High dimensional stochastic problems: some examples and some approximation techniques

T. Lelièvre
Joint work with
S. Boyaval, E. Cancès, V. Ehrlacher, J. Infante Acevedo, C. Le Bris and Y. Maday

Financial support from Michelin is acknowledged.

CERMICS, Ecole des Ponts ParisTech & MicMac project-team, INRIA.

Plan

1. Three examples

2. Two algorithms
   - Greedy algorithms for high dimensional problems
   - Variance reduction and the reduced basis method

Plan

Molecular dynamics

One crucial quantity in Molecular Dynamics is the commitor function:

$$\varphi(\mathbf{x}) = \mathbb{P}(\tau_B^{\mathbf{x}} < \tau_A^{\mathbf{x}})$$

where

$$\mathbf{X}_t^{\mathbf{x}} = \mathbf{x} - \int_0^t \nabla U(\mathbf{X}_s)\, ds + \sqrt{2\beta^{-1}}\mathbf{W}_t$$

and $\tau_A^{\mathbf{x}} = \inf\{t > 0,\, \mathbf{X}_t^{\mathbf{x}} \in A\}$, $\tau_B^{\mathbf{x}} = \inf\{t > 0,\, \mathbf{X}_t^{\mathbf{x}} \in B\}$. U is the interacting potential of N particles with position $\mathbf{X}_t^{\mathbf{x}} \in \mathbb{R}^{3N}$.

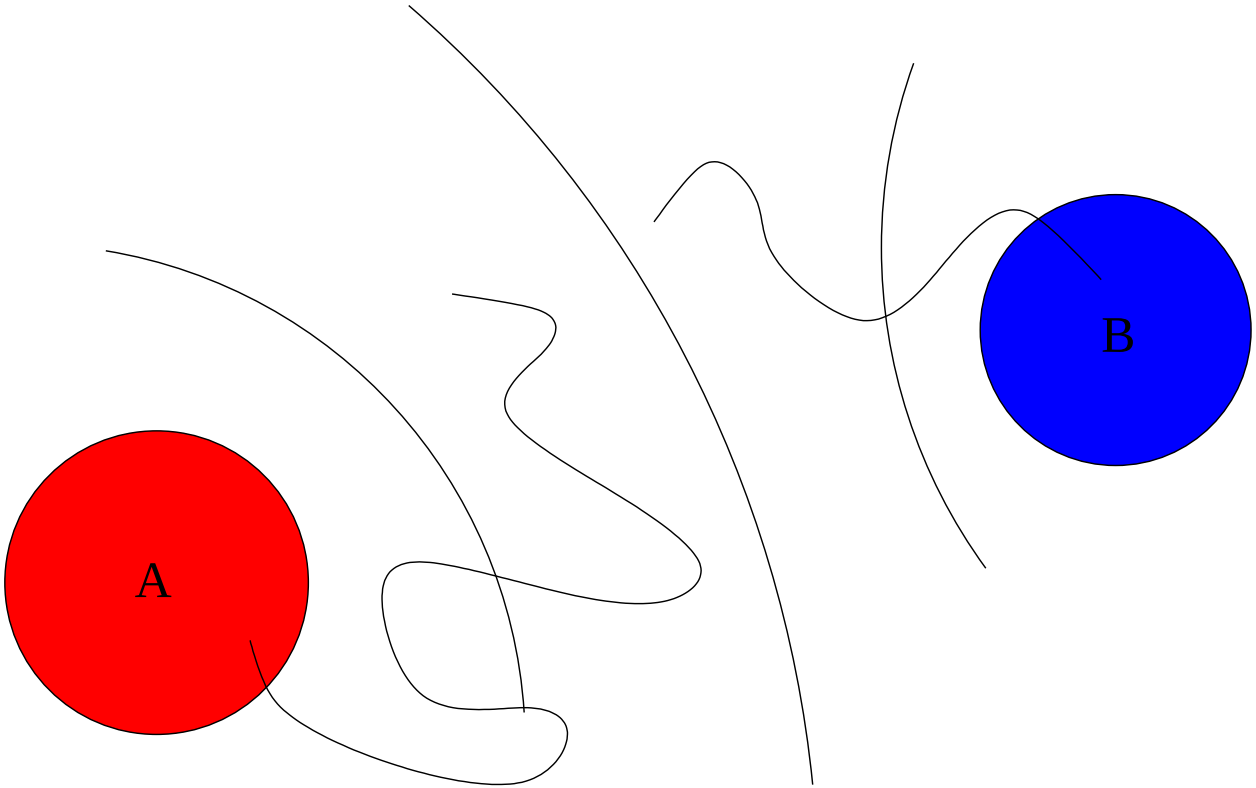
The level sets of φ yield a natural foliation of the configurational space, and in particular

$$\{\mathbf{x},\, \varphi(\mathbf{x}) = 1/2\}$$

is the $(1/2)$ iso-commitor surface, which defines, in some sense, the transition states.

The function φ is the most natural definition of a a one-dimensional index of reactive paths between A and B.

Molecular dynamics



Molecular dynamics

Variational formulation:

$$\varphi = \arg\min_{\phi \in V} \int_{\mathbb{R}^{3N}} |\nabla \phi|^2 \exp(-\beta U)$$

where

$$V = \left\{ \phi, \int \left(\phi^2 + |\nabla \phi|^2\right) \exp(-\beta U) < \infty, \ \phi = 0 \text{ on } A, \ \phi = 1 \text{ on } B \right\}.$$

Typically, $\mathbf{X}_t \in \mathbb{R}^{3N}$ with $N \simeq 10^6$. How to compute φ ?

Remark: In practice, one would like an approximation of the form:

$$\varphi(\mathbf{x}) \simeq \widetilde{\varphi}(\xi_1(\mathbf{x}), \ldots, \xi_d(\mathbf{x}))$$

with $d \ll N$, and given functions ξ_i's. Actually, finding the ξ_i's is part of the question !

Uncertainty quantification

How the uncertainty on parameters (input) propagate to the solution (output) ?

Prototypical example: almost surely,

$$\begin{cases} -\operatorname{div}_x(a(T,x)\nabla_x u(T,x)) = f(T,x) \text{ on } \Omega, \\ u(T,x) = 0 \text{ on } \partial\Omega, \end{cases}$$

where $x \in \Omega \subset \mathbb{R}^n$ and $T = (T_1, ..., T_p)$ are random variables. The total dimension $d = n + p$ may be large...

The problem is particularly interesting for *nonlinear* problems.

Uncertainty quantification

Different questions, with different techniques:

- Sampling: build efficiently ensembles of i.i.d. $u(T,x)$'s
- Estimate probability of rare events: $\mathbb{P}(F(u(T,x)) \geq F_0)$ [Reliability]
- Compute an average response $\mathbb{E}(u(T,x))$ and analyze the sensitivity of the output in a neighborhood. [Sensitivity]

In any cases, it is interesting to get a rough approximation (response surface) of the function

$$(t,x) \in \mathbb{R}^{n+p} \mapsto u(t,x).$$

This can be used for variance reduction techniques, for example (FORM, SORM,...).

Uncertainty quantification

Variational form:

$$u = \arg\min_{v \in V} \mathbb{E}\left(\frac{1}{2} \int_\Omega a(T,x) \left| \nabla_x v(T,x) \right|^2 dx - \int_\Omega f(T,x)\, v(T,x)\, dx \right)$$

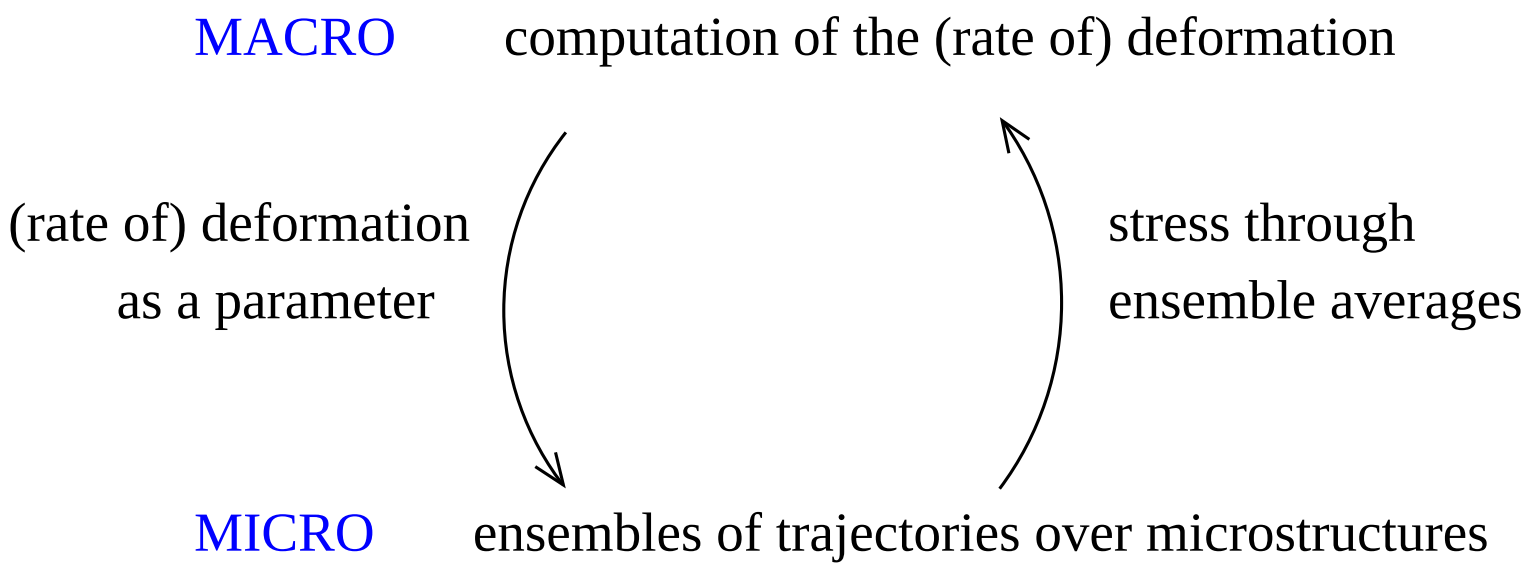
where

$$V = \left\{ v(t,x),\ \mathbb{E}\left(\int_\Omega \left(|\nabla_x v(T,x)|^2 + |v(T,x)|^2 \right)\ dx \right) < \infty, \right.$$
$$\left. \text{and } v(t,\cdot) = 0 \text{ on } \partial\Omega \right\}.$$

How to compute (an approximation of) u ?

Remark: Very important practical aspects: calibration (law of T), and validation (comparison with experiments).

Multiscale modelling

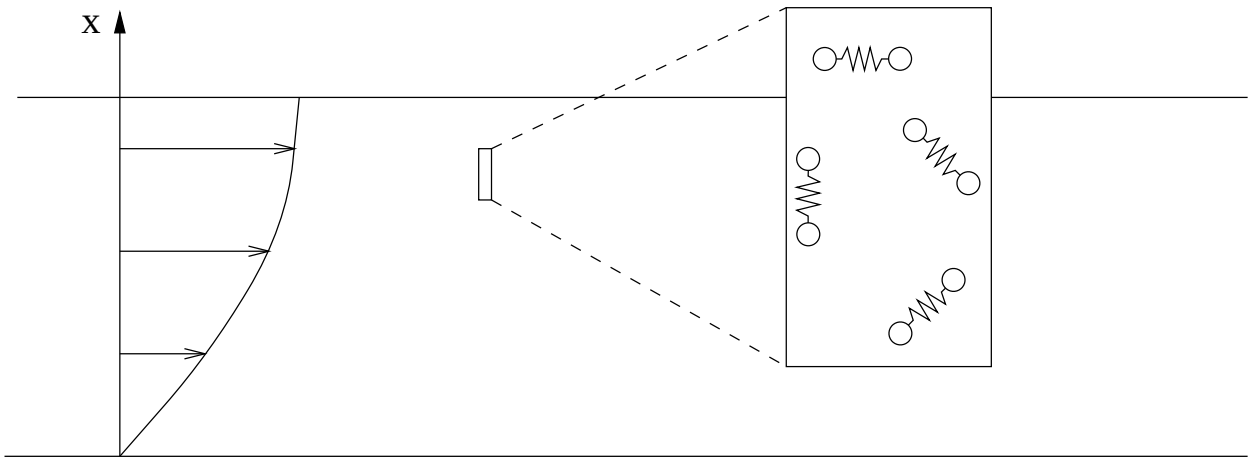


In many contexts, one has to compute many times (different positions, different times) and for many values of a parameter, some averages. How to do this efficiently ?

Multiscale modelling

A specific example: micro-macro models for polymeric fluids:

$$\begin{cases} \mathrm{Re}\left(\dfrac{\partial u}{\partial t} + u \cdot \nabla u\right) = -\nabla p + (1-\epsilon)\Delta u + \mathrm{div}\,\tau \\ \mathrm{div}\, u = 0 \\ \tau = \dfrac{\epsilon}{\mathrm{We}} \mathbb{E}\left(X_t \otimes F(X_t) - \mathrm{Id}\right) \\ dX_t + u \cdot \nabla X_t dt = \left(\nabla u X_t - \dfrac{1}{2\mathrm{We}} F(X_t)\right) dt + \dfrac{1}{\sqrt{\mathrm{We}}} dW_t \end{cases}$$



Multiscale modelling

In discretizations based on coupled Monte-Carlo-finite element methods, one has to compute the expected values:

$$\mathbb{E}(X_t \otimes F(X_t))$$

where X_t satisfies the SDE:

$$dX_t + u \cdot \nabla_x X_t \, dt = \left(\nabla u X_t - \frac{1}{2\mathrm{We}} F(X_t) \right) dt + \frac{1}{\sqrt{\mathrm{We}}} dW_t$$

at many macroscopic points and for many values of the velocity gradient.

How to do this efficiently ?

Other examples

Other contexts where high-dimensional functions naturally appear:

- Homogenization,
- Quantum Mechanics,
- Finance,
- ...

Plan

1. Three examples

2. Two algorithms
   - Greedy algorithms for high dimensional problems
   - Variance reduction and the reduced basis method

The greedy algorithm

The aim is to solve solve high-dimensional PDE. The approach has been proposed recently by: (i) Chinesta *et al.* to solve high-dimensional Fokker-Planck equations in the context of kinetic models for polymers and (ii) Nouy *et al* in the context of UQ. [Inspired by previous works by P. Ladevèze, in the 80's.]

It is related to:

- so-called Greedy Algorithms introduced in nonlinear approximation theory: [Temlyakov, Acta Numerica 2008] (Cohen, DeVore, Mallat, Avellaneda, ...)
- the best rank-n approximation of operators: [Kolda, Bader, SIAM Review 2009] (Hackbusch, Beylkin, Mohlenkamp, ...).

The technique consists in:

- approximating the solution by a sum of tensor products of low-dimensional functions,
- and looking iteratively for the best tensor product (greedy feature).

The greedy algorithm

Let us consider for simplicity the case of tensor products of only two spaces: $u(x_1, x_2) \in V$ is approximated by $\sum_{k \geq 1} r_k \otimes s_k$, where $(r_k, s_k) \in V_1 \times V_2$ and $r_k \otimes s_k(x_1, x_2) = r_k(x_1) s_k(x_2)$.
The algorithm and (almost) all the convergence results generalize to the case of tensor products of more than two functions:
$u \simeq \sum_{k \geq 1} r_k^1 \otimes r_k^2 \otimes \ldots \times r_k^d$.

Let us consider a functional $\mathcal{E} : V \to \mathbb{R}$ with a unique global minimizer that we want to compute:

$$u = \operatorname*{argmin}_{v \in V} \mathcal{E}(v).$$

The so-called greedy algorithm writes:

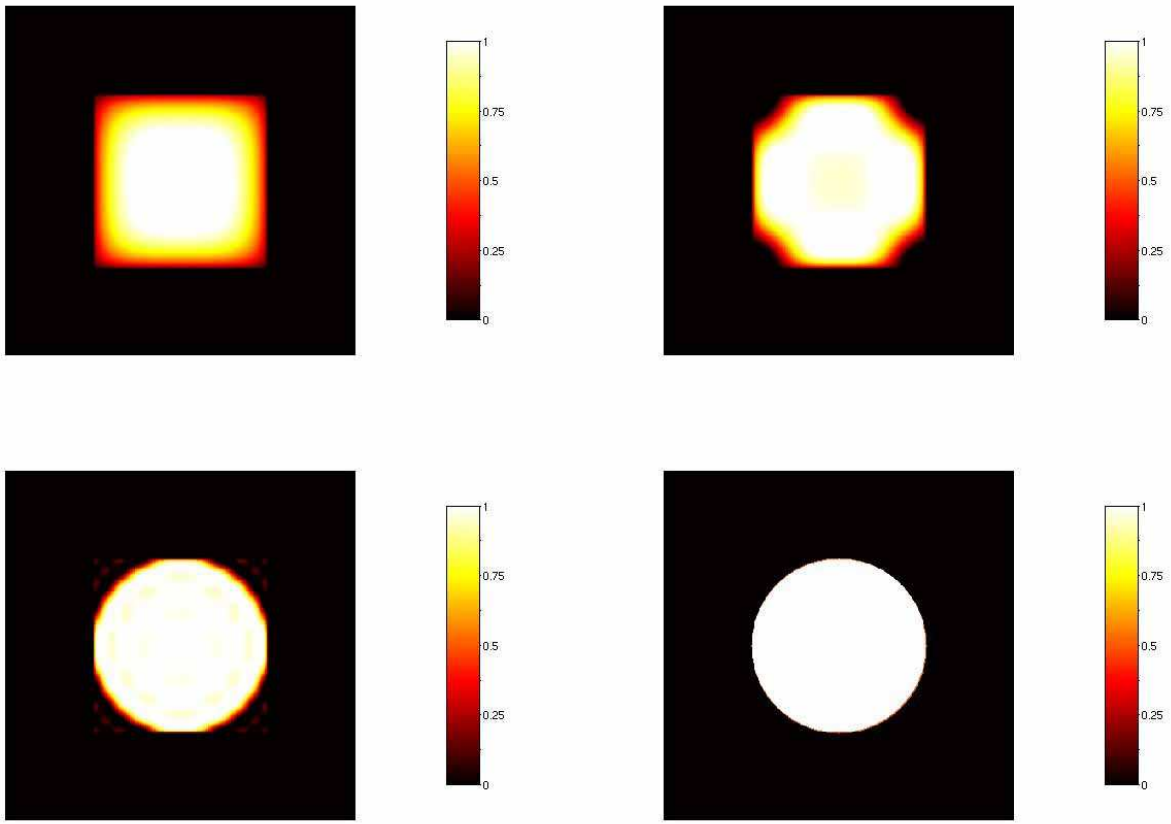
$$(r_n, s_n) \in \operatorname*{argmin}_{r \in V_1, s \in V_2} \mathcal{E}\left(\sum_{k=1}^{n-1} r_k \otimes s_k + r \otimes s \right).$$

We of course assume that V, V_1 and V_2 are Hilbert spaces such that $\mathrm{Vect}\{r \otimes s, r \in V_1, s \in V_2\} \subset V$ is dense.

The greedy algorithm: a toy example

$$\mathcal{E}(v) = \int_{[0,1]^2} |v - 1_{ball}|^2,$$

where $V_1 = V_2 = L^2(0,1)$. Approximations after 1, 2, 5, 60 iterations.



The greedy algorithm: examples

Recall the examples:

- commitor function:

$$\mathcal{E}(\phi) = \int_{\mathbb{R}^{3N}} |\nabla \phi|^2 \exp(-\beta V)$$

and
$V = \left\{\phi, \int \left(\phi^2 + |\nabla \phi|^2\right) \exp(-\beta U) < \infty,\ \phi = 0 \text{ on } A,\ \phi = 1 \text{ on } B\right\}.$

- uncertainty quantification:

$$\mathcal{E}(v) = \mathbb{E}\left(\frac{1}{2}\int_\Omega a(T,x)|\nabla_x v(T,x)|^2\,dx - \int_\Omega f(T,x)v(T,x)\,dx\right)$$

where
$V = \left\{v,\ \mathbb{E}\left(\int_\Omega \left(|\nabla_x v(T,\cdot)|^2 + |v(T,\cdot)|^2\right)\right) < \infty, v(t,\cdot) = 0 \text{ on } \partial\Omega\right\}$
and typically $V_1 = \{v_1(t),\ \mathbb{E}(|v_1(T)|^2) < \infty\}$, $V_2 = H_0^1(\Omega)$.

The greedy algorithm

To solve the minization problem, one typically considers the associated Euler-Lagrange equation: find $(r_n, s_n) \in V_1 \times V_2$ such that for all $(\delta r, \delta s) \in V_1 \times V_2$,

$$\left\langle \nabla \mathcal{E} \left(\sum_{k=1}^{n-1} r_k \otimes s_k + r_n \otimes s_n \right), r_n \otimes \delta s + \delta r \otimes s_n \right\rangle = 0.$$

The original problem (optimization over V) has exponential complexity in terms of the number of Hilbert spaces (say d). Each step of the greedy algorithm has linear complexity with respect to d.

Question: Let us denote

$$u_n = \sum_{k=1}^{n} r_k \otimes s_k.$$

Does u_n converge to u ?

The greedy algorithm

A typical convergence result [E. Cancès, V. Ehrlacher, TL, 2010]: Assume $\mathcal{E}$ is α-convex with Lipscihitz gradient on bounded sets of V, then

- the iterations are well-defined ((r_n, s_n) exists and is non-zero iff $u_{n-1} \neq u$),
- strong convergence: $\lim_{n\to\infty} u_n = u$ in V.

For (much) more, see the talk by V. Ehrlacher this afternoon.

Greedy algorithms

Future works:

- How to modify the algorithm to treat non-symmetric problems ?
- How to obtain a rate of convergence for nonlinear infinite dimensional cases ?
- How to reduce the number of terms generated by the algorithm ?

Plan

1. Three examples

2. Two algorithms
   - Greedy algorithms for high dimensional problems
   - Variance reduction and the reduced basis method

A stochastic reduced basis technique

Recall the third example. The question is: how to efficiently compute $\mathbb{E}(Z^\lambda)$ for many values of λ, using a Monte Carlo procedure ?

Idea: combine a variance reduction technique, together with the reduced basis technique.

Variance reduction by control variate

Here, λ is fixed. Instead of approximating $\mathbb{E}(Z)$, approximate

$$\mathbb{E}(Z - \alpha Y) = \mathbb{E}(Z)$$

where Y is a zero-mean random variable, and α is a deterministic parameter to be chosen. We want

$$\mathrm{Var}(Z - \alpha Y) \ll \mathrm{Var}(Z).$$

Optimal $\alpha = \alpha^* = \frac{\mathrm{Covar}(Z,Y)}{\mathrm{Var}(Y)}$ so that
$\mathrm{Var}(Z - \alpha^* Y) = \mathrm{Var}(Z)(1 - \rho(Z,Y)^2)$ with

$$\rho(Z,Y) = \frac{\mathrm{Covar}(Z,Y)}{\sqrt{\mathrm{Var}(Y)\mathrm{Var}(Z)}} \in [-1,1].$$

Best case: $|\rho(Z,Y)| = 1$ (*i.e.* $Y = Z - \mathbb{E}(Z)$!).

Worst case: $\rho(Z,Y) = 0$ (uncorrelated Y and Z).

How to build a correlated random variable Y with zero mean ?

A stochastic reduced basis technique

Algorithm:

- In an offline stage, compute *exact* control variates

$$Y^{\lambda_i} = Z^{\lambda_i} - \mathbb{E}(Z^{\lambda_i})$$

  for well chosen values of the λ_i's, $1 \leq i \leq I$.

- In the online stage, for a new value of the parameter λ, approximate the control variate as

$$\widetilde{Y}^I = \sum_{i=1}^{I} \alpha_i^* Y^{\lambda_i}$$

  where the (α_i^*) are solution to a least square problem:

$$(\alpha_i^*) = \arg \inf_{(\alpha_i)} \mathrm{Var} \left(Z^\lambda - \sum_{i=1}^{I} \alpha_i Y^{\lambda_i} \right).$$

Remark: In practice, all the expectation values are approximated by empirical means over i.i.d. realizations.

A stochastic reduced basis technique

How to choose the λ_i's in the offline stage ? By using a greedy procedure over a (large) discrete set Λ_{trial} of parameters: once $(\lambda_1, \ldots, \lambda_i)$ have been chosen, set

$$\lambda_{i+1} = \arg \max_{\lambda \in \Lambda_{\text{trial}}} \text{Var} \left(Z^\lambda - \sum_{j=1}^{i} \alpha_j^* Y^{\lambda_j} \right)$$

where $(\alpha_j^*) = \arg \inf_{(\alpha_j)} \text{Var} \left(Z^\lambda - \sum_{j=1}^{i} \alpha_j Y^{\lambda_j} \right)$.

For (much) more, see the talk by S. Boyaval this afternoon.

A stochastic reduced basis technique

Questions:

- What can be said on the overall efficiency of the algorithm ?
- How to choose Λ_{trial} ?

Personal conclusions

- Non-linear approximation techniques should be used to efficiently tackle high-dimensional problems (greedy techniques, mesh adaptation, reduced basis, ...).
- Deterministic approaches should be used in combination with stochastic approaches to reduce computational costs (typically using variance reduction techniques).

References

- A. Ammar, B. Mokdad, F. Chinesta, and R. Keunings. A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modeling of complex fluids. J. Non-Newtonian Fluid Mech., 139:153-176, 2006.
- S. Boyaval and TL, A variance reduction method for parametrized stochastic differential equations using the reduced basis paradigm, Communications in Mathematical Sciences, 8(3), 735-762, 2010.
- E. Cancès, V. Ehrlacher and TL. Convergence of a greedy algorithm for high-dimensional convex nonlinear problems, http://hal.archives-ouvertes.fr/hal-00469622, to appear in M3AS.
- T.G. Kolda and B.W. Bader, Tensor Decompositions and Applications, SIAM Rev., 51(3), 455-500, 2009.
- C. Le Bris, TL and Y. Maday. Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations, Constructive Approximation 30(3):621-651, 2009.
- A. Nouy, Recent Developments in Spectral Stochastic Methods for the Numerical Solution of Stochastic Partial Differential Equations, Archives of Computational Methods in Engineering 16:251-285, 2009.
- A. Nouy and O. Le Maitre. Generalized spectral decomposition for stochastic nonlinear problems, J. Comput. Phys. 228(1):202-235, 2009.
- V.N. Temlyakov. Greedy approximation. Acta Numerica, 17:235-409, 2008.