Parametric Quantities, their Representations and Factorisations, and Inverse Identifications Methods

Hermann G. Matthies

Bojana V. Rosić, Alexander Litvinenko, Oliver Pajonk

Institut für Wissenschaftliches Rechnen
TU Braunschweig

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

Overview

1. Parametric problems
2. System properties identification
3. Stochastic forward problem
4. Bayesian update through projection
5. Computational issues
6. Examples

Parametric problems

For each ω in a parameter set Ω, let $r(\omega)$ be an element belonging to some problem in a Hilbert space $\mathcal{V}$ (for simplicity).

With $r : \Omega \to \mathcal{V}$, denote $\mathcal{U} = \overline{\mathrm{span}}\, r(\Omega) = \overline{\mathrm{span}}\, \mathrm{im}\, r$.

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

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

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

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 kernel

$$\varkappa(\omega_1, \omega_2) = \langle r(\omega_1) | r(\omega_2) \rangle_{\mathcal{U}} \in \mathbb{R}^{\Omega \times \Omega}$$

Reproducing property:

$$\forall \phi \in \mathcal{R} : \ \langle \varkappa(\omega, \cdot) | \phi(\cdot) \rangle_{\mathcal{R}} = \phi(\omega) =: \langle \delta_\omega, \phi \rangle_{\mathcal{R}^* \times \mathcal{R}}.$$

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

Choose CONS $\{\varphi_m\}_{m \in \mathbb{N}}$ in $\mathcal{R}$: $R^{-1} = \sum_m w_m \otimes \varphi_m$, with $R w_m = \varphi_m$.
Let $Q_{\mathcal{R}} : \ell_2 \ni \boldsymbol{a} = (a_1, a_2, \dots) \mapsto \sum_m a_m \varphi_m \in \mathcal{R}$.

$\Rightarrow$ tensor representation $R^{-1} \circ Q_{\mathcal{R}} : \ell_2 \ni \boldsymbol{a} \mapsto \sum_m a_m w_m \in \mathcal{U}$

‘Correlation’

Assume scalar product $\langle\cdot|\cdot\rangle_Q$ on $\mathbb{R}^\Omega \to$ Hilbert space $\mathcal{Q}$.
If (Ω, μ) is a measure space, take $\mathcal{Q} = L_2(\Omega, \mu)$.

Define self-adjoint and positive definite ‘correlation’ operator C in $\mathcal{U}$ by

$$u, v \in \mathcal{U}: \ \langle Cu|v\rangle_{\mathcal{U}} = \langle Ru|Rv\rangle_{\mathcal{Q}} = \langle\ (\langle u|r(\cdot)\rangle_{\mathcal{U}}|\langle r(\cdot)|v\rangle_{\mathcal{U}})\ \rangle_{\mathcal{Q}}.$$

$C = R^* R$; [If $\mathcal{Q} = L_2(\Omega)$: $C = \int_\Omega r(\omega) \otimes r(\omega)\, \mu(\mathrm{d}\omega)$.]

$\Rightarrow$ has spectrum $\sigma(C) \subseteq \mathbb{R}_+$.

Spectral decomposition with projectors E_λ

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

Spectral decomposition

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

$$Cu = \sum_m \lambda_m \sum_n^{\text{mult.}\lambda_m} \langle v_m^n | u \rangle_{\mathcal{U}} \, v_m^n = \sum_m \lambda_m \sum_n^{\text{mult.}\lambda_m} (v_m^n \otimes v_m^n) \, u.$$

If $\sigma(C) \neq \sigma_p(C)$: generalised eigenvectors v_λ and Gelfand triplets (rigged Hilbert spaces) for the continuous spectrum:

$$Cu = \sum_n^{\text{max mult.}} \int_{\mathbb{R}_+} \lambda \, (v_\lambda^n \otimes v_\lambda^n) \, u \, \varrho_n(\mathrm{d}\lambda).$$

Representation as sum / integral of rank-1 operators.
Numerical approximation will give a sum. Assumed from now on.

Singular value decomposition

C unitarily equivalent to multiplication operator M_k, with $k \geq 0$:

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

This connects to the singular value decomposition (SVD) of $R = SM_k^{1/2}V^*$, with a (here) unitary S.

With $\sqrt{\lambda_m}\, s_m := Rv_m \in \mathcal{R}$:

$$(Ru)(\omega) = \langle r(\omega)|u\rangle_{\mathcal{U}} = \sum_m \sqrt{\lambda_m}\, \langle v_m|u\rangle_{\mathcal{U}}\, s_m(\omega)$$

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

Model reduction possible by truncating the sum.

Karhunen-Loève Expansion

For partly continuous spectrum we get

$$r(\omega) = \sum_{n}^{\text{max mult.}} \int_{\mathbb{R}_+} \sqrt{\lambda}\, \langle v_\lambda^n, u \rangle\, s_\lambda^n(\omega)\, \varrho_n(\mathrm{d}\lambda)$$

With approximation or only point spectrum

$$r(\omega) = \sum_m \sqrt{\lambda_m}\, s_m(\omega) v_m, \quad r \in L_2(\Omega) \otimes \mathcal{U}.$$

This is the Karhunen-Loève-expansion, due to the SVD.

A sum of rank-1 operators / tensors.

Observe that r is linear in the s_m.

A representation of r, model reduction possible by truncation of sum.

Kernel spectral decomposition

For $\phi, \psi \in \mathcal{Q}$ we have also $\langle R^*\phi | R^*\psi\rangle_{\mathcal{U}}$; to compute R^*,
for $\psi \in \mathcal{Q}$ define an operator $\hat{C} = RR^*$ on $\mathcal{Q} = [L_2(\Omega)]$ by

$$(\hat{C}\psi)(\omega_1) := \langle \varkappa(\omega_1, \cdot) | \psi(\cdot)\rangle_{\mathcal{Q}} \quad [= \int_\Omega \varkappa(\omega_1, \omega_2)\, \psi(\omega_2)\, \mu(\mathrm{d}\omega_2)\,].$$

$$\langle R^*\phi | R^*\psi\rangle_{\mathcal{U}} = \langle \phi | \hat{C}\psi\rangle_{\mathcal{Q}} \, [= \iint_{\Omega\times\Omega} \phi(\omega_1)\, \varkappa(\omega_1, \omega_2)\, \psi(\omega_2)\, \mu(\mathrm{d}\omega_1)\, \mu(\mathrm{d}\omega_2).\,]$$

Eigenvalue problem for $\hat{C}$ gives (Mercer's theorem)

$$\varkappa(\omega_1, \omega_2) = \sum_m \lambda_m\, s_m(\omega_1) s_m(\omega_2),$$

$\{s_m\}$ is CONS in $\mathcal{Q}\, [\, = L_2(\Omega)\,]$, $\{\sqrt{\lambda_m}\, s_m\}$ is CONS in $\mathcal{R}$.

$$R^*\phi = \sum_m \sqrt{\lambda_m}\, v_m \langle s_m | \phi\rangle_{\mathcal{Q}}, \quad R^{-1}\phi = \sum_m \lambda_m^{-1/2}\, v_m \langle s_m | \phi\rangle_{\mathcal{Q}}.$$

Factorisations

R^* (or truncation) now serves as a representation. This is a factorisation of C, let $C = B^*B$ be an arbitrary one. Some possible ones:

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

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 $B : \mathcal{U} \to \mathcal{H}$, let $\{e_k\}$ be CONS in $\mathcal{H}$.

Unitary $Q : \ell_2 \ni \boldsymbol{a} = (a_1, \ldots, a_n, \ldots) \mapsto \sum_k a_k e_k \in \mathcal{H}$.

Let $\tilde{r}(\boldsymbol{a}) := B^*Q\boldsymbol{a} := \tilde{R}^*\boldsymbol{a}$, i.e. $\tilde{R}^* : \ell_2 \to \mathcal{U}$. Then

$$\tilde{R}^*\tilde{R} = (B^*Q)(Q^*B) = B^*B = C.$$

Integral decompositions

More decompositions and representations possible via $\hat{C}$. Let
$\varkappa(\omega_1, \omega_2) = \int_X g(\omega_1, x) g(\omega_2, x)\, \nu(\mathrm{d}x)$.

Set $g_m(x) := \langle g(\cdot, x) | s_m \rangle_{\mathcal{Q}}$ to give

$$p : X \ni x \mapsto p(x) := \sum_m \lambda_m^{1/2} g_m(x) v_m = R^* g(\cdot, x) \in \mathcal{U},$$

We can arrange $\mathcal{U} = \overline{\mathrm{span}}\, \mathrm{im}\, r = \overline{\mathrm{span}}\, \mathrm{im}\, p$.

Then $p(x)$ gives a representation over X:
define $\hat{R}^* : L_2(X, \nu) \to \mathcal{U}$

$$\hat{R}^* : L_2(X, \nu) \ni f \mapsto \hat{R}^* f := \int_Y p(x) f(x)\, \nu(\mathrm{d}x) \in \mathcal{U},$$

$$\Rightarrow C = \hat{R}^* \hat{R}$$

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 together with R^{-1}.
- The spectral decomposition over $\sigma(C)$ or via $VM_k^{1/2}$.
- The Karhunen-Loève expansion based on SVD via R^*.
- Other multiplicative decompositions, such as $C = B^*B$.
- The kernel decompositions and representation on $L_2(X, \nu)$.

Choice depends on what is wanted / needed.

Examples and interpretations

- If $\mathcal{V}$ is a space of centred RVs, r is a random field / stochastic process indexed by Ω, kernel $\varkappa(\omega_1, \omega_2)$ is covariance function.
- If in this case $\Omega = \mathbb{R}^d$ and moreover $\varkappa(\omega_1, \omega_2) = c(\omega_1 - \omega_2)$ (stationary process / homogeneous field), then diagonalisation V is real Fourier transform, typically $\sigma(C)_p = \emptyset \Rightarrow$ need Gelfand triplets.
- If μ is a probability measure ($\mu(\Omega) = 1$), and r is a centred $\mathcal{V}$-valued RV, then C is the covariance.
- If $\Omega = \{1, 2, \ldots, n\}$ and $\mathcal{R} = \mathbb{R}^n$, then $\varkappa$ is the Gram matrix of the vectors $r_1, \ldots, r_n$.
- If $\Omega = [0, T]$ and $r(\omega)$ is the response of a dynamical system, then R^* leads to proper orthogonal decomposition (POD).

Further factorisation

We have found representations in $\mathcal{U} \otimes \mathcal{S}$, where

$$\mathcal{S} = \mathcal{R}, L_2(\Omega), L_2(\sigma(C)), \bigoplus_n L_2(\mathbb{R}, \varrho_n), \ell_2, \mathcal{H}, L_2(X), \ldots$$

Combinations may occur, so that $\mathcal{S} = \mathcal{S}_I \otimes \mathcal{S}_{II} \otimes \mathcal{S}_{III} \otimes \ldots$

This was only a basic decomposition.

Often the problem allows $\mathcal{U} = \bigotimes_k \mathcal{U}_k$.

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

Synopsis of Bayesian inference

Unknown quantities are uncertain, modelled as random.
This can be considered as a model of our state of knowledge.

After some new information (an observation, a measurement),
our model has to be made consistent with the new information,
i.e. we are looking for conditional probabilities.

The idea is to change our present model by just so much —
as little as possible — so that it becomes consistent.

For this we have to predict — with our present knowledge / model —
the probability of all possible observations and
compare with the actual observation.

Setting for the identification process

General idea:

We observe / measure a system, whose structure we know in principle.
The system behaviour depends on some quantities (parameters),
which we do not know $\Rightarrow$ uncertainty.

We model (uncertainty in) our knowledge in a Bayesian setting:
as a probability distribution on the parameters.

We start with what we know a priori, then perform a measurement.
This gives new information, to update our knowledge (identification).

Update in probabilistic setting works with conditional probabilities
$\Rightarrow$ Bayes's theorem.

Repeated measurements lead to better identification.

Mathematical formulation I

Consider operator equation, physical system modelled by A:

$$A(u) = f \qquad u \in \mathcal{U},\ f \in \mathcal{F},$$

$$\Leftrightarrow \forall v \in \mathcal{U} : \qquad \langle A(u), v\rangle = \langle f, v\rangle,$$

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

Solution operator: $u = S(f)$, inverse of A.

Operator depends on parameters $q \in \mathcal{Q}$, hence state u is also function of q:

$$A(u; q) = f \quad \Rightarrow \quad u = S(f; q).$$

Measurement operator Y with values in $\mathcal{Y}$:

$$y = Y(q; u) = Y(q, S(f; q)).$$

Mathematical formulation II

For given f, measurement y is just a function of q.
This function is usually not invertible $\Rightarrow$ ill-posed problem,
measurement y does not contain enough information.

In Bayesian framework state of knowledge modelled in a probabilistic way,
parameters q are uncertain, and assumed as random.

Bayesian setting allows updating / sharpening of information
about q when measurement is performed.

The problem of updating distribution—state of knowledge of q
becomes well-posed.

Can be applied successively, each new measurement y and
forcing f —may also be uncertain—will provide new information.

Reminder of Bayes's theorem

Assume that A is an event where we want more information, and that B is a possible observation. If the conditional probability $\mathbb{P}(A|B) = \mathbb{P}(A)$, in other words $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, then A and B are independent, i.e. B contains no information regarding A. Otherwise $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. As also $\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A)$:

$$\Longrightarrow \qquad \mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)}{\mathbb{P}(B)}\mathbb{P}(A).$$

Sir Harold Jeffreys: Bayes's theorem "is to the theory of probability what Pythagoras's theorem is to geometry".

Model with uncertainties

For simplicity assume that $\mathcal{Q}$ is a Hilbert space, and $q(\omega)$ has finite variance — $\|q\|_{\mathcal{Q}} \in \mathcal{S} := L_2(\Omega)$, so that

$$q \in L_2(\Omega, \mathcal{Q}) \cong \mathcal{Q} \otimes L_2(\Omega) = \mathcal{Q} \otimes \mathcal{S} =: \mathscr{Q}.$$

System model is now

$$A(u(\omega); q(\omega)) = f(\omega) \qquad \text{a.s. in } \omega \in \Omega,$$

state $u = u(\omega)$ becomes $\mathcal{U}$-valued random variable (RV), element of a tensor space $\mathscr{U} = \mathcal{U} \otimes \mathcal{S}$.

As variational statement:

$$\forall v \in \mathscr{U} : \quad \mathbb{E}\left(\langle A(u(\cdot); q(\cdot)), v\rangle\right) = \mathbb{E}\left(\langle f(\cdot), v\rangle\right) =: \langle\langle f, v\rangle\rangle.$$

Leads to well-posed stochastic PDE (SPDE).

Measurement

With state $u \in \mathscr{U} = \mathcal{U} \otimes \mathcal{S}$ a RV, the quantity to be measured

$$z(\omega) = Y(q(\omega), u(\omega))) + \epsilon(\omega) \in \mathscr{Y} := \mathcal{Y} \otimes \mathcal{S}$$

is also uncertain—a random variable—plus a random error ϵ.
This is the predicted new measurement,
whereas the observation gives $\hat{y} \in \mathcal{Y}$.

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.

Kolmogorov starts from conditional expectation $\mathbb{E}\left(\cdot | M_z\right)$,
from this conditional probability via $\mathbb{P}(I_q | M_z) = \mathbb{E}\left(\chi_{I_q} | M_z\right)$.

Important points I

The probability measure $\mathbb{P}$ is not the object of desire.
It is the distribution of q, a measure on $\mathcal{Q}$—push forward of $\mathbb{P}$:

$$q_*\mathbb{P}(\mathcal{E}) := \mathbb{P}(q^{-1}(\mathcal{E})) \quad \text{for measurable} \quad \mathcal{E} \subseteq \mathcal{Q}.$$

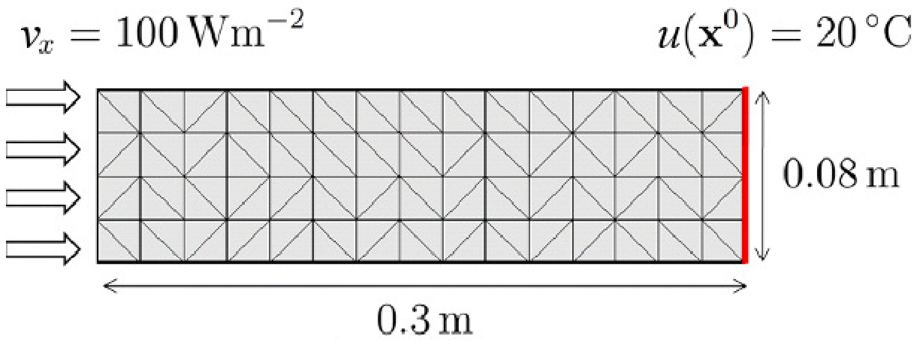
Bayes's original formula changes $\mathbb{P}$, leaves q as is.
Kolmogorov's conditional expectation changes q, leaves $\mathbb{P}$ as is.
In both cases the update is a new $q_*\mathbb{P}$.

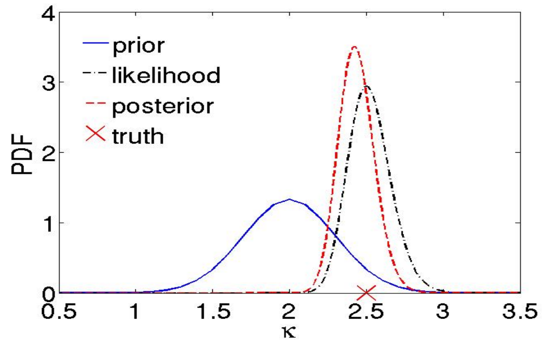
$\mathbb{P}$ (a probability measure) is on positive part of unit sphere,
whereas q is free in a vector space.

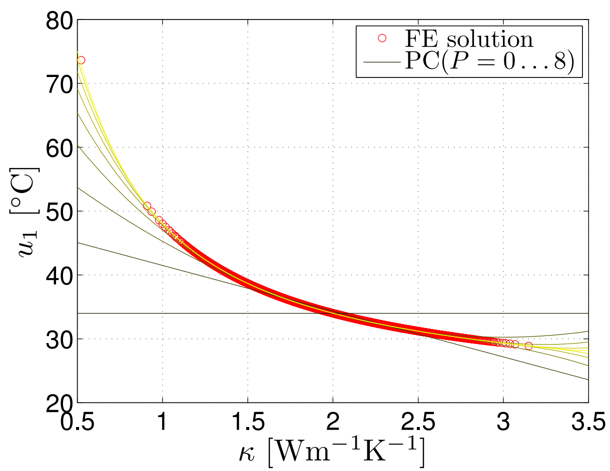
This will allow the use of (multi-)linear algebra
and tensor approximations.

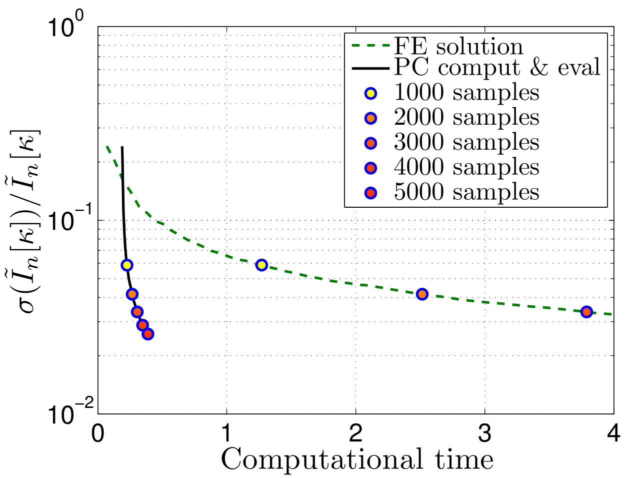
Example A — linear heat flow (MCMC)

Constant unknown conductivity, solved by 100 000 Markov chain Monte Carlo (MCMC) samples.







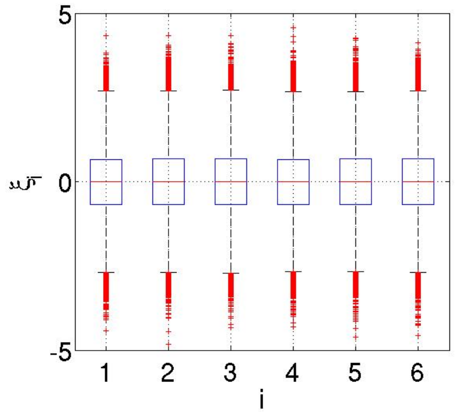


Comparison proxy model with pure FE.

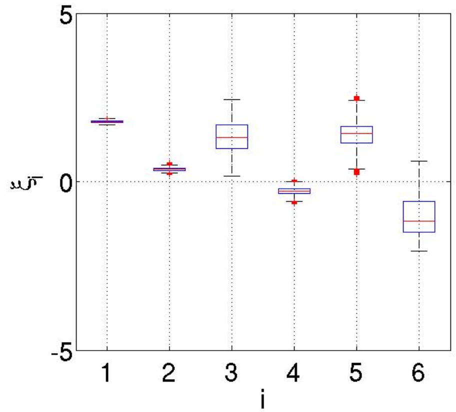
Example B — non-linear heat flow (MCMC)

Conductivity as random field, 1000 MCMC samples.

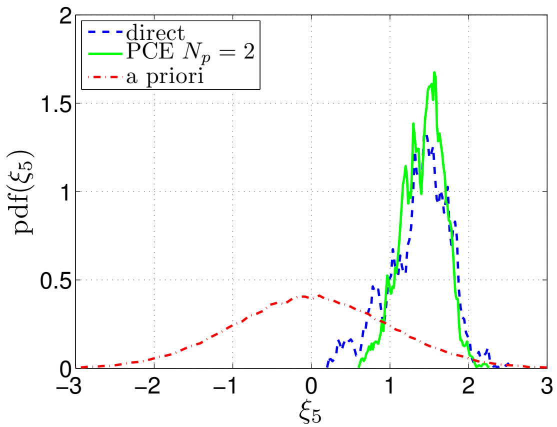
initial



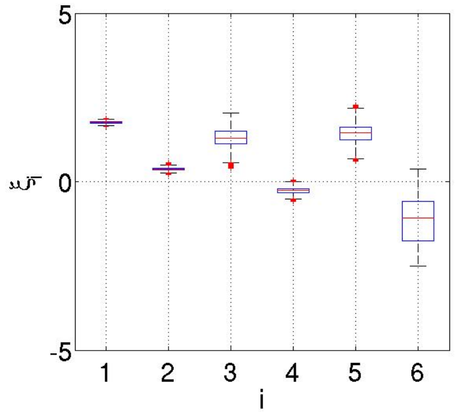
final MCMC-FE



pdf



final MCMC-PCE



Update

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

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

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

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

Reduction of variance—Pythagoras: $\|q\|^2_{L_2} = \|q - q_a\|^2_{L_2} + \|q_a\|^2_{L_2}$

Doob-Dynkin: $\mathscr{Q}_\infty = \{\varphi \in \mathscr{Q} \ : \ \varphi = \phi \circ z, \phi \text{ measurable } \}$

Important points II

Identification process:

- Use forward problem $A(u(\omega); q(\omega)) = f(\omega)$ to forecast new state $u_f(\omega)$ and measurement $z(\omega) = Y(q(\omega), u_f(\omega))) + \epsilon(\omega)$.
- Perform minimisation of loss function to obtain update map / filter.
- Use innovation in inverse problem from measurement $\hat{y}$ to update forecast q_f to obtain assimilated (updated) q_a with update map.
- All operations in vector space, use tensor approximations throughout.

Case with Prior Information

Here we have a prior estimate $q_f(\omega)$ (forecast)
obtained by minimising over $\mathscr{Q}_f$
and measurements z generating as before via Y a subspace $\mathscr{Q}_\infty \subset \mathscr{Q}$.

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

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 $q_i = P_{\mathscr{Q}_i} q$?

Approximation

Minimising loss $\varPhi$ equivalent to orthogonality: find $\phi \in L_0(\mathcal{Y}, \mathcal{Q})$

$$\forall v \in \mathscr{Q}_\infty : \quad \langle\!\langle \mathrm{D}_{q_a}\varPhi(q_a(\phi)), v \rangle\!\rangle_{L_2} = \langle\!\langle q - q_a, v \rangle\!\rangle_{L_2} = 0,$$

$$\Leftrightarrow \mathrm{D}_\phi \varPhi := \mathrm{D}_{q_a}\varPhi \circ \mathrm{D}_\phi q_a = 0 \text{ with } q_a(\phi) := \phi(z).$$

Approximation of $\mathscr{Q}_\infty$: take $\mathscr{Q}_n \subset \mathscr{Q}_\infty$

$$\mathscr{Q}_n := \{\varphi \in \mathscr{Q} \ : \ \varphi = \psi_n \circ z, \ \psi_n \text{ a } n^{\text{th}} \text{ degree polynomial}\}$$

i.e. $\varphi = {}^0H + {}^1H\, z + \cdots + {}^kH\, z^{\vee k} + \cdots + {}^nH\, z^{\vee n}$,

where ${}^kH \in \mathscr{L}_s^k(\mathcal{Y}, \mathcal{Q})$ is symmetric and k-linear; $z^{\vee k} := \mathrm{Sym}(z^{\otimes k})$.

With $q_a(\phi) = q_a(({}^0H, \dots, {}^kH, \dots, {}^nH)) = \sum_{k=0}^n {}^kH\, z^{\vee k} = P_{\mathscr{Q}_n} q$,
orthogonality implies

$$\forall \ell = 0, \dots, n : \quad \mathrm{D}_{({}^\ell H)}\varPhi(q_a({}^0H, \dots, {}^kH, \dots, {}^nH)) = 0$$

Determining the n-th degree Bayesian update

Theorem: For each $n \geq 0$, with the abbreviations

$$\langle p \otimes v^{\vee k} \rangle := \mathbb{E}\left(p \otimes v^{\vee k}\right) = \int_{\Omega} p(\omega) \otimes v(\omega)^{\vee k}\, \mathbb{P}(\mathrm{d}\omega),$$

and ${}^{k}H\langle z^{\vee(\ell+k)}\rangle := \langle z^{\vee \ell} \otimes ({}^{k}H\, z^{\vee k})\rangle = \mathbb{E}\left(z^{\vee \ell} \otimes ({}^{k}H\, z^{\vee k})\right)$,

we have for the unknowns $({}^{0}H, \dots, {}^{k}H, \dots, {}^{n}H)$

$$\begin{array}{llllll}
\ell = 0: & {}^{0}H & \cdots + {}^{k}H\langle z^{\vee k}\rangle & \cdots + {}^{n}H\langle z^{\vee n}\rangle & = & \langle q\rangle, \\
\ell = 1: & {}^{0}H\langle z\rangle & \cdots + {}^{k}H\langle z^{\vee(1+k)}\rangle & \cdots + {}^{n}H\langle z^{\vee(1+n)}\rangle & = & \langle q \otimes z\rangle, \\
\vdots & & \cdots & \vdots & & \vdots \\
\ell = n: & {}^{0}H\langle z^{\vee n}\rangle & \cdots + {}^{k}H\langle z^{\vee(n+k)}\rangle & \cdots + {}^{n}H\langle z^{\vee(2n)}\rangle & = & \langle q \otimes z^{\vee n}\rangle
\end{array}$$

a linear system with symmetric positive definite Hankel operator matrix $(\langle z^{\vee(\ell+k)}\rangle)_{\ell,k}$.

Bayesian update in components

For short $\forall \ell = 0, \ldots, n$:

$$\sum_{k=0}^{n} {}^{k}H \langle z^{\vee(\ell+k)} \rangle = \langle q \otimes z^{\vee \ell} \rangle,$$

For finite dimensional spaces, or after discretisation,
in components (or à la Penrose in 'symbolic index' notation):
let $q = (q^m)$, $z = (z^\jmath)$, and ${}^kH = ({}^kH^m_{\jmath_1 \cdots \jmath_k})$, then:

$$\forall \ell = 0, \ldots, n; \; \jmath_1 \leq \ldots \leq \jmath_\ell \leq \ldots \leq \jmath_{\ell+k} \leq \ldots \leq \jmath_{\ell+n}$$

$$\langle z^{\jmath_1} \cdots z^{\jmath_\ell} \rangle \left({}^0H^m\right) + \cdots + \langle z^{\jmath_1} \cdots z^{\jmath_{\ell+1}} \cdots z^{\jmath_{\ell+k}} \rangle \left({}^kH^m_{\jmath_{\ell+1} \cdots \jmath_{\ell+k}}\right) +$$

$$\cdots + \langle z^{\jmath_1} \cdots z^{\jmath_{\ell+1}} \cdots z^{\jmath_{\ell+n}} \rangle \left({}^nH^m_{\jmath_{\ell+1} \cdots \jmath_{\ell+n}}\right) = \langle q^m z^{\jmath_1} \cdots z^{\jmath_\ell} \rangle.$$

(Einstein summation convention used)
matrix does not depend on m—it is identically block diagonal.

Special cases

For $n = 0$ (constant functions) $\Rightarrow$ $q_a = {}^0H = \langle q \rangle \quad (= \mathbb{E}(q))$.

For $n = 1$ the approximation is with affine functions

$$\begin{aligned} {}^0H \quad + {}^1H\langle z \rangle \quad &= \langle q \rangle \\ {}^0H\langle z \rangle + {}^1H\langle z \otimes z \rangle &= \langle q \otimes z \rangle \end{aligned}$$

$\Longrightarrow$ (remember that $[\mathrm{cov}_{q,z}] = \langle q \otimes z \rangle - \langle q \rangle \otimes \langle z \rangle$)

$$\begin{aligned} {}^0H = \quad & \langle q \rangle - {}^1H\langle z \rangle \\ {}^1H(\langle z \otimes z \rangle - \langle z \rangle \otimes \langle z \rangle) = & \langle q \otimes z \rangle - \langle q \rangle \otimes \langle z \rangle \end{aligned}$$

$\Rightarrow {}^1H = [\mathrm{cov}_{q,z}][\mathrm{cov}_{z,z}]^{-1}$ (Kalman's solution),
${}^0H = \langle q \rangle - [\mathrm{cov}_{q,z}][\mathrm{cov}_{z,z}]^{-1}\langle z \rangle$,
and finally

$$q_a = {}^0H + {}^1H\,z = \langle q \rangle + [\mathrm{cov}_{q,z}][\mathrm{cov}_{z,z}]^{-1}(z - \langle z \rangle).$$

Simplification $n = 1$

The case $n = 1$—linear functions, projecting onto $\mathscr{Q}_1$—is well known:

this is the linear minimum variance estimate $\hat{q}_a$.

Theorem: (Generalisation of Gauss-Markov)

$$^1q_a(\omega) = q_f(\omega) + K(\hat{y} - z(\omega)),$$

where the linear Kalman gain operator $K := {}^1H : \mathscr{Y} \to \mathscr{Q}$ is

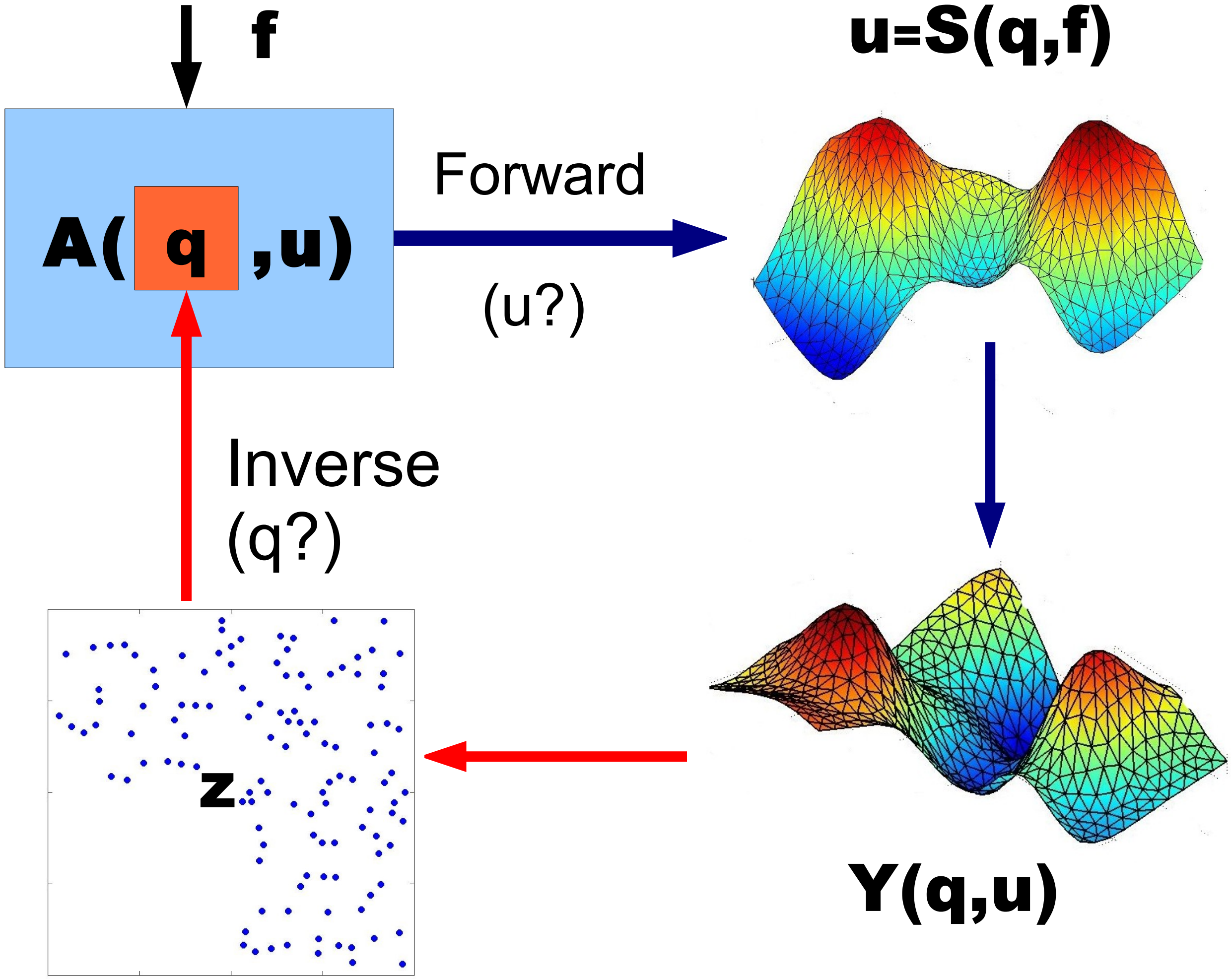
$$K := [\mathrm{cov}_{q_f,y}]\big([\mathrm{cov}_{y,y}] + [\mathrm{cov}_{\epsilon,\epsilon}]\big)^{-1} \text{ and } z(\omega) = Y(q_f(\omega)) + \epsilon(\omega).$$

Or in tensor space $q \in \mathscr{Q} = \mathcal{Q} \otimes \mathcal{S}$: $^1q_a = q_f + (K \otimes I)(\hat{y} - z)$

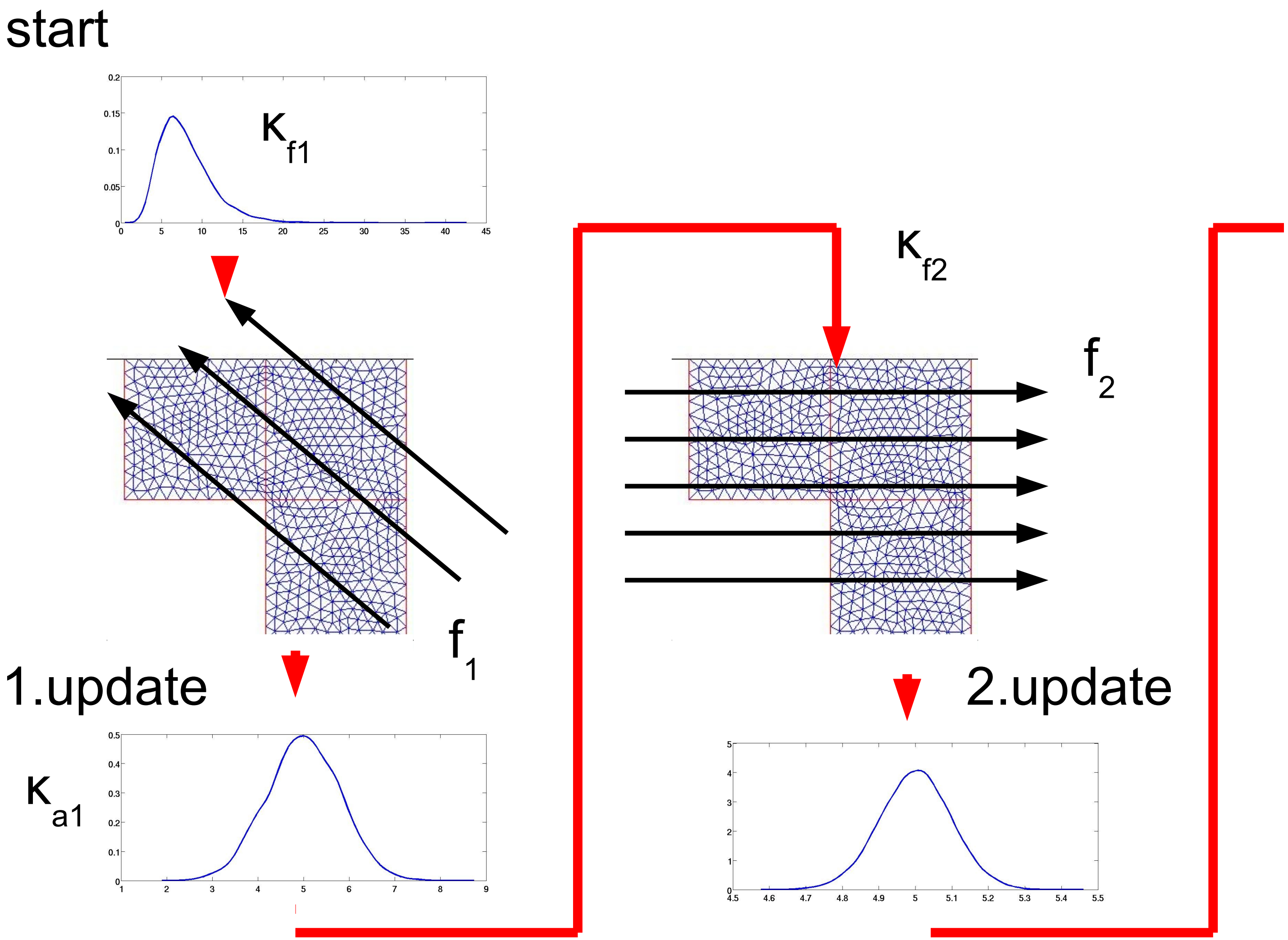
Classical Kalman filter is low order part of this update.

e.g. $[\mathrm{cov}_{q_a,q_a}] = [\mathrm{cov}_{q_f,q_f}] - K[\mathrm{cov}_{q_f,y}]^T$

Schematic representation



Sequential updating



Computational issues

For linear systems and Gaussian noise $\Rightarrow$ analytical Kalman filter.

Otherwise Monte Carlo simulation (MCS) for forward problem, Markov chains (MCMC) or particle filters for update via measures.

Or forward problem via MCS, theorem (Kalman) on MCS ensemble, covariances from ensemble $\Rightarrow$ ensemble Kalman (EnKF) filter.

Here: forward problem with stochastic Galerkin / projection / collocation, update by projection of theorem on stochastic Galerkin basis.

Two ingredients are needed:

1. a forward solver, to predict $z(\omega)$,
2. a way to evaluate and apply the update / Kalman gain.

Discretisation

Spatial and temporal discretisation of forward problem leads to:

$$\boldsymbol{A}(\boldsymbol{u}(\omega);\boldsymbol{q}(\omega)) = \boldsymbol{f}(\omega) \text{ and } \boldsymbol{z}(\omega) = \boldsymbol{Y}(\boldsymbol{q}_f(\omega), \boldsymbol{S}(\boldsymbol{f}(\omega), \boldsymbol{q}_f(\omega))) + \boldsymbol{\epsilon}(\omega),$$

where e.g. $\boldsymbol{u}(\omega) \in \mathcal{U}_h \subset \mathcal{U}$ (semi-discrete problem).

Update on discretisation: ${}^1\boldsymbol{q}_a(\omega) = \boldsymbol{q}_f(\omega) + \boldsymbol{K}(\hat{\boldsymbol{y}} - \boldsymbol{z}(\omega))$,
with Kalman matrix $\boldsymbol{K} = \mathrm{cov}(\boldsymbol{q}_f, \boldsymbol{y})\big(\mathrm{cov}(\boldsymbol{y},\boldsymbol{y}) + \mathrm{cov}(\boldsymbol{\epsilon},\boldsymbol{\epsilon})\big)^{-1}$
In tensor product $\mathcal{Q}_h \otimes \mathcal{S}$ the Kalman operator is $\boldsymbol{K} \otimes I$.

Stochastic discretisation $\mathcal{S}_k \subset \mathcal{S}$ with Galerkin projector $\Pi : \mathcal{S} \to \mathcal{S}_k$ via "spectral stochastic" ansatz (Wiener's polynomial chaos expansion—PCE) with Hermite polynomials $H_\alpha(\omega) := H_\alpha(\boldsymbol{\theta}(\omega))$:
$\boldsymbol{u}(\omega) = \sum_{\alpha\in\mathcal{J}} \boldsymbol{u}^\alpha H_\alpha(\omega)$ and similarly for $\boldsymbol{q}(\omega)$, $\boldsymbol{y}(\omega)$, and $\boldsymbol{z}(\omega)$.

$\boldsymbol{K}$ computed analytically, e.g. $\mathrm{cov}(\boldsymbol{q}_f, \boldsymbol{y}) = \sum_{\alpha>0} \alpha!\, \boldsymbol{q}_f^\alpha (\boldsymbol{y}^\alpha)^T$.

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

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

Update on $\mathcal{Q}_h \otimes \mathcal{S}_k$: $\quad {}^1\mathbf{q}_a = \mathbf{q}_f + (\boldsymbol{K} \otimes \boldsymbol{I})\big(\hat{\mathbf{y}} - \mathbf{z}\big).$

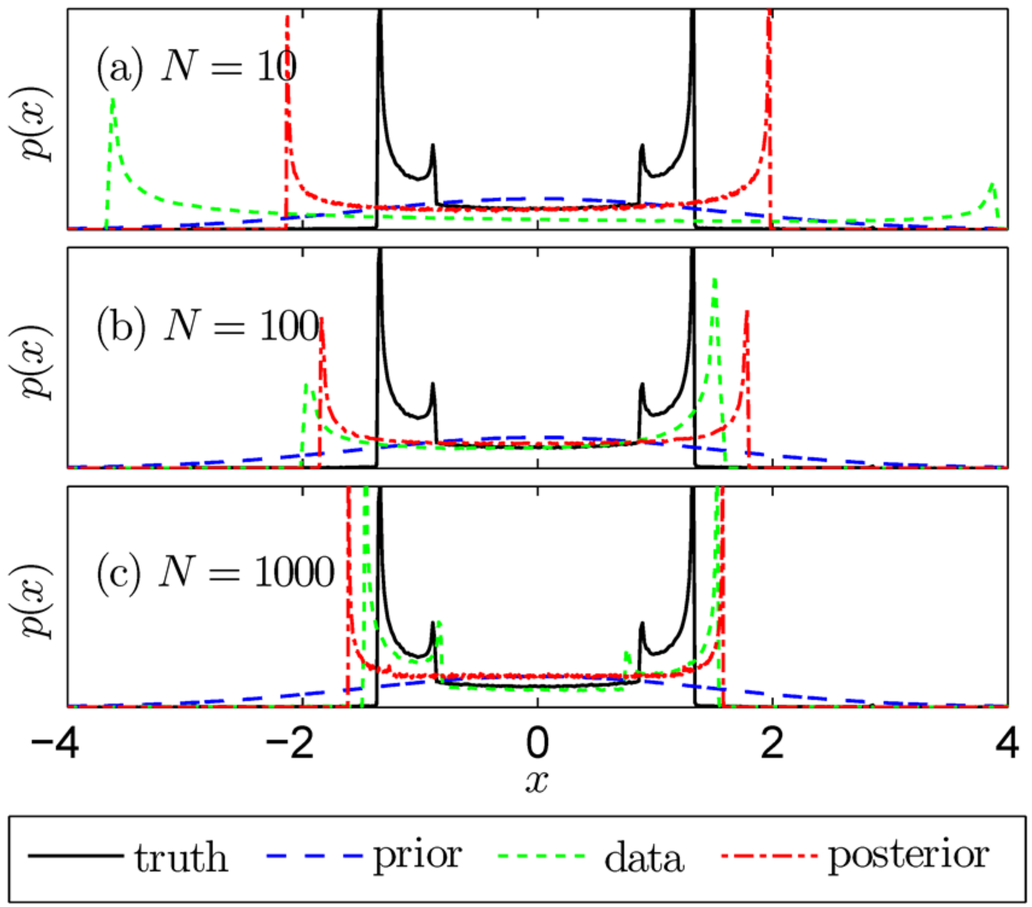
Forward problem and update benefit from low-rank /sparse approximation, e.g. $\mathbf{q} \approx \sum_j \boldsymbol{p}_j \otimes \boldsymbol{s}_j$.
Further tensor factorisation $\mathcal{Q}_h \otimes \mathcal{S}_k = \mathcal{Q}_h \otimes (\bigotimes_m \mathcal{S}_{k,m})$—another story.

Example 1: multi-modal distribution

Setup: Scalar RV x with 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

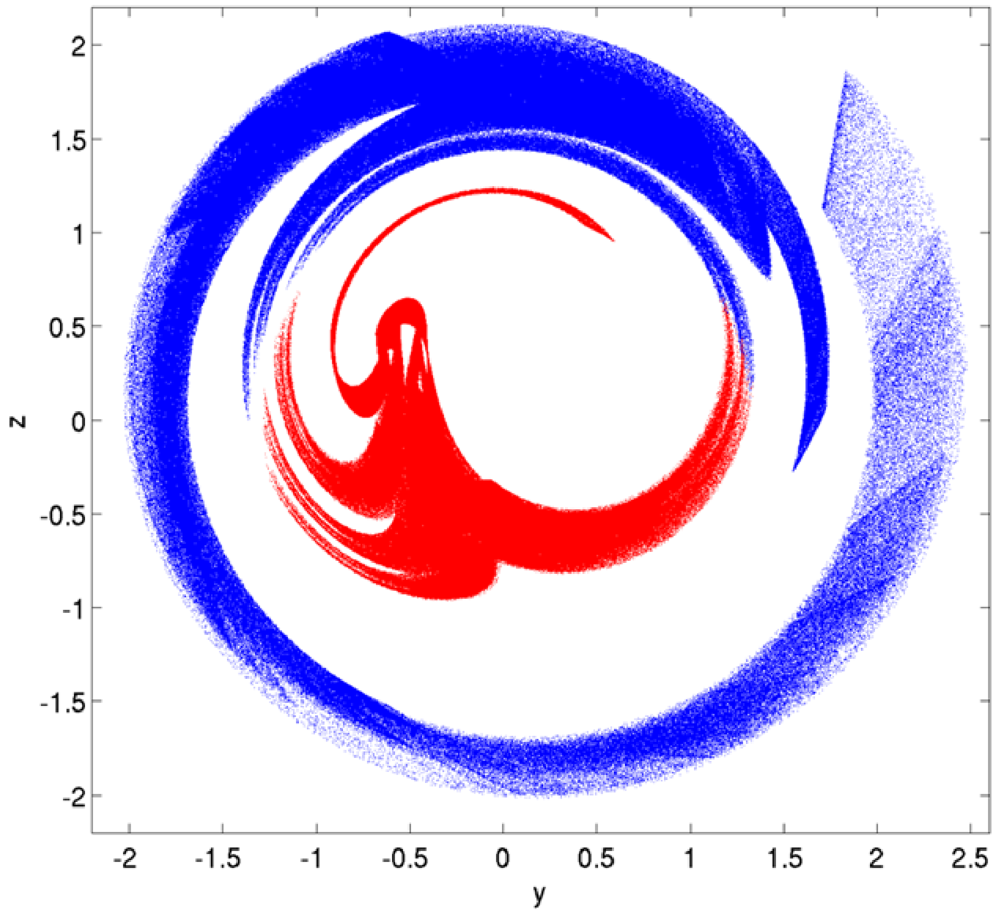
Setup: Non-linear, chaotic system

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

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.

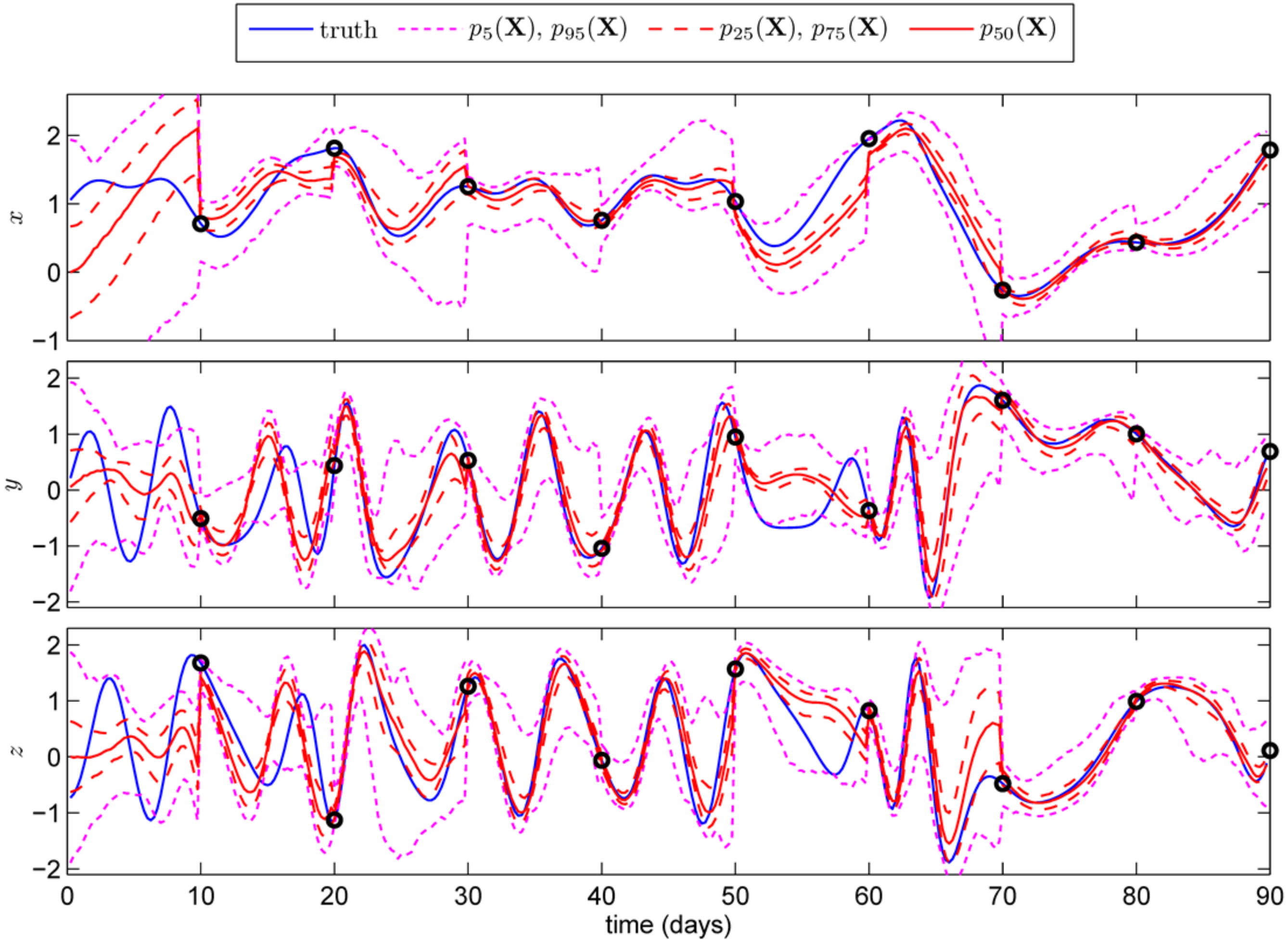


Poincaré cut for $x = 1$.

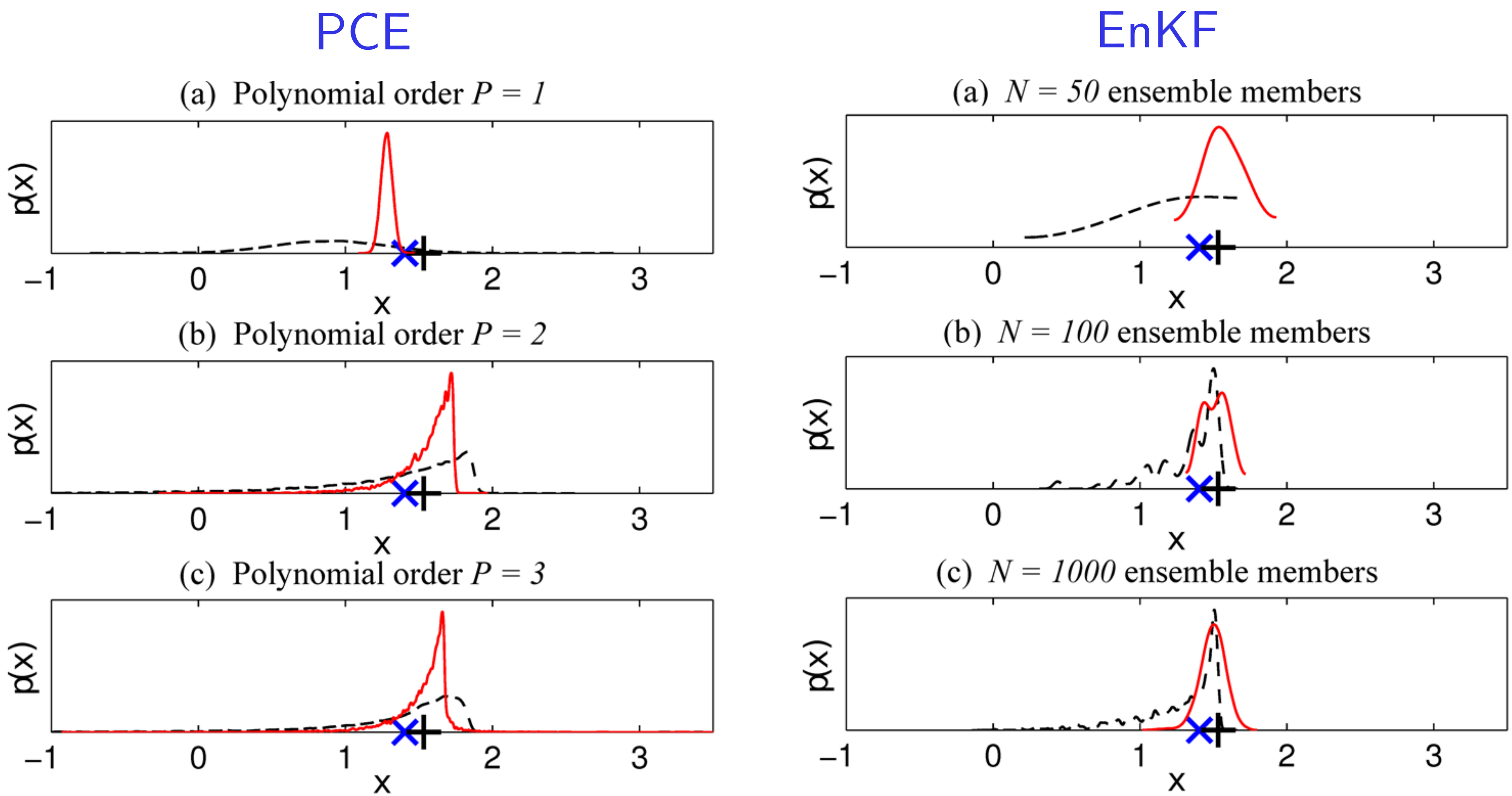
Example 2: Lorenz-84 PCE representation

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

Skewed structure clearly visible, preserved by updates.



Example 2: Lorenz-84 non-Gaussian identification



truth × **measurement +** posterior *prior*

Example 3: Diffusion

Model example diffusion with unknown diffusion coefficient,

$$A(u) = -\nabla_x \cdot (\kappa(x,\omega)\nabla_x u(x,\omega)) = f(x,\omega).$$

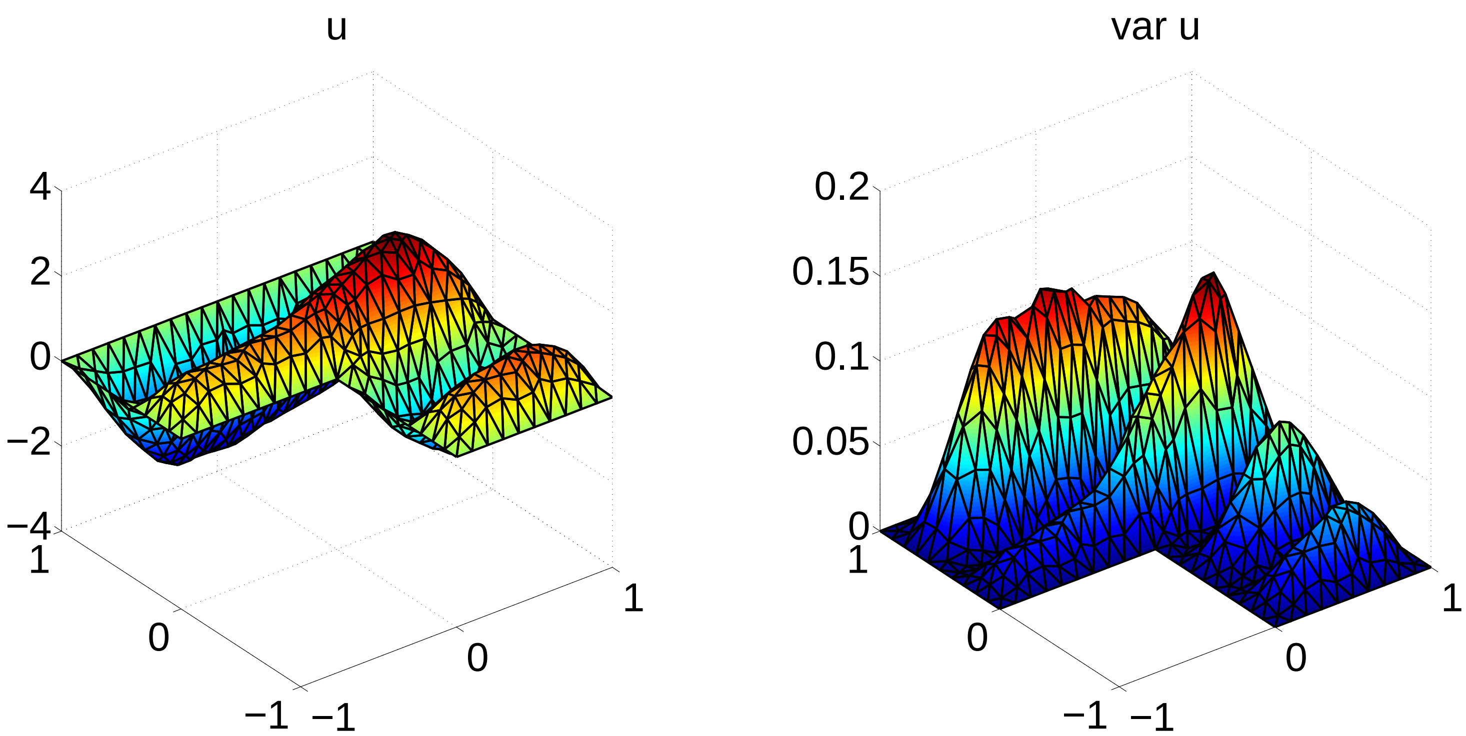
Fully discrete form of forward problem:

$$\mathbf{A}(\mathbf{u}) = \left(\sum_j \boldsymbol{A}_j \otimes \mathit{\Delta}^j \right) \mathbf{u} = \mathbf{f}.$$

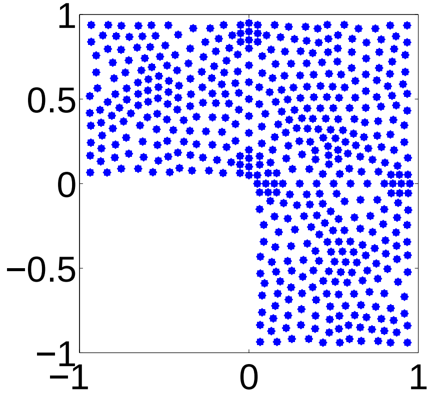
The unknown parameter is $q = \log \kappa$, as $\kappa > 0$, and hence is not free (is in a cone) in a vector space.

The measurement $\mathbf{y} = \mathbf{Y}(\mathbf{q}, \mathbf{u})$ is local averaging around some points.

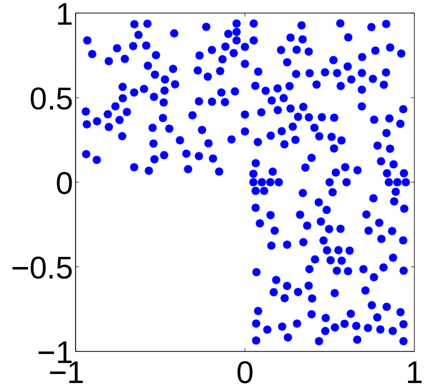
Example forward solution



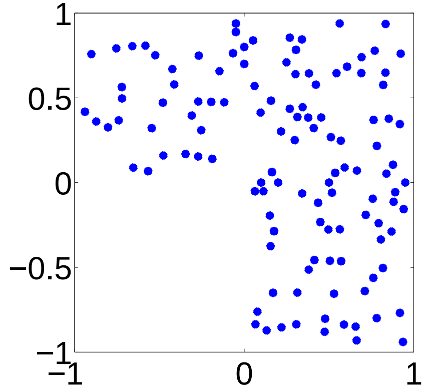
Measurement patches



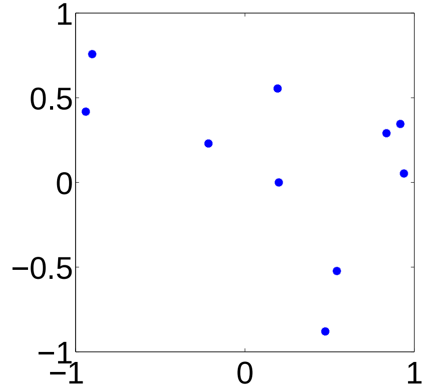
447 measurement patches



239 measurement patches



120 measurement patches



10 measurement patches

Convergence of update

Different truths:

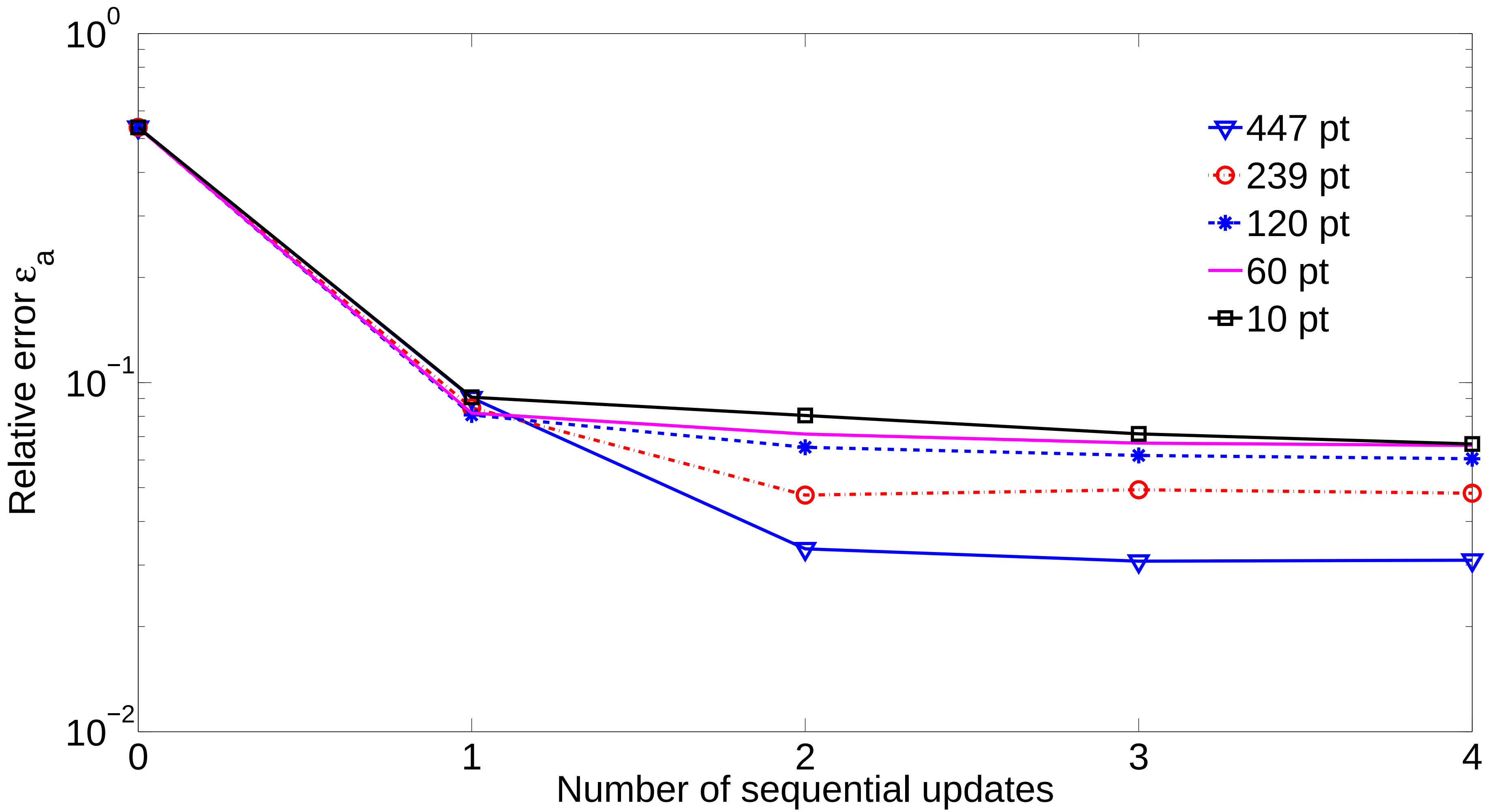
$$\kappa_t = 2, \quad \kappa_t = 2 + 0.3(x+y), \quad \kappa_t = 2.2 - 0.1(x^2+y^2).$$

Experiment	# patches	ϵ_p	1st	2nd	3rd	4th
1	447	0.45	0.08	0.04	0.03	0.03
2	239	0.45	0.08	0.05	0.05	0.04
3	120	0.45	0.07	0.06	0.05	0.05
4	10	0.45	0.13	0.08	0.07	0.07

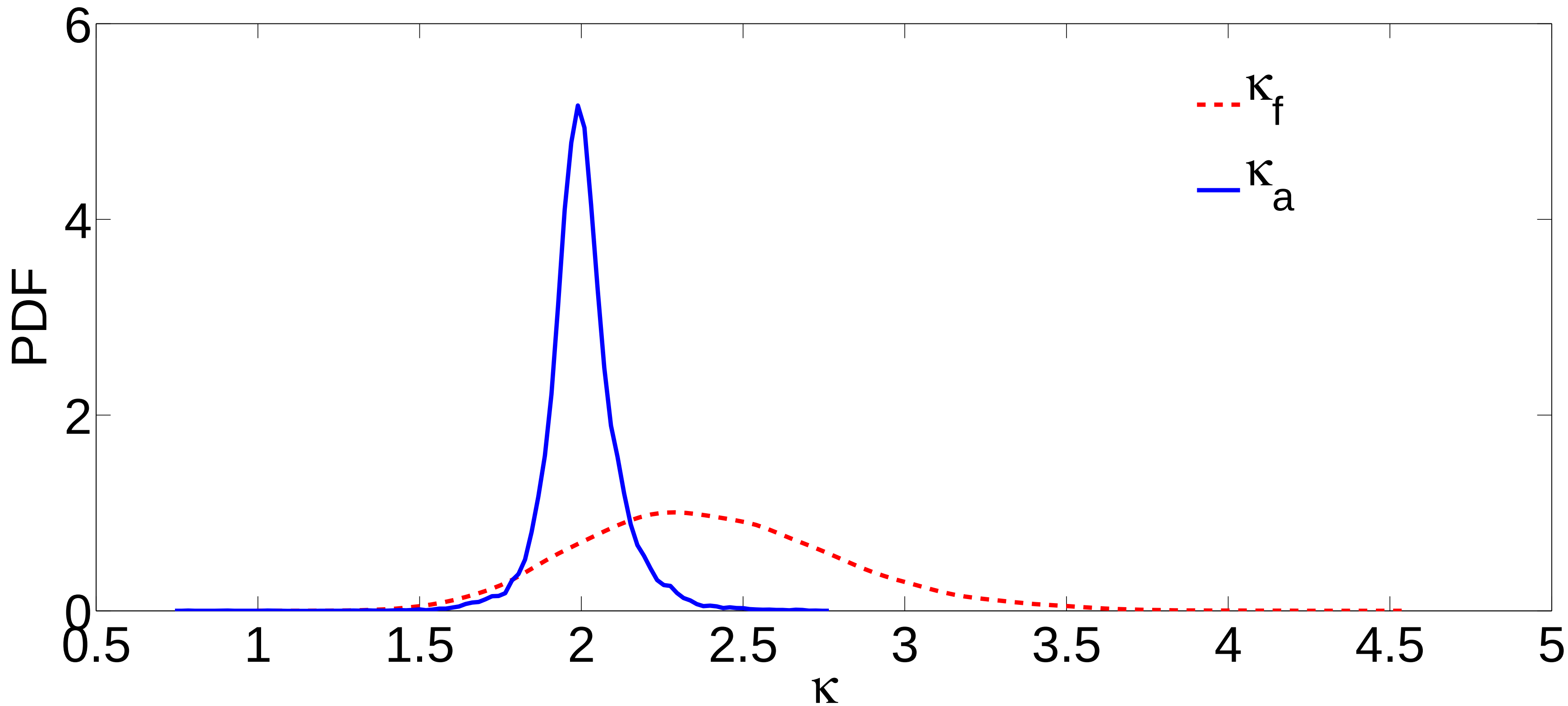
"Constant truth": Decay of relative error ϵ_a in each experiment.

Definition of error : $$\epsilon_a = \frac{\|\kappa_a - \kappa_t\|_{L_2}}{\|\kappa_t\|_{L_2}}.$$

Convergence plot of updates



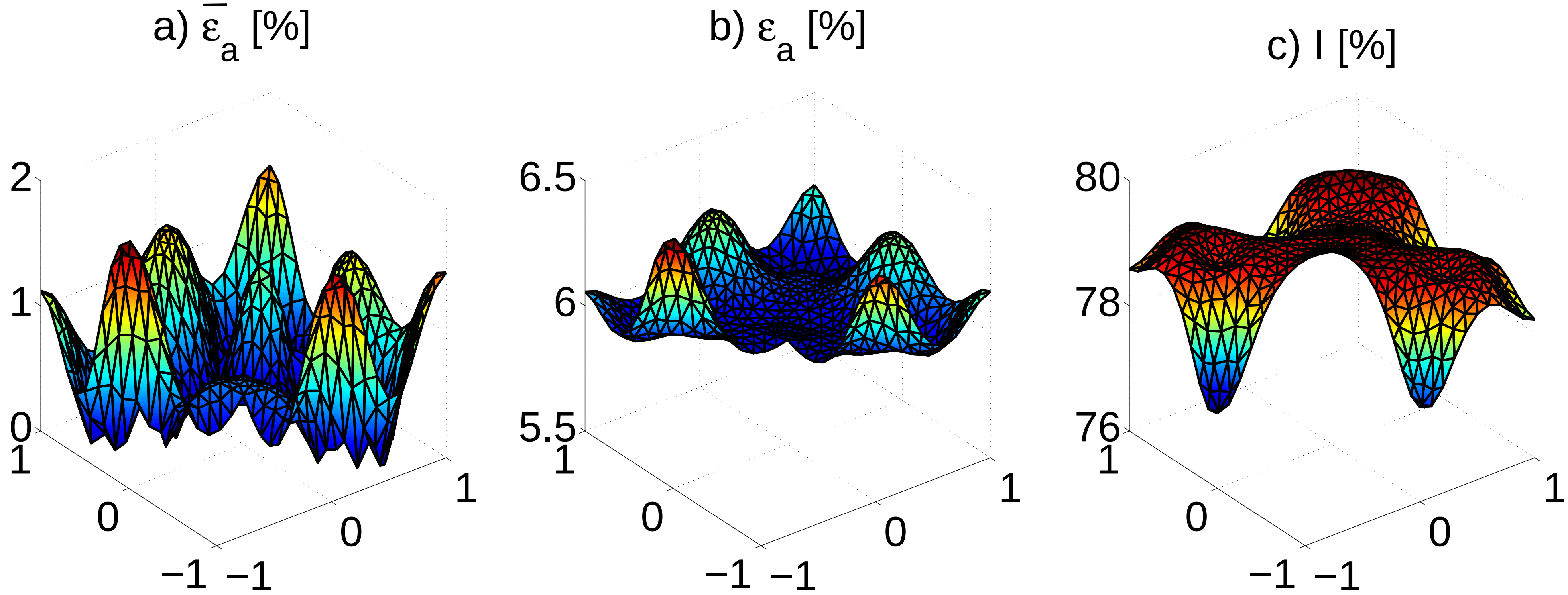
Forecast and Assimilated pdfs



Forecast and assimilated probability density functions (pdfs) for κ at a point where $\kappa_t = 2$.

Computations with constant, linear, quadratic, random draw "truth".

Accuracy constant truth



Elasto-plastic body with uncertainty

Let $u = (v, \varepsilon_p, \nu) \in \mathscr{H} = \mathscr{U} \times \mathscr{P} \times \mathscr{N}$ be the state variable (also random variables) of an elasto-plastic body, $a(\cdot, \cdot)$ the stored-energy bilinear form, $\mathscr{K}$ the elastic domain. Then find $u \in H^1([0,T], \mathscr{H})$ and $u^* \in H^1([0,T], \mathscr{H}^*)$ such that

$$\forall z \in \mathscr{H} : \quad a(u(t), z) + \langle\!\langle \dot{u}(t), z \rangle\!\rangle = \langle\!\langle f(t), z \rangle\!\rangle,$$

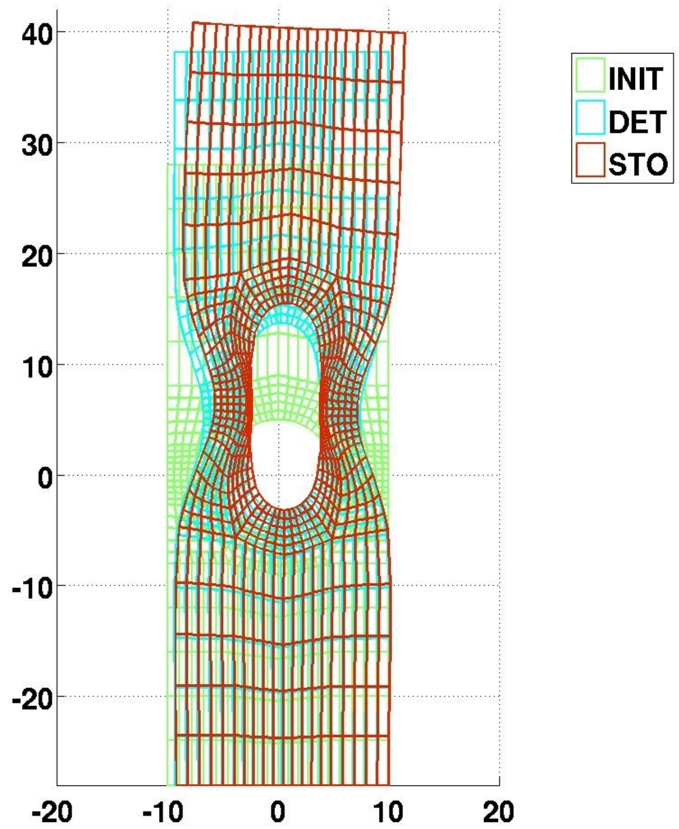
$$\forall z^* \in \mathscr{K} : \quad \langle\!\langle \dot{u}(t), z^* - u^*(t) \rangle\!\rangle \leq 0.$$

Spatial and stochastic discretisation leads to: find $\mathbf{u}(t) = (\mathbf{v}(t), \boldsymbol{\varepsilon}_p(t), \boldsymbol{\nu}(t))$ and $\mathbf{u}^*(t)$ such that

$$\mathbf{A}\mathbf{u}(t) + \dot{\mathbf{u}}(t) = \mathbf{f}(t),$$

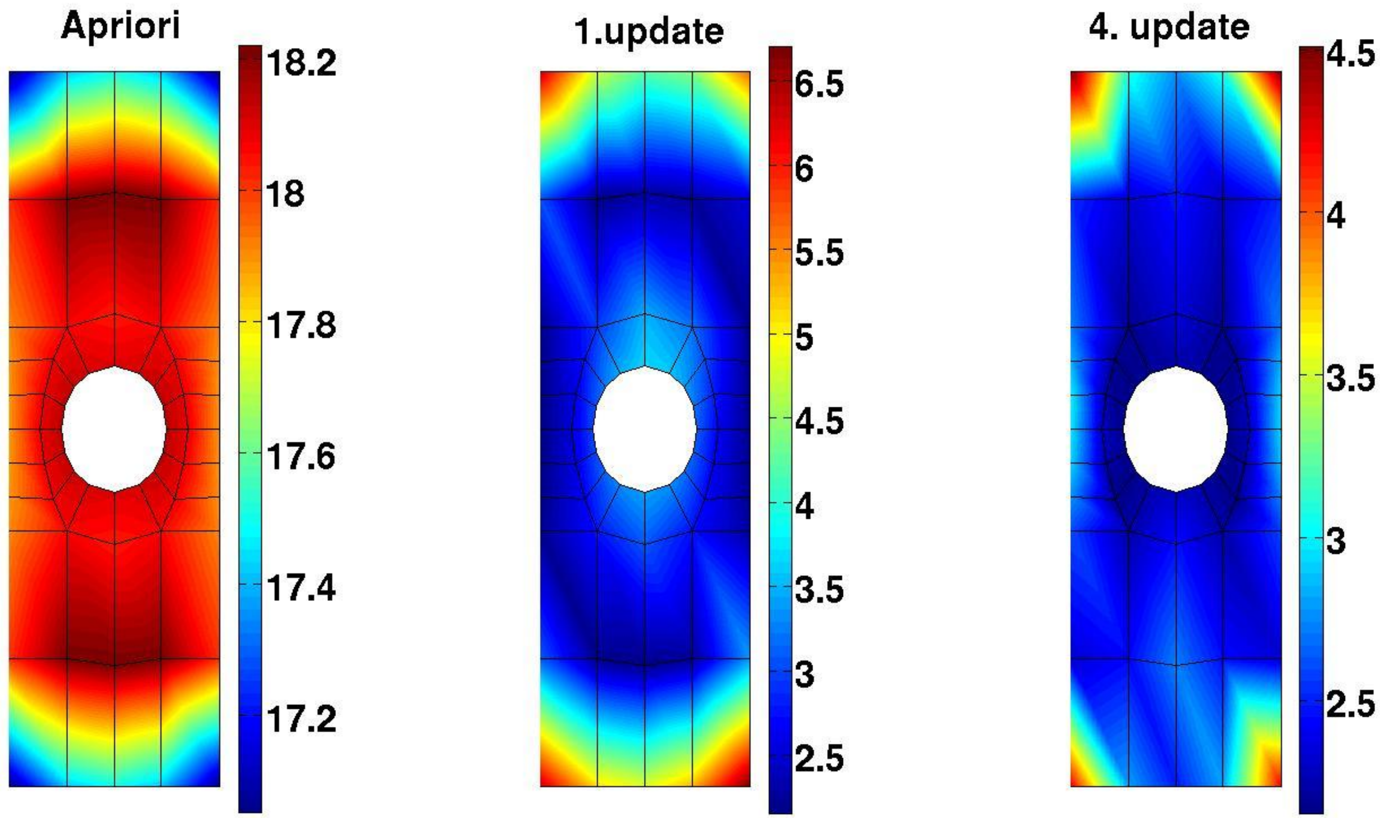
$$\forall \mathbf{z}^* \in \mathscr{K}_{hk} : \quad \langle\!\langle \dot{\mathbf{u}}(t), \mathbf{z}^* - \mathbf{u}^*(t) \rangle\!\rangle \leq 0.$$

Example: plate with hole



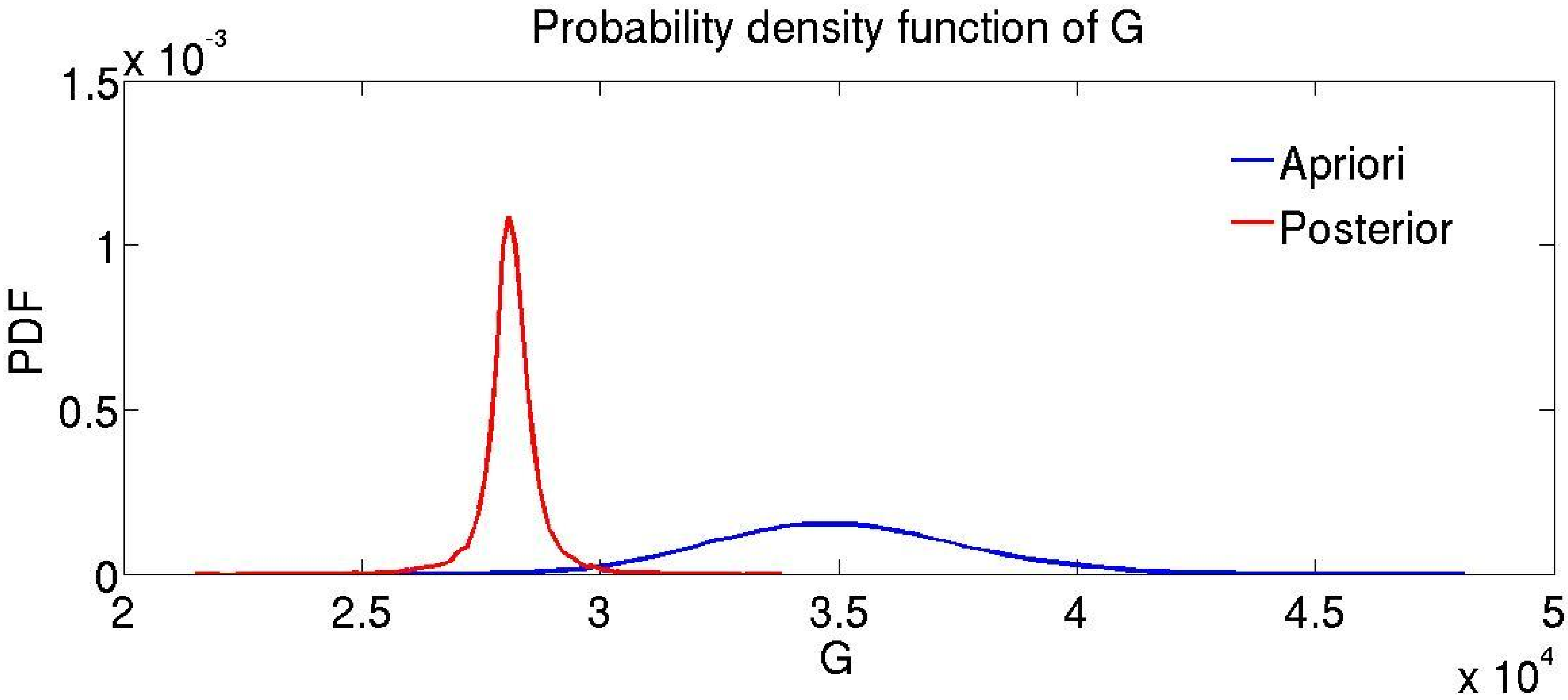
Forward problem: the comparison of the mean values of the total displacement for deterministic, initial and stochastic configuration

Relative variance of shear modulus estimate



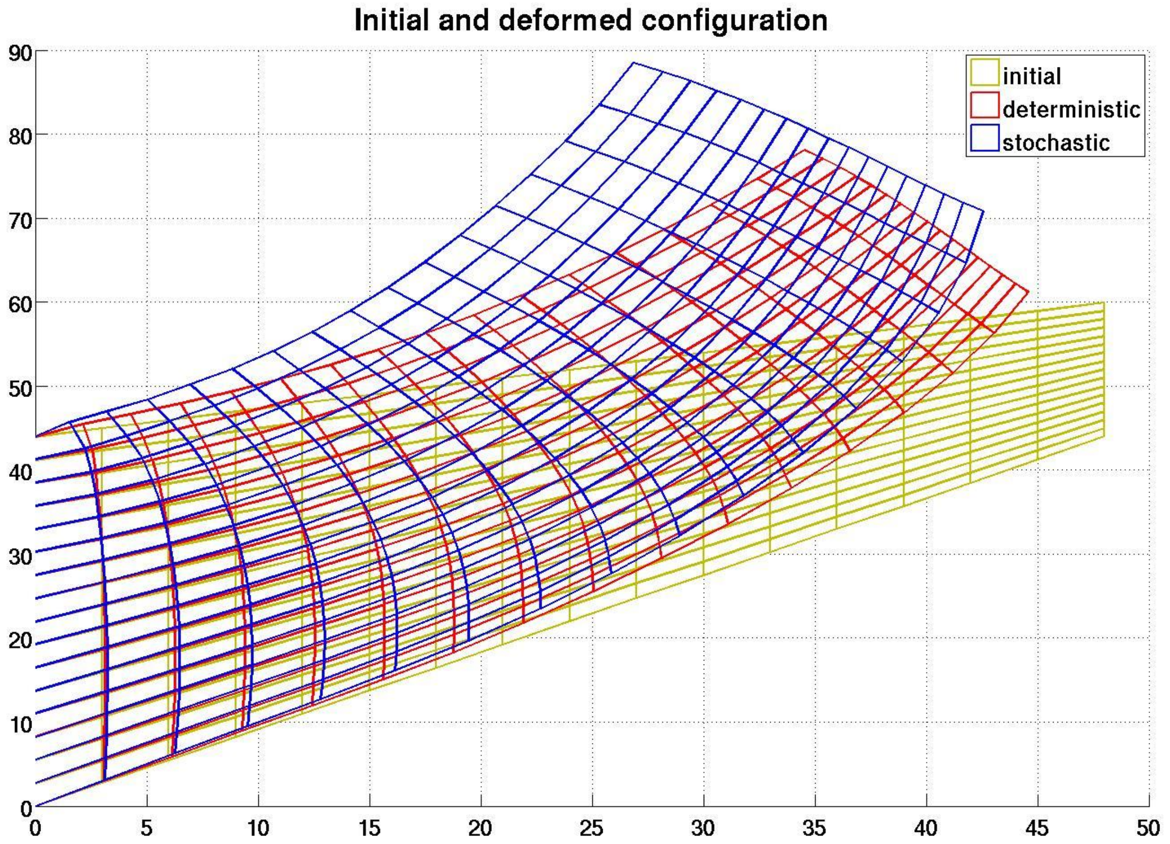
Relative RMSE of variance [%] after 4th update in 10% equally distributed measurment points

Probability density shear modulus



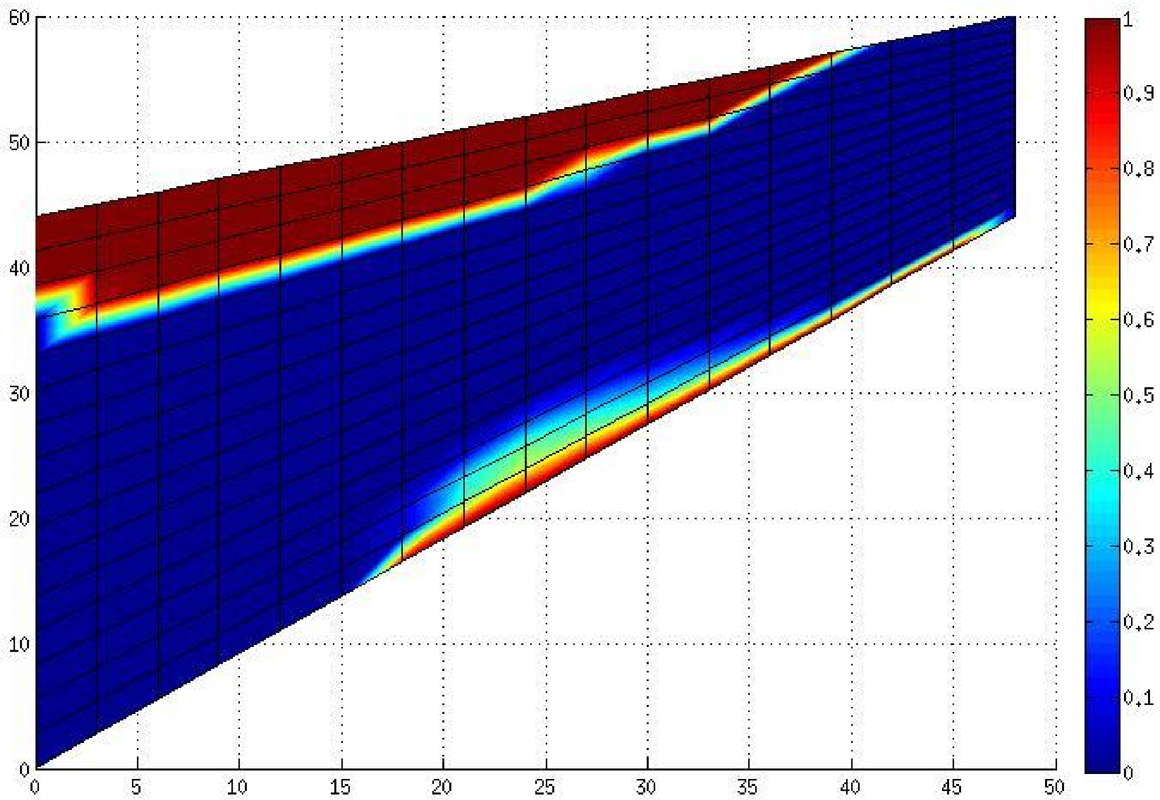
Comparison of prior and posterior distribution

Cook's membrane



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

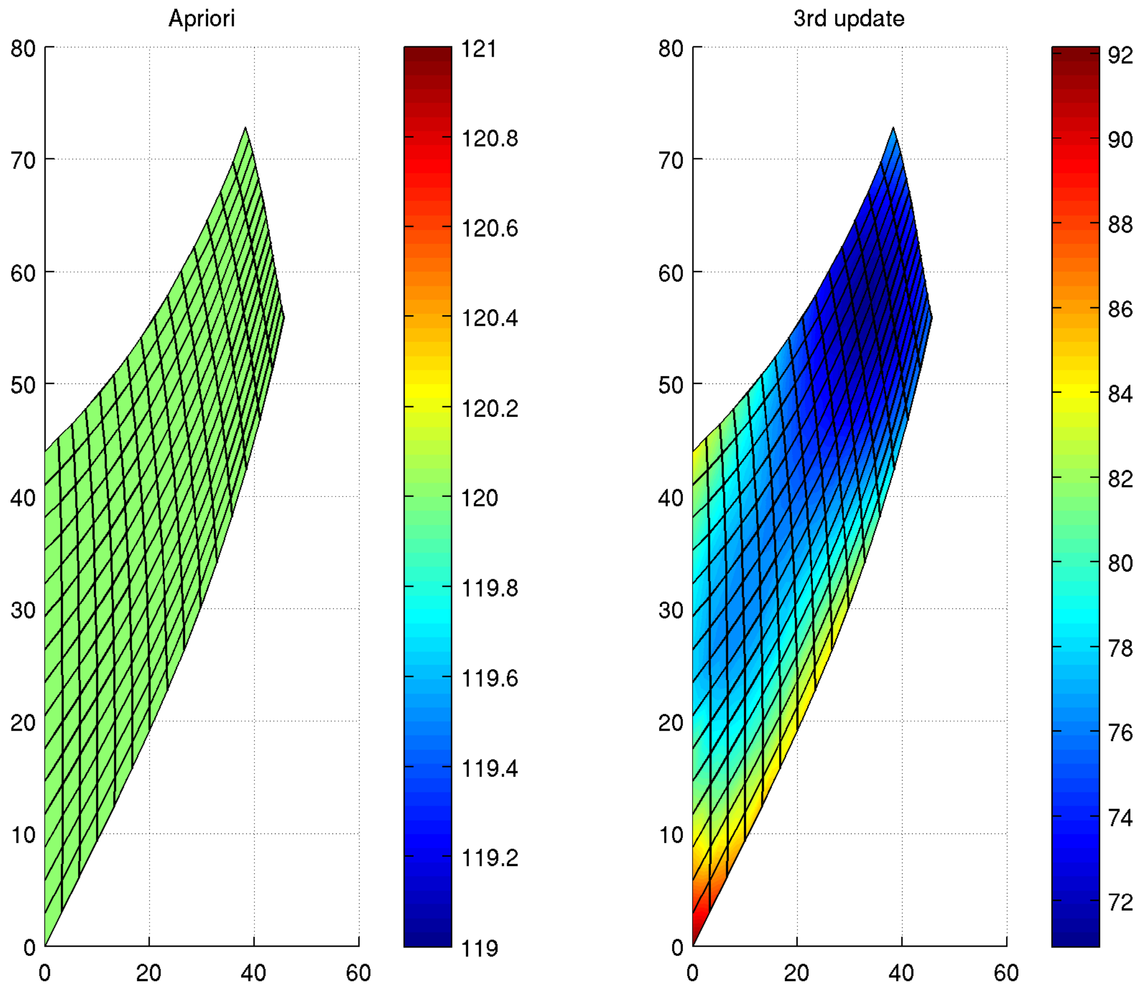
Exceedence probability



Forward problem: probability exceedance for shear stress under criteria

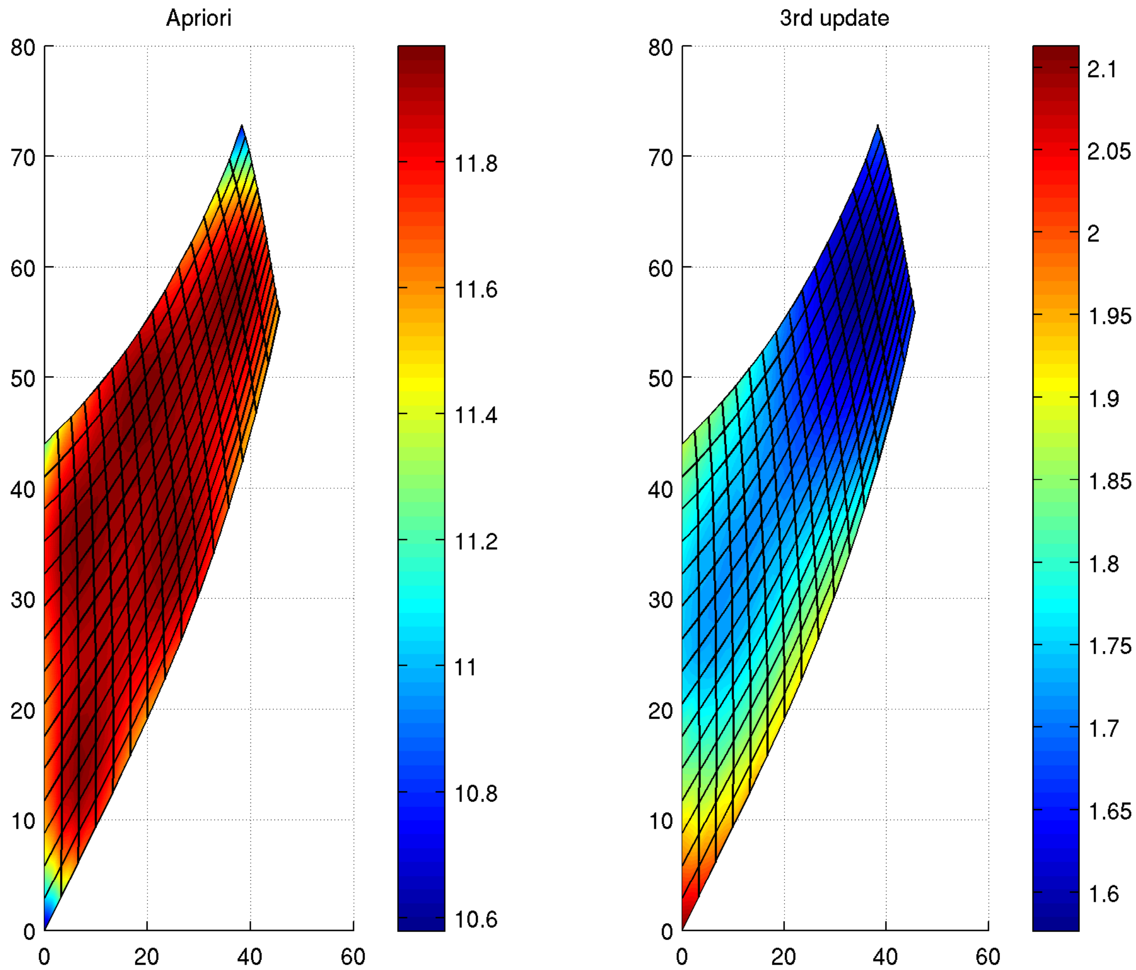
$$|\sigma_{xy}| > 2$$

Update shear modulus—mean



Change of mean of shear modulus from apriori to 3rd update

Update shear modulus—variance



Change of variance of shear modulus from apriori to 3rd update

Conclusion

1. Tensor representation linked with factorisations of C
2. Bayes's theorem can be used for system identification
3. Bayesian update is a projection
4. Bayesian update can be done on spectral expansion
5. Needs no Monte Carlo
6. Works on highly nonlinear examples like elasto-plasticity