Numerical Methods for PDE Constrained Optimization with Uncertain Data

Organised by
Matthias Heinkenschloss, Houston
Volker Schulz, Trier

27 January – 2 February 2013

Table of Contents

Alexander Litvinenko (joint with Hermann G. Matthies)

Sampling and Low-Rank Tensor Approximations ....................... 3

Hermann G. Matthies (joint with A. Litvinenko, B. V. Rosić, A. Kučerová,
J. Sýkora, O. Pajonk)

Stochastic Setting for Inverse Identification Problems ............... 6
Abstracts

Sampling and Low-Rank Tensor Approximations

ALEXANDER LITVINENKO
(joint work with Hermann G. Matthies)

Computing of a response surface (also called surrogate model) can be very helpful in the area of uncertainty quantification. The idea is to approximate unknown solution function (which is expensive to evaluate) by a cheap surrogate (e.g. a polynomial) which is easy to evaluate. In this work we show how to compute and then how to update such low-rank response surface on the fly with a linear complexity [4, 3]. Later on the constructed surrogate model is used for generation of a large sample which is necessary for computing different statistics of the solution, such as mean, variance, error bars, cumulative distribution function and histograms.

After the response surface (RS) is constructed from few samples, we use new data (snapshots) and the residual to update it. A motivation for this idea comes from the fact that in many software packages it is impossible or very difficult to change the code, but it is possible to access the residual.

Numerical examples [3, 4] show that in the case of a smooth solution one can save around 80% of computing cost (i. e. the response surface is a good preconditioner), but if the solution is discontinuous then the response surface produces a poor approximation and the full computation is necessary.

1. Data compression

Let \( v_i \in \mathbb{R}^n \), \( i = 1, \ldots, Z \), be the solution vectors (snapshots), where \( Z \) is a number of stochastic realisations of the solution and \( n \) number of degrees of freedom in the physical space. Let us build from all these vectors the matrix \( W = (v_1, \ldots, v_Z) \in \mathbb{R}^{n \times Z} \) and consider the approximation (with accuracy \( \varepsilon \))

\[
W \approx W_k = AB^T, \quad \|W - W_k\| < \varepsilon, \quad k \ll \min\{n, Z\}.
\]

To compute factors \( A \in \mathbb{R}^{n \times k} \) and \( B \in \mathbb{R}^{Z \times k} \) we compute the truncated singular value decomposition \( W_k = U_k \Sigma_k V_k^T \), where \( U_k \in \mathbb{R}^{n \times k} \) contains the first \( k \) columns of \( U \), \( V_k \in \mathbb{R}^{Z \times k} \) contains the first \( k \) columns of \( V \) and \( \Sigma_k \in \mathbb{R}^{k \times k} \) contains the \( k \)-biggest singular values of \( \Sigma \). We denote \( A = U_k \Sigma_k \) and \( B = V_k \).

Denote the \( i \)-th column of matrix \( B^T \) by \( b_i \in \mathbb{R}^k \) be given. Eq. 2 and Eq. 3 show how to compute the mean \( \bar{v} \in \mathbb{R}^n \) and the covariance without multiplying \( A \) and \( B \):

\[
\bar{v} = \frac{1}{Z} \sum_{i=1}^{Z} v_i \approx \bar{v}_k = \frac{1}{Z} \sum_{i=1}^{Z} A \cdot b_i = A \bar{b}.
\]

The computational complexity is \( O(k(Z + n)) \), besides \( O(nZ) \) for usual dense data format. By definition, the covariance matrix is \( C = \frac{1}{Z-1} W_c W_c^T \). Now, the
covariance matrix can be approximated like
\begin{equation}
C \approx \frac{1}{Z-1} U_k \Sigma_k \Sigma_k^T U_k^T = \frac{1}{Z-1} AA^T.
\end{equation}

The variance of the solution vector (i.e. the diagonal of the covariance matrix in Eq. 3) can be computed with the complexity $O(k^2(Z + n))$.

**Lemma:** Let $\|W - W_k\|_2 \leq \epsilon$, and $\mu_k$ be a rank-$k$ approximation of the mean $\mu$. Then a) $\|\mu - \mu_k\| \leq \frac{\epsilon}{\sqrt{Z}}$ b) $\|C - C_k\| \leq \frac{1}{Z-1} \epsilon^2$.

For the proof see [4, 3].

Suppose $W_k$ is given. Suppose also that matrix $W' \in \mathbb{R}^{n \times m}$ contains new $m$ solution vectors. For a small $m$, computing the factors $C \in \mathbb{R}^{n \times k}$ and $D \in \mathbb{R}^{m \times k}$ such that $W' \approx CD^T$ is not expensive. Now our aim is to compute with a linear complexity the rank-$k$ approximation of $W_{new} := [W W'] \in \mathbb{R}^{n \times (Z + m)}$. To do this, we build two concatenated matrices $A_{new} := [A C] \in \mathbb{R}^{n \times 2k}$ and $B_{new} = \text{blockdiag}[B^T D^T] \in \mathbb{R}^{2k \times (Z + m)}$. Note that matrices $A_{new}$ and $B_{new}$ have rank $2k$. To truncate rank from $2k$ to $k$ we use the QR-algorithm with the linear complexity $O((n + Z)k^2 + k^3)$ [1, 4, 3].

2. Response surface and its low-rank approximation

Let $v(x, \theta)$ be the solution (or a functional of the solution). It can be pressure, density, velocity, lift, drag etc. $v(x, \theta)$ can be approximated in a set of new independent Gaussian random variables (polynomial chaos expansions (PCE) of Wiener [6])
\begin{equation}
v(x, \theta(\omega)) \approx \sum_{\beta \in J_{M,p}} v_\beta(x) H_\beta(\theta) = [...v_\beta(x) ][...H_\beta(\theta) ...]^T,
\end{equation}
where $\theta(\omega) = (\theta_1(\omega), ..., \theta_M(\omega))$, $v_\beta(x)$ are coefficients, $H_\beta(\theta)$ the multivariate Hermite polynomials, $J_{M,p}$ a finite multi-index subset of infinite dimensional multi-index set and $p$ the maximal polynomial order $H_\beta(\theta)$. Since Hermite polynomials are orthogonal, the coefficients $v_\beta(x)$ can be computed by projection:
\[
v_\beta(x) = \frac{1}{\beta!} \int_{\mathbb{R}^d} H_\beta(\theta) v(x, \theta) \mathbb{P}(d\theta) \approx \frac{1}{\beta!} \sum_{i=1}^{n_q} H_\beta(\theta_i) v(x, \theta_i) w_i,
\]
where the multidimensional integral over $\Theta$ is computed approximately, for example, on a sparse Gauss-Hermite grid where $\theta_i$ are quadrature points, $w_i$ corresponding weights and $n_q$ is the number of quadrature points. Using the rank $k$ approximation $AB^T$ of the set of realisations, obtain
\begin{equation}
v_\beta(x) = \frac{1}{\beta!} [v(x, \theta_1), ..., v(x, \theta_{n_q})] \cdot [H_\beta(\theta_1) w_1, ..., H_\beta(\theta_{n_q}) w_{n_q}]^T \approx AB^T c_\beta,
\end{equation}
where vector $c_\beta := \frac{1}{\beta!} [H_\beta(\theta_1) w_1, ..., H_\beta(\theta_{n_q}) w_{n_q}]^T$. The matrix of all PCE coefficients will be $[...v_\beta(x) ...] = AB^T [...c_\beta ...]$, $\beta \in J_{M,p}$. Put all together, obtain a low-rank representation of RS
\begin{equation}
v(x, \theta) \approx AB^T [...c_\beta ...][...H_\beta(\theta) ...]^T.
\end{equation}
3. Update of Response Surface via Computing the Residual

Assume that we already approximated the unknown solution by a response surface like in Eq. 6. The following algorithm updates the given response surface.

Algorithm: (Update of the response surface)

1. Take the next point \( \theta_{n_{q}+1} \) and evaluate the response surface Eq. 6 in this point. Let \( u(x, \theta_{n_{q}+1}) \) be the obtained solution.
2. Compute residual \( ||r(u(x, \theta_{n_{q}+1}))|| \). Only if \( ||r|| \) is large solve expensive deterministic problem. It can be also used, e. g., as a start value in an iterative process.
3. Update \( A, B^T, [...c_{3}...] \) and go to (1).

In the best case we never solve the deterministic problem again. In the worst case we must solve the deterministic problem for each \( \theta_{n_{q}+1}, i = 1, 2, ... \). The numerical results [4, 3] (solution is smooth, no shock) show that with this algorithm one can reduce e. g. the number of needed iterations (for non-linear Navier Stokes) from 10000 to 2000. If the solution is discontinuous (e. g. with shock) then our response surface is a very poor approximation and the produced value can not be used as a good start point in iterations.

Future plans: In this work we split the spatial part from the stochastic part via SVD. The next aim is to split high-dimensional stochastic part and represent/approximate it in a low-rank tensor format and to keep this format during the further postprocessing. We have successfully tried this idea in [2] for the elliptic problems with uncertain coefficients.

REFERENCES

Stochastic Setting for Inverse Identification Problems

Hermann G. Matthies
(joint work with A. Litvinenko, B. V. Rosić, A. Kučerová, J. Sýkora, O. Pajonk)

In trying to predict the behaviour of physical systems, one is often confronted with the fact that some parameters which characterise the system may only be incompletely known, or in other words they are uncertain. We want to identify these parameters (denote by $q$) through observations or measurement of the response of the system [2, 5, 6].

Such an identification can be approached in different ways. One way is to measure the difference between observed and predicted system output and try to find parameters such that this difference is minimised, this optimisation approach leads to regularisation procedures. Here we take the view that our lack of knowledge or uncertainty of the actual value of the parameters can be described in a Bayesian way through a probabilistic model. The unknown parameter is then modelled as a random variable — also called the prior model — and additional information on the system through measurement or observation changes the probabilistic description to the so-called posterior model. Bayesian setting allows updating / sharpening of information about $q$ when measurement is performed.

It is well-known that such a Bayesian update (BU) is in fact a conditional expectation, and this is the basis of the presented method. As the Bayesian update may be numerically very demanding, we accelerate this update through methods based on functional approximation or spectral representation of stochastic problems [3, 1], in particular Wiener's so-called homogeneous or polynomial chaos expansion — which are polynomials in independent Gaussian random variables and which can also be used numerically in a Galerkin procedure.

The idea presented in the talk is extension (non-linear BU in contrast to linear BU) of ideas presented in [6, 5]. In [5] we present a fully deterministic method to compute sequential updates for stochastic state estimates of dynamic models from noisy measurements. It does not need any assumptions about the type of distribution for either data or measurement — in particular it does not have to assume any of them as Gaussian. The implementation is based on a polynomial chaos expansion of the stochastic variables of the model — however, any other orthogonal basis would do. We use a minimum variance estimator that combines an a priori state estimate and noisy measurements in a Bayesian way. For computational purposes, the update equation is projected onto a finite-dimensional PCE-subspace. The resulting Kalman-type update formula for the PCE coefficients can be efficiently computed solely within the PCE. As it does not rely on sampling, the method is deterministic and fast. The original Kalman filter is shown to be a low-order special case of the presented method. For the numerical illustration we perform a bi-modal identification using noisy measurements. Additionally, we provide numerical experiments by applying it to the well known Lorenz-84 model and compare it to a related method, the ensemble Kalman filter.
In [6] we present a sampling-free approach to a probabilistic interpretation of an inverse problem in which unknown coefficient (e.g. conductivity or permeability field) is represented by a random field. The arising stochastic forward problem is solved through stochastic Galerkin method [4, 8]. The forward solution is used to forecast the measurement. The update of the prior is a projection of the minimum variance estimator from linear Bayesian updating onto the polynomial chaos basis. With the help of such representation the probabilistic identification problem is cast in a polynomial chaos expansion setting and the Bayes’ linear form of updating. By introducing the Hermite algebra this becomes a direct, purely algebraic way of computing the posterior, which is comparatively inexpensive to evaluate. In addition, we show that the well-known Kalman filter is the low order part of this update. The proposed method is tested on a stationary diffusion equation with prescribed source terms, characterised by an uncertain conductivity parameter which is then identified from limited and noisy data obtained by a measurement of the diffusing quantity.

We remind that in [6, 5] we used the linear Bayesian update. The new idea is to use an non-linear Bayesian update. The Bayesian update of order $n$ looks as follow

$$q_a(\cdot) = q_a(0H, \ldots, kH, \ldots, nH) = \sum_{k=0}^{n} kHz^{\otimes k},$$

where $q_a(\cdot)$ the PCE coefficients of the value of interest (e.g. conductivity field), $kH$ - $k$-th non-linear Bayesian update coefficient and $z$ - the difference between the noisy measurements and the numerical forecast of these measurements. We note that coefficients $kH, k = 0..n$, are unknown and should be computed first. To compute $kH$ we set up the system, which comes from the corresponding minimization problem (see [2]) and solve it. The solution is the vector $(0H, \ldots, kH, \ldots, nH)$ of unknown coefficients. Since dimension of the system is large, solving such system is a non-trivial task. The computational cost is $O(n^\alpha N^\beta L^\gamma)$, where $L$ the number of PCE coefficients, $N$ number of degrees of freedom in the physical space and $\alpha, \beta, \gamma > 1$. In our implementation all matrix elements, the right-hand side as well as unknowns of this system are tensors. This is a reason why an efficient low-rank tensor arithmetic in necessary here.

We underline that the main idea here was to do the (linear as well as non-linear) Bayesian update directly on the polynomial chaos expansion of the value of interest without any sampling. This idea has appeared independently (only as a linear update) in [7] as a variant of the Kalman filter.

References


*Reporter: Volker Schulz*