Exploiting Sparsity in Bayesian Inverse Problems of Parametric Operator Equations

Claudia Schillings, Christoph Schwab

MascotNum

Workshop on Computer Experiments and Meta-models for Uncertainty Quantification

April 23rd-25th 2014, ETH Zürich

SAM

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

European Research Council
Established by the European Commission

research supported by the European Research Council under FP7 Grant AdG247277

Outline

Inverse Problem

Physical Model

$$G(u) \to \delta$$

- u parameter vector / parameter function
- G the forward map modelling the physical process
- δ result / observations

Forward Problem

Find the output δ for given parameters u

$\rightarrow$ **well-posed**

Inverse Problem

Find the parameters u from (noisy) observations δ

$\rightarrow$ **ill-posed**

Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

- u parameter vector / parameter function
- G the forward map modelling the physical process
- $\mathcal{O}$ bounded, linear observation operator
- $\mathcal{G}$ uncertainty-to-observation map, $\mathcal{G} = \mathcal{O} \circ G$
- δ noisy observations
- η observational noise

Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Deterministic optimization problem

$$\min_u \frac{1}{2}\|\delta - \mathcal{G}(u)\|^2 + R(u)$$

- $\|\delta - \mathcal{G}(u)\|$ potential / data misfit
- R regularization term

Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Deterministic optimization problem

$$\min_u \frac{1}{2}\|\delta - \mathcal{G}(u)\|^2 + R(u)$$

- **Large-scale, deterministic optimization problem**
- **No quantification of the uncertainty in the unknown $\mathbf{u}$**
- **Proper choice of the regularization term $\mathbf{R}$**

Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Bayesian inverse problem

$$\delta = \mathcal{G}(u) + \eta$$

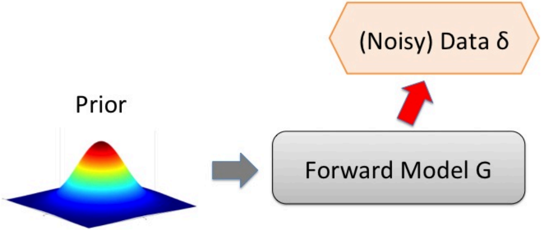
- u, η, δ random variables / fields
- Goal of computation: moments of system quantities under the posterior w.r. to noisy data δ

Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Bayesian inverse problem

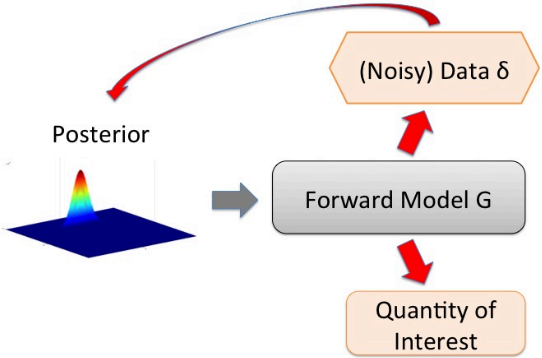


Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Bayesian inverse problem



Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Bayesian inverse problem

$$\delta = \mathcal{G}(u) + \eta$$

- **Quantification of uncertainty in $\mathbf{u}$ and system quantities**
- **Well-posedness of the inverse problem**
- **Incorporation of prior knowledge on the uncertain data $\mathbf{u}$**
- **Need of efficient approximations of the posterior**

Bayesian Inverse Problems (Stuart 2010)

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Goal: Efficient estimation of system quantities from noisy observations

- Infinite-dimensional parameter space
- Fast convergence by exploiting sparsity of the underlying forward problem
- Suitable for application to a broad class of forward problems

Bayesian Inverse Problems (Stuart 2010)

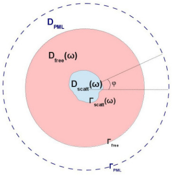
Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Goal: Efficient estimation of system quantities from noisy observations

- Infinite-dimensional parameter space
- Fast convergence by exploiting sparsity of the underlying forward problem
- Suitable for application to a broad class of forward problems

UQ in Nano Optics



Bayesian Inverse Problems (Stuart 2010)

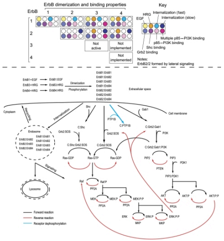
Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

Goal: Efficient estimation of system quantities from noisy observations

UQ in Biochemical Networks

- Infinite-dimensional parameter space
- Fast convergence by exploiting sparsity of the underlying forward problem
- Suitable for application to a broad class of forward problems



Source: Chen et al.

Bayesian Inverse Problems (Stuart 2010)

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

- X separable Banach space
- $G : X \mapsto \mathcal{X}$ the forward map

Abstract Operator Equation

$$\text{Given } u \in X, \text{ find } q \in \mathcal{X} : \quad A(u;q) = F(u) \quad \text{in} \quad \mathcal{Y}'$$

with $A(u;\cdot) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$, $F : X \mapsto \mathcal{Y}'$, $\mathcal{X}$, $\mathcal{Y}$ reflexive Banach spaces,
$\mathfrak{a}(v,w) :=_{\mathcal{Y}} \langle w, Av \rangle_{\mathcal{Y}'}$ $\forall v \in \mathcal{X}, w \in \mathcal{Y}$ corresponding bilinear form

- $\mathcal{O} : \mathcal{X} \mapsto \mathbb{R}^K$ bounded, linear observation operator
- $\mathcal{G} : X \mapsto \mathbb{R}^K$ uncertainty-to-observation map, $\mathcal{G} = \mathcal{O} \circ G$
- $\eta \in \mathbb{R}^K$ the observational noise ($\eta \sim \mathcal{N}(0, \Gamma)$)

Bayesian Inverse Problems (Stuart 2010)

Find the unknown data $u \in X$ from noisy observations

$$\delta = \mathcal{G}(u) + \eta,$$

- X separable Banach space
- $G : X \mapsto \mathcal{X}$ the forward map
- $\mathcal{O} : \mathcal{X} \mapsto \mathbb{R}^K$ bounded, linear observation operator
- $\mathcal{G} : X \mapsto \mathbb{R}^K$ uncertainty-to-observation map, $\mathcal{G} = \mathcal{O} \circ G$
- $\eta \in \mathbb{R}^K$ the observational noise ($\eta \sim \mathcal{N}(0, \Gamma)$)

Least squares potential $\Phi : X \times \mathbb{R}^K \to \mathbb{R}$

$$\Phi(u; \delta) := \frac{1}{2}\Big((\delta - \mathcal{G}(u))^\top \Gamma^{-1} (\delta - \mathcal{G}(u))\Big)$$

Reformulation of the forward problem with unknown stochastic input data as an infinite dimensional, parametric deterministic problem

Bayesian Inverse Problems (Stuart 2010)

Parametric representation of the unknown u

$$u = u(\boldsymbol{y}) := \langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j \in X$$

- $\boldsymbol{y} = (y_j)_{j \in \mathbb{J}}$ iid sequence of real-valued random variables $y_j \sim \mathcal{U}[-1, 1]$
- $\langle u \rangle, \psi_j \in X$
- $\mathbb{J}$ finite or countably infinite index set

Prior measure on the uncertain input data

$$\mu_0(d\boldsymbol{y}) := \bigotimes_{j \in \mathbb{J}} \frac{1}{2} \lambda_1(dy_j) \ .$$

- $(U, \mathcal{B}) = \left([-1, 1]^{\mathbb{J}}, \ \bigotimes_{j \in \mathbb{J}} \mathcal{B}^1[-1, 1] \right)$ measurable space

Bayesian Inverse Problem

Theorem (ChS and Stuart 2011)

Assume that $\mathcal{G}(u)\Big|_{u=\langle u\rangle+\sum_{j\in\mathbb{J}} y_j\psi_j}$ is bounded and continuous.

Then $\mu^\delta(d\boldsymbol{y})$, the distribution of $\boldsymbol{y} \in U$ given δ, is absolutely continuous with respect to $\mu_0(d\boldsymbol{y})$, and

$$\frac{d\mu^\delta}{d\mu_0}(\boldsymbol{y}) = \frac{1}{Z}\Theta(\boldsymbol{y})$$

with the parametric Bayesian posterior Θ given by

$$\Theta(\boldsymbol{y}) = \exp\big(-\Phi(u;\delta)\big)\Big|_{u=\langle u\rangle+\sum_{j\in\mathbb{J}} y_j\psi_j},$$

and the normalization constant

$$Z = \int_U \Theta(\boldsymbol{y})\mu_0(d\boldsymbol{y}) \ .$$

Bayesian Inverse Problem

Expectation of a *Quantity of Interest* $\phi : X \to S$

$$\mathbb{E}^{\mu^\delta}[\phi(u)] = Z^{-1} \int_U \exp\big(-\Phi(u;\delta)\big)\phi(u)\Big|_{u=\langle u\rangle + \sum_{j\in\mathbb{J}} y_j\psi_j} \mu_0(d\boldsymbol{y}) =: Z'/Z$$

with $Z = \mathbb{E}^{\mu^\delta}[1] = \int_U \exp(-\frac{1}{2}\left((\delta - \mathcal{G}(u))^\top \Gamma^{-1}(\delta - \mathcal{G}(u))\right))\mu_0(d\boldsymbol{y})$.

- Reformulation of the forward problem with unknown stochastic input data as an infinite dimensional, parametric deterministic problem
- Parametric, deterministic representation of the derivative of the posterior measure with respect to the prior μ_0
- Approximation of Z' and Z to compute the expectation of QoI under the posterior given data δ

Efficient algorithm to approximate the conditional expectations given the data with dimension-independent rates of convergence

Bayesian Inverse Problem

Expectation of a *Quantity of Interest* $\phi : X \to S$

$$\mathbb{E}^{\mu^\delta}[\phi(u)] = Z^{-1} \int_U \exp\big(-\Phi(u;\delta)\big)\phi(u)\Big|_{u=\langle u\rangle + \sum_{j\in\mathbb{J}} y_j\psi_j} \mu_0(d\boldsymbol{y}) =: Z'/Z$$

with $Z = \mathbb{E}^{\mu^\delta}[1] = \int_U \exp(-\frac{1}{2}\left((\delta - \mathcal{G}(u))^\top \Gamma^{-1}(\delta - \mathcal{G}(u))\right))\mu_0(d\boldsymbol{y})$.

Exploiting sparsity in the parametric operator equation

- Parameters belonging to a specified sparsity class
- Analytic regularity of the parametric, deterministic Bayesian posterior
- Parametric, deterministic Bayesian posterior belongs to the same sparsity class

$\rightarrow$ **Sparsity of Legendre pce + dimension-independent convergence rates for Smolyak integration algorithms**

$(\boldsymbol{b}, p, \epsilon)$-Analyticity

$(\boldsymbol{b}, p, \epsilon) : 1$ (well-posedness)

For each $\boldsymbol{y} \in U$, there exists a unique realization $u(\boldsymbol{y}) \in X$ and a unique solution $q(\boldsymbol{y}) \in \mathcal{X}$ of the forward problem. This solution satisfies the a-priori estimate

$$\forall \boldsymbol{y} \in U : \quad \|q(\boldsymbol{y})\|_{\mathcal{X}} \leq C_0 \,.$$

$(\boldsymbol{b}, p, \epsilon) : 2$ (analyticity)

There exist $0 < p < 1$ and $\boldsymbol{b} = (b_j)_{j \in \mathbb{J}} \in \ell^p(\mathbb{J})$ such that for $\epsilon > 0$, there exist $C_\epsilon > 0$ and $\rho = (\rho_j)_{j \in \mathbb{J}}$ of poly-radii $\rho_j > 1$ such that

$$\sum_{j \in \mathbb{J}} (\rho_j - 1) b_j \leq \epsilon \,,$$

and $U \ni \boldsymbol{y} \mapsto q(\boldsymbol{y}) \in \mathcal{X}$ admits an analytic continuation to the open polyellipse $\mathcal{E}_\rho := \prod_{j \in \mathbb{J}} \mathcal{E}_{\rho_j} \subset \mathbb{C}^{\mathbb{J}}$ with

$$\forall \boldsymbol{z} \in \mathcal{E}_\rho : \quad \|q(\boldsymbol{z})\|_{\mathcal{X}} \leq C_\epsilon \,.$$

$(\boldsymbol{b}, p, \epsilon)$-Analyticity of Parametric Operator Families

$$u \in X : A(u; q) = F(u) \quad q \in \mathcal{X}$$

Assumption A1

For $\epsilon > 0$ and some $0 < p < 1$, there exists a positive sequence $\boldsymbol{b} = (b_j)_{j\geq 1} \in \ell^p(\mathbb{N})$, such that for any sequence $\rho := (\rho_j)_{j\geq 1}$ $\rho := (\rho_j)_{j\geq 1}$ with $\rho_j > 1,\ \sum_{j\in\mathbb{J}}(\rho_j - 1)b_j \leq \epsilon$, $\mathfrak{a}$ and F are holomorphic in $\mathcal{E}_\rho$.

Assumption A2

The holomorphic extensions satisfy the uniform continuity conditions

$$\sup_{w\in\mathcal{Y}\setminus\{0\}} \frac{|f(z;w)|}{\|w\|_{\mathcal{Y}}} \leq M, \qquad \sup_{v\in\mathcal{X}\setminus\{0\}, w\in\mathcal{Y}\setminus\{0\}} \frac{|\mathfrak{a}(z;v,w)|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \leq R,$$

with $M < \infty, f$ corresponding linear form of F.

Assumption A3

There hold the uniform inf-sup conditions:

$$\inf_{v\in\mathcal{X}\setminus\{0\}} \sup_{w\in\mathcal{Y}\setminus\{0\}} \frac{|\mathfrak{a}(z;v,w)|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \geq r \quad \text{und} \quad \inf_{w\in\mathcal{Y}\setminus\{0\}} \sup_{v\in\mathcal{X}\setminus\{0\}} \frac{|\mathfrak{a}(z;v,w)|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \geq r$$

with $0 < r \leq R < \infty$.

$(\boldsymbol{b}, p, \epsilon)$-Analyticity of Parametric Operator Families

$$u \in X : A(u;q) = F(u) \quad q \in \mathcal{X}$$

Assumption A1 $\mathfrak{a}$ and F holomorphic in $\mathcal{E}_\rho$

Assumption A2

$$\sup_{w \in \mathcal{Y}\setminus\{0\}} \frac{|f(z;w)|}{\|w\|_{\mathcal{Y}}} \leq M, \qquad \sup_{v \in \mathcal{X}\setminus\{0\}, w \in \mathcal{Y}\setminus\{0\}} \frac{|\mathfrak{a}(z;v,w)|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \leq R$$

Assumption A3

$$\inf_{v \in \mathcal{X}\setminus\{0\}} \sup_{w \in \mathcal{Y}\setminus\{0\}} \frac{|\mathfrak{a}(z;v,w)|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \geq r \quad \text{und} \quad \inf_{w \in \mathcal{Y}\setminus\{0\}} \sup_{v \in \mathcal{X}\setminus\{0\}} \frac{|\mathfrak{a}(z;v,w)|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \geq r$$

Theorem (Cohen, Chkifa, ChS 2013)

Unter Assumptions **A1 - A3**, $\mathcal{A}(u;q) = A(u;q) - F(u)$ satisfies the $(\boldsymbol{b}, p, \epsilon)$-holomorphy assumptions.

Sparsity of the Forward Solution

Theorem (Chkifa, Cohen, DeVore and ChS)

Assume that the parametric forward solution map $q(\boldsymbol{y})$ admits a $(\boldsymbol{b}, p, \epsilon)$-analytic extension to the poly-ellipse $\mathcal{E}_\rho \subset \mathbb{C}^{\mathbb{J}}$.

- The Legendre series converges unconditionally,

$$q(\boldsymbol{y}) = \sum_{\nu \in \mathcal{F}} q_\nu^P P_\nu(\boldsymbol{y}) \quad \text{in } L^\infty(U, \mu_0; \mathcal{X})$$

  with Legendre polynomials $P_k(1) = 1,\ \|P_k\|_{L^\infty(-1,1)} = 1\ , \quad k = 0, 1, ...$

- There exists a p-summable, monotone envelope $\boldsymbol{q} = \{\boldsymbol{q}_\nu\}_{\nu \in \mathcal{F}}$,
  i.e. $\boldsymbol{q}_\nu := \sup_{\mu \geq \nu} \|q_\nu^P\|_{\mathcal{X}}$ with $C(p, \boldsymbol{q}) := \|\boldsymbol{q}\|_{\ell^p(\mathcal{F})} < \infty$.
  and monotone $\Lambda_N^P \subset \mathcal{F}$ corresponding to the N largest terms of $\boldsymbol{q}$ with

$$\sup_{\boldsymbol{y} \in U} \left\| q(\boldsymbol{y}) - \sum_{\nu \in \Lambda_N^P} q_\nu^P P_\nu(\boldsymbol{y}) \right\|_{\mathcal{X}} \leq C(p, \boldsymbol{q}) N^{-(1/p-1)} \ .$$

Sparsity of the Posterior

Theorem (ClS and ChS 2013)

Assume that the forward solution map $U \ni \boldsymbol{y} \mapsto q(\boldsymbol{y})$ is $(\boldsymbol{b}, p, \epsilon)$–analytic for some $0 < p < 1$.
Then the Bayesian posterior $\Theta(\boldsymbol{y})$ is, as a function of the parameter $\boldsymbol{y}$, likewise $(\boldsymbol{b}, p, \epsilon)$–analytic, with the same p and the same ϵ.

Sketch of proof

- Establish holomorphy of the complex extension Θ on the poly-ellipse $\mathcal{E}_\rho \subset \mathbb{C}^{\mathbb{J}}$
- Derive bounds on the modulus of the posterior

$$\sup_{\boldsymbol{z} \in \mathcal{E}_\rho} |\Theta(\boldsymbol{z})| \leq \exp\left(\sup_{\boldsymbol{z} \in \mathcal{E}_\rho} \frac{1}{2} \mathrm{Im}\left(\mathcal{G}(u(\boldsymbol{z}))\right)^\top \Gamma^{-1} \mathrm{Im}\left(\mathcal{G}(u(\boldsymbol{z}))\right) \right)$$

Sparsity of the Posterior

Theorem (ClS and ChS 2013)

Assume that the forward solution map $U \ni \boldsymbol{y} \mapsto q(\boldsymbol{y})$ is $(\boldsymbol{b}, p, \epsilon)$–analytic for some $0 < p < 1$.
Then the Bayesian posterior $\Theta(\boldsymbol{y})$ is, as a function of the parameter $\boldsymbol{y}$, likewise $(\boldsymbol{b}, p, \epsilon)$–analytic, with the same p and the same ϵ.

N-term Approximation Results

$$\sup_{\boldsymbol{y} \in U} \left\| \Theta(\boldsymbol{y}) - \sum_{\nu \in \Lambda_N^P} \Theta_\nu^P P_\nu(\boldsymbol{y}) \right\|_{\mathcal{X}} \leq N^{-s} \|\boldsymbol{\theta}^P\|_{\ell_m^p(\mathcal{F})}, \quad s := \frac{1}{p} - 1 \; .$$

Adaptive Smolyak quadrature algorithm with convergence rates depending only on the summability of the parametric operator

Sparsity of the Posterior

Theorem (ClS and ChS 2013)

Assume that the forward solution map $U \ni \boldsymbol{y} \mapsto q(\boldsymbol{y})$ is $(\boldsymbol{b}, p, \epsilon)$–analytic for some $0 < p < 1$.
Then the Bayesian posterior $\Theta(\boldsymbol{y})$ is, as a function of the parameter $\boldsymbol{y}$, likewise $(\boldsymbol{b}, p, \epsilon)$–analytic, with the same p and the same ϵ.

Examples

- Parametric initial value ODEs (Hansen & ChS; Vietnam J. Math. 2013)
- Affine-parametric, linear operator equations (ClS & ChS; 2013)
- Semilinear elliptic PDEs (Hansen & ChS; Math. Nachr. 2013)
- Elliptic multiscale problems (Hoang & ChS; Analysis and Applications 2012)
- Nonaffine holomorphic-parametric, nonlinear problems (Cohen, Chkifa & ChS; 2013)

Univariate Quadrature

Univariate quadrature operators of the form

$$Q^k(\mathtt{g}) = \sum_{i=0}^{n_k} w_i^k \cdot \mathtt{g}(z_i^k)$$

with $\mathtt{g} : [-1,1] \mapsto \mathbb{R}$.

- $(Q^k)_{k\geq 0}$ sequence of univariate quadrature formulas
- $(z_j^k)_{j=0}^{n_k} \subset [-1,1]$ with $z_j^k \in [-1,1]\,, \forall j,k$ and $z_0^k = 0\,, \forall k$ quadrature points
- $w_j^k\,, 0 \leq j \leq n_k,\ \forall k \in \mathbb{N}_0$ quadrature weights

Assumption

(i) $(I - Q^k)(g_k) = 0\,, \quad \forall g_k \in \mathbb{P}_k = \mathrm{span}\{y^j : j \in \mathbb{N}_0, j \leq k\}$
with $I(g_k) = \int_{[-1,1]} g_k(y)\lambda_1(dy)$

(ii) $w_j^k > 0\,, \qquad 0 \leq j \leq n_k,\ \forall k \in \mathbb{N}_0$.

Univariate Quadrature

Univariate quadrature operators of the form

$$Q^k(\mathrm{g}) = \sum_{i=0}^{n_k} w_i^k \cdot \mathrm{g}(z_i^k)$$

with $\mathrm{g} : [-1,1] \mapsto \mathbb{R}$.

- $(Q^k)_{k\geq 0}$ sequence of univariate quadrature formulas
- $(z_j^k)_{j=0}^{n_k} \subset [-1,1]$ with $z_j^k \in [-1,1]\,, \forall j,k$ and $z_0^k = 0\,, \forall k$ quadrature points
- $w_j^k\,, 0 \leq j \leq n_k,\ \forall k \in \mathbb{N}_0$ quadrature weights

Univariate quadrature difference operator

$$\Delta_j = Q^j - Q^{j-1}, \qquad j \geq 0$$

with $Q^{-1} = 0$ and $z_0^0 = 0, w_0^0 = 1$.

Univariate Quadrature

Univariate quadrature operators of the form

$$Q^k(\mathrm{g}) = \sum_{i=0}^{n_k} w_i^k \cdot \mathrm{g}(z_i^k)$$

with $\mathrm{g} : [-1,1] \mapsto \mathbb{R}$.

- $(Q^k)_{k\geq 0}$ sequence of univariate quadrature formulas
- $(z_j^k)_{j=0}^{n_k} \subset [-1,1]$ with $z_j^k \in [-1,1]\,, \forall j,k$ and $z_0^k = 0\,, \forall k$ quadrature points
- $w_j^k\,, 0 \leq j \leq n_k,\ \forall k \in \mathbb{N}_0$ quadrature weights

Univariate quadrature operator rewritten as telescoping sum

$$Q^k = \sum_{j=0}^{k} \Delta_j$$

with $\mathcal{Z}^k = \{z_j^k : 0 \leq j \leq n_k\} \subset [-1,1]$ set of points corresponding to Q^k.

Sparse Quadrature Operator

Tensorized multivariate operators

$$\mathcal{Q}_\nu = \bigotimes_{j\geq 1} Q^{\nu_j}\,, \qquad \Delta_\nu = \bigotimes_{j\geq 1} \Delta_{\nu_j}$$

with associated set of multivariate points $\mathcal{Z}^\nu = \times_{j\geq 1}\mathcal{Z}^{\nu_j} \in U$.

- If $\nu = 0_\mathcal{F}$, then $\Delta_\nu g = Q^\nu g = g(z_{0_\mathcal{F}}) = g(0_\mathcal{F})$
- If $0_\mathcal{F} \neq \nu \in \mathcal{F}$, with $\hat{\nu} = (\nu_j)_{j\neq i}$

$$Q^\nu g = Q^{\nu_i}(t \mapsto \bigotimes_{j\geq 1} Q^{\hat{\nu}_j} g_t)\,, \qquad i \in \mathbb{I}_\nu$$

and

$$\Delta_\nu g = \Delta_{\nu_i}(t \mapsto \bigotimes_{j\geq 1} \Delta_{\hat{\nu}_j} g_t)\,, \qquad i \in \mathbb{I}_\nu\,,$$

for $g \in \mathcal{Z}$, g_t is the function defined on $\mathcal{Z}^\mathbb{N}$ by
$g_t(\hat{y}) = g(y), y = (\ldots, y_{i-1}, t, y_{i+1}, \ldots)\,, i > 1$ and $y = (t, y_2, \ldots)\,, i = 1$

Sparse Quadrature Operator

For any finite monotone set $\Lambda \subset \mathcal{F}$, the quadrature operator is defined by

$$\mathcal{Q}_\Lambda = \sum_{\nu \in \Lambda} \Delta_\nu = \sum_{\nu \in \Lambda} \bigotimes_{j \geq 1} \Delta_{\nu_j}$$

with associated collocation grid

$$\mathcal{Z}_\Lambda = \cup_{\nu \in \Lambda} \mathcal{Z}^\nu \, .$$

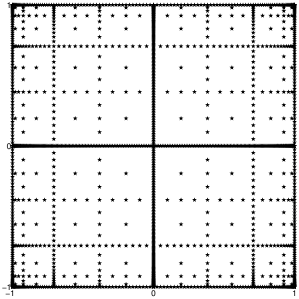
Sparse Quadrature Operator

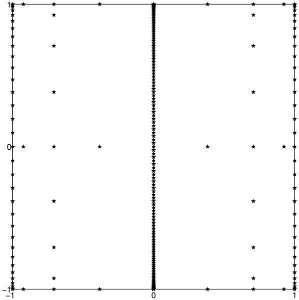
For any finite monotone set $\Lambda \subset \mathcal{F}$, the quadrature operator is defined by

$$\mathcal{Q}_\Lambda = \sum_{\nu \in \Lambda} \Delta_\nu = \sum_{\nu \in \Lambda} \bigotimes_{j \geq 1} \Delta_{\nu_j}$$

with associated collocation grid

$$\mathcal{Z}_\Lambda = \cup_{\nu \in \Lambda} \mathcal{Z}^\nu \,.$$





Convergence Rates for Adaptive Smolyak Integration

Theorem

Assume that the forward solution map $U \ni \boldsymbol{y} \mapsto q(\boldsymbol{y})$ is $(\boldsymbol{b}, p, \epsilon)$–analytic for some $0 < p < 1$.

Then there exists a sequence $(\Lambda_N)_{N\geq 1}$ of monotone index sets $\Lambda_N \subset \mathcal{F}$ such that $\#\Lambda_N \leq N$ and

$$|I[\Theta] - \mathcal{Q}_{\Lambda_N}[\Theta]| \leq C^1 N^{-s} ,$$

with $s = 1/p - 1$, $I[\Theta] = \int_U \Theta(y)\mu_0(d\boldsymbol{y})$ and,

$$\|I[\Psi] - \mathcal{Q}_{\Lambda_N}[\Psi]\|_{\mathcal{X}} \leq C^2 N^{-s} , \qquad s = \frac{1}{p} - 1 .$$

with $I[\Psi] = \int_U \Psi(y)\mu_0(d\boldsymbol{y})$, $C^1, C^2 > 0$ independent of N.

Remark: SAME index sets Λ_N for BOTH, Z' and Z.

ClS and ChS Sparsity in Bayesian Inversion of Parametric Operator Equations, 2013.

Convergence Rates for Adaptive Smolyak Integration

Sketch of proof

- Relating the quadrature error with the Legendre coefficients

$$|I(\Theta) - \mathcal{Q}_\Lambda(\Theta)| \leq 2 \cdot \sum_{\nu \notin \Lambda} \gamma_\nu |\theta_\nu^P|$$

  and

$$\|I(\Psi) - \mathcal{Q}_\Lambda(\Psi)\|_\mathcal{X} \leq 2 \cdot \sum_{\nu \notin \Lambda} \gamma_\nu \|\psi_\nu^P\|_\mathcal{X}$$

  for any monotone set $\Lambda \subset \mathcal{F}$, where $\gamma_\nu := \prod_{j \in \mathbb{J}} (1 + \nu_j)^2$.

- $(\gamma_\nu |\theta_\nu^P|)_{\nu \in \mathcal{F}} \in l_m^p(\mathcal{F})$ and $(\gamma_\nu \|\psi_\nu^P\|_\mathcal{X})_{\nu \in \mathcal{F}} \in l_m^p(\mathcal{F})$.

$\Rightarrow \exists$ sequence $(\Lambda_N)_{N \geq 1}$ of monotone sets $\Lambda_N \subset \mathcal{F}$, $\#\Lambda_N \leq N$, such that the Smolyak quadrature converges with order $1/p - 1$.

Adaptive Construction of the Monotone Index Set

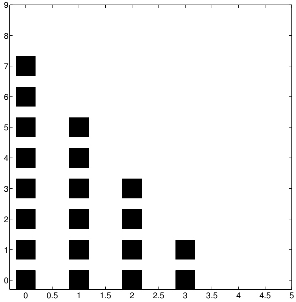
Successive identification of the N largest contributions

$$|\Delta_\nu(\Theta)| = |\bigotimes_{j\geq 1} \Delta_{\nu_j}(\Theta)|, \quad \nu \in \mathcal{F}$$

Adaptive Construction of the Monotone Index Set

Successive identification of the N largest contributions

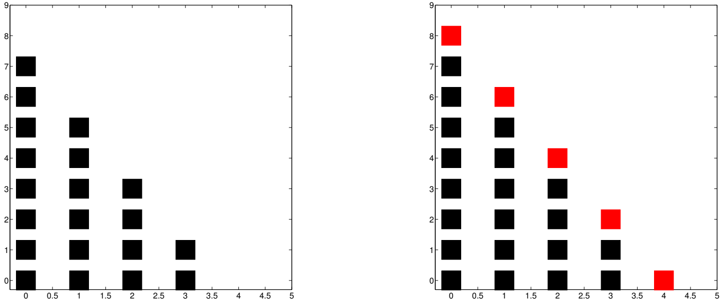
$$|\Delta_\nu(\Theta)| = |\bigotimes_{j\geq 1} \Delta_{\nu_j}(\Theta)|, \quad \nu \in \mathcal{F}$$



Adaptive Construction of the Monotone Index Set

Successive identification of the N largest contributions

$$|\Delta_\nu(\Theta)| = |\bigotimes_{j\geq 1} \Delta_{\nu_j}(\Theta)|\,, \quad \nu \in \mathcal{F}$$



Adaptive Construction of the Monotone Index Set

Successive identification of the N largest contributions

$$|\Delta_\nu(\Theta)| = |\bigotimes_{j\geq 1} \Delta_{\nu_j}(\Theta)|, \quad \nu \in \mathcal{F}$$

$\rightarrow$ A. Chkifa, A. Cohen and ChS. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs, 2012.

Reduced set of neighbors

$$\mathcal{N}(\Lambda) := \{\nu \notin \Lambda : \nu - e_j \in \Lambda, \forall j \in \mathbb{I}_\nu \text{ and } \nu_j = 0\,, \forall j > j(\Lambda) + 1\}$$

with $j(\Lambda) = \max\{j : \nu_j > 0$ for some $\nu \in \Lambda\}$, $\mathbb{I}_\nu = \{j \in \mathbb{N} : \nu_j \neq 0\} \subset \mathbb{N}$.

Adaptive Construction of the Monotone Index Set

```
function ASG
    Set $\Lambda_1 = \{0\}, k = 1$ and compute $\Delta_0(\Theta)$.
    Determine the reduced set of neighbors $\mathcal{N}(\Lambda_1)$.
    Compute $\Delta_\nu(\Theta), \forall \nu \in \mathcal{N}(\Lambda_1)$.
    while $\sum_{\nu \in \mathcal{N}(\Lambda_k)} |\Delta_\nu(\Theta)| > tol$ do
        Select $\nu \in \mathcal{N}(\Lambda_k)$ with largest $|\Delta_\nu|$ and set $\Lambda_{k+1} = \Lambda_k \cup \{\nu\}$.
        Determine the reduced set of neighbors $\mathcal{N}(\Lambda_{k+1})$.
        Compute $\Delta_\nu(\Theta), \forall \nu \in \mathcal{N}(\Lambda_{k+1})$.
        Set $k = k + 1$.
    end while
end function
```

T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature, *Computing*, 2003

Numerical Experiments

Model parametric parabolic problem

$$\begin{aligned} \partial_t q(t,x) - \mathsf{div}(u(x)\nabla q(t,x)) &= 100 \cdot tx \qquad && (t,x) \in T \times D\,, \\ q(0,x) &= 0 && x \in D\,, \\ q(t,0) = q(t,1) &= 0 && t \in T \end{aligned}$$

with

$$u(x,y) = \langle u \rangle + \sum_{j=1}^{64} y_j \psi_j\,, \text{where } \langle u \rangle = 1 \text{ and } \psi_j = \alpha_j \chi_{D_j}$$

where $D_j = [(j-1)\frac{1}{64}, j\frac{1}{64}]$, $y = (y_j)_{j=1,\ldots,64}$ and $\alpha_j = \frac{0.9}{j^\zeta}$, $\zeta = 2, 3, 4$.

- Finite element method using continuous, piecewise linear ansatz functions in space, backward Euler scheme in time
- Uniform mesh with meshwidth $h_T = h_D = 2^{-11}$
- LAPACK's DPTSV routine

Numerical Experiments

Find the unknown data u for given (noisy) data δ,

$$\delta = \mathcal{G}(u) + \eta\,,$$

Expectation of interest Z'/Z

$$\begin{aligned} Z' &= \int_U \exp\big(-\Phi(u;\delta)\big)\phi(u)\Big|_{u=\langle u\rangle+\sum_{j=1}^{64} y_j\psi_j}\mu_0(dy) \\ Z &= \int_U \exp\big(-\Phi(u;\delta)\big)\Big|_{u=\langle u\rangle+\sum_{j=1}^{64} y_j\psi_j}\mu_0(dy) \end{aligned}$$

- Observation operator $\mathcal{O}$ consists of system responses at K observation points in $T \times D$ at $t_i = \frac{i}{2^{N_{K,T}}}, i = 1, \ldots, 2^{N_{K,T}} - 1$, $x_j = \frac{j}{2^{N_{K,D}}}, k = 1, \ldots, 2^{N_{K,D}} - 1$, $o_k(\cdot,\cdot) = \delta(\cdot - t_k)\delta(\cdot - x_k)$ with $K = 3,\ N_{K,D} = 2,\ N_{K,T} = 1$
- $\mathcal{G} : X \to \mathbb{R}^K$, with $\phi(u) = G(u)$
- $\eta = (\eta_j)_{j=1,\ldots,K}$ iid with $\eta_j \sim \mathcal{N}(0,1)$

Numerical Experiments

Quadrature points

- Clenshaw-Curtis (CC)

$$
\begin{aligned}
z_j^k &= -cos\left(\frac{\pi j}{n_k - 1}\right), j = 0, \ldots, n_k - 1, \text{if } n_k > 1 \text{ and} \\
z_0^k &= 0, \text{if } n_k = 1
\end{aligned}
$$

with $n_0 = 1$ and $n_k = 2^k + 1$, for $k \geq 1$

- $\mathfrak{R}$-Leja sequence (RL)

Numerical Experiments

Quadrature points

- Clenshaw-Curtis (CC)
- $\mathfrak{R}$-Leja sequence (RL)
  projection on $[-1,1]$ of a Leja sequence for the complex unit disk initiated at i

$$
\begin{aligned}
z_0^k &= 0\,, z_1^k = 1\,, z_2^k = -1\,, \text{if } j = 0,1,2 \text{ and} \\
z_j^k &= \mathfrak{R}(\hat{z}), \text{ with } \hat{z} = \operatorname*{argmax}_{|z|\le 1} \prod_{l=1}^{j-1} |z - z_l^k|\,, j = 3,\ldots,n_k, \text{if } j \text{ odd}\,, \\
z_j^k &= -z_{j-1}^k\,, j = 3,\ldots,n_k, \text{if } j \text{ even}\,,
\end{aligned}
$$

with $n_k = 2\cdot k + 1,$ for $k \ge 0$

J.-P. Calvi and M. Phung Van. On the Lebesgue constant of Leja sequences for the unit disk and its applications to multivariate interpolation *Journal of Approximation Theory*, 2011.

J.-P. Calvi and M. Phung Van. Lagrange interpolation at real projections of Leja sequences for the unit disk *Proceedings of the American Mathematical Society*, 2012.

A. Chkifa. On the Lebesgue constant of Leja sequences for the unit disk *Journal of Approximation Theory*, 2013.

Normalization Constant Z

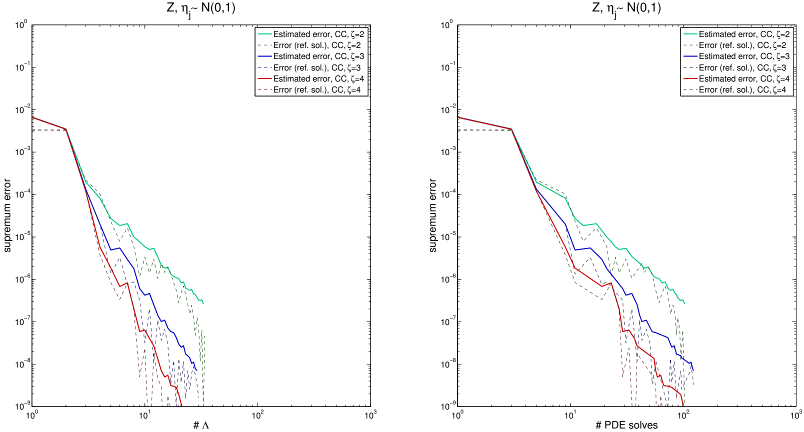


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the normalization constant Z w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence CC with $K = 3$, $\eta \sim \mathcal{N}(0,1)$ and with $\zeta = 2,3,4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Normalization Constant Z

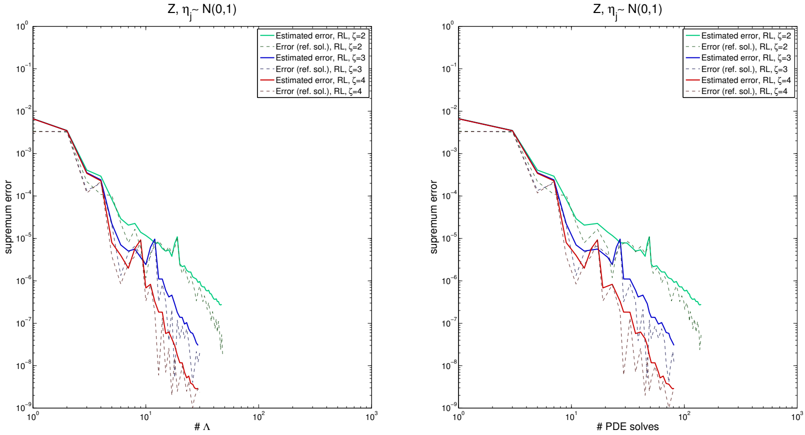


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the normalization constant Z w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence RL with $K = 3$, $\eta \sim \mathcal{N}(0,1)$ and with $\zeta = 2, 3, 4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Quantity Z'

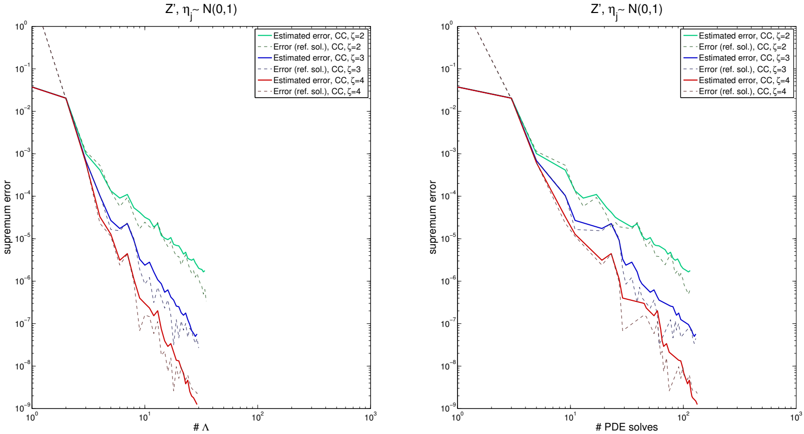


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the quantity Z' w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence CC with $K = 3$, $\eta \sim \mathcal{N}(0,1)$ and with $\zeta = 2, 3, 4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Quantity Z'

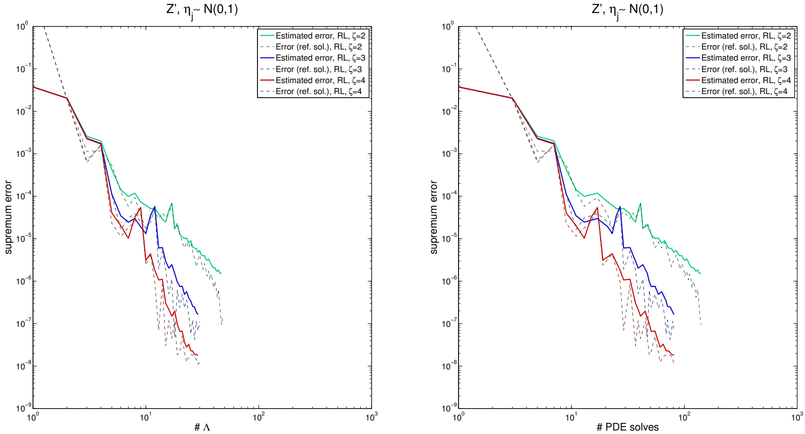


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the quantity Z' w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence RL with $K = 3$, $\eta \sim \mathcal{N}(0,1)$ and with $\zeta = 2, 3, 4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Numerical Experiments

Model parametric parabolic problem

$$\begin{aligned} \partial_t q(t,x) - \mathrm{div}(u(x)\nabla q(t,x)) &= 100 \cdot tx \qquad && (t,x) \in T \times D\,, \\ q(0,x) &= 0 && x \in D\,, \\ q(t,0) = q(t,1) &= 0 && t \in T \end{aligned}$$

with

$$u(x,y) = \langle u \rangle + \sum_{j=1}^{128} y_j \psi_j\,, \text{where } \langle u \rangle = 1 \text{ and } \psi_j = \alpha_j \chi_{D_j}$$

where $D_j = [(j-1)\frac{1}{128}, j\frac{1}{128}]$, $y = (y_j)_{j=1,\dots,128}$ and $\alpha_j = \frac{0.6}{j^\zeta}\,, \zeta = 2, 3, 4$.

- Finite element method using continuous, piecewise linear ansatz functions in space, backward Euler scheme in time
- Uniform mesh with meshwidth $h_T = h_D = 2^{-11}$
- LAPACK's DPTSV routine

Normalization Constant Z (128 parameters)

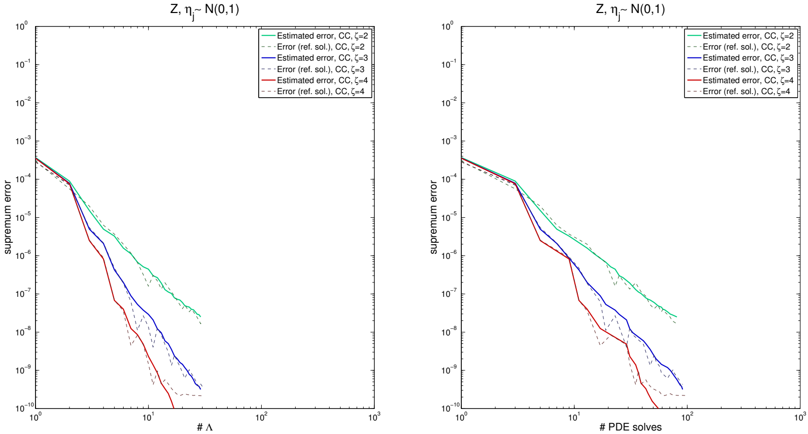


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the normalization constant Z w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence CC with $K = 3$, $\eta \sim \mathcal{N}(0, 1)$ and with $\zeta = 2, 3, 4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Normalization Constant Z (128 parameters)

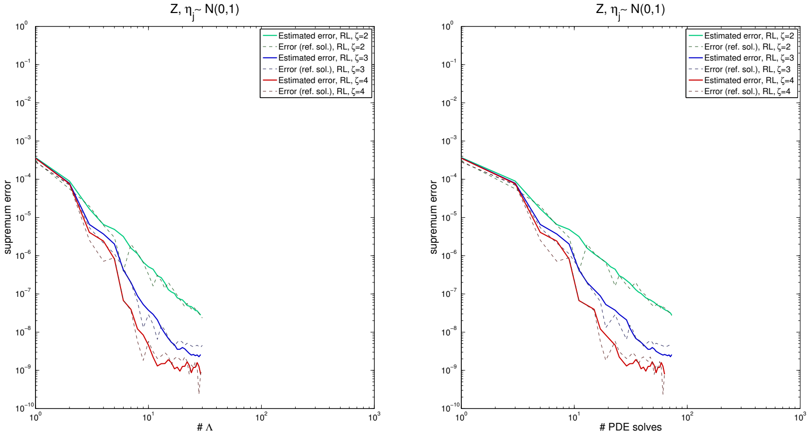


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the normalization constant Z w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence RL with $K = 3$, $\eta \sim \mathcal{N}(0, 1)$ and with $\zeta = 2, 3, 4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Quantity Z' (128 parameters)

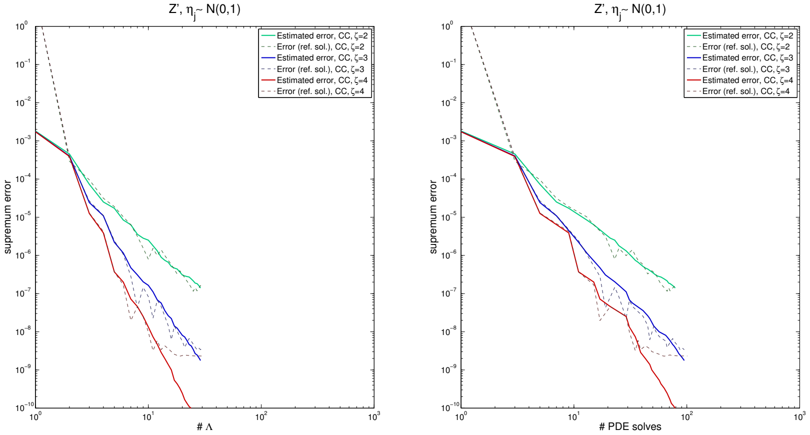


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the quantity Z' w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence CC with $K = 3$, $\eta \sim \mathcal{N}(0,1)$ and with $\zeta = 2, 3, 4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Quantity Z' (128 parameters)

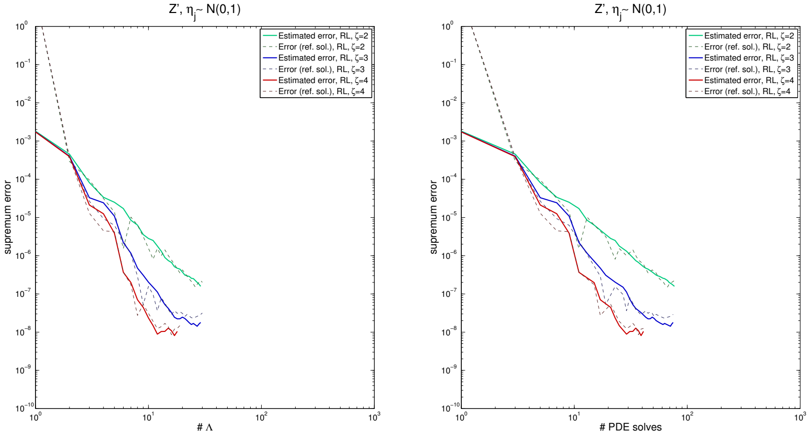


Figure: Comparison of the estimated error and actual error. Plots of error w.r. to reference solutions of the quantity Z' w.r. to the cardinality of the index set Λ_N (l.) and w.r. to the PDE solves needed (r.) based on the sequence RL with $K = 3$, $\eta \sim \mathcal{N}(0,1)$ and with $\zeta = 2, 3, 4$, $h_T = h_D = 2^{-11}$ for the reference and the adaptively computed solution.

Numerical Experiments

Model parametric elliptic problem

$$-\mathrm{div}(u\nabla q) = f \quad \text{in } D := [0,1]\,,\ q = 0 \quad \text{in } \partial D\,,$$

with $f(x) = 100 \cdot x$ and

$$\ln(u(x,y)) = \sum_{j=1}^{32} \frac{0.1}{(2j)^{\zeta}} \cos(2j\pi x) y_{2j} + \frac{0.1}{(2j-1)^{\zeta}} \sin((2j-1)\pi x) y_{2j-1}\,,$$

where $y = (y_j)_{j=1,\dots,64}$ are independently normally distributed, ie. $y_j \sim \mathcal{N}(0,1)$, and $\zeta = 2, 3, 4$.

- Finite element method using continuous, piecewise linear ansatz functions in space, backward Euler scheme in time
- Uniform mesh with meshwidth $h_D = 2^{-11}$
- LAPACK's DPTSV routine

Numerical Experiments

Find the unknown data u for given (noisy) data δ,

$$\delta = \mathcal{G}(u) + \eta\,,$$

Expectation of interest Z'/Z

$$\begin{aligned} Z' &= \int_U \exp\big(-\Phi(u;\delta)\big)\phi(u)\Big|_{u=\langle u\rangle+\sum_{j=1}^{64} y_j\psi_j}\mu_0(dy) \\ Z &= \int_U \exp\big(-\Phi(u;\delta)\big)\Big|_{u=\langle u\rangle+\sum_{j=1}^{64} y_j\psi_j}\mu_0(dy) \end{aligned}$$

- Observation operator $\mathcal{O}$ consists of system responses at K observation points in D at $x_j = \frac{j}{2^{N_{K,D}}}, k = 1, \ldots, 2^{N_{K,D}} - 1$, $o_k(\cdot) = \delta(\cdot - x_k)$ with $K = 3,\ N_{K,D} = 2$
- $\mathcal{G} : X \to \mathbb{R}^K$, with $\phi(u) = G(u)$
- $\eta = (\eta_j)_{j=1,\ldots,K}$ iid with $\eta_j \sim \mathcal{N}(0,1)$

Numerical Experiments

Quadrature points

- Gauss-Hermite (GH)

$$\begin{aligned} W(x) &= e^{-x^2}, -\infty < x < \infty \\ H_{j+1} &= 2xH_j - 2jH_{j-1}, \quad H_{-1} = 0, H_0 = 1 \end{aligned}$$

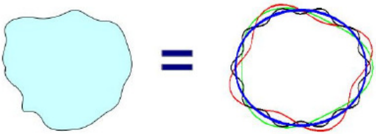
with $n_0 = 1$ and $n_k = 2^k + 1$, for $k \geq 1$

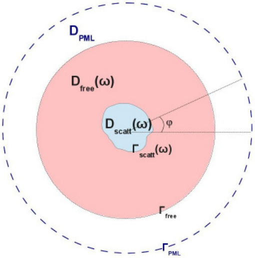
Uncertainty Quantification in Nano Optics

Goal: Quantification of the influence of defects in fabrication process on the optical response of nano structures

- Propagation of plane wave and its interaction with scatterer described by Helmholtz equation (2D).
- Stochastic shape of the scatterer

$$0 < \rho_{\min} \leq \rho(\omega, \phi) \leq \rho_{\max}, \quad \omega \in \Omega,\ \phi \in [0, 2\pi)$$



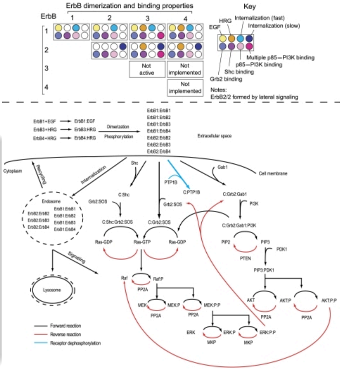


Collaboration with Ralf Hiptmair, Laura Scarabosio

SAM ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

High Dimensional Initial Value Problem

- Mass action models with uncertain reaction rates
- Biochemical reaction pathways with uncertain reaction rate constants
- Chemical reaction cascades with uncertain reaction rate constants

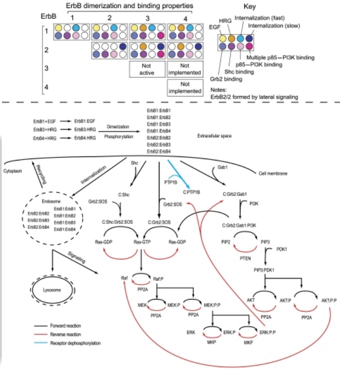


Source: Chen et al., Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data

Goal of computation:
Approximation of system characteristics on the entire, possibly infinite dimensional parameter space

High Dimensional Initial Value Problem

- Mass action models with uncertain reaction rates
- Biochemical reaction pathways with uncertain reaction rate constants
- Chemical reaction cascades with uncertain reaction rate constants



Source: Chen et al.

Collaboration with the research group of J. Stelling

D-BSSE
Department of Biosystems Science and Engineering

High Dimensional Initial Value Problem

Given $x_0(\boldsymbol{y}) \in \mathcal{S}$, $T = [0,1]$, $U = [-1,1]^{\mathbb{N}}$, find $X(t, x_0; \boldsymbol{y}) : T \times \mathcal{S} \times U \to \mathcal{S}$ such that

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= f(t, X; \boldsymbol{y}) \\ &= f_0(t, X) + \sum_{j \geq 1} y_j f_j(t, X) \end{aligned}$$

with $X(0; \boldsymbol{y}) = x_0\,, \;\; 0 \leq t \leq 1\,, \;\; \forall \boldsymbol{y} = (y_j)_{j \geq 1} \in U$

- $\mathcal{S}$ state space (separable and reflexive Banach space)

Affine parameter dependence of the right hand side

Mass action models in computational biology

Stoichiometry with uncertain reaction rate constants

Conclusions and Outlook

- New class of sparse, adaptive quadrature methods for Bayesian inverse problems for a broad class of operator equations
- Dimension-independent convergence rates depending only on the summability of the parametric operator
- Numerical confirmation of the predicted convergence rates

Conclusions and Outlook

- New class of sparse, adaptive quadrature methods for Bayesian inverse problems for a broad class of operator equations
- Dimension-independent convergence rates depending only on the summability of the parametric operator
- Numerical confirmation of the predicted convergence rates

- Gaussian priors and lognormal coefficients
- Adaptive control of the discretization error of the forward problem with respect to the expected significance of its contribution to the Bayesian estimate
- Efficient treatment of large sets of data δ and small observation noise variance Γ

References

- Chkifa A 2013 On the Lebesgue constant of Leja sequences for the unit disk *Journal of Approximation Theory* **166** 176-200
- Chkifa A, Cohen A and Schwab C 2013 High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs *Foundations of Computational Mathematics* 1-33
- Cohen A, DeVore R and Schwab C 2011 Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs *Analysis and Applications* **9** 1-37
- Gerstner T and Griebel M 2003 Dimension-adaptive tensor-product quadrature *Computing* **71** 65-87
- Hansen M and Schwab C 2013 Sparse adaptive approximation of high dimensional parametric initial value problems *Vietnam J. Math.* **1** 1-35
- Hansen M, Schillings C and Schwab C 2012 Sparse approximation algorithms for high dimensional parametric initial value problems *Proceedings 5th Conf. High Performance Computing, Hanoi*
- Schillings C and Schwab C 2013 Sparse, adaptive Smolyak algorithms for Bayesian inverse problems *Inverse Problems* **29** 065011
- Schillings C and Schwab C 2013 Sparsity in Bayesian inversion of parametric operator equations *(accepted for publication in Inverse Problems)*
- Schillings C 2013 A Note on Sparse, Adaptive Smolyak Quadratures for Bayesian Inverse Problems *Oberwolfach Report* **10/1** 239-293
- Schwab C and Schillings C Sparse Quadrature Approach to Bayesian Inverse Problems *SAM Report 2013-27 (in review)*
- Schwab C and Stuart A M 2012 Sparse deterministic approximation of Bayesian inverse problems *Inverse Problems* **28** 045003