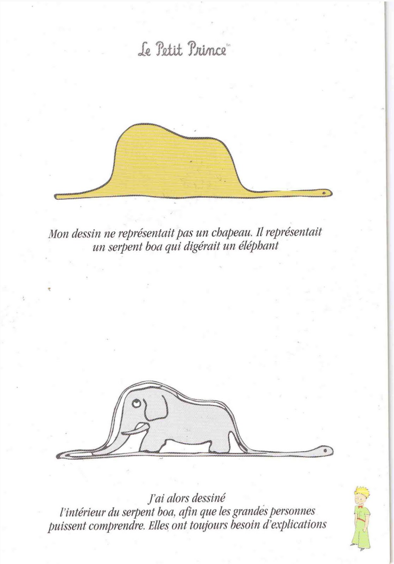
Summary

Summary



Antoine de Saint-Exupéry,
The Little Prince, 1943

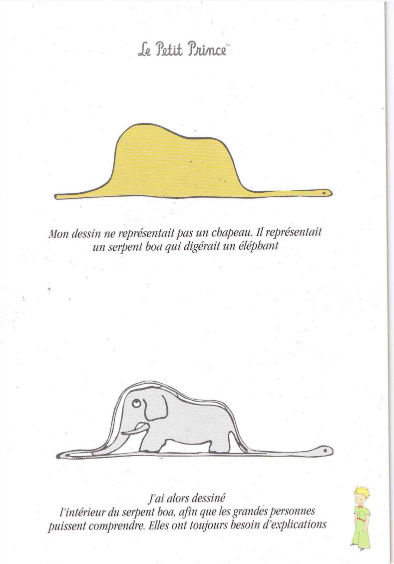
Summary



Antoine de Saint-Exupéry,
The Little Prince, 1943

- We appear to have swallowed an elephant…

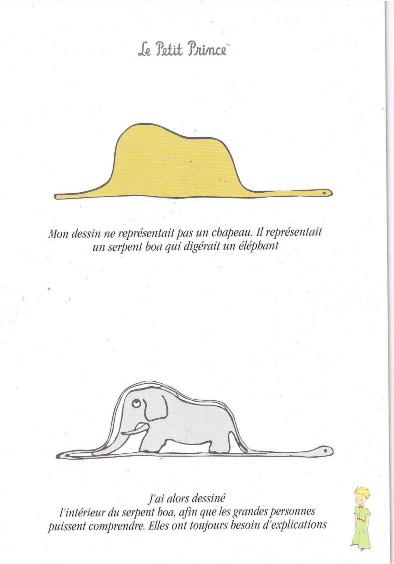
SUMMARY



Antoine de Saint-Exupéry,
The Little Prince, 1943

- We appear to have swallowed an elephant…
- …in the sense that there is a lot of PN activity going on, the SAMSI WG is just part of it, and more is coming.

Summary



Antoine de Saint-Exupéry,
The Little Prince, 1943

- We appear to have swallowed an elephant…
- …in the sense that there is a lot of PN activity going on, the SAMSI WG is just part of it, and more is coming.
- The field is coming into mathematical and statistical maturity.
  - Proof-of-concept examples ✓
  - Rigorous analysis and underpinnings ✓
  - "Killer apps" ✓/✗
  - Reference implementations ✗

Thank You!

References I

C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2): 697–725, 2009. doi:10.1214/07-AOS574.

M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3): 1139–1160, 2003.

F.-X. Briol, J. Cockayne, O. Teymur, W. W. Yoo, M. Schober, and P. Hennig. Contributed discussion on article by Chkrebtii, Campbell, Calderhead, and Girolami. *Bayesian Anal.*, 11(4):1285–1293, 2016. doi:10.1214/16-BA1017A.

O. A. Chkrebtii. Adaptive grid designs for state-space probabilistic ODE solvers, In preparation.

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

J. Cockayne, C. J. Oates, and M. Girolami. A Bayesian conjugate gradient method, 2018. arXiv:1801.05242.

P. R. Conrad, M. Girolami, S. Särkkä, A. M. Stuart, and K. C. Zygalakis. Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Stat. Comput.*, 27(4), 2016. doi:10.1007/s11222-016-9671-0.

M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 27, 2013. doi:10.1088/0266-5611/29/9/095017.

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.

V. Dukic and D. M. Bortz. Uncertainty quantification using probabilistic numerics: application to models in mathematical epidemiology. *Inv. Probl. Sci. Engrng*, 26(2):223–232, 2018. doi:10.1080/17415977.2017.1312364.

T. Helin and M. Burger. Maximum *a posteriori* probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 22, 2015. doi:10.1088/0266-5611/31/8/085009.

P. Hennig, H. Kersting, and T. J. Sullivan. Convergence rates of Gaussian ODE filters, In preparation.

T. Karvonen, C. J. Oates, and S. Särkkä. A Bayes–Sard cubature method, 2018. arXiv:1804.03016.

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

E. B. Le, A. Myers, T. Bui-Thanh, and Q. P. Nguyen. A data-scalable randomized misfit approach for solving large-scale PDE-constrained inverse problems. *Inverse Probl.*, 33(6):065003, 2017. doi:10.1088/1361-6420/aa6cbd.

H. C. Lie and T. J. Sullivan. Equivalence of weak and strong modes of measures on topological vector spaces, 2017. arXiv:1708.02516.

H. C. Lie, A. M. Stuart, and T. J. Sullivan. Strong convergence rates of probabilistic integrators for ordinary differential equations, 2017a. arXiv:1703.03680.

H. C. Lie, T. J. Sullivan, and A. L. Teckentrup. Random forward models and log-likelihoods in Bayesian inverse problems, 2017b. arXiv:1712.05717.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.*, 19(4):1574–1609, 2008. doi:10.1137/070704277.

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

C. J. Oates, S. Niederer, A. Lee, F.-X. Briol, and M. Girolami. Probabilistic models for integration error in assessment of functional cardiac models. In *Advances in Neural Information Processing Systems*, volume 30. 2017b. arXiv:1606.06841.

H. Owhadi and C. Scovel. Universal scalable robust solvers from computational information games and fast eigenspace adapted multiresolution analysis, 2017. arXiv:1703.10761.

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

J. Rathinavel and F. J. Hickernell. Automatic Bayesian cubature, In preparation.

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

F. Schäfer, T. J. Sullivan, and H. Owhadi. Compression, inversion, and approximate PCA of dense kernel matrices at near-linear computational complexity, 2017. arXiv:1706.02205.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2009. doi:10.1137/1.9780898718751.

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.

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

A. M. Stuart and A. L. Teckentrup. Posterior consistency for Gaussian process approximations of Bayesian posterior distributions. *Math. Comp.*, 87(310):721–753, 2018. doi:10.1090/mcom/3244.

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.

L. N. Trefethen. Is Gauss quadrature better than Clenshaw–Curtis? *SIAM Rev.*, 50(1):67–87, 2008. doi:10.1137/060659831.

X. Xi, F.-X. Briol, and M. Girolami. Bayesian quadrature for multiple related integrals, 2018. arXiv:1801.04153.