Parametric Uncertainty Computations with Tensor Product Representations

Hermann G. Matthies

A. Litvinenko, M. Meyer, O. Pajonk, B. Rosić, E. Zander

Institute of Scientific Computing, TU Braunschweig
Brunswick, Germany

wire@tu-bs.de

http://www.wire.tu-bs.de

$Id: 10_Paris.tex,v 7.4 2011/08/02 03:21:18 hgm Exp $

Overview

1. Parameter dependent problems

2. Decompositions and factorisations

3. Formulation in tensor product spaces

4. Examples

5. Model reduction and sparse representation

6. Bayesian updating, inverse problems

7. Examples and Conclusion

Mathematical formulation I

Consider operator equation, physical system modelled by A, depending on quantity q:

$$A(q;u) = f \qquad u \in \mathcal{V},\ f \in \mathcal{F},$$

$$\Leftrightarrow \quad \forall v \in \mathcal{V}: \quad a(q;u;v) = \langle A(q;u), v\rangle = \langle f, v\rangle,$$

$\mathcal{V}$ — space of states, $\mathcal{F} = \mathcal{V}^*$ — dual space of actions / forcings.

Variant: $A(\varsigma(q);u) = f$, dependence on a function $\varsigma(q)$, such that parameter $p \in \mathcal{P}$ may be

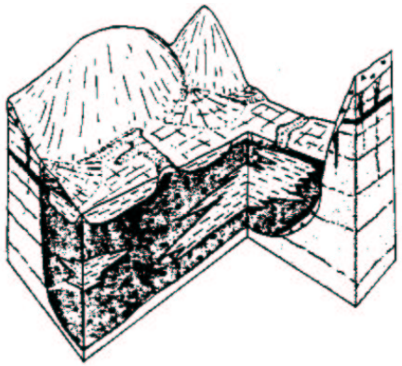
$$p = q \quad | \quad p = (q, f) \quad | \quad p = (q, f, u_0) \quad | \quad p = (\varsigma(q), \dots) \dots$$

General formulation—non-linear operator, semi-linear form:

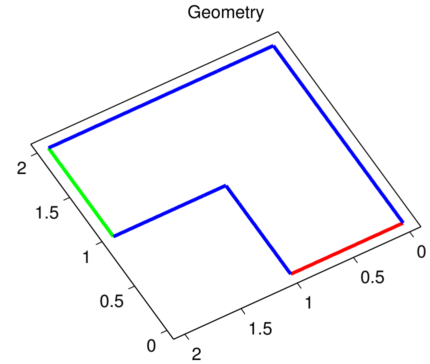
$$A(p;u) = f \quad \Leftrightarrow \quad \forall v \in \mathcal{V}: \quad a(p;u;v) = \langle A(p;u), v\rangle = \langle f, v\rangle.$$

Want to describe $A(p,\cdot)$, $f(p)$, $\varsigma(p)$, or $u(p) \longrightarrow r(p)$.
In the end desired quantities of interest (QoI) $\Psi_\iota(p, u(p))$.

Problem with parameters—diffusion SPDE



Aquifer



2D model domain $\mathcal{G}$

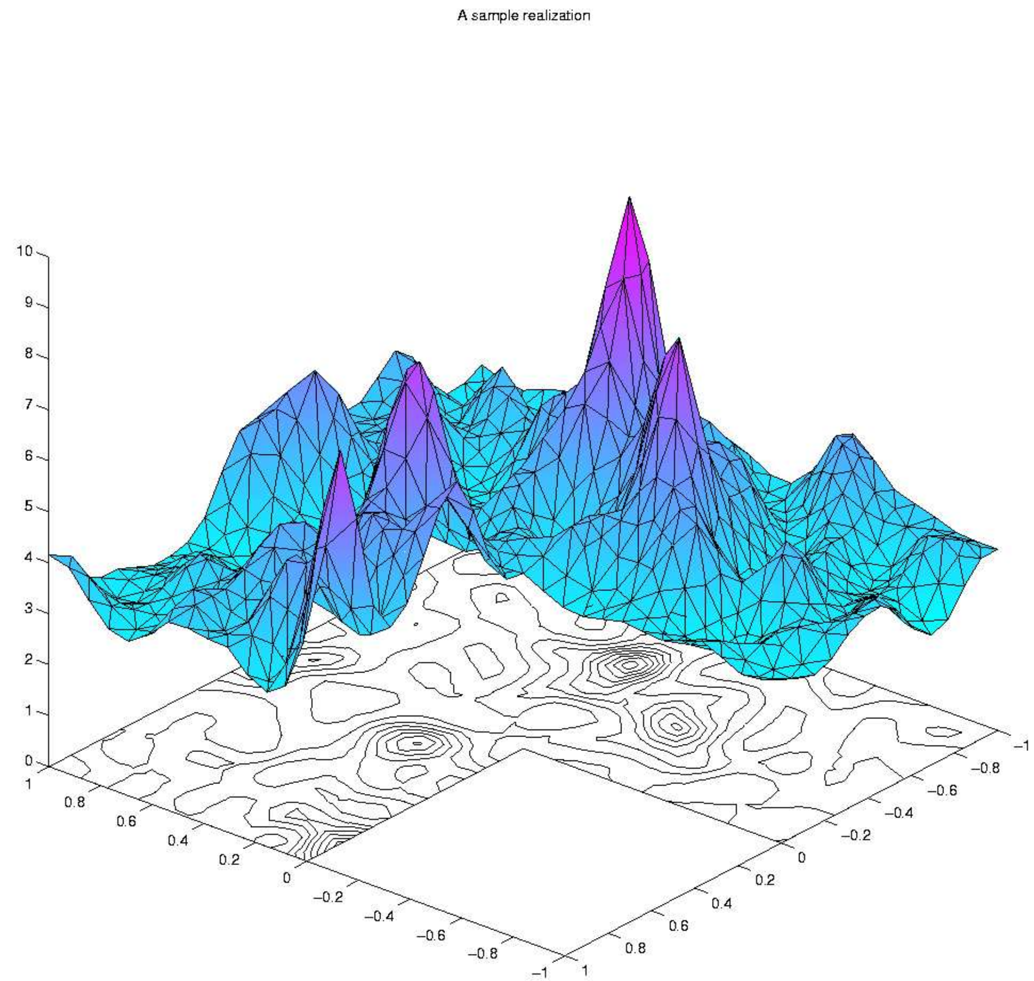
Simple stationary model of groundwater flow with parameters

$$-\nabla \cdot (\kappa(x) \cdot \nabla u(x)) = f(x) \qquad x \in \mathcal{G} \subset \mathbb{R}^d \qquad \& \text{ b.c.}$$

Parameters from modelling epistemic / aleatoric uncertainty or design.

Specific values of parameter p are realisations of κ, f, or b.c.
This involves an infinite (at first sight uncountable) real functions (random variables—RVs)

Realisation of κ



Parametric problems

For each p in a parameter set $\mathcal{P}$, let $r(p)$ be an element in a Hilbert space $\mathcal{Z}$ (for simplicity).

With $r : \mathcal{P} \to \mathcal{Z}$, denote $\mathcal{U} = \overline{\operatorname{span}}\, r(\mathcal{P}) = \overline{\operatorname{span}}\, \operatorname{im} r$.

What we are after: other representations of r or $\mathcal{U} = \overline{\operatorname{span}}\, \operatorname{im} r$.

To each function $r : \mathcal{P} \to \mathcal{U}$ corresponds a linear map $R : \mathcal{U} \to \tilde{\mathcal{R}}$:

$$R : \mathcal{U} \ni u \mapsto \langle r(\cdot) | u \rangle_{\mathcal{U}} \in \tilde{\mathcal{R}} = \operatorname{im} R \subset \mathbb{R}^{\mathcal{P}}.$$

By construction R is injective. Use this to make $\tilde{\mathcal{R}}$ a pre-Hilbert space:

$$\forall \phi, \psi \in \tilde{\mathcal{R}} : \ \langle \phi | \psi \rangle_{\mathcal{R}} := \langle R^{-1} \phi | R^{-1} \psi \rangle_{\mathcal{U}}.$$

R^{-1} is unitary on completion $\mathcal{R}$.

RKHS and classification

$\mathcal{R}$ is a reproducing kernel Hilbert space —RKHS— with symmetric kernel

$$\varkappa(p_1, p_2) = \langle r(p_1) | r(p_2) \rangle_{\mathcal{U}} \in \mathbb{R}^{\mathcal{P}\times\mathcal{P}}; \qquad \forall p \in \mathcal{P}: \ \varkappa(p, \cdot) \in \mathcal{R},$$

and $\overline{\mathrm{span}}\{\varkappa(\cdot, p) \mid p \in \mathcal{P}\} = \mathcal{R}$.

Reproducing property:

$$\forall \phi \in \mathcal{R}: \ \langle \varkappa(p, \cdot) | \phi(\cdot) \rangle_{\mathcal{R}} = \phi(p).$$

In other settings (classification, machine learning, SVM), when different subsets of $\mathcal{P}$ have to be classified, the space $\mathcal{U}$ and the map $r : \mathcal{P} \to \mathcal{U}$ is not given, but can be freely chosen.

It is then called the feature map.
The whole procedure is called the kernel trick.

Examples

The function $r(p)$ may be

- function(s) describing some system— $A(p; u) = f$.
- a random field / process as input to some system— $A(\varsigma(p); u) = f$.
- the solution / state of some system depending on the above— $u(p)$.
- the (non-linear) operator $A(p; \cdot)$ / bi-linear / semi-linear form $a(p; \cdot; \cdot)$ determining some system.

One special case is when the parameter is a random quantity.
Methods can be inspired from this model.

Representation on RKHS

In contrast to $\mathcal{P}$ (just some set), $\mathcal{R}$ is a vector space.

Assume that $\mathcal{R}$ is separable, choose complete orthonormal system (CONS) $\{y_m\}_m$ such that $\overline{\text{span}}\{y_1, y_2, \ldots\} = \mathcal{R}$.

Set $u_m = R^{-1} y_m \in \mathcal{U}$ then $r(p) = \sum_m y_m(p) u_m$ (linear in y_m).

We find that $r \in \mathcal{U} \otimes \mathcal{R}$, and
$R = \sum_m u_m \otimes y_m$ and $R^{-1} = \sum_m y_m \otimes u_m$.

But choice of CONS is arbitrary.

Let $Q_{\mathcal{R}} : \ell_2 \ni \boldsymbol{a} = (a_1, a_2, \ldots) \mapsto \sum_m a_m y_m \in \mathcal{R}$ —a unitary map.
Then $R^{-1} \circ Q_{\mathcal{R}}$ (unitary) represents $\mathcal{U}$, linear in $\boldsymbol{a}$ $\longrightarrow$ $r \in \mathcal{U} \otimes \ell_2$.
We are looking for representations on other vector spaces.

'Correlation'

If there is another inner product $\langle\cdot|\cdot\rangle_{\mathcal{Q}}$ on a subspace $\mathcal{Q} \subset \mathbb{R}^{\mathcal{P}}$, (e.g. if $(\mathcal{P}, \mu)$ is measure space, set $\mathcal{Q} := L_2(\mathcal{P}, \mu)$) a linear map $C := R^* R$ —the 'correlation' operator—is defined by

$$\forall u, v \in \mathcal{U};\ \langle Cu, v\rangle_{\mathcal{U}' \times \mathcal{U}} = \langle Ru|Rv\rangle_{\mathcal{Q}}; \qquad R^* \text{ w.r.t. } \mathcal{Q}.$$

$$\Big(\text{ In case } \mathcal{Q} = L_2(\mathcal{P}, \mu): \quad C = \int_{\mathcal{P}} r(p) \otimes r(p)\, \mu(\mathrm{d}p) \Big)$$

It is self-adjoint and positive definite $\to$ has spectrum $\sigma(C) \subseteq \mathbb{R}_+$.

Spectral decomposition with projectors

E_λ on $\lambda \in \sigma(C) = \sigma_p(C) \cup \sigma_c(C)$

$$Cu = \int_0^\infty \lambda\, \mathrm{d}E_\lambda u = \sum_{\lambda_m \in \sigma_p(C)} \lambda_m \langle v_m | u\rangle_{\mathcal{U}}\, v_m + \int_{\sigma_c(C)} \lambda\, \mathrm{d}E_\lambda u.$$

(Assume simple spectrum for simplicity ;-)

Spectral decomposition

Often C has a pure point spectrum (e.g. C or C^{-1} compact)
$\Rightarrow$ last integral vanishes, i.e. $\sigma(C) = \sigma_p(C)$:

$$Cu = \sum_m \lambda_m \langle v_m | u \rangle v_m = \sum_m \lambda_m \, (v_m \otimes v_m) \, u.$$

If $\sigma(C)_c \neq \emptyset$ need generalised eigenvectors v_λ
and Gel'fand triplets (rigged Hilbert spaces) for the continuous spectrum:

$$\int_{\sigma_c(C)} \lambda \, \mathrm{d}E_\lambda u = \int_{\sigma_c(C)} \lambda \, (v_\lambda \otimes v_\lambda) \, u \, \varrho(\mathrm{d}\lambda).$$

$$\Rightarrow \quad Cu = \sum_{\lambda_m \in \sigma_p(C)} \lambda_m \, (v_m \otimes v_m) \, u + \int_{\sigma_c(C)} \lambda \, (v_\lambda \otimes v_\lambda) \, u \, \varrho(\mathrm{d}\lambda).$$

Representation as sum / integral of rank-1 operators.

Singular value decomposition

Another spectral decomposition: C unitarily equiv. to multiplication operator M_k on $L_2(X)$

$$C = VM_kV^* = (VM_k^{1/2})(VM_k^{1/2})^*, \text{ with } M_k^{1/2} = M_{\sqrt{k}},$$

spectrum $\sigma(C)$ is (ess.) range of $k : X \to \mathbb{R}$, hence $k(x) \geq 0$ a.e. $x \in X$.

This connects to the singular value decomposition (SVD) of $R = SM_k^{1/2}V^*$, with a (here) unitary $S \longrightarrow r \in \mathcal{U} \otimes L_2(X)$.

$$\text{With } \sqrt{\lambda_m}\, s_m := Rv_m : \qquad R = \sum_m \sqrt{\lambda_m}\; (v_m \otimes s_m)\,.$$

A sum / integral of rank-1 operators.

Model reduction

For purely discrete spectrum we get $r \in \mathcal{U} \otimes \mathcal{Q}$

$$r(p) = \sum_m \sqrt{\lambda_m}\, s_m(p) v_m.$$

This is Karhunen-Loève-expansion, due to SVD $\longrightarrow$ $r \in \mathcal{U} \otimes L_2(\sigma(C))$.

A sum of rank-1 operators / tensors. Corresponds to

$$R^* = \sum_m \sqrt{\lambda_m}\, (s_m \otimes v_m)\,.$$

Observe that r is linear in the "coordinates" $\sqrt{\lambda_m}\, s_m$,
e.g. necessary for offline part in reduced basis method (RBM).

A representation of r, model reduction possible by truncation of sum,
weighted by singular values $\sqrt{\lambda_m}$.

Factorisations / re-parametrisations

R^* serves as representation. This is a factorisation of $C = R^*R$.
Some other possible ones:

$$C = R^*R = (VM_k^{1/2})(VM_k^{1/2})^* = C^{1/2}C^{1/2} = B^*B,$$

where $C = B^*B$ is an arbitrary one.

Each factorisation leads to a representation—all unitarily equivalent.
(When C is a matrix, a favourite is Cholesky: $C = LL^*$).

Assume that $C = B^*B$ and $B : \mathcal{U} \to \mathcal{H} \longrightarrow r \in \mathcal{U} \otimes \mathcal{H}$.

Analogous results / factorisations / representations follow from considering $\hat{C} := RR^* : \mathcal{Q} \to \mathcal{Q}$.
Also known as kernel decompositions, usually integral transforms.

Representations

We have seen several ways to represent the solution space
by a—hopefully—simpler space.

These can all be used for model reduction, choosing a smaller subspace.

- The RKHS-representation on $\mathcal{R}$ together with R^{-1}.
- The Karhunen-Loève expansion on $\mathcal{Q}$ via R^* (SVD).
- The spectral decomposition over $L_2(\sigma(C))$ or via $VM_k^{1/2}$ on $L_2(X)$.
- Other multiplicative decompositions, such as $C = B^*B$ on $\mathcal{H}$.
- Analogous: The kernel decompositions and representation based on kernel $\varkappa$ or $\hat{C} = RR^*$ lead to integral transforms.

Choice depends on what is wanted / needed.
Notion of measure / probability measure on $\mathcal{P}$ was not needed.

Examples and interpretations

- If $\mathcal{V}$ is a space of centred RVs, r is a random field / stochastic process indexed by $\mathcal{P}$, kernel $\varkappa(p_1, p_2)$ is covariance function.
- If in this case $\mathcal{P} = \mathbb{R}^d$ and moreover $\varkappa(p_1, p_2) = c(p_1 - p_2)$ (stationary process / homogeneous field), then diagonalisation V is real Fourier transform, typically $\sigma_p(C) = \emptyset \Rightarrow$ need Gel'fand triplets.
- If μ is a probability measure on $\mathcal{P} = \Omega$ ($\mu(\Omega) = 1$), and r is a centred $\mathcal{V}$-valued RV, then C is the covariance operator.
- If $\mathcal{P} = \{1, 2, \dots, n\}$ and $\mathcal{R} = \mathbb{R}^n$, then $\varkappa$ is the Gram matrix of the vectors $r_1, \dots, r_n$.
- If $\mathcal{P} = [0, T]$ and $r(t)$ is the response of a dynamical system, then R^* leads to proper orthogonal decomposition (POD).

Further decomposition

We have found representations $r \in \mathscr{W} := \mathcal{U} \otimes \mathcal{S}$, where

$$\mathcal{S} = \mathcal{R}, \ell_2, \mathcal{Q}, L_2(\sigma(C)), L_2(X), L_2(Z), \ldots$$

This was only a basic decomposition, as combinations may occur, so that $\mathcal{S} = \mathcal{S}_I \otimes \mathcal{S}_{II} \otimes \mathcal{S}_{III} \otimes \ldots$
Often the problem allows $\mathcal{U} = \bigotimes_k \mathcal{U}_k$, e.g. $\mathcal{U} = \mathcal{U}_x \otimes \mathcal{U}_t$.

Or the parameters allow $\mathcal{S} = \bigotimes_j \mathcal{S}_j$.
In case of random fields / stochastic processes

$$\mathcal{S} = L_2(\Omega) \cong \bigotimes_j L_2(\Omega_j) \cong L_2(\mathbb{R}^{\mathbb{N}}, \Gamma) \cong \bigotimes_{k=1}^{\infty} L_2(\mathbb{R}, \Gamma_1) \ldots$$

So $\mathscr{W} = \mathcal{U} \otimes \mathcal{S} \cong \left(\bigotimes_j \mathcal{U}_j\right) \otimes \left(\bigotimes_k \mathcal{S}_{I,k}\right) \otimes \left(\bigotimes_m \mathcal{S}_{II,m}\right) \otimes \ldots$

Example: $\mathcal{U}_t \otimes \mathcal{U}_x \otimes \mathcal{S}_1 \otimes \mathcal{S}_2 \ni v = \sum_{i,j,k,m} v_{i,j}^{k,m} \varphi_i(t) \phi_j(x) X_k(\omega_1) X_m(\omega_2)$.

Important Points I

- Aim is to replace parameter set $\mathcal{P}$ through a vector space $\mathcal{S}$, and to *represent / emulate / generate response surface / surrogate(proxy) model / (interpolate)* approximate $r(p)$.
- A function $r : \mathcal{P} \to \mathcal{U}$ generates linear map $R : \mathcal{U} \to \mathbb{R}^{\mathcal{P}}$ $\longrightarrow$ linear functional analysis / RKHS-representation.
- With Hilbert subspace $\mathcal{Q} \subset \mathbb{R}^{\mathcal{P}}$ it defines 'correlation' $C = R^*R$ or $\hat{C} = RR^*$ $\longrightarrow$ spectral decomposition / SVD / POD.
- Other factorisations $C = BB^*$ give rise to other representations.
- One may view $r \in \mathscr{W} = \mathcal{U} \otimes \mathcal{S}$ in a tensor product space. This is both theoretically and computationally advantageous.
- Not necessarily required: (probability) measures.

Model on Tensor Product

With $A(p,u) = f$, one finds state $u(p)$ is $\mathcal{V}$-valued function,
it lives in a tensor space $\mathscr{W} = \mathcal{V} \otimes \mathcal{S}$.
Variational statement: $\forall w = v \otimes s \in \mathscr{W} = \mathcal{V} \otimes \mathcal{S}$:

$$\langle A(\cdot, u(\cdot)) - f \,|\, w \rangle_{\mathscr{W}} := \langle\, \langle A(\cdot, u(\cdot)) - f \,|\, v \rangle_{\mathcal{U}} \,|\, s \rangle_{\mathcal{S}} = 0.$$

May allow to show that problem is well-posed on $\mathscr{W} = \mathcal{V} \otimes \mathcal{S}$.

Usual semi-discretisation on finite dimensional $\mathcal{V}_N \subset \mathcal{V}$:

$$\boldsymbol{A}(p, \boldsymbol{u}(p)) = \boldsymbol{f}, \qquad p \in \mathcal{P}.$$

Choose $\{\boldsymbol{v}_n\}_{n=1}^N$ as basis in $\mathcal{V}_N$, then $\boldsymbol{u}(\cdot) \in \mathcal{V}_N \otimes \mathcal{S}$:

$$\boldsymbol{u}(p) = \sum_{n=1}^{N} v_n(p) \boldsymbol{v}_n.$$

Discretisation of Parameter Representation $\mathcal{S}$

Need to discretise (usually infinite dimensional) $\mathcal{S} \subset \mathbb{R}^{\mathcal{P}}$.
Special but important case is when $\mathcal{P} = (\Omega, \mathbb{P})$ is probability space
and $r(p)$ is a $\mathcal{U}$-valued random variable (RV).

Possible representations / discretisations are:

- Samples: the best known representation, i.e. $\{r(p_1), r(p_2), \ldots\}$, e.g. Monte Carlo $\{r(\omega_1), r(\omega_2), \ldots\}$ for RVs.
- Distribution of r in case of RVs (or measure space $(\mathcal{P}, \mu)$). This is the push-forward measure $r_*(\mu)$ on $\mathcal{U}$.
- Moments of r in case of RVs, like $\mathbb{E}\left(r^{\otimes k}\right)$ (mean, covariance, ...).
- Functional representation: function of other (known) functions, $r(p) = \hat{r}(\varsigma_1(p), \varsigma_2(p), \ldots) = \hat{r}(\boldsymbol{\varsigma})$. For RV r function of (known) RVs, e.g. Wiener's polynomial chaos $r(\omega) = \hat{r}(\theta_1(\omega)), \theta_2(\omega), \ldots) =: \hat{r}(\boldsymbol{\theta})$.

Solution by Functional Approximation

Choose finite dimensional subspace $\mathcal{S}_B \subset \mathcal{S}$ with basis $\{X_\beta\}_{\beta=1}^B$, make ansatz for each $v_n(p) \approx \sum_\beta u_n^\beta X_\beta(p)$, giving

$$\boldsymbol{u}(p) = \sum_{n,\beta} u_n^\beta X_\beta(\omega) \boldsymbol{v}_n = \sum_{n,\beta} u_n^\beta X_\beta(\omega) \otimes \boldsymbol{v}_n.$$

Solution is in tensor product $\mathscr{W}_{N,B} := \mathcal{V}_N \otimes \mathcal{S}_B \subset \mathcal{V} \otimes \mathcal{S} = \mathscr{W}$.

Parametric state $\boldsymbol{u}(p)$ represented by tensor $\mathbf{u} = \mathbf{u}_N^B := \{u_n^\beta\}_{n=1,\dots,N}^{\beta=1,\dots,B}$, determined by Galerkin conditions—weighted residua:

$$\forall X_\beta, \boldsymbol{v}_n : \qquad \langle \boldsymbol{A}(\cdot, \boldsymbol{u}(\cdot)) - \boldsymbol{f} \,|\, \boldsymbol{v}_n \otimes X_\beta \rangle_{\mathscr{W}} = 0,$$

$$\text{giving} \qquad \mathbf{A}(\mathbf{u}) = \mathbf{f}.$$

A large $(N \times B)$ coupled system, but may be solved non-intrusively.

Discretisation — model reduction

On continuous level discretisation is choice of subspace

$$\mathscr{W}_{N,B} := \mathcal{V}_N \otimes \mathcal{S}_B \subset \mathcal{V} \otimes \mathcal{S} =: \mathscr{W}$$

and—important for computation—good basis in it.

On discrete level reduced models find sub-manifold $\mathscr{W}_R \subset \mathscr{W}_{N,B}$ with smaller dimensionality $\dim \mathscr{W}_R = R \ll N \times B = \dim \mathscr{W}_{N,B}$. They can work on $\mathcal{S}_B$ or $\mathcal{V}_N$, or both.

Different approaches to choose reduced model:

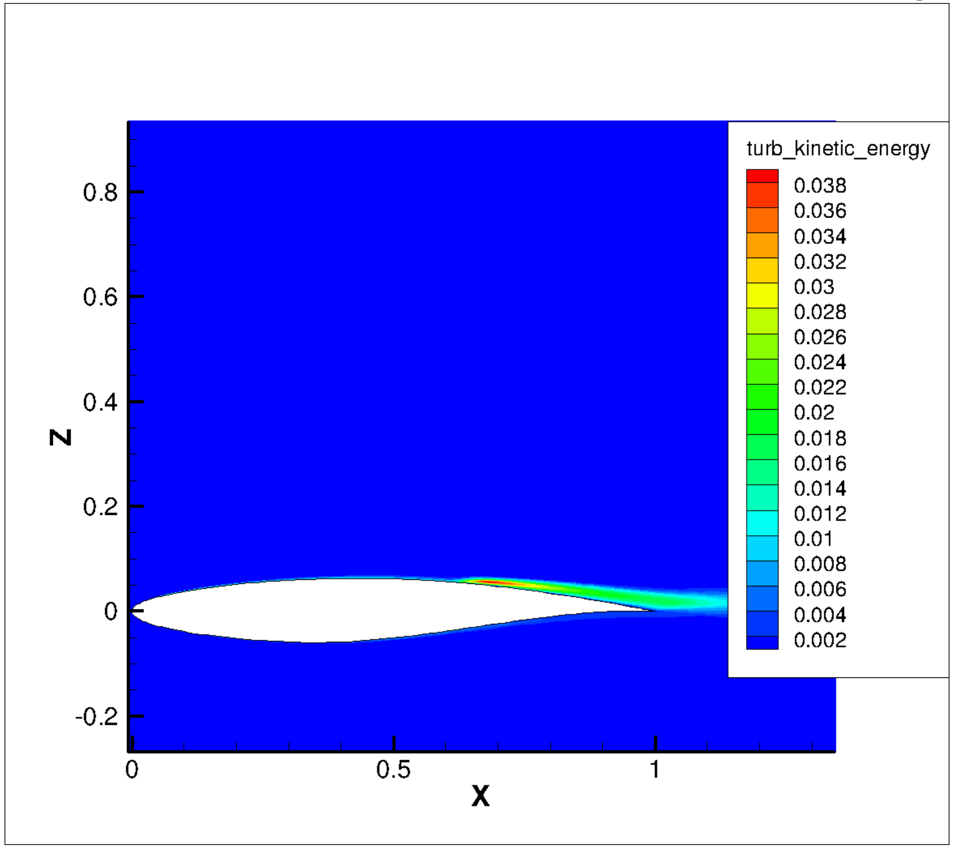
- Before solution process (e.g. modal projection, reduced basis method).
- After solution process (essentially data compression).
- During solution, computing solution and reduction simultaneously.

Here we use low-rank approximations: $\mathbf{u} \approx \sum_{r=1}^{R} \boldsymbol{y}_r \otimes \boldsymbol{g}^r$.

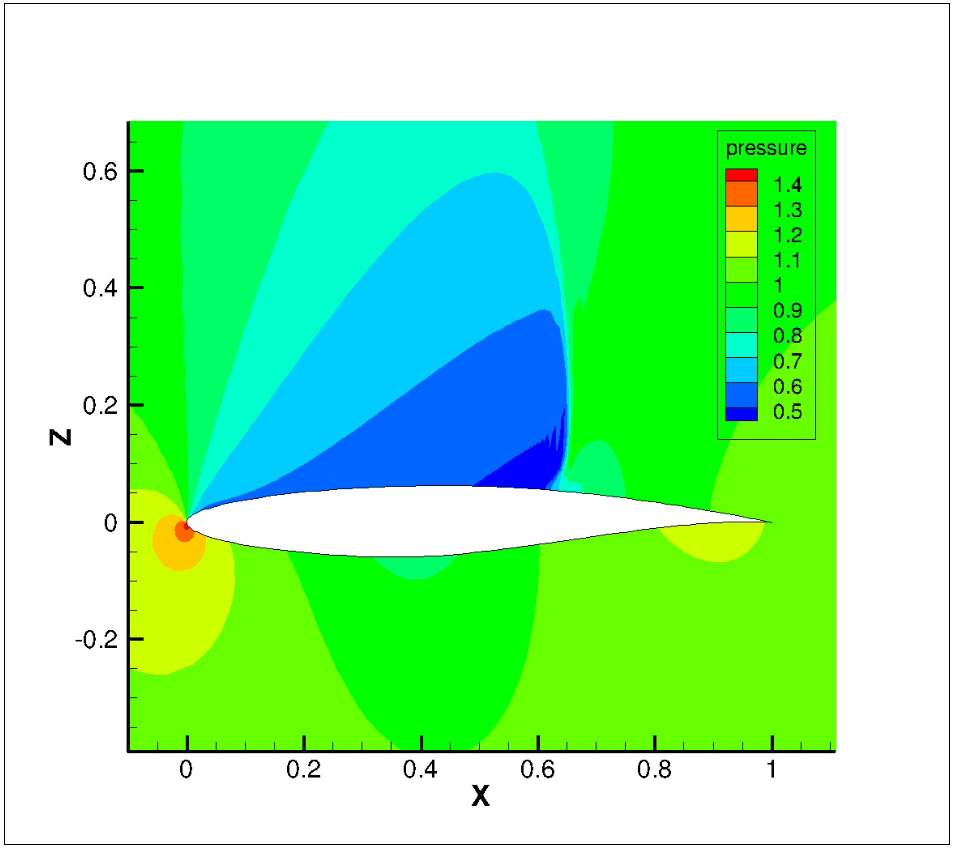
Use in UQ-MC sampling / colocation I

Example: Compressible RANS-flow around RAE air-foil.

Sample solution



turbulent kinetic energy



pressure

Use in UQ-MC sampling / colocation II

Inflow and air-foil shape uncertain.
Data compression achieved by updated SVD:
Made from 600 samples, SVD is updated every 10 samples.

$$N = 260,000 \qquad Z = 600$$

Updated SVD: Relative errors, memory requirements:

rank R	pressure	turb. kin. energy	memory [MB]
10	1.9×10^{-2}	4.0×10^{-3}	21
20	1.4×10^{-2}	5.9×10^{-3}	42
50	5.3×10^{-3}	1.5×10^{-4}	104

Full tensor $\in \mathbb{R}^{260000 \times 600}$ would cost 10 GB of storage.

Use in Galerkin method

Solution process to obtain co-efficients for coupled problem

$$\mathbf{u}_{k+1} = \mathbf{\Phi}(\mathbf{u}_k), \qquad \mathbf{u} \in \mathscr{W}_{N,B} := \mathcal{U}_N \otimes \mathcal{S}_B \subset \mathcal{U} \otimes \mathcal{S} = \mathscr{W}$$

(with contraction $\varrho < 1$) may be written as tensorised mapping

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \mathbf{\Xi}(\mathbf{u}_k) = \mathbf{u}_k - \left(\sum_{m=1}^{M} \boldsymbol{Y}_m \otimes \boldsymbol{G}^m \right)(\mathbf{u}_k).$$

How to find low-rank $\mathbf{u}^R = \sum_{j=1}^{R} \boldsymbol{y}_j \otimes \boldsymbol{g}^j \in \mathscr{W}_R \subset \mathscr{W}_{N,B} \subset \mathscr{W}$?

- PGD—proper generalised decomposition: build $\mathbf{u}^{j+1}$ from $\mathbf{u}^j$ by e.g. greedy algorithm (alternating least squares) alternating solutions on $\mathcal{U}_N$ and $\mathcal{S}_B$.
- low-rank iteration: start with $\mathbf{u}_0^R = \sum_{j=1}^{R} \boldsymbol{y}_{0,j} \otimes \boldsymbol{g}^{0,j}$, keep it like that in iteration (truncated / perturbed iteration saves on computation).

Perturbed low-rank iteration

$$\mathbf{u}_1 = \sum_{j=1}^{R_0} \boldsymbol{y}_{0,j} \otimes \boldsymbol{g}^{0,j} - \sum_{m=1}^{M} \boldsymbol{Y}_m(\mathbf{u}_0) \otimes \boldsymbol{G}^m(\mathbf{u}_0).$$

Rank of $\mathbf{u}_{k+1}$ grows by M.

Possible for pre-conditioned linear iteration,

and modified-, full-, inexact- and quasi-Newton iteration.

If iteration and rank-truncation $\mathbf{T}_\epsilon$ are alternated, rank stays low.

$$\hat{\mathbf{u}}_{k+1} = \mathbf{u}_k - \boldsymbol{\Xi}(\mathbf{u}_k), \quad \mathbf{u}_{k+1} = \mathbf{T}_\epsilon(\hat{\mathbf{u}}_{k+1}) \quad \text{with} \quad \|\mathbf{T}_\epsilon(\mathbf{v}) - \mathbf{v}\| \leq \epsilon.$$

Theorem: [Hackbusch, Tyrtyshnikov] super-linearly (or linearly $\varrho < 1/2$) originally convergent process converges to stagnation range 2ϵ.

Theorem: [Zander, HGM] all originally convergent processes converge, if $\varrho > 0$ (linear) to stagnation range $\epsilon/(1-\varrho)$.

Diffusion SPDE and variational form

Solution $u(x,\omega)$ is sought in tensor product space

$$\mathscr{W} := \mathcal{V} \otimes \mathcal{S} = \mathring{H}^1(\mathcal{G}) \otimes L_2(\Omega).$$

Variational formulation: find $u \in \mathscr{W}$ such that $\forall\, w = w \otimes s \in \mathscr{W}$:

$$\mathsf{a}(w,u) := \mathbb{E}\left(\int_{\mathcal{G}} \nabla_x w(x,\omega) \cdot \kappa(x,\omega) \cdot \nabla_x u(x,\omega)\,\mathrm{d}x\right)$$

$$= \mathbb{E}\left(\int_{\mathcal{G}} w(x,\omega) f(x,\omega)\,\mathrm{d}x\right) =: \langle\!\langle v, f \rangle\!\rangle.$$

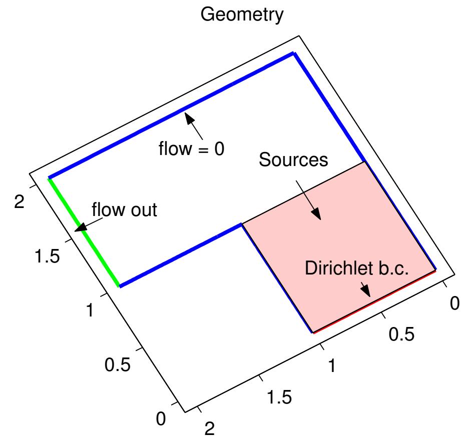
Lax-Milgram lemma $\to$ existence, uniqueness, and well-posedness.

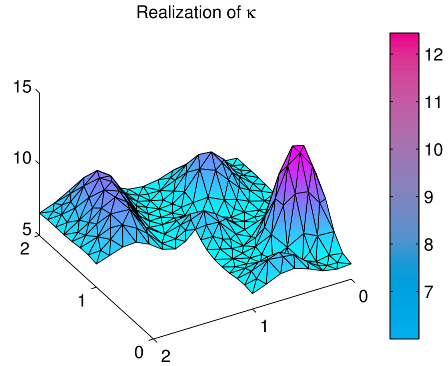
Galerkin discretisation on $\mathscr{W}_{B,N} = \mathcal{V}_N \otimes \mathcal{S}_B \subset \mathcal{V} \otimes \mathcal{S} = \mathscr{W}$ leads to

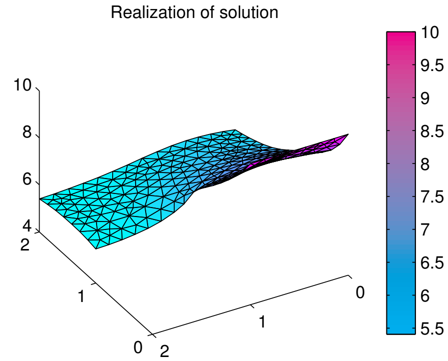
$$\mathbf{A}\,\mathbf{u} = \left(\sum_{m=1}^{M} \xi_m\, \boldsymbol{A}_m \otimes \boldsymbol{\Delta}^{(m)}\right) \mathbf{u} = \mathbf{f}.$$

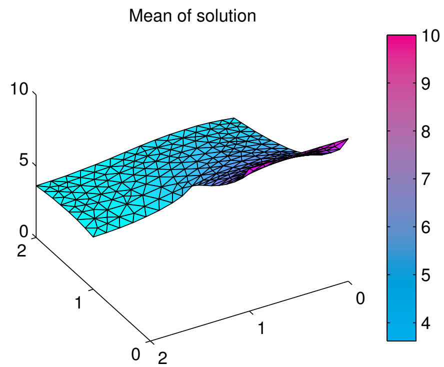
Céa's lemma $\to$ Galerkin converges.

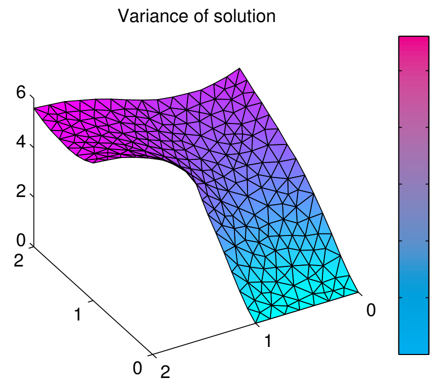
Example solution

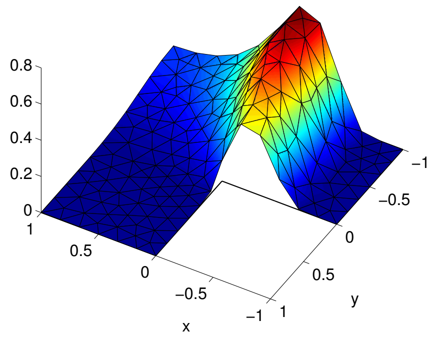








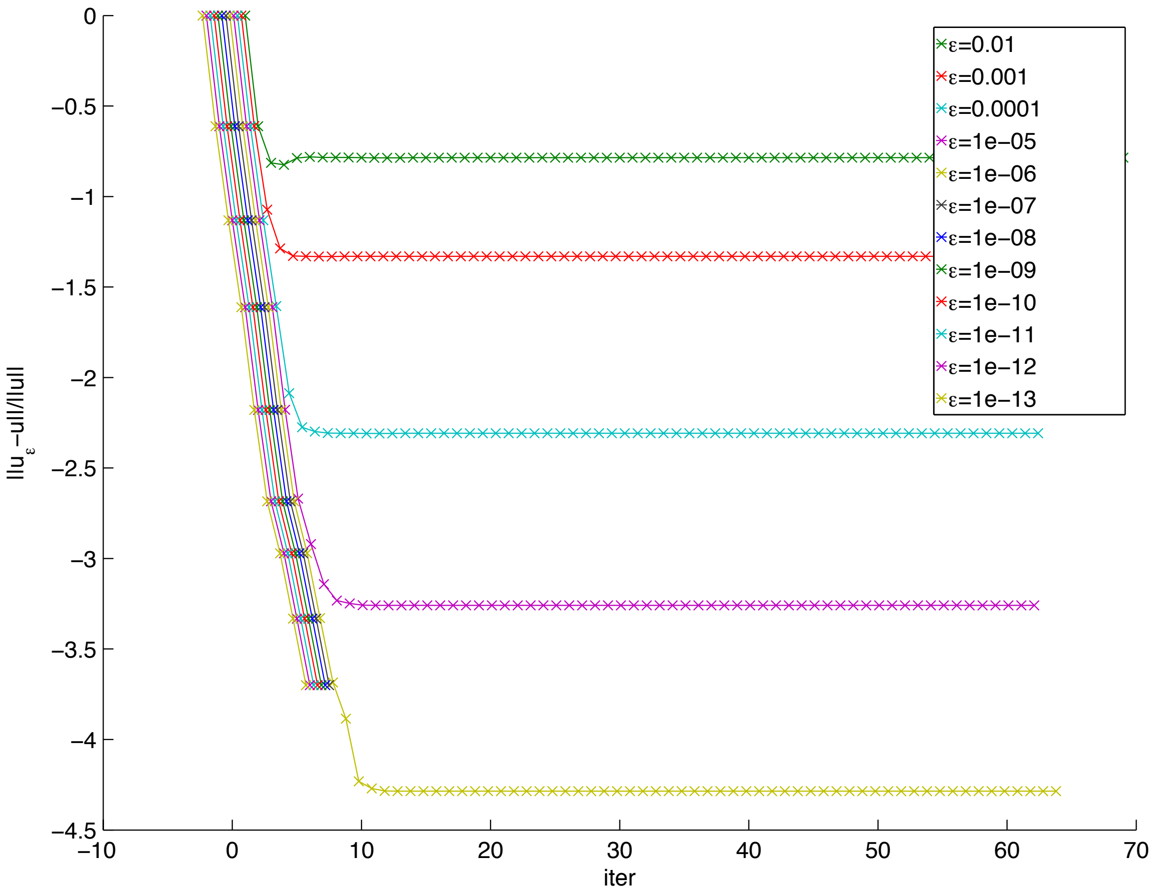




$\Pr\{u(x) > 8\}$

Iteration accuracy

Convergence of truncated iteration. $N \times B \approx 10^8$ on 2GB Laptop.



Important points II

- Discretisation may use tensor product.
- Discretised system equation may be written in tensorised form.
- Solver may be represented in tensorised form.
- To view $u \in \mathscr{W} = \mathcal{U} \otimes \mathcal{S}$ in a tensor product space, allows to show well-posedness and Galerkin convergence.
- Tensor representation allows a sparse approximation of dense quantities via low-rank approximation.
- Allows large savings in computation and storage.

Inverse problem—updating

Quantity $q(p) \in \mathcal{Q}$ is unknown / uncertain, measurement operator $y = Y(q; u) = Y(q, u(f; q))$ for observations $z = y + \varepsilon$ with random error ε to determine / update q. Function not invertible $\Rightarrow$ ill-posed problem, obseravtion z does not contain enough information.

In Bayesian framework state of knowledge modelled in a probabilistic way, parameters q are uncertain, and assumed as random.
Updating the distribution—state of knowledge of q is well-posed.

Classically, Bayes's theorem gives conditional probability

$$\mathbb{P}(I_q|M_z) = \frac{\mathbb{P}(M_z|I_q)}{\mathbb{P}(M_z)}\mathbb{P}(I_q);$$

expectation with this posterior measure is conditional expectation.

Modern approach starts from conditional expectation $\mathbb{E}\left(\cdot|M_z\right)$ on $\mathcal{S} = L_2(\Omega, \mathbb{P}, \mathfrak{A})$, from this $\mathbb{P}(I_q|M_z) = \mathbb{E}\left(\chi_{I_q}|M_z\right)$.

Update

Definition: conditional expectation is defined as orthogonal projection onto the subspace $L_2(\Omega, \mathbb{P}, \sigma(z))$:

$$\mathbb{E}(q|\sigma(z)) := P_{\mathscr{Q}_n} q = \text{argmin}_{\tilde{q} \in L_2(\Omega, \mathbb{P}, \sigma(z))} \|q - \tilde{q}\|^2_{L_2}$$

The subspace $\mathscr{Q}_n := L_2(\Omega, \mathbb{P}, \sigma(z))$ represents the available information, the estimate minimises the function $\|q - (\cdot)\|^2$ over $\mathscr{Q}_n$. More general loss functions than mean square error are possible.

The update, also called the assimilated value $q_a(\omega) := P_{\mathscr{Q}_n} q = \mathbb{E}(q|\sigma(z))$, a function of z, is a $\mathcal{Q}$-valued RV and represents new state of knowledge after the measurement.

$\Rightarrow$ Pythagoras $\|q\|^2_{L_2} = \|q - q_a\|^2_{L_2} + \|q_a\|^2_{L_2}$ shows reduction of variance.

Case with prior information

Here we have prior information $\mathscr{Q}_f$ and prior estimate $q_f(\omega)$ (forecast) and measurements z generating a subspace $\mathscr{Y}_0 \subset \mathscr{Y}$, and via Y a subspace $\mathscr{Q}_0 \subset \mathscr{Q}$.

We now need projection onto $\mathscr{Q}_n = \mathscr{Q}_f + \mathscr{Q}_0$, with reformulation as an orthogonal direct sum: $\mathscr{Q}_n = \mathscr{Q}_f + \mathscr{Q}_0 = \mathscr{Q}_f \oplus (\mathscr{Q}_0 \cap \mathscr{Q}_f^{\perp}) = \mathscr{Q}_f \oplus \mathscr{Q}_i$.

The update / conditional expectation / assimilated value is the orthogonal projection

$$q_a = q_f + P_{\mathscr{Q}_i} q = q_f + q_i,$$

where q_i is the innovation.

How can one compute q_a or $q_i = P_{\mathscr{Q}_i} q$?

Simplification

The RV $P_{\mathscr{Q}_i} q$ is a function of the measurement z.

For simplicity do not consider subspace $\mathscr{Q}_0$ generated by all measurable functions of z, but only linearly generated $\mathscr{Q}_\ell$. This gives linear minimum variance estimate $\hat{q}_a$.

Theorem: (Generalisation of Gauss-Markov)

$$\hat{q}_a(\omega) = q_f(\omega) + K(z(\omega) - y_f(\omega)),$$

where the linear Kalman gain operator $K : \mathscr{Y} \to \mathscr{Q}$ is

$$K := \mathrm{cov}(q_f, y)\big(\mathrm{cov}(y, y) + \mathrm{cov}(\epsilon, \epsilon)\big)^{-1}.$$

(The normal Kalman filter is a special case.)

Or in tensor space $q \in \mathscr{Q} = \mathcal{Q} \otimes \mathcal{S}$—works well for low-rank rep.:

$$\hat{q}_a = q_f + (K \otimes I)(z - y_f).$$

Update

On semi-discretisation, stochastic discretisation is

$$I \otimes \Pi : \mathcal{Q}_h \otimes \mathcal{S} \to \mathcal{Q}_h \otimes \mathcal{S}_k.$$

It commutes with $\boldsymbol{K} \otimes I$, so the update equation (projection / conditional expectation) may be projected on the fully discrete space.

With $\mathbf{u} := [\ldots, \boldsymbol{u}^\alpha, \ldots] \in \mathcal{Q}_h \otimes \mathcal{S}_k$ the forward problem is

$$\mathbf{A}(\mathbf{u};\mathbf{q}) = \mathbf{f} \text{ and } \mathbf{y}_f = \mathbf{Y}(\mathbf{q}_f, \mathbf{S}(\mathbf{f}, \mathbf{q}_f)) \in \mathcal{Y}_h \otimes \mathcal{S}_k.$$

Update on $\mathcal{Q}_h \otimes \mathcal{S}_k$: $\quad \hat{\mathbf{q}}_a = \mathbf{q}_f + (\boldsymbol{K} \otimes \boldsymbol{I})\big(\mathbf{z} - \mathbf{y}_f\big).$

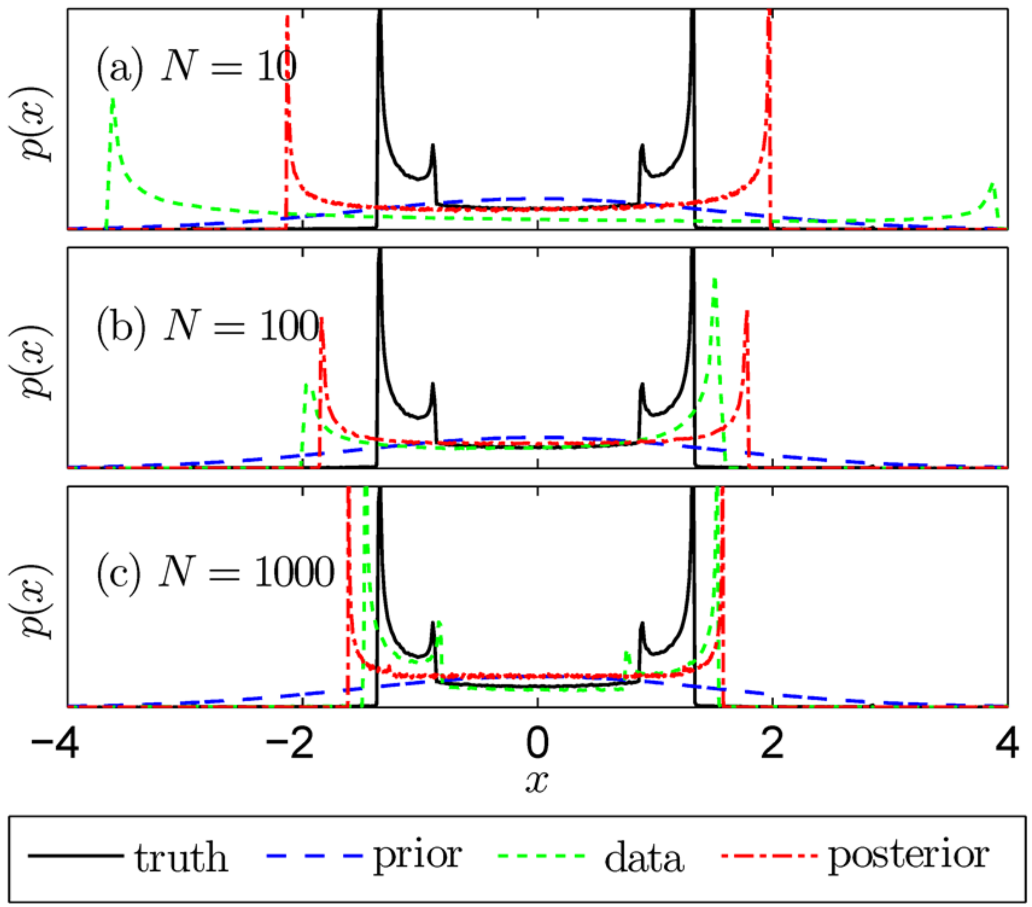
Forward problem and update benefit from low-rank / sparse approximation, e.g. $\mathbf{q} \approx \sum_j \boldsymbol{p}_j \otimes \boldsymbol{s}_j$.

Example 1: Identification of bi-modal dist

Setup: Scalar RV x with non-Gaussian bi-modal "truth" $p(x)$; Gaussian prior; Gaussian measurement errors.

Aim: Identification of $p(x)$.

10 updates of $N = 10, 100, 1000$ measurements.



Example 2: Lorenz-84 chaotic model

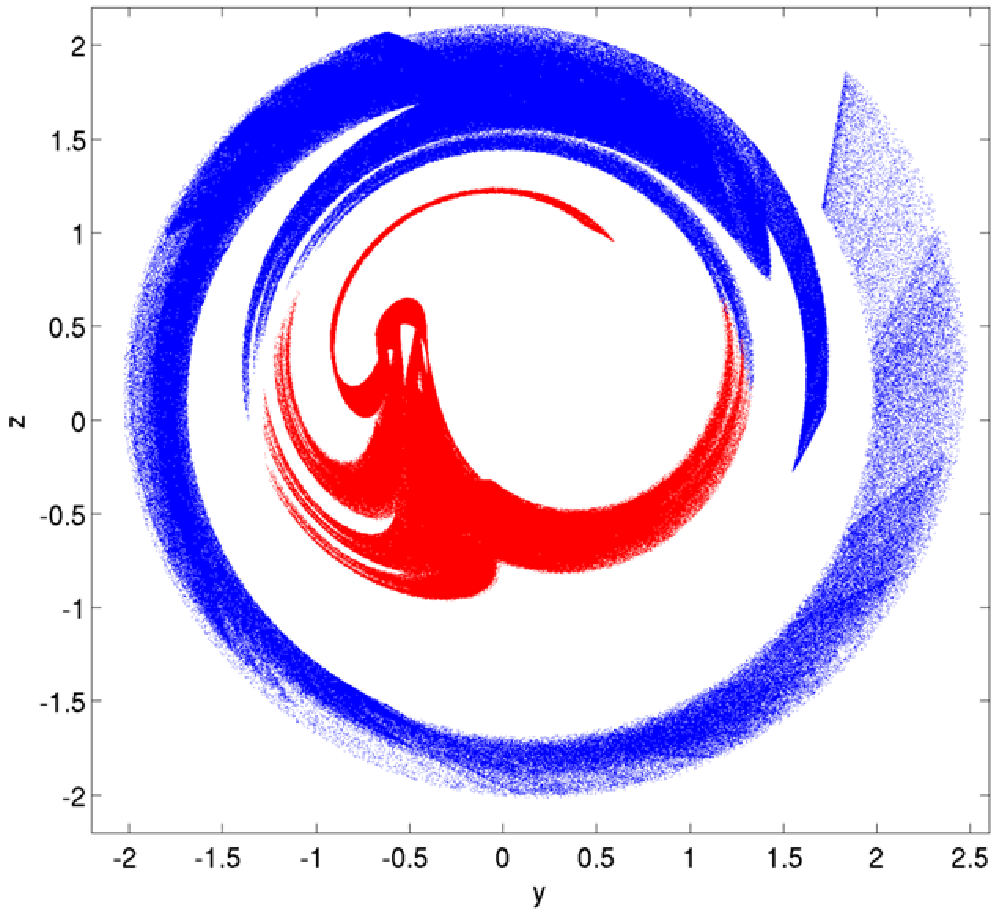
Setup: Non-linear, chaotic system

$$\dot{u} = f(u), \; u = [x, y, z]$$

Small uncertainties in initial conditions u_0 have large impact.

Aim: Sequentially identify state u_t.

Methods: PCE representation and PCE updating and sampling representation and (Ensemble Kalman Filter) EnKF updating.

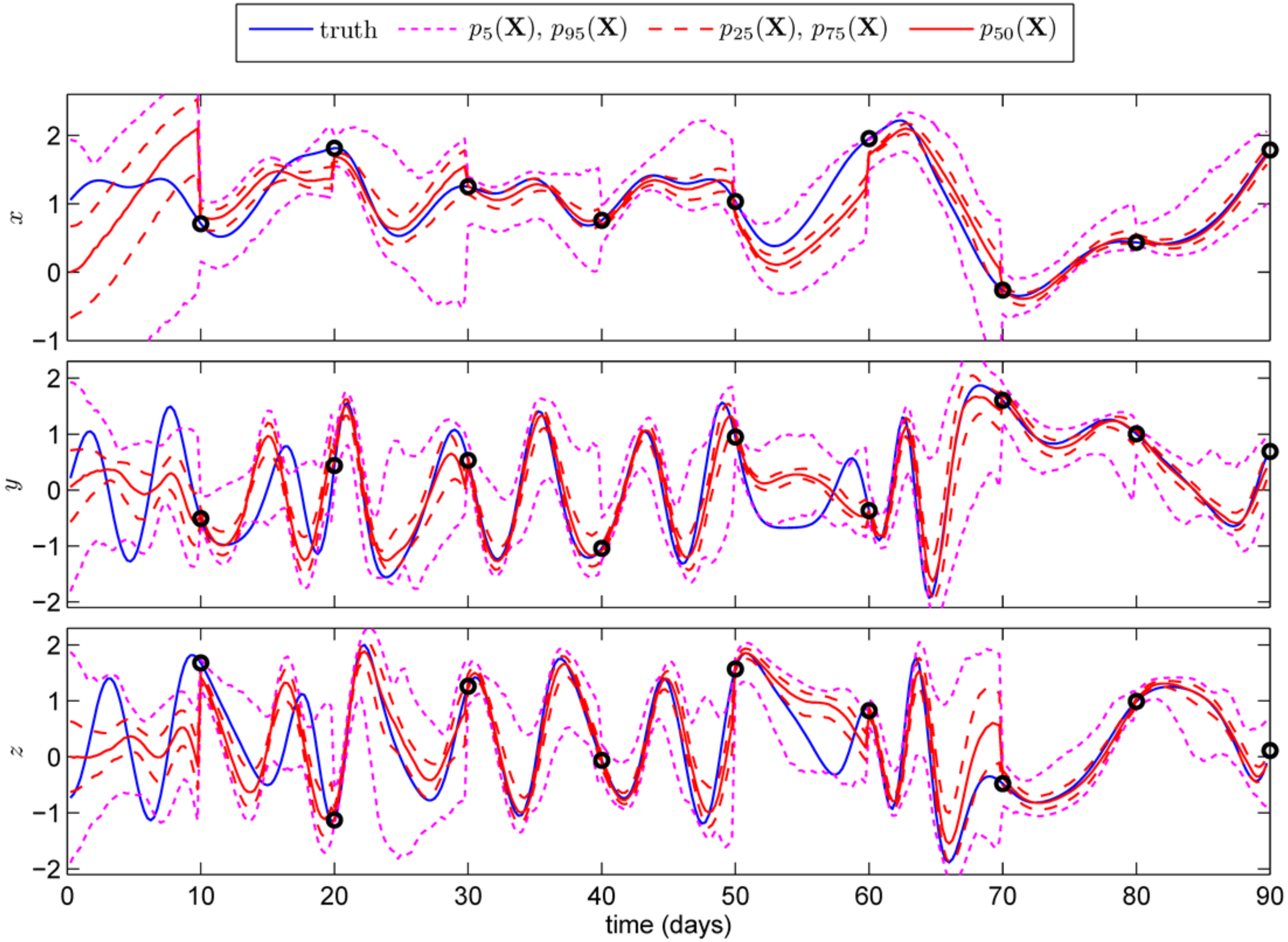


Poincaré cut for $x = 1$.

Example 2: Lorenz-84 PCE representation

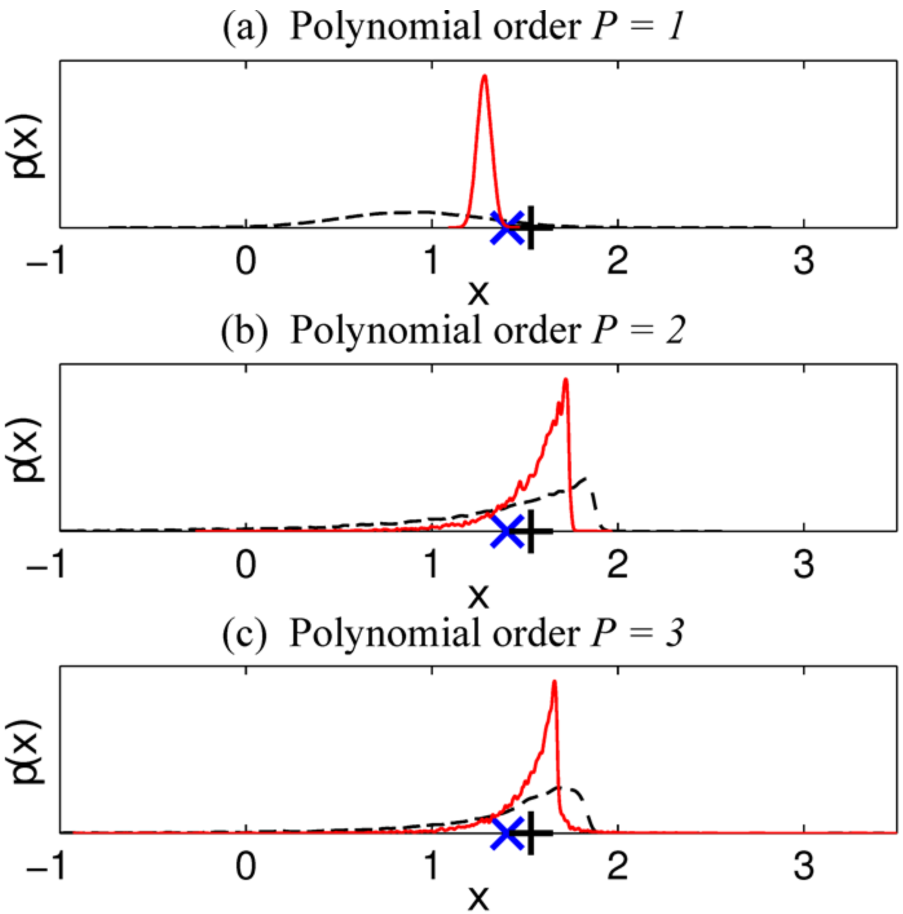
PCE: Variance reduction and shift of mean at update points.

Skewed structure clearly visible, preserved by updates.

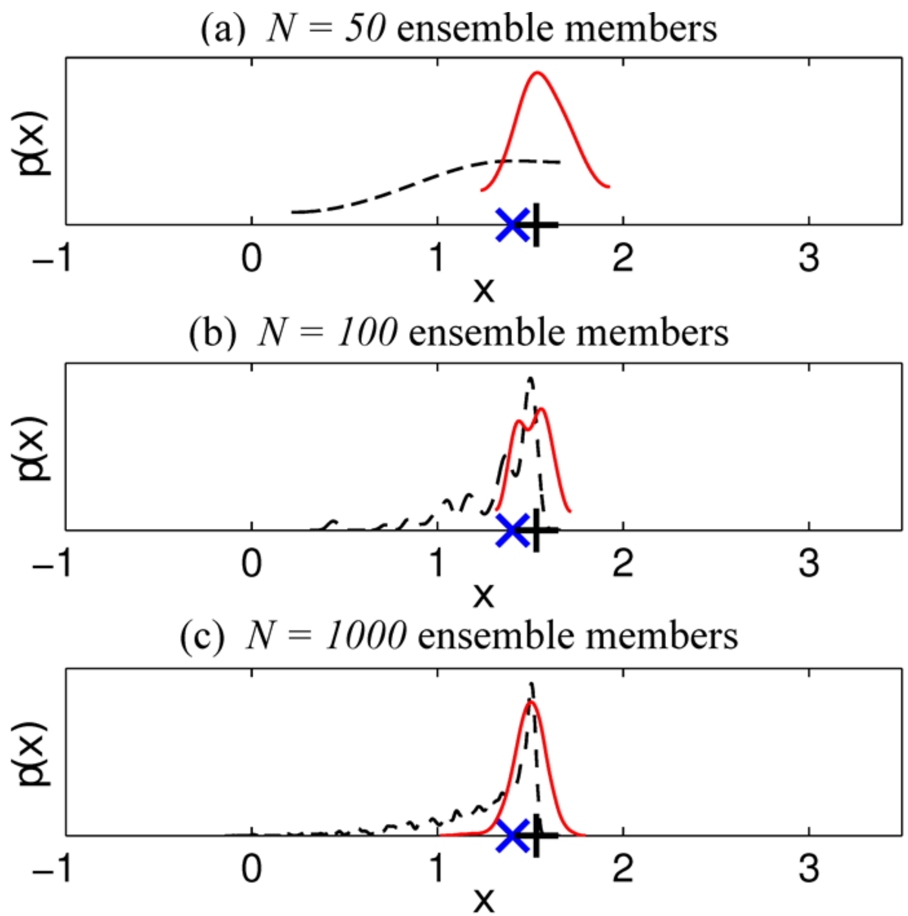


Example 2: Lorenz-84 non-Gaussian identification

PCE

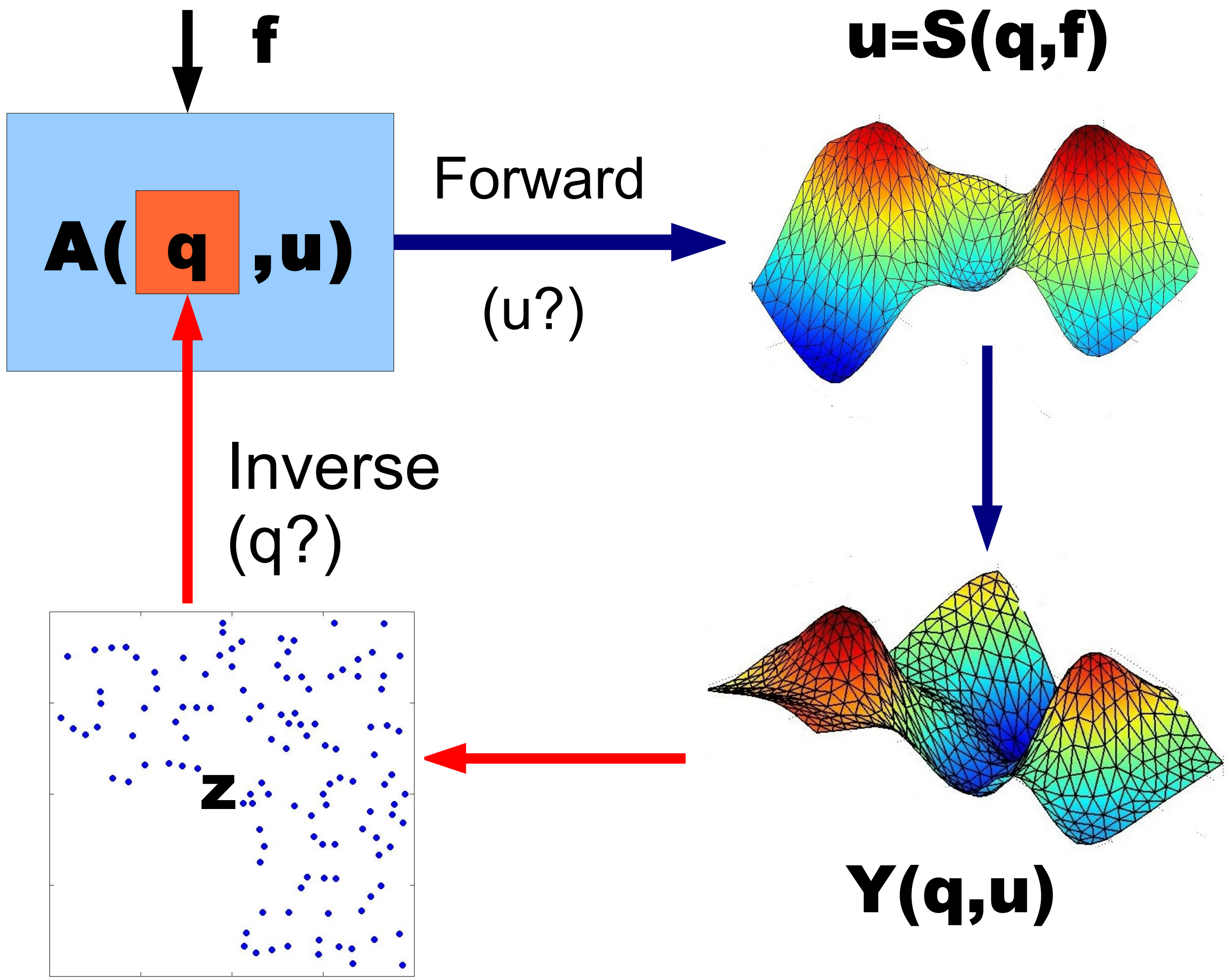


EnKF

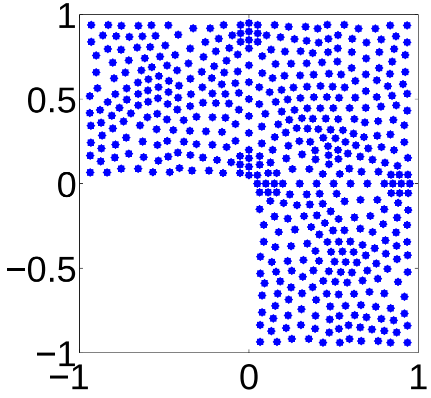


truth $\times$ **measurement** $+$ posterior *prior*

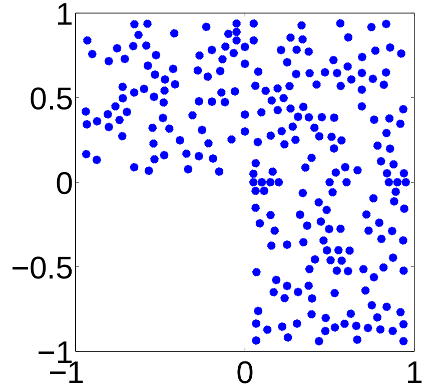
Example 3: diffusion—schematic representation



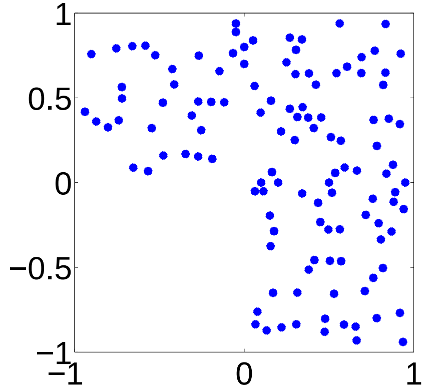
Measurement patches



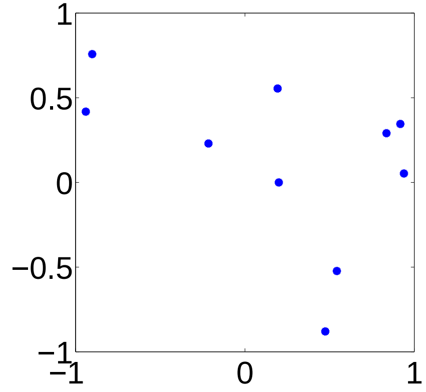
447 measurement patches



239 measurement patches

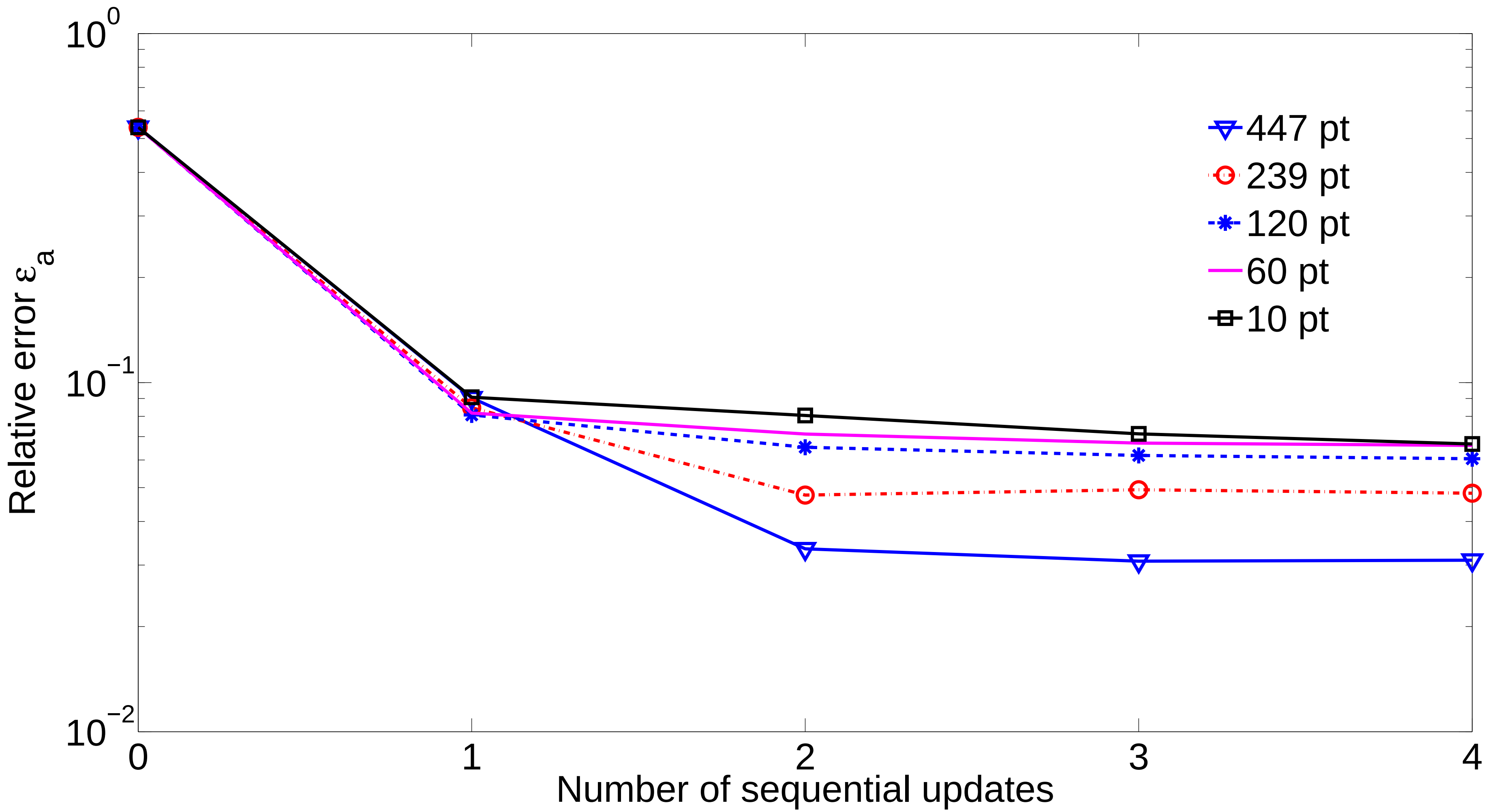


120 measurement patches

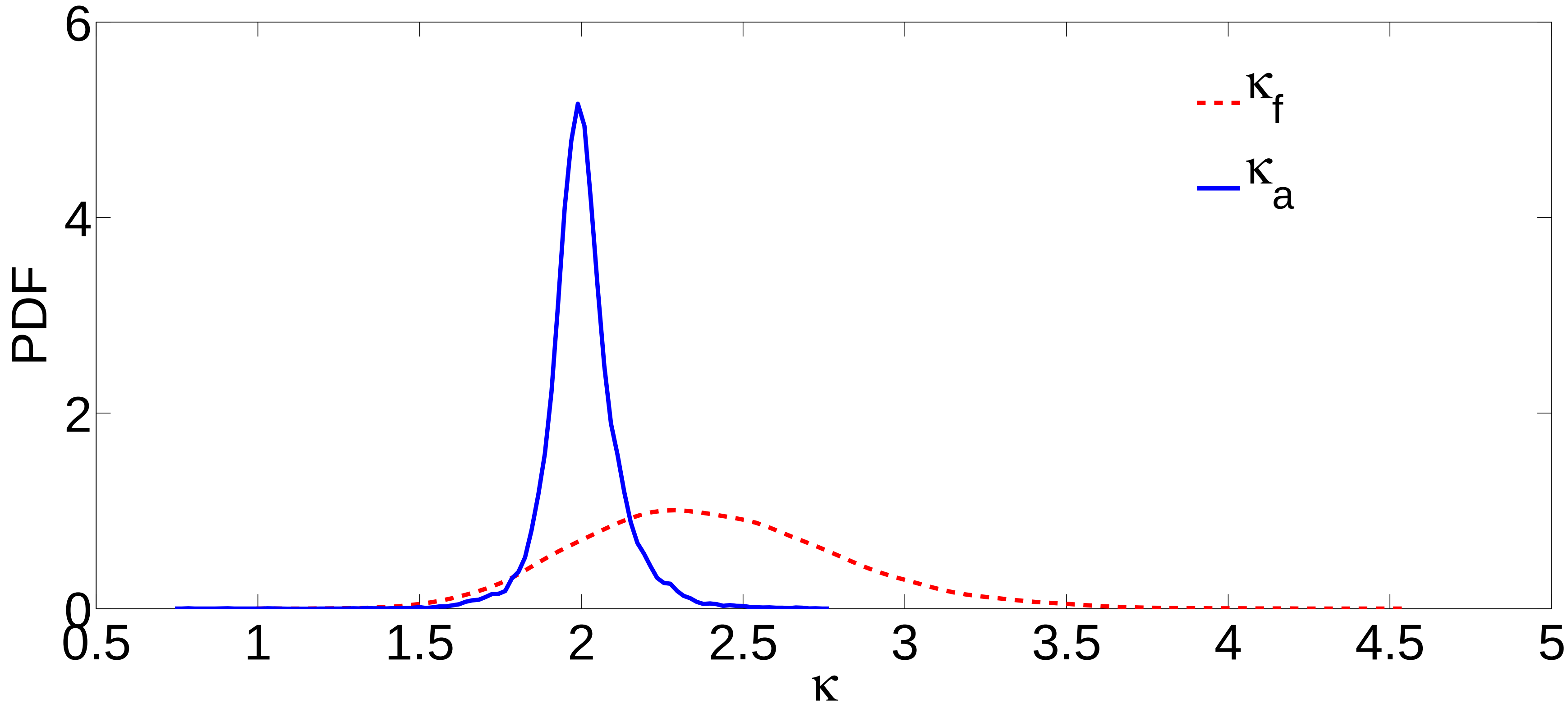


10 measurement patches

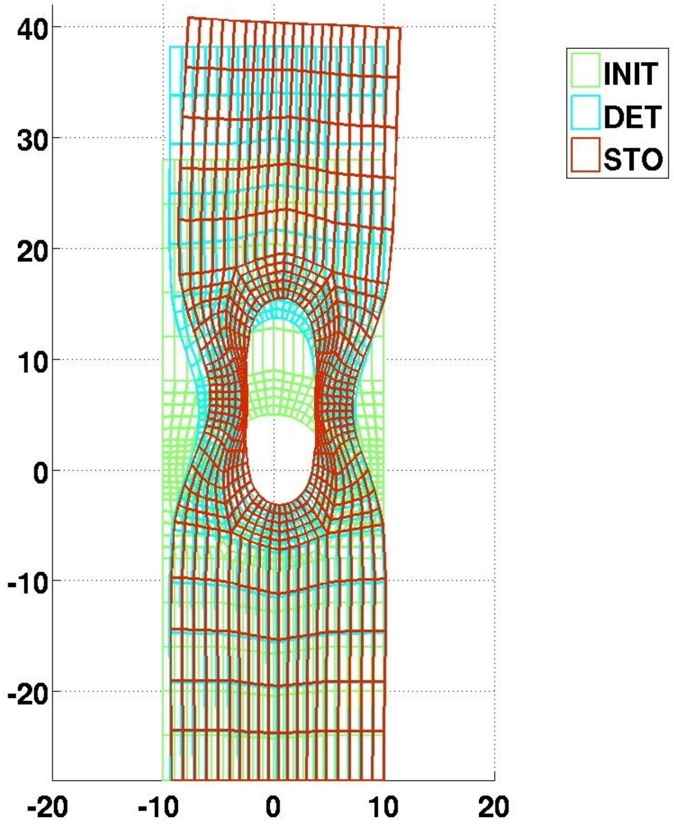
Convergence plot of updates



Forecast and Assimilated pdfs

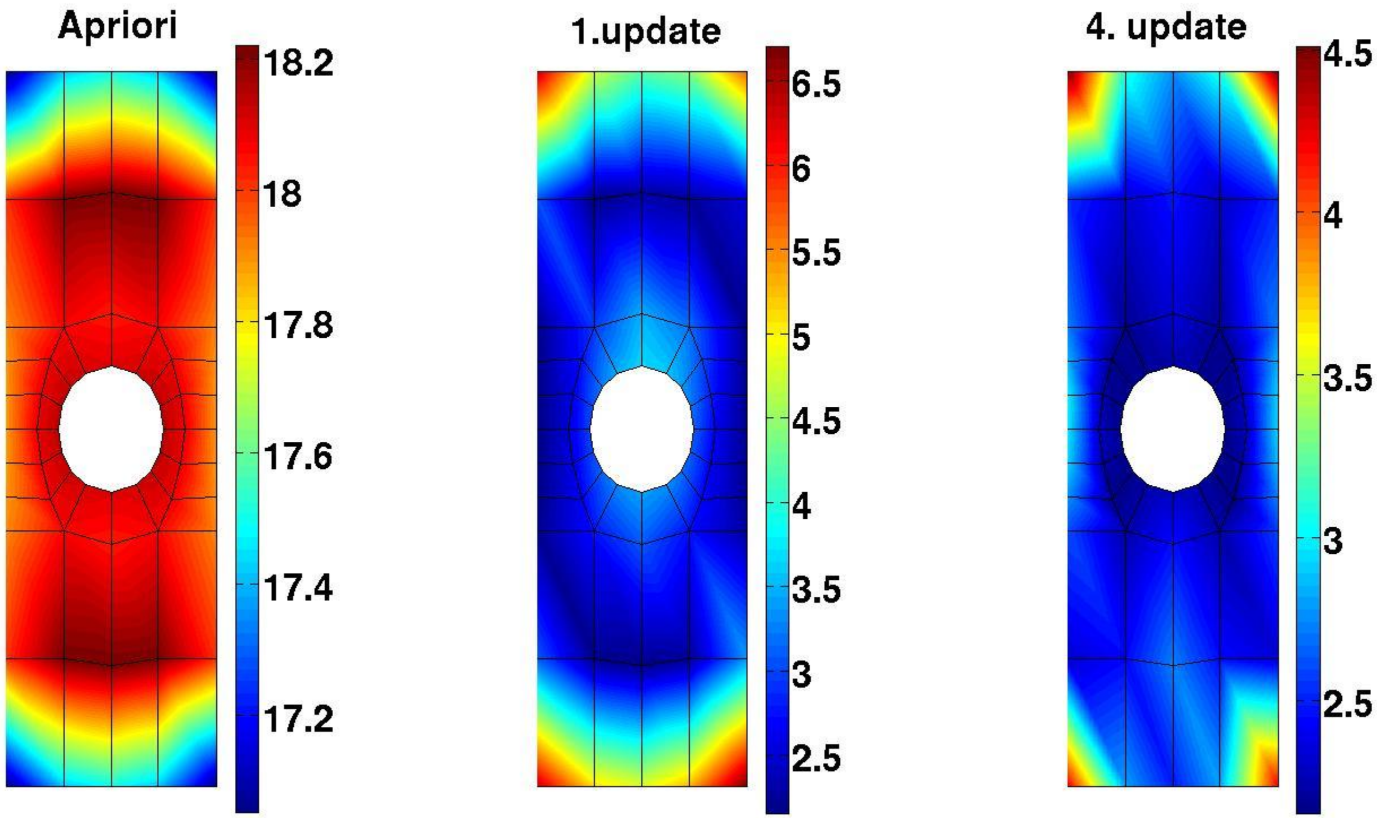


Example 4: Elasto-plastic plate with hole



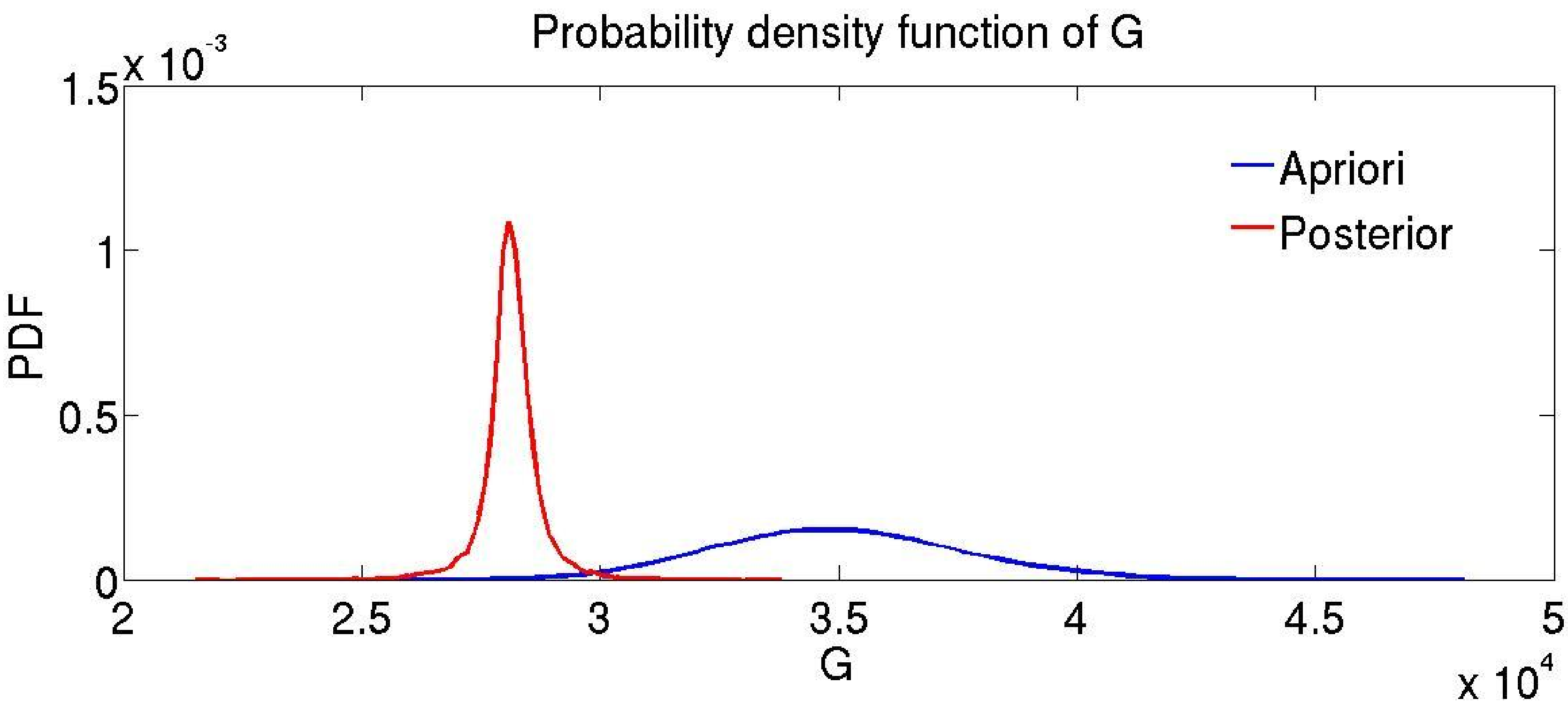
Forward problem: the comparison of the mean values of the total displacement fo r deterministic, initial and stochastic configuration

Relative variance of shear modulus estimate



Relative RMSE of variance [%] after 4th update in 10% equally distributed m easurment points

Probability density shear modulus



Comparison of prior and posterior distribution

Conclusion

- Parametric models lead to factorisations / representations in tensor product form.
- Sparse low-rank tensor products save storage and computation in sampling and functional approximation.
- Works also for non-linear non-Gaussian problems and solvers.
- Bayesian update is a projection, needs no Monte Carlo.
- Compatible with low-rank and spectral representation.
- Works on non-smooth non-Gaussian examples.