Probabilistic Numerical Methods

Chris. J. Oates
Newcastle University
Alan Turing Institute

June 2017 @ Dobbiaco

Download link: `oates.work/dobbiaco`

Conspirators

Mark Girolami
Imperial & ATI

Jon Cockayne
Warwick

F-X Briol
Warwick

Tim Sullivan
F.U. Berlin

Philipp Hennig
MPI Tuebingen

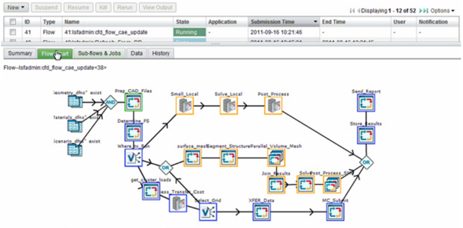
Mike Osborne
Oxford

Dino Sejdinovic
Oxford

Andrew Stuart
Caltech

Motivation: Computational Pipelines

Numerical analysis for the "drag and drop" era of computational pipelines:



[Fig: IBM High Performance Computation]

The sophistication and scale of modern computer models creates an urgent need to better understand the propagation and accumulation of numerical error within arbitrary - often large - pipelines of computation, so that "numerical risk" to end-users can be controlled.

Consider numerical solution for $x \in \mathcal{X}$ of the Poisson equation

$$\begin{aligned} -\Delta x &= f && \text{in } D \\ x &= g && \text{on } \partial D \end{aligned}$$

based on (noiseless) information of the form

$$A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_m) \\ x(t_{m+1}) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} f(t_1) \\ \vdots \\ f(t_m) \\ g(t_{m+1}) \\ \vdots \\ g(t_n) \end{bmatrix}, \qquad \{t_i\}_{i=1}^{m} \in D, \quad \{t_i\}_{i=m+1}^{d} \in \partial D.$$

This is an ill-posed inverse problem and must be regularised.

The onus is on us to establish principled statistical foundations that are general.

Motivation: Solution of Poisson's Equation

Consider numerical solution for $x \in \mathcal{X}$ of the Poisson equation

$$\begin{aligned} -\Delta x &= f && \text{in } D \\ x &= g && \text{on } \partial D \end{aligned}$$

based on (noiseless) information of the form

$$A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_m) \\ x(t_{m+1}) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} f(t_1) \\ \vdots \\ f(t_m) \\ g(t_{m+1}) \\ \vdots \\ g(t_n) \end{bmatrix}, \qquad \{t_i\}_{i=1}^{m} \in D, \quad \{t_i\}_{i=m+1}^{d} \in \partial D.$$

This is an ill-posed inverse problem and must be regularised.

The onus is on us to establish principled statistical foundations that are general.

Motivation: Solution of Poisson's Equation

Consider numerical solution for $x \in \mathcal{X}$ of the Poisson equation

$$\begin{aligned} -\Delta x &= f && \text{in } D \\ x &= g && \text{on } \partial D \end{aligned}$$

based on (noiseless) information of the form

$$A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_m) \\ x(t_{m+1}) \\ \vdots \\ x(t_n) \end{bmatrix} = \begin{bmatrix} f(t_1) \\ \vdots \\ f(t_m) \\ g(t_{m+1}) \\ \vdots \\ g(t_n) \end{bmatrix}, \qquad \{t_i\}_{i=1}^{m} \in D, \quad \{t_i\}_{i=m+1}^{d} \in \partial D.$$

This is an ill-posed inverse problem and must be regularised.

The onus is on us to establish principled statistical foundations that are general.

Insight: Numerical Analysis as Bayesian Inversion

The *Bayesian* approach, popularised in Stuart (2010), can be used:

- a *prior* measure P_x is placed on $\mathcal{X}$
- a *posterior* measure $P_{x|a}$ is defined as the "restriction of P_x to those functions $x \in \mathcal{X}$ for which

$$A(x) = a \qquad \text{e.g.} \quad A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_n) \end{bmatrix} = a$$

is satisfied" (to be formalised).

$\implies$ Principled and general uncertainty quantification for numerical methods.

Insight: Numerical Analysis as Bayesian Inversion

The *Bayesian* approach, popularised in Stuart (2010), can be used:

- a *prior* measure P_x is placed on $\mathcal{X}$
- a *posterior* measure $P_{x|a}$ is defined as the "restriction of P_x to those functions $x \in \mathcal{X}$ for which

$$A(x) = a \qquad \text{e.g.} \quad A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_n) \end{bmatrix} = a$$

is satisfied" (to be formalised).

$\Longrightarrow$ Principled and general uncertainty quantification for numerical methods.

The *Bayesian* approach, popularised in Stuart (2010), can be used:

- a *prior* measure P_x is placed on $\mathcal{X}$
- a *posterior* measure $P_{x|a}$ is defined as the "restriction of P_x to those functions $x \in \mathcal{X}$ for which

$$A(x) = a \qquad \text{e.g.} \quad A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_n) \end{bmatrix} = a$$

is satisfied" (to be formalised).

$\Longrightarrow$ Principled and general uncertainty quantification for numerical methods.

Insight: Numerical Analysis as Bayesian Inversion

The *Bayesian* approach, popularised in Stuart (2010), can be used:

- a *prior* measure P_x is placed on $\mathcal{X}$
- a *posterior* measure $P_{x|a}$ is defined as the "restriction of P_x to those functions $x \in \mathcal{X}$ for which

$$A(x) = a \qquad \text{e.g.} \quad A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_n) \end{bmatrix} = a$$

is satisfied" (to be formalised).

$\Longrightarrow$ Principled and general uncertainty quantification for numerical methods.

The *Bayesian* approach, popularised in Stuart (2010), can be used:

- a *prior* measure P_x is placed on $\mathcal{X}$
- a *posterior* measure $P_{x|a}$ is defined as the "restriction of P_x to those functions $x \in \mathcal{X}$ for which

$$A(x) = a \qquad \text{e.g.} \quad A(x) = \begin{bmatrix} -\Delta x(t_1) \\ \vdots \\ -\Delta x(t_n) \end{bmatrix} = a$$

is satisfied" (to be formalised).

$\Longrightarrow$ Principled and general uncertainty quantification for numerical methods.

The Agenda

Part I

1. First Job: Elicit the Abstract Structure
2. Second Job: Review of Classical Numerical Analysis
3. Third Job: Discuss Choice of P_x

Part II

4. Fourth Job: Check Well-Defined, Existence and Uniqueness
5. Fifth Job: Algorithms to Access $P_{x|a}$

Part III

4. Sixth Job: Analysis of the Gaussian Case
5. Seventh Job: Solution of Integrals, in Detail

The Agenda

Part IV

4. Eighth Job: Solution of PDEs
5. Ninth Job: Characterise Optimal Information

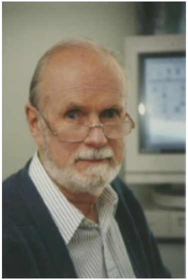
Part V

4. Tenth Job: Extension to More Challenging Integrals
5. Eleventh Job: Non-Bayesian Methods?

Part VI

4. Twelfth Job: Introduction to Graphical Models
5. Thirteenth Job: Pipelines of Computation

Part I

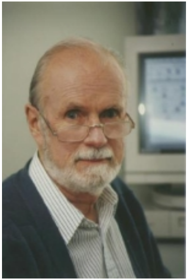


Tests of Probabilistic Models for Propagation of Roundoff Errors

T. E. HULL, University of Toronto; J. R. SWENSON, New York University (Ed: J. Traub)

Communications of the ACM, 9(2):108–113, 1966.

> *In any prolonged computation it is generally assumed that the accumulated effect of roundoff errors is in some sense statistical. The purpose of this paper is to give precise descriptions of certain probabilistic models for roundoff error, and then to describe a series of experiments for testing the validity of these models. It is concluded that the models are in general very good. Discrepancies are both rare and mild. The test techniques can also be used to experiment with various types of special arithmetic.*



Tests of Probabilistic Models for Propagation of Roundoff Errors

T. E. HULL, University of Toronto; J. R. SWENSON, New York University (Ed: J. Traub)

Communications of the ACM, 9(2):108–113, 1966.

> *In any prolonged computation it is generally assumed that the accumulated effect of roundoff errors is in some sense statistical. The purpose of this paper is to give precise descriptions of certain probabilistic models for roundoff error, and then to describe a series of experiments for testing the validity of these models. It is concluded that the models are in general very good. Discrepancies are both rare and mild. The test techniques can also be used to experiment with various types of special arithmetic.*

First Job: Elicit the Abstract Structure

Abstractly, consider an unobserved state variable $x \in \mathcal{X}$ together with:

- A *quantity of interest*, denoted $Q(x) \in \mathcal{Q}$
- An *information operator*, denoted $x \mapsto A(x) \in \mathcal{A}$. $(\dim(\mathcal{A}) = n < \infty)$

Examples:

Task	$Q(x)$	$A(x)$
Integration	$\int x(t)\nu(\mathrm{d}t)$	$\{x(t_i)\}_{i=1}^n$
Optimisation	$\arg\max x(t)$	$\{x(t_i)\}_{i=1}^n$
Solution of Poisson Eqn	$x(\cdot)$	$\{-\Delta x(t_i)\}_{i=1}^m \cup \{x(t_i)\}_{i=m+1}^n$

Abstractly, consider an unobserved state variable $x \in \mathcal{X}$ together with:

- A *quantity of interest*, denoted $Q(x) \in \mathcal{Q}$
- An *information operator*, denoted $x \mapsto A(x) \in \mathcal{A}$. $(\dim(\mathcal{A}) = n < \infty)$

Examples:

Task	$Q(x)$	$A(x)$
Integration	$\int x(t)\nu(\mathrm{d}t)$	$\{x(t_i)\}_{i=1}^{n}$
Optimisation	$\arg\max x(t)$	$\{x(t_i)\}_{i=1}^{n}$
Solution of Poisson Eqn	$x(\cdot)$	$\{-\Delta x(t_i)\}_{i=1}^{m} \cup \{x(t_i)\}_{i=m+1}^{n}$

Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $T_\# \mu$ denote the "pushforward" measure, st $(T_\# \mu)(S) = \mu(T^{-1}(S))$.

		Classical Numerical Method	Probabilistic Numerical Method
Inputs	Assumed	e.g. smoothness	$P_x \in \mathcal{P}_\mathcal{X}$
	Information	$a \in \mathcal{A}$	$a \in \mathcal{A}$
Output		$b(a) \in \mathcal{Q}$	$B(P_x, a) \in \mathcal{P}_\mathcal{Q}$

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_{x|a}$.

Abstract Structure

Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $T_\# \mu$ denote the "pushforward" measure, st $(T_\# \mu)(S) = \mu(T^{-1}(S))$.

		Classical Numerical Method	Probabilistic Numerical Method
Inputs	Assumed	e.g. smoothness	$P_x \in \mathcal{P}_\mathcal{X}$
	Information	$a \in \mathcal{A}$	$a \in \mathcal{A}$
Output		$b(a) \in \mathcal{Q}$	$B(P_x, a) \in \mathcal{P}_\mathcal{Q}$

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_{x|a}$.

Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $T_\# \mu$ denote the "pushforward" measure, st $(T_\# \mu)(S) = \mu(T^{-1}(S))$.

		Classical Numerical Method	Probabilistic Numerical Method
Inputs	Assumed	e.g. smoothness	$P_x \in \mathcal{P}_\mathcal{X}$
	Information	$a \in \mathcal{A}$	$a \in \mathcal{A}$
Output		$b(a) \in \mathcal{Q}$	$B(P_x, a) \in \mathcal{P}_\mathcal{Q}$

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_{x|a}$.

Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $T_\# \mu$ denote the "pushforward" measure, st $(T_\# \mu)(S) = \mu(T^{-1}(S))$.

		Classical Numerical Method	Probabilistic Numerical Method
Inputs	Assumed	e.g. smoothness	$P_x \in \mathcal{P}_\mathcal{X}$
	Information	$a \in \mathcal{A}$	$a \in \mathcal{A}$
Output		$b(a) \in \mathcal{Q}$	$B(P_x, a) \in \mathcal{P}_\mathcal{Q}$

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_{x|a}$.

Abstract Structure

Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $T_\# \mu$ denote the "pushforward" measure, st $(T_\# \mu)(S) = \mu(T^{-1}(S))$.

		Classical Numerical Method	Probabilistic Numerical Method
Inputs	Assumed	e.g. smoothness	$P_x \in \mathcal{P}_\mathcal{X}$
	Information	$a \in \mathcal{A}$	$a \in \mathcal{A}$
Output		$b(a) \in \mathcal{Q}$	$B(P_x, a) \in \mathcal{P}_\mathcal{Q}$

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_{x|a}$.

Abstract Structure

Let $\mathcal{P}_\bullet$ denote the set of distributions on $\bullet$.

Let $T_\# \mu$ denote the "pushforward" measure, st $(T_\# \mu)(S) = \mu(T^{-1}(S))$.

		Classical Numerical Method	Probabilistic Numerical Method
Inputs	Assumed	e.g. smoothness	$P_x \in \mathcal{P}_\mathcal{X}$
	Information	$a \in \mathcal{A}$	$a \in \mathcal{A}$
Output		$b(a) \in \mathcal{Q}$	$B(P_x, a) \in \mathcal{P}_\mathcal{Q}$

A Probabilistic Numerical Method is Bayesian iff $B(P_x, a) = Q_\# P_{x|a}$.

The Grand Plan

The grand plan of these lectures is to study the object

$$B(P_x, a) = Q_{\#} P_{x|a}$$

of a Bayesian probabilistic numerical method in detail.

But, before we jump in, we will first review some background on classical numerical analysis and information-based complexity of numerical methods.

The Grand Plan

The grand plan of these lectures is to study the object

$$B(P_x, a) = Q_{\#} P_{x|a}$$

of a Bayesian probabilistic numerical method in detail.

But, before we jump in, we will first review some background on classical numerical analysis and information-based complexity of numerical methods.

Second Job: Review of Classical Numerical Analysis

Consider a (classical) numerical method

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

for instance the trapezoidal rule

$$b(\{x(t_i)\}_{i=1}^{n}) = \sum_{i=1}^{n-1} (t_{i+1} - t_i) \frac{x(t_{i+1}) - x(t_i)}{2}.$$

In what sense should the performance of this method be assessed?

Typical considerations in numerical analysis:

1. Order of convergence
2. Numerical stability (e.g. floating point error propagation)

In the case of the trapezoidal rule, these are fairly dull.

Consider a (classical) numerical method

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

for instance the trapezoidal rule

$$b(\{x(t_i)\}_{i=1}^n) = \sum_{i=1}^{n-1} (t_{i+1} - t_i) \frac{x(t_{i+1}) - x(t_i)}{2}.$$

In what sense should the performance of this method be assessed?

Typical considerations in numerical analysis:

1. Order of convergence
2. Numerical stability (e.g. floating point error propagation)

In the case of the trapezoidal rule, these are fairly dull.

Consider a (classical) numerical method

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

for instance the trapezoidal rule

$$b(\{x(t_i)\}_{i=1}^{n}) = \sum_{i=1}^{n-1} (t_{i+1} - t_i) \frac{x(t_{i+1}) - x(t_i)}{2}.$$

In what sense should the performance of this method be assessed?

Typical considerations in numerical analysis:

1. Order of convergence
2. Numerical stability (e.g. floating point error propagation)

In the case of the trapezoidal rule, these are fairly dull.

Perhaps more interesting questions are raised in Information-Based Complexity:

Three core frameworks of information-based complexity:

1. "Worst-case" (minimise the maximal error)
2. "Average-case" (minimise the average error)
3. "Probabilistic" (minimise the cost required to achieve low error with high probability)

N.B. The third framework has (arguably) little to do with Probabilistic Numerics (as we will see in Part IV). But, to avoid confusion of the terminology, we won't discuss this framework further.

Perhaps more interesting questions are raised in Information-Based Complexity:

Three core frameworks of information-based complexity:

1. "Worst-case" (minimise the maximal error)
2. "Average-case" (minimise the average error)
3. "Probabilistic" (minimise the cost required to achieve low error with high probability)

N.B. The third framework has (arguably) little to do with Probabilistic Numerics (as we will see in Part IV). But, to avoid confusion of the terminology, we won't discuss this framework further.

Perhaps more interesting questions are raised in Information-Based Complexity:

Three core frameworks of information-based complexity:

1. "Worst-case" (minimise the maximal error)
2. "Average-case" (minimise the average error)
3. "Probabilistic" (minimise the cost required to achieve low error with high probability)

N.B. The third framework has (arguably) little to do with Probabilistic Numerics (as we will see in Part IV). But, to avoid confusion of the terminology, we won't discuss this framework further.

Worst Case Error

To set up the worst-case analysis, we need to restrict to a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and introduce a loss function $L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$.

Then define the *worst case error* of the method $M = (A, b)$:

$$e_{\mathrm{WCE}}(M) = \sup_{\|x\|_{\mathcal{X}} \leq 1} L(b(A(x)), Q(x))$$

Can consider minimisation of $e_{\mathrm{WCE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\operatorname*{arg\,inf}_{b:\mathcal{A}\to\mathcal{Q}} e_{\mathrm{WCE}}(M)$$

Such methods are "worst case optimal" for the given information operator A.

e.g. for $\|x\|_{\mathcal{X}} = (\int x(t)^2 \mathrm{d}t)^{1/2}$ and $L(q, q') = (q - q')^2$, the trapezium rule is worst case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

Worst Case Error

To set up the worst-case analysis, we need to restrict to a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and introduce a loss function $L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$.

Then define the *worst case error* of the method $M = (A, b)$:

$$e_{\text{WCE}}(M) = \sup_{\|x\|_{\mathcal{X}} \leq 1} L(b(A(x)), Q(x))$$

Can consider minimisation of $e_{\text{WCE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\arg\inf_{b:\mathcal{A}\to\mathcal{Q}} e_{\text{WCE}}(M)$$

Such methods are "worst case optimal" for the given information operator A.

e.g. for $\|x\|_{\mathcal{X}} = (\int x(t)^2 \mathrm{d}t)^{1/2}$ and $L(q, q') = (q - q')^2$, the trapezium rule is worst case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

To set up the worst-case analysis, we need to restrict to a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and introduce a loss function $L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$.

Then define the *worst case error* of the method $M = (A, b)$:

$$e_{\mathrm{WCE}}(M) \quad = \quad \sup_{\|x\|_{\mathcal{X}} \leq 1} L(b(A(x)), Q(x))$$

Can consider minimisation of $e_{\mathrm{WCE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\arg\inf_{b:\mathcal{A}\to\mathcal{Q}} e_{\mathrm{WCE}}(M)$$

Such methods are "worst case optimal" for the given information operator A.

e.g. for $\|x\|_{\mathcal{X}} = (\int x(t)^2 \mathrm{d}t)^{1/2}$ and $L(q, q') = (q - q')^2$, the trapezium rule is worst case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

Worst Case Error

To set up the worst-case analysis, we need to restrict to a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and introduce a loss function $L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$.

Then define the *worst case error* of the method $M = (A, b)$:

$$e_{\text{WCE}}(M) = \sup_{\|x\|_{\mathcal{X}} \leq 1} L(b(A(x)), Q(x))$$

Can consider minimisation of $e_{\text{WCE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\arg\inf_{b:\mathcal{A}\to\mathcal{Q}} e_{\text{WCE}}(M)$$

Such methods are "worst case optimal" for the given information operator A.

e.g. for $\|x\|_{\mathcal{X}} = (\int x(t)^2 \mathrm{d}t)^{1/2}$ and $L(q, q') = (q - q')^2$, the trapezium rule is worst case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

Worst Case Error

To set up the worst-case analysis, we need to restrict to a normed space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ and introduce a loss function $L : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$.

Then define the *worst case error* of the method $M = (A, b)$:

$$e_{\mathrm{WCE}}(M) \quad = \quad \sup_{\|x\|_{\mathcal{X}} \leq 1} L(b(A(x)), Q(x))$$

Can consider minimisation of $e_{\mathrm{WCE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\operatorname*{arg\,inf}_{b:\mathcal{A}\to\mathcal{Q}} e_{\mathrm{WCE}}(M)$$

Such methods are "worst case optimal" for the given information operator A.

e.g. for $\|x\|_{\mathcal{X}} = (\int x(t)^2 \mathrm{d}t)^{1/2}$ and $L(q, q') = (q - q')^2$, the trapezium rule is worst case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

To set up the average-case analysis, we instead need to restrict to a measurable space $(\mathcal{X}, \Sigma_{\mathcal{X}})$ and introduce a distribution $P_x \in \mathcal{P}_{\mathcal{X}}$.

Then define the *average case error* of the method $M = (A, b)$:

$$e_{\text{ACE}}(M) \quad = \quad \int L(b(A(x)), Q(x)) \, \mathrm{d}P_x$$

Can consider minimisation of $e_{\text{ACE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\arg\inf_{b:\mathcal{A}\to\mathcal{Q}} \; e_{\text{ACE}}(M)$$

Such methods are "average case optimal" for the information operator A.

e.g. for P_x the Weiner process and $L(q, q') = (q - q')^2$, the trapezium rule is average case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

Average Case Error

To set up the average-case analysis, we instead need to restrict to a measurable space $(\mathcal{X}, \Sigma_{\mathcal{X}})$ and introduce a distribution $P_x \in \mathcal{P}_{\mathcal{X}}$.

Then define the *average case error* of the method $M = (A, b)$:

$$e_{\text{ACE}}(M) = \int L(b(A(x)), Q(x)) \, \mathrm{d}P_x$$

Can consider minimisation of $e_{\text{ACE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\arg\inf_{b:\mathcal{A}\to\mathcal{Q}} e_{\text{ACE}}(M)$$

Such methods are "average case optimal" for the information operator A.

e.g. for P_x the Weiner process and $L(q, q') = (q - q')^2$, the trapezium rule is average case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

To set up the average-case analysis, we instead need to restrict to a measurable space $(\mathcal{X}, \Sigma_{\mathcal{X}})$ and introduce a distribution $P_x \in \mathcal{P}_{\mathcal{X}}$.

Then define the *average case error* of the method $M = (A, b)$:

$$e_{\text{ACE}}(M) \quad = \quad \int L(b(A(x)), Q(x)) \, \mathrm{d}P_x$$

Can consider minimisation of $e_{\text{ACE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\operatorname*{arg\,inf}_{b:\mathcal{A}\to\mathcal{Q}} \; e_{\text{ACE}}(M)$$

Such methods are "average case optimal" for the information operator A.

e.g. for P_x the Weiner process and $L(q, q') = (q - q')^2$, the trapezium rule is average case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

To set up the average-case analysis, we instead need to restrict to a measurable space $(\mathcal{X}, \Sigma_{\mathcal{X}})$ and introduce a distribution $P_x \in \mathcal{P}_{\mathcal{X}}$.

Then define the *average case error* of the method $M = (A, b)$:

$$e_{\text{ACE}}(M) \quad = \quad \int L(b(A(x)), Q(x)) \, \mathrm{d}P_x$$

Can consider minimisation of $e_{\text{ACE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\operatorname*{arg\,inf}_{b:\mathcal{A}\to\mathcal{Q}} \; e_{\text{ACE}}(M)$$

Such methods are "average case optimal" for the information operator A.

e.g. for P_x the Weiner process and $L(q, q') = (q - q')^2$, the trapezium rule is average case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

Average Case Error

To set up the average-case analysis, we instead need to restrict to a measurable space $(\mathcal{X}, \Sigma_{\mathcal{X}})$ and introduce a distribution $P_x \in \mathcal{P}_{\mathcal{X}}$.

Then define the *average case error* of the method $M = (A, b)$:

$$e_{\text{ACE}}(M) \quad = \quad \int L(b(A(x)), Q(x)) \, \mathrm{d}P_x$$

Can consider minimisation of $e_{\text{ACE}}(M)$ over the choice of $b : \mathcal{A} \to \mathcal{Q}$:

$$\underset{b:\mathcal{A}\to\mathcal{Q}}{\arg\inf} \; e_{\text{ACE}}(M)$$

Such methods are "average case optimal" for the information operator A.

e.g. for P_x the Weiner process and $L(q, q') = (q - q')^2$, the trapezium rule is average case optimal for $A(x) = [x(t_1), \ldots, x(t_n)]$ (modulo technical details - see Part IV).

Questions

Coincidence that the trapezoidal rule is both worst case optimal and average case optimal?

Closely related to Probabilistic Numerics?

Well, both involve a choice for P_x at least...

Questions

Coincidence that the trapezoidal rule is both worst case optimal and average case optimal?

Closely related to Probabilistic Numerics?

Well, both involve a choice for P_x at least...

Coincidence that the trapezoidal rule is both worst case optimal and average case optimal?

Closely related to Probabilistic Numerics?

Well, both involve a choice for P_x at least...

Third Job: Discuss Choice of P_x

Motivation: Beyond Gaussian Processes

Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a Banach space (i.e. a complete normed vector space; in this case over $\mathbb{R}$) equipped with a *Schauder basis* $\{\phi_i\}_{i=1}^{\infty}$. i.e. for each $x \in \mathcal{X}$ there exists a unique sequence $\alpha \in \mathbb{R}^{\infty}$ such that

$$x(\cdot) = \sum_{i=1}^{\infty} \alpha_i \phi_i(\cdot).$$

It will be further assumed that the basis is *normalised*, meaning that $\|\phi_i\|_{\mathcal{X}} = 1$ for all $i \in \mathbb{N}$.

Key Idea: Randomise the coefficients $\alpha \sim P_{\omega}$ and consider the push-forward $P_x = T_{\#}P_{\omega}$ where $T\alpha = \sum_{i=1}^{\infty} \alpha_i \phi_i$.

Question 1: How to select the basis elements ϕ_i?

Question 2: How to select the distribution of the coefficients α_i?

Motivation: Beyond Gaussian Processes

Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a Banach space (i.e. a complete normed vector space; in this case over $\mathbb{R}$) equipped with a *Schauder basis* $\{\phi_i\}_{i=1}^{\infty}$. i.e. for each $x \in \mathcal{X}$ there exists a unique sequence $\alpha \in \mathbb{R}^{\infty}$ such that

$$x(\cdot) \quad = \quad \sum_{i=1}^{\infty} \alpha_i \phi_i(\cdot).$$

It will be further assumed that the basis is *normalised*, meaning that $\|\phi_i\|_{\mathcal{X}} = 1$ for all $i \in \mathbb{N}$.

Key Idea: Randomise the coefficients $\alpha \sim P_{\omega}$ and consider the push-forward $P_x = T_{\#}P_{\omega}$ where $T\alpha = \sum_{i=1}^{\infty} \alpha_i \phi_i$.

Question 1: How to select the basis elements ϕ_i?

Question 2: How to select the distribution of the coefficients α_i?

Motivation: Beyond Gaussian Processes

Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a Banach space (i.e. a complete normed vector space; in this case over $\mathbb{R}$) equipped with a *Schauder basis* $\{\phi_i\}_{i=1}^{\infty}$. i.e. for each $x \in \mathcal{X}$ there exists a unique sequence $\alpha \in \mathbb{R}^{\infty}$ such that

$$x(\cdot) = \sum_{i=1}^{\infty} \alpha_i \phi_i(\cdot).$$

It will be further assumed that the basis is *normalised*, meaning that $\|\phi_i\|_{\mathcal{X}} = 1$ for all $i \in \mathbb{N}$.

Key Idea: Randomise the coefficients $\alpha \sim P_{\omega}$ and consider the push-forward $P_x = T_{\#} P_{\omega}$ where $T\alpha = \sum_{i=1}^{\infty} \alpha_i \phi_i$.

Question 1: How to select the basis elements ϕ_i?

Question 2: How to select the distribution of the coefficients α_i?

Prior Construction

Motivation: Beyond Gaussian Processes

Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a Banach space (i.e. a complete normed vector space; in this case over $\mathbb{R}$) equipped with a *Schauder basis* $\{\phi_i\}_{i=1}^{\infty}$. i.e. for each $x \in \mathcal{X}$ there exists a unique sequence $\alpha \in \mathbb{R}^{\infty}$ such that

$$x(\cdot) \quad = \quad \sum_{i=1}^{\infty} \alpha_i \phi_i(\cdot).$$

It will be further assumed that the basis is *normalised*, meaning that $\|\phi_i\|_{\mathcal{X}} = 1$ for all $i \in \mathbb{N}$.

Key Idea: Randomise the coefficients $\alpha \sim P_\omega$ and consider the push-forward $P_x = T_{\#} P_\omega$ where $T\alpha = \sum_{i=1}^{\infty} \alpha_i \phi_i$.

Question 1: How to select the basis elements ϕ_i?

Question 2: How to select the distribution of the coefficients α_i?

Motivation: Beyond Gaussian Processes

Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a Banach space (i.e. a complete normed vector space; in this case over $\mathbb{R}$) equipped with a *Schauder basis* $\{\phi_i\}_{i=1}^{\infty}$. i.e. for each $x \in \mathcal{X}$ there exists a unique sequence $\alpha \in \mathbb{R}^{\infty}$ such that

$$x(\cdot) = \sum_{i=1}^{\infty} \alpha_i \phi_i(\cdot).$$

It will be further assumed that the basis is *normalised*, meaning that $\|\phi_i\|_{\mathcal{X}} = 1$ for all $i \in \mathbb{N}$.

Key Idea: Randomise the coefficients $\alpha \sim P_\omega$ and consider the push-forward $P_x = T_\# P_\omega$ where $T\alpha = \sum_{i=1}^{\infty} \alpha_i \phi_i$.

Question 1: How to select the basis elements ϕ_i?

Question 2: How to select the distribution of the coefficients α_i?

Motivation: Beyond Gaussian Processes

Let $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ be a Banach space (i.e. a complete normed vector space; in this case over $\mathbb{R}$) equipped with a *Schauder basis* $\{\phi_i\}_{i=1}^{\infty}$. i.e. for each $x \in \mathcal{X}$ there exists a unique sequence $\alpha \in \mathbb{R}^{\infty}$ such that

$$x(\cdot) = \sum_{i=1}^{\infty} \alpha_i \phi_i(\cdot).$$

It will be further assumed that the basis is *normalised*, meaning that $\|\phi_i\|_{\mathcal{X}} = 1$ for all $i \in \mathbb{N}$.

Key Idea: Randomise the coefficients $\alpha \sim P_\omega$ and consider the push-forward $P_x = T_{\#}P_\omega$ where $T\alpha = \sum_{i=1}^{\infty} \alpha_i \phi_i$.

Question 1: How to select the basis elements ϕ_i?

Question 2: How to select the distribution of the coefficients α_i?

Selection of Basis Elements ϕ_i

Some insight is provided in the case where $\mathcal{X}$ is a reproducing kernel Hilbert space (see Part III) by Mercer's theorem:

Let $k(t, t')$ be a symmetric positive definite kernel on $\mathcal{X}$. If

$$\int \sqrt{k(t,t)} \mathrm{d}\nu(t) \quad < \quad \infty$$

then there exist $\{\psi_i\}_{i=1}^{\infty} \subset L^2(\nu)$ and $\{\lambda_i\}_{i=1}^{\infty} \subset [0, \infty)$ such that

$$k(t, t') \quad = \quad \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i(t') \qquad \text{("kernel trick")}.$$

Moreover the $\{\lambda_i^{1/2} \psi_i\}_{i=1}^{n}$ form an orthonormal basis of $\mathcal{X}$.

So could, for instance, use $\phi_i = \lambda_i^{1/2} \psi_i$ to build a Schauder basis for $\mathcal{X}$.

But we could equally take a basis of orthogonal polynomials, wavelets, etc. Next, we only assume that we are provided with a normalised Schauder basis.

Selection of Basis Elements ϕ_i

Some insight is provided in the case where $\mathcal{X}$ is a reproducing kernel Hilbert space (see Part III) by Mercer's theorem:

Let $k(t, t')$ be a symmetric positive definite kernel on $\mathcal{X}$. If

$$\int \sqrt{k(t,t)}\mathrm{d}\nu(t) \quad < \quad \infty$$

then there exist $\{\psi_i\}_{i=1}^{\infty} \subset L^2(\nu)$ and $\{\lambda_i\}_{i=1}^{\infty} \subset [0, \infty)$ such that

$$k(t, t') \quad = \quad \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i(t') \qquad \text{("kernel trick").}$$

Moreover the $\{\lambda_i^{1/2} \psi_i\}_{i=1}^{n}$ form an orthonormal basis of $\mathcal{X}$.

So could, for instance, use $\phi_i = \lambda_i^{1/2} \psi_i$ to build a Schauder basis for $\mathcal{X}$.

But we could equally take a basis of orthogonal polynomials, wavelets, etc. Next, we only assume that we are provided with a normalised Schauder basis.

Selection of Basis Elements ϕ_i

Some insight is provided in the case where $\mathcal{X}$ is a reproducing kernel Hilbert space (see Part III) by Mercer's theorem:

Let $k(t, t')$ be a symmetric positive definite kernel on $\mathcal{X}$. If

$$\int \sqrt{k(t,t)} \mathrm{d}\nu(t) \quad < \quad \infty$$

then there exist $\{\psi_i\}_{i=1}^{\infty} \subset L^2(\nu)$ and $\{\lambda_i\}_{i=1}^{\infty} \subset [0, \infty)$ such that

$$k(t, t') \quad = \quad \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i(t') \qquad \text{("kernel trick").}$$

Moreover the $\{\lambda_i^{1/2} \psi_i\}_{i=1}^{n}$ form an orthonormal basis of $\mathcal{X}$.

So could, for instance, use $\phi_i = \lambda_i^{1/2} \psi_i$ to build a Schauder basis for $\mathcal{X}$.

But we could equally take a basis of orthogonal polynomials, wavelets, etc. Next, we only assume that we are provided with a normalised Schauder basis.

Selection of Basis Elements ϕ_i

Some insight is provided in the case where $\mathcal{X}$ is a reproducing kernel Hilbert space (see Part III) by Mercer's theorem:

Let $k(t, t')$ be a symmetric positive definite kernel on $\mathcal{X}$. If

$$\int \sqrt{k(t,t)} \mathrm{d}\nu(t) \quad < \quad \infty$$

then there exist $\{\psi_i\}_{i=1}^{\infty} \subset L^2(\nu)$ and $\{\lambda_i\}_{i=1}^{\infty} \subset [0, \infty)$ such that

$$k(t, t') \quad = \quad \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i(t') \qquad \text{(“kernel trick”)}.$$

Moreover the $\{\lambda_i^{1/2} \psi_i\}_{i=1}^{n}$ form an orthonormal basis of $\mathcal{X}$.

So could, for instance, use $\phi_i = \lambda_i^{1/2} \psi_i$ to build a Schauder basis for $\mathcal{X}$.

But we could equally take a basis of orthogonal polynomials, wavelets, etc. Next, we only assume that we are provided with a normalised Schauder basis.

Selection of Basis Elements ϕ_i

Some insight is provided in the case where $\mathcal{X}$ is a reproducing kernel Hilbert space (see Part III) by Mercer's theorem:

Let $k(t, t')$ be a symmetric positive definite kernel on $\mathcal{X}$. If

$$\int \sqrt{k(t,t)} \mathrm{d}\nu(t) \quad < \quad \infty$$

then there exist $\{\psi_i\}_{i=1}^{\infty} \subset L^2(\nu)$ and $\{\lambda_i\}_{i=1}^{\infty} \subset [0, \infty)$ such that

$$k(t, t') \quad = \quad \sum_{i=1}^{\infty} \lambda_i \psi_i(t) \psi_i(t') \qquad \text{("kernel trick")}.$$

Moreover the $\{\lambda_i^{1/2} \psi_i\}_{i=1}^{n}$ form an orthonormal basis of $\mathcal{X}$.

So could, for instance, use $\phi_i = \lambda_i^{1/2} \psi_i$ to build a Schauder basis for $\mathcal{X}$.

But we could equally take a basis of orthogonal polynomials, wavelets, etc. Next, we only assume that we are provided with a normalised Schauder basis.

Let $\{\phi_i\}_{i=1}^{\infty}$ be a (normalised) Schauder basis.

Consider a decomposition

$$\alpha_i = \gamma_i u_i$$

where γ_i are fixed and u_i are random; independent and identically distributed.

When does the summation

$$\sum_{i=1}^{\infty} \alpha_i \phi_i \quad \left(= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i \right)$$

converge in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$?

NB: The Karhunen-Loève expansion corresponds to $\gamma_i \equiv 1$, $\phi_i = \lambda_i^{1/2} \psi_i$ and $u_i \sim N(0,1)$. This clearly does not converge in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$!

Distribution of the Coefficients α_i

Let $\{\phi_i\}_{i=1}^{\infty}$ be a (normalised) Schauder basis.

Consider a decomposition

$$\alpha_i = \gamma_i \, u_i$$

where γ_i are fixed and u_i are random; independent and identically distributed.

When does the summation

$$\sum_{i=1}^{\infty} \alpha_i \phi_i \quad \left(= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i \right)$$

converge in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$?

NB: The Karhunen-Loève expansion corresponds to $\gamma_i \equiv 1$, $\phi_i = \lambda_i^{1/2} \psi_i$ and $u_i \sim N(0,1)$. This clearly does not converge in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$!

Distribution of the Coefficients α_i

Let $\{\phi_i\}_{i=1}^{\infty}$ be a (normalised) Schauder basis.

Consider a decomposition

$$\alpha_i = \gamma_i \, u_i$$

where γ_i are fixed and u_i are random; independent and identically distributed.

When does the summation

$$\sum_{i=1}^{\infty} \alpha_i \phi_i \quad \left(= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i \right)$$

converge in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$?

NB: The Karhunen-Loève expansion corresponds to $\gamma_i \equiv 1$, $\phi_i = \lambda_i^{1/2} \psi_i$ and $u_i \sim N(0,1)$. This clearly does not converge in $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$!

Distribution of the Coefficients α_i

Let $\{\phi_i\}_{i=1}^{\infty}$ be a (normalised) Schauder basis.

Consider a decomposition

$$\alpha_i \quad = \quad \gamma_i \, u_i$$

where γ_i are fixed and u_i are random; independent and identically distributed.

When does the summation

$$\sum_{i=1}^{\infty} \alpha_i \phi_i \quad \left(= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i \right)$$

converge in $(\mathcal{X}, \| \cdot \|_{\mathcal{X}})$?

NB: The Karhunen-Loève expansion corresponds to $\gamma_i \equiv 1$, $\phi_i = \lambda_i^{1/2} \psi_i$ and $u_i \sim N(0, 1)$. This clearly does <u>not</u> converge in $(\mathcal{X}, \| \cdot \|_{\mathcal{X}})$!

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$
\begin{aligned}
(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} &= \left\| \sum_{i=M+1}^{N} \alpha_i \phi_i \right\|_{\mathcal{X}} \\
&\leq \sum_{i=M+1}^{N} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= \sum_{i=M+1}^{N} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&= \sum_{i=M+1}^{\infty} |\gamma_i| \rightarrow 0 \text{ as } M \rightarrow \infty \quad (\text{def'n of } \ell^1)
\end{aligned}
$$

$\implies (x^N)_{N=1}^{\infty}$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$
\begin{aligned}
(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} &= \left\| \sum_{i=M+1}^N \alpha_i \phi_i \right\|_{\mathcal{X}} \\
&\leq \sum_{i=M+1}^N |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= \sum_{i=M+1}^N |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&= \sum_{i=M+1}^\infty |\gamma_i| \to 0 \text{ as } M \to \infty \quad (\text{def'n of } \ell^1)
\end{aligned}
$$

$\implies (x^N)_{N=1}^\infty$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$
\begin{aligned}
(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} &= \left\| \sum_{i=M+1}^{N} \alpha_i \phi_i \right\|_{\mathcal{X}} \\
&\leq \sum_{i=M+1}^{N} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= \sum_{i=M+1}^{N} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&= \sum_{i=M+1}^{\infty} |\gamma_i| \to 0 \text{ as } M \to \infty \quad (\text{def'n of } \ell^1)
\end{aligned}
$$

$\implies (x^N)_{N=1}^{\infty}$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} = \left\| \sum_{i=M+1}^{N} \alpha_i \phi_i \right\|_{\mathcal{X}}$$

$$\leq \sum_{i=M+1}^{N} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1}$$

$$= \sum_{i=M+1}^{N} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1}$$

$$= \sum_{i=M+1}^{\infty} |\gamma_i| \to 0 \text{ as } M \to \infty \quad (\text{def'n of } \ell^1)$$

$\implies (x^N)_{N=1}^{\infty}$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$
\begin{aligned}
(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} &= \left\| \sum_{i=M+1}^{N} \alpha_i \phi_i \right\|_{\mathcal{X}} \\
&\leq \sum_{i=M+1}^{N} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= \sum_{i=M+1}^{N} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&= \sum_{i=M+1}^{\infty} |\gamma_i| \to 0 \text{ as } M \to \infty \quad (\text{def'n of } \ell^1)
\end{aligned}
$$

$\implies (x^N)_{N=1}^{\infty}$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$
\begin{aligned}
(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} &= \left\| \sum_{i=M+1}^{N} \alpha_i \phi_i \right\|_{\mathcal{X}} \\
&\leq \sum_{i=M+1}^{N} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= \sum_{i=M+1}^{N} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&= \sum_{i=M+1}^{\infty} |\gamma_i| \to 0 \text{ as } M \to \infty \quad (\text{def'n of } \ell^1)
\end{aligned}
$$

$\implies (x^N)_{N=1}^{\infty}$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$
\begin{aligned}
(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} &= \left\| \sum_{i=M+1}^{N} \alpha_i \phi_i \right\|_{\mathcal{X}} \\
&\leq \sum_{i=M+1}^{N} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= \sum_{i=M+1}^{N} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&= \sum_{i=M+1}^{\infty} |\gamma_i| \to 0 \text{ as } M \to \infty \quad (\text{def'n of } \ell^1)
\end{aligned}
$$

$\implies (x^N)_{N=1}^{\infty}$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Well-defined? Let $x^N = \sum_{i=1}^N \alpha_i \phi_i$.

$$
\begin{aligned}
(N > M) \quad \|x^N - x^M\|_{\mathcal{X}} &= \left\| \sum_{i=M+1}^{N} \alpha_i \phi_i \right\|_{\mathcal{X}} \\
&\leq \sum_{i=M+1}^{N} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= \sum_{i=M+1}^{N} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&= \sum_{i=M+1}^{\infty} |\gamma_i| \to 0 \text{ as } M \to \infty \quad (\text{def'n of } \ell^1)
\end{aligned}
$$

$\implies (x^N)_{N=1}^{\infty}$ Cauchy $\implies$ converges to a limit in the Banach space $\mathcal{X}$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- Boundedness: Consider the specific choice $\|x\|_{\mathcal{X}} = \|x\|_\infty = \sup_t |x(t)|$. Then we have that

$$
\begin{aligned}
x(t) &\geq -\sum_{i=1}^{\infty} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= -\sum_{i=1}^{\infty} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&\geq -\sum_{i=1}^{\infty} |\gamma_i| \\
&= -\|\gamma\|_1 < \infty \quad \text{(def'n of } \ell^1\text{)}.
\end{aligned}
$$

- Similarly $x(t) \leq \|\gamma\|_1$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- **Boundedness:** Consider the specific choice $\|x\|_{\mathcal{X}} = \|x\|_\infty = \sup_t |x(t)|$. Then we have that

$$
\begin{aligned}
x(t) &\geq -\sum_{i=1}^{\infty} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= -\sum_{i=1}^{\infty} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&\geq -\sum_{i=1}^{\infty} |\gamma_i| \\
&= -\|\gamma\|_1 < \infty \quad (\text{def'n of } \ell^1).
\end{aligned}
$$

- Similarly $x(t) \leq \|\gamma\|_1$.

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Boundedness: Consider the specific choice $\|x\|_{\mathcal{X}} = \|x\|_\infty = \sup_t |x(t)|$. Then we have that

$$
\begin{aligned}
x(t) &\geq -\sum_{i=1}^{\infty} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= -\sum_{i=1}^{\infty} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&\geq -\sum_{i=1}^{\infty} |\gamma_i| \\
&= -\|\gamma\|_1 < \infty \quad \text{(def'n of } \ell^1\text{)}.
\end{aligned}
$$

- Similarly $x(t) \leq \|\gamma\|_1$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Boundedness: Consider the specific choice $\|x\|_{\mathcal{X}} = \|x\|_\infty = \sup_t |x(t)|$. Then we have that

$$\begin{aligned} x(t) &\geq -\sum_{i=1}^{\infty} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\ &= -\sum_{i=1}^{\infty} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\ &\geq -\sum_{i=1}^{\infty} |\gamma_i| \\ &= -\|\gamma\|_1 \; < \; \infty \quad (\text{def'n of } \ell^1). \end{aligned}$$

- Similarly $x(t) \leq \|\gamma\|_1$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Boundedness: Consider the specific choice $\|x\|_{\mathcal{X}} = \|x\|_\infty = \sup_t |x(t)|$. Then we have that

$$
\begin{aligned}
x(t) &\geq -\sum_{i=1}^{\infty} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\
&= -\sum_{i=1}^{\infty} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\
&\geq -\sum_{i=1}^{\infty} |\gamma_i| \\
&= -\|\gamma\|_1 < \infty \quad \text{(def'n of } \ell^1\text{)}.
\end{aligned}
$$

- Similarly $x(t) \leq \|\gamma\|_1$.

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- Boundedness: Consider the specific choice $\|x\|_{\mathcal{X}} = \|x\|_\infty = \sup_t |x(t)|$. Then we have that

$$\begin{aligned} x(t) &\geq -\sum_{i=1}^{\infty} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\ &= -\sum_{i=1}^{\infty} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\ &\geq -\sum_{i=1}^{\infty} |\gamma_i| \\ &= -\|\gamma\|_1 \ < \ \infty \quad (\text{def'n of } \ell^1). \end{aligned}$$

- Similarly $x(t) \leq \|\gamma\|_1$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Boundedness: Consider the specific choice $\|x\|_{\mathcal{X}} = \|x\|_\infty = \sup_t |x(t)|$. Then we have that

$$\begin{aligned} x(t) &\geq -\sum_{i=1}^{\infty} |\alpha_i| \underbrace{\|\phi_i\|_{\mathcal{X}}}_{=1} \\ &= -\sum_{i=1}^{\infty} |\gamma_i| \underbrace{|u_i|}_{\leq_P 1} \\ &\geq -\sum_{i=1}^{\infty} |\gamma_i| \\ &= -\|\gamma\|_1 < \infty \quad \text{(def'n of } \ell^1\text{)}. \end{aligned}$$

- Similarly $x(t) \leq \|\gamma\|_1$.

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Distribution of marginals: Fix t. Then

$$\begin{aligned} x(t) &= \sum_{i=1}^{\infty} \alpha_i \phi_i(t) \\ &= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i(t) \\ &= \sum_{i=1}^{\infty} \underbrace{[\gamma_i \phi_i(t)]}_{\text{weight}} \underbrace{u_i}_{\sim U[-1,1]} \\ &\sim \text{"weighted Irwin-Hall"} \end{aligned}$$

but note that it is bounded (unlike, e.g. for a Gaussian process).

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1, 1]$.

- Distribution of marginals: Fix t. Then

$$
\begin{aligned}
x(t) &= \sum_{i=1}^{\infty} \alpha_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \underbrace{[\gamma_i \phi_i(t)]}_{\text{weight}} \underbrace{u_i}_{\sim U[-1,1]} \\
&\sim \text{"weighted Irwin-Hall"}
\end{aligned}
$$

but note that it is bounded (unlike, e.g. for a Gaussian process).

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Distribution of marginals: Fix t. Then

$$
\begin{aligned}
x(t) &= \sum_{i=1}^{\infty} \alpha_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \underbrace{[\gamma_i \phi_i(t)]}_{\text{weight}} \underbrace{u_i}_{\sim U[-1,1]} \\
&\sim \text{“weighted Irwin-Hall”}
\end{aligned}
$$

but note that it is bounded (unlike, e.g. for a Gaussian process).

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Distribution of marginals: Fix t. Then

$$
\begin{aligned}
x(t) &= \sum_{i=1}^{\infty} \alpha_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \underbrace{[\gamma_i \phi_i(t)]}_{\text{weight}} \underbrace{u_i}_{\sim U[-1,1]} \\
&\sim \text{“weighted Irwin-Hall”}
\end{aligned}
$$

but note that it is bounded (unlike, e.g. for a Gaussian process).

Distribution of the Coefficients $\alpha_i = \gamma_i u_i$

Example: The *uniform* prior takes $\gamma \in \ell^1$ and $u_i \sim U[-1,1]$.

- Distribution of marginals: Fix t. Then

$$
\begin{aligned}
x(t) &= \sum_{i=1}^{\infty} \alpha_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \gamma_i u_i \phi_i(t) \\
&= \sum_{i=1}^{\infty} \underbrace{[\gamma_i \phi_i(t)]}_{\text{weight}} \underbrace{u_i}_{\sim U[-1,1]} \\
&\sim \text{“weighted Irwin-Hall”}
\end{aligned}
$$

but note that it is bounded (unlike, e.g. for a Gaussian process).

Exercise: The Gaussian Case

Example: Consider functions $x : \mathbb{R}^d \to \mathbb{R}$. The *Gaussian* prior has coefficients $\alpha_i = \gamma_i u_i$ where $\gamma_i \asymp i^{-\frac{s}{d}}$ and $u_i \sim N(0,1)$, for some $s > 0$.

- Let $\{\phi_i\}_{i=1}^{\infty}$ be orthonormal in a Hilbert space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$.
- Let $x = \sum_{i=1}^{\infty} \alpha_i \phi_i$ and consider a norm $\|x\|_{\mathcal{X},t}^2 = \sum_{i=1}^{\infty} i^{\frac{2t}{d}} \alpha_i^2$ for some $t > 0$. (This is known as a "Hilbert scale" of $\mathcal{X}$.)
- Question: For which values of s, t does $x = \sum_{i=1}^{\infty} \gamma_i u_i \phi_i$ exist as a $L_P^2(\mathcal{X}, \|\cdot\|_{\mathcal{X},t})$ limit?
- i.e. For which values of s, t is $\mathbb{E}[\|x\|_{\mathcal{X},t}^2] < \infty$?

Exercise: The Gaussian Case

Example: Consider functions $x : \mathbb{R}^d \to \mathbb{R}$. The *Gaussian* prior has coefficients $\alpha_i = \gamma_i u_i$ where $\gamma_i \asymp i^{-\frac{s}{d}}$ and $u_i \sim N(0,1)$, for some $s > 0$.

- Let $\{\phi_i\}_{i=1}^{\infty}$ be orthonormal in a Hilbert space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$.
- Let $x = \sum_{i=1}^{\infty} \alpha_i \phi_i$ and consider a norm $\|x\|_{\mathcal{X},t}^2 = \sum_{i=1}^{\infty} i^{\frac{2t}{d}} \alpha_i^2$ for some $t > 0$. (This is known as a "Hilbert scale" of $\mathcal{X}$.)
- Question: For which values of s, t does $x = \sum_{i=1}^{\infty} \gamma_i u_i \phi_i$ exist as a $L_P^2(\mathcal{X}, \|\cdot\|_{\mathcal{X},t})$ limit?
- i.e. For which values of s, t is $\mathbb{E}[\|x\|_{\mathcal{X},t}^2] < \infty$?

Example: Consider functions $x : \mathbb{R}^d \to \mathbb{R}$. The *Gaussian* prior has coefficients $\alpha_i = \gamma_i u_i$ where $\gamma_i \asymp i^{-\frac{s}{d}}$ and $u_i \sim N(0,1)$, for some $s > 0$.

- Let $\{\phi_i\}_{i=1}^{\infty}$ be orthonormal in a Hilbert space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$.
- Let $x = \sum_{i=1}^{\infty} \alpha_i \phi_i$ and consider a norm $\|x\|_{\mathcal{X},t}^2 = \sum_{i=1}^{\infty} i^{\frac{2t}{d}} \alpha_i^2$ for some $t > 0$. (This is known as a "Hilbert scale" of $\mathcal{X}$.)
- Question: For which values of s, t does $x = \sum_{i=1}^{\infty} \gamma_i u_i \phi_i$ exist as a $L_P^2(\mathcal{X}, \|\cdot\|_{\mathcal{X},t})$ limit?
- i.e. For which values of s, t is $\mathbb{E}[\|x\|_{\mathcal{X},t}^2] < \infty$?

Exercise: The Gaussian Case

Example: Consider functions $x : \mathbb{R}^d \to \mathbb{R}$. The *Gaussian* prior has coefficients $\alpha_i = \gamma_i u_i$ where $\gamma_i \asymp i^{-\frac{s}{d}}$ and $u_i \sim N(0,1)$, for some $s > 0$.

- Let $\{\phi_i\}_{i=1}^{\infty}$ be orthonormal in a Hilbert space $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$.
- Let $x = \sum_{i=1}^{\infty} \alpha_i \phi_i$ and consider a norm $\|x\|_{\mathcal{X},t}^2 = \sum_{i=1}^{\infty} i^{\frac{2t}{d}} \alpha_i^2$ for some $t > 0$. (This is known as a "Hilbert scale" of $\mathcal{X}$.)
- Question: For which values of s, t does $x = \sum_{i=1}^{\infty} \gamma_i u_i \phi_i$ exist as a $L_P^2(\mathcal{X}, \|\cdot\|_{\mathcal{X},t})$ limit?
- i.e. For which values of s, t is $\mathbb{E}[\|x\|_{\mathcal{X},t}^2] < \infty$?

Example: Consider functions $x : \mathbb{R}^d \to \mathbb{R}$. The *Gaussian* prior has coefficients $\alpha_i = \gamma_i u_i$ where $\gamma_i \asymp i^{-\frac{s}{d}}$ and $u_i \sim N(0, 1)$, for some $s > 0$.

- Let $\{\phi_i\}_{i=1}^{\infty}$ be orthonormal in a Hilbert space $(\mathcal{X}, \| \cdot \|_{\mathcal{X}})$.
- Let $x = \sum_{i=1}^{\infty} \alpha_i \phi_i$ and consider a norm $\|x\|_{\mathcal{X},t}^2 = \sum_{i=1}^{\infty} i^{\frac{2t}{d}} \alpha_i^2$ for some $t > 0$. (This is known as a "Hilbert scale" of $\mathcal{X}$.)
- Question: For which values of s, t does $x = \sum_{i=1}^{\infty} \gamma_i u_i \phi_i$ exist as a $L_P^2(\mathcal{X}, \| \cdot \|_{\mathcal{X},t})$ limit?
- i.e. For which values of s, t is $\mathbb{E}[\|x\|_{\mathcal{X},t}^2] < \infty$?

Conclusion

In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- This enables us to go beyond Gaussian processes - but demands additional technical details.

END OF PART I

Conclusion

In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- This enables us to go beyond Gaussian processes - but demands additional technical details.

END OF PART I

Conclusion

In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- This enables us to go beyond Gaussian processes - but demands additional technical details.

END OF PART I

Conclusion

In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- This enables us to go beyond Gaussian processes - but demands additional technical details.

END OF PART I

Conclusion

In Part I it has been argued that:

- The onus is on us to establish principled statistical foundations that are general.
- The Bayesian approach to inverse problems, popularised in Stuart [2010], provides such a framework.
- This enables us to go beyond Gaussian processes - but demands additional technical details.

END OF PART I