King Abdullah University of Science and Technology KAUST

NON-LINEAR APPROXIMATION OF BAYESIAN UPDATE

Alexander Litvinenko[1], Hermann G. Matthies[2], Elmar Zander[2]

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology | Center for Uncertainty Quantification

http://sri-uq.kaust.edu.sa/

[1]Extreme Computing Research Center, KAUST,
[2]Institute of Scientific Computing, TU Braunschweig, Brunswick, Germany

KAUST

Figure : KAUST campus, 7 years old, approx. 7000 people (include 1700 kids), 100 nations, 36 km^2.

Stochastic Numerics Group at KAUST

Advances in Uncertainty Quantification Methods, Algorithms and Applications (UQAW 2015)

January 6 – 9, 2015
9:00 a.m. – 5:00 p.m.
Level 0 auditorium, between Al-Jazri and Al-Kindi (buildings 4 and 5)

WORKSHOP TOPICS

1- Uncertainty Quantification Methods and Algorithms
2- Verification and Validation
3- Experimental Design
4- Applications to Problems in Computational Science, Engineering, Networks and the Environment

ORGANIZERS

RAUL TEMPONE
Director, Center for Uncertainty Quantification in Computational Science & Engineering. Associate Professor, Computer, Electrical, and Mathematical Sciences and Engineering Division, KAUST

OMAR KNIO
Deputy Director, Center for Uncertainty Quantification in Computational Science & Engineering. Professor, Computer, Electrical and Mathematical Sciences and Engineering Division, KAUST

For more information contact:
http://sri-uq.kaust.edu.sa
Raul.Tempone@kaust.edu.sa; Omar.Knio@kaust.edu.sa
& Alexander.Litvinenko@kaust.edu.sa

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

A BIG picture

1. Computing the full Bayesian update is very expensive (MCMC is expensive)
2. Look for a cheap surrogate (linear, quadratic, cubic,... approx.)
3. Kalman filter is a particular case
4. Do Bayesian update of Polynomial Chaos Coefficients! (not probability densities!)
5. Consider non-Gaussian cases

History

1. O. Pajonk, B. V. Rosic, A. Litvinenko, and H. G. Matthies, A Deterministic Filter for Non-Gaussian Bayesian Estimation, Physica D: Nonlinear Phenomena, Vol. 241(7), pp. 775-788, 2012.
2. B. V. Rosic, A. Litvinenko, O. Pajonk and H. G. Matthies, Sampling Free Linear Bayesian Update of Polynomial Chaos Representations, J. of Comput. Physics, Vol. 231(17), 2012 , pp 5761-5787
3. A. Litvinenko and H. G. Matthies, Inverse problems and uncertainty quantification `http://arxiv.org/abs/1312.5048`, 2013
4. H. G. Matthies, E. Zander, B.V. Rosic, A. Litvinenko, Parameter estimation via conditional expectation - A Bayesian Inversion, accepted to AMOS-D-16-00015, 2016.
5. H. G. Matthies, E. Zander, B.V. Rosic, A. Litvinenko, Bayesian parameter estimation via filtering and functional approximation, Tehnicki vjesnik 23, 1(2016), 1-17.

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 ⇒ 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

⇒ Bayes's theorem.

Repeated measurements lead to better identification.

Mathematical setup

Consider

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

where S is solution operator.
Operator depends on parameters $q \in \mathcal{Q}$,
hence state $u \in \mathcal{U}$ is also function of q:
Measurement operator Y with values in $\mathcal{Y}$:

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

Examples of measurements:
$y(\omega) = \int_{\mathcal{D}_0} u(\omega, x) dx$, or u in few points

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.

With state u a RV, the quantity to be measured

$$y(\omega) = Y(q(\omega), u(\omega)))$$

is also uncertain, a random variable.
Noisy data: $\hat{y} + \epsilon(\omega)$, where $\hat{y}$ is the "true" value and a random error ϵ.
Forecast of the measurement: $z(\omega) = y(\omega) + \epsilon(\omega)$.
Classically, Bayes's theorem gives conditional probability

$$\mathbb{P}(I_q|M_z) = \frac{\mathbb{P}(M_z|I_q)}{\mathbb{P}(M_z)}\mathbb{P}(I_q) \quad (\text{or } \pi_q(q|z) = \frac{p(z|q)}{Z_s}p_q(q))$$

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

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

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

The subspace $\mathscr{Q}_\infty := L_2(\Omega, \mathbb{P}, \sigma(z))$ represents the available information.

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

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

Multivariate Hermite polynomials were used to approximate random fields/stochastic processes with Gaussian random variables. According to Cameron and Martin theorem PCE expansion converges in the L_2 sense.

Let $Y(x, \boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, ..., \theta_M, ...)$, is approximated:

$$Y(x, \boldsymbol{\theta}) = \sum_{\beta \in \mathcal{J}_{m,p}} H_\beta(\boldsymbol{\theta}) Y_\beta(x), \quad |\mathcal{J}_{m,p}| = \frac{(m+p)!}{m!p!},$$

$H_\beta(\boldsymbol{\theta}) = \prod_{k=1}^{M} h_{\beta_k}(\theta_k)$,

$$Y_\beta(x) = \frac{1}{\beta!} \int_\Theta H_\beta(\boldsymbol{\theta}) Y(x, \boldsymbol{\theta}) \, \mathbb{P}(d\boldsymbol{\theta}).$$

$$Y_\beta(x) \approx \frac{1}{\beta!} \sum_{i=1}^{N_q} H_\beta(\boldsymbol{\theta}_i) Y(x, \boldsymbol{\theta}_i) w_i.$$

Look for φ such that $q(\xi) = \varphi(z(\xi))$, $z(\xi) = y(\xi) + \varepsilon(\omega)$:

$$\varphi \approx \tilde{\varphi} = \sum_{\alpha \in \mathcal{J}_p} \varphi_\alpha \Phi_\alpha(z(\xi)) \tag{1}$$

and minimize $\|q(\xi) - \tilde{\varphi}(z(\xi))\|^2$, where Φ_α are polynomials (e.g. Hermite, Laguerre, Chebyshev or something else). Taking derivatives with respect to φ_α:

$$\frac{\partial}{\partial \varphi_\alpha} \langle q(\xi) - \tilde{\varphi}(z(\xi)), q(\xi) - \tilde{\varphi}(z(\xi)) \rangle = 0 \quad \forall \alpha \in \mathcal{J}_p \tag{2}$$

Inserting representation for $\tilde{\varphi}$, obtain

Numerical computation of NLBU

$$\frac{\partial}{\partial \varphi_\alpha}\mathbb{E}\left(q^2(\xi) - 2\sum_{\beta\in\mathcal{J}} q\varphi_\beta\Phi_\beta(z) + \sum_{\beta,\gamma\in\mathcal{J}} \varphi_\beta\varphi_\gamma\Phi_\beta(z)\Phi_\gamma(z)\right)$$

$$= 2\mathbb{E}\left(-q\Phi_\alpha(z) + \sum_{\beta\in\mathcal{J}} \varphi_\beta\Phi_\beta(z)\Phi_\alpha(z)\right)$$

$$= 2\left(\sum_{\beta\in\mathcal{J}} \mathbb{E}\left[\Phi_\beta(z)\Phi_\alpha(z)\right]\varphi_\beta - \mathbb{E}\left[q\Phi_\alpha(z)\right]\right) = 0 \quad \forall\alpha\in\mathcal{J}$$

$$\mathbb{E}\left[\Phi_\beta(z)\Phi_\alpha(z)\right]\varphi_\beta = \mathbb{E}\left[q\Phi_\alpha(z)\right]$$

Now, rewriting the last sum in a matrix form, obtain the linear system of equations ($=: A$) to compute coefficients φ_β:

$$\begin{pmatrix} \cdots & \cdots & \cdots \\ \vdots & \mathbb{E}\left[\Phi_\alpha(z(\xi))\Phi_\beta(z(\xi))\right] & \vdots \\ \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} \vdots \\ \varphi_\beta \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \mathbb{E}\left[q(\xi)\Phi_\alpha(z(\xi))\right] \\ \vdots \end{pmatrix},$$

where $\alpha, \beta \in \mathcal{J}$, A is of size $|\mathcal{J}| \times |\mathcal{J}|$.

Using the same quadrature rule of order q for each element of A, we can write

$$A = \mathbb{E}\left[\Phi_{\mathcal{J}_\alpha}(z(\xi))\Phi_{\mathcal{J}_\beta}(z(\xi))^T\right] \approx \sum_{i=1}^{N^A} w_i^A \Phi_{\mathcal{J}_\alpha}(z_i)\Phi_{\mathcal{J}_\beta}(z_i)^T, \quad (3)$$

where (w_i^A, ξ_i) are weights and quadrature points, $z_i := z(\xi_i)$ and $\Phi_{\mathcal{J}_\alpha}(z_i) := (...\Phi_\alpha(z(\xi_i))....)^T$ is a vector of length $|\mathcal{J}_\alpha|$.

$$b = \mathbb{E}\left[q(\xi)\Phi_{\mathcal{J}_\alpha}(z(\xi))\right] \approx \sum_{i=1}^{N^b} w_i^b q(\xi_i)\Phi_{\mathcal{J}_\alpha}(z_i), \quad (4)$$

where $\Phi_{\mathcal{J}_\alpha}(z(\xi_i)) := (..., \Phi_\alpha(z(\xi_i)), ...), \alpha \in \mathcal{J}_\alpha$.

We can write the Eq. 15 with the right-hand side in Eq. 4 in the compact form:

$$[\Phi_A]\,[\mathrm{diag}(...w_i^A...)]\,[\Phi_A]^T \begin{pmatrix} \vdots \\ \varphi_\beta \\ \vdots \end{pmatrix} = [\Phi_b] \begin{pmatrix} w_0^b q(\xi_0) \\ \ldots \\ w_{N^b}^b q(\xi_{N^b}) \end{pmatrix} \quad (5)$$

$[\Phi_A] \in \mathbb{R}^{\mathcal{J}_\alpha \times N^A}$, $[\mathrm{diag}(...w_i^A...)] \in \mathbb{R}^{N^A \times N^A}$, $[\Phi_b] \in \mathbb{R}^{\mathcal{J}_\alpha \times N^b}$, $[w_0^b q(\xi_0)...w_{N^b}^b q(\xi_{N^b})] \in \mathbb{R}^{N^b}$.
Solving Eq. 5, obtain vector of coefficients $(...\varphi_\beta...)^T$ for all β.
Finally, the assimilated parameter q_a will be

$$q_a = q_f + \tilde{\varphi}(\hat{y}) - \tilde{\varphi}(z), \quad (6)$$

$z(\xi) = y(\xi) + \varepsilon(\omega)$, $\tilde{\varphi} = \sum_{\beta \in \mathcal{J}_p} \varphi_\beta \Phi_\beta(z(\xi))$

Example 1: φ does not exist in the Hermite basis

Assume $z(\xi) = \xi^2$ and $q(\xi) = \xi^3$. The normalized PCE coefficients are $(1, 0, 1, 0)$
$(\xi^2 = 1 \cdot H_0(\xi) + 0 \cdot H_1(\xi) + 1 \cdot H_2(\xi) + 0 \cdot H_3(\xi))$
and $(0, 3, 0, 1)$
$(\xi^3 = 0 \cdot H_0(\xi) + 3 \cdot H_1(\xi) + 0 \cdot H_2(\xi) + 1 \cdot H_3(\xi))$.
For such data the mapping φ does not exist. The matrix A is close to singular.
Support of Hermite polynomials (used for Gaussian RVs) is $(-\infty, \infty)$.

Example 2: φ does exist in the Laguerre basis

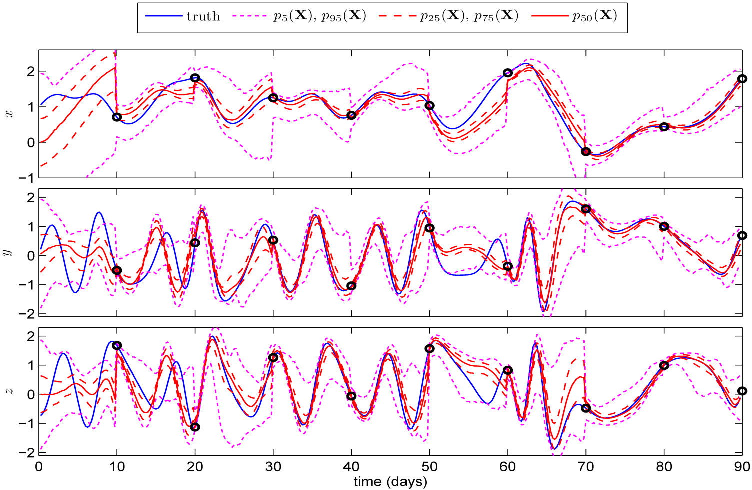
Assume $z(\xi) = \xi^2$ and $q(\xi) = \xi^3$.
The normalized gPCE coefficients are $(2, -4, 2, 0)$ and $(6, -18, 18, -6)$.
For such data the mapping mapping φ of order 8 and higher produces a very accurate result.
Support of Laguerre polynomials (used for Gamma RVs) is $[0, \infty)$.

Is a system of ODEs. Has chaotic solutions for certain parameter values and initial conditions.

$$\dot{x} = \sigma(\omega)(y - x)$$
$$\dot{y} = x(\rho(\omega) - z) - y$$
$$\dot{z} = xy - \beta(\omega)z$$

Initial state $q_0(\omega) = (x_0(\omega), y_0(\omega), z_0(\omega))$ are uncertain.

Solving in $t_0, t_1, ..., t_{10}$, Noisy Measur. $\rightarrow$ UPDATE, solving in $t_{11}, t_{12}, ..., t_{20}$, Noisy Measur. $\rightarrow$ UPDATE,...



Trajectories of x,y and z in time. After each update (new information coming) the uncertainty drops. (O. Pajonk)

Lorenz-84 Problem

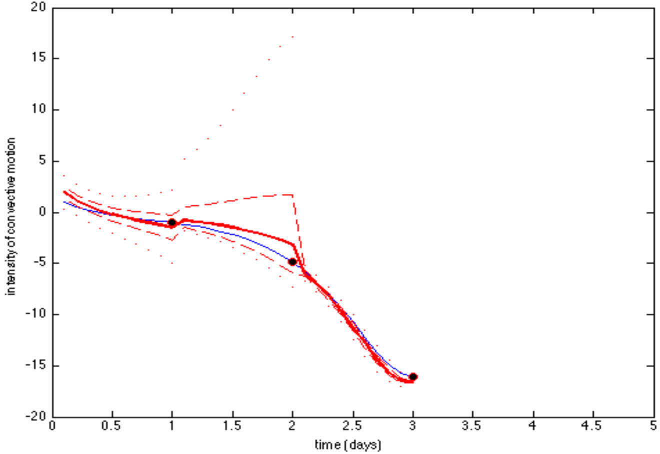


Figure : Partial state trajectory with uncertainty and three updates

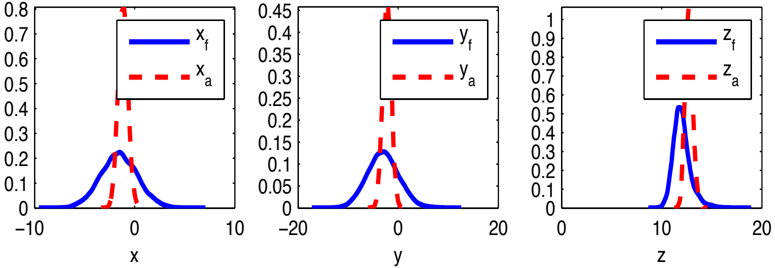


Figure : NLBU: Linear measurement $(x(t), y(t), z(t))$: prior and posterior after one update

Lorenz-84 Problem

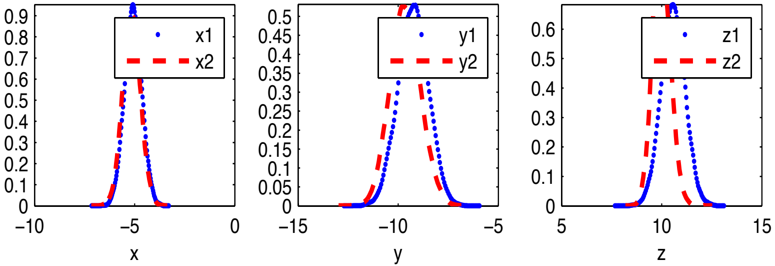


Figure : Linear measurement: Comparison posterior for LBU and NLBU after second update

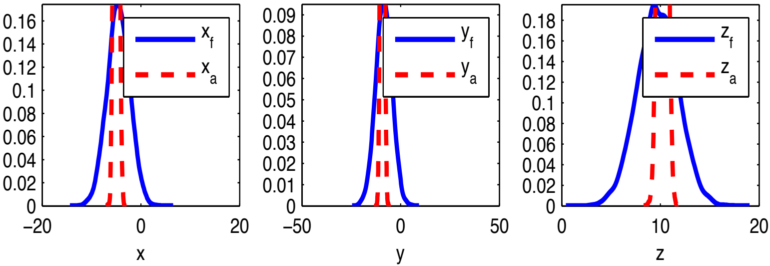


Figure : Quadratic measurement $(x(t)^2, y(t)^2, z(t)^2)$: Comparison of a priori and a posterior for NLBU

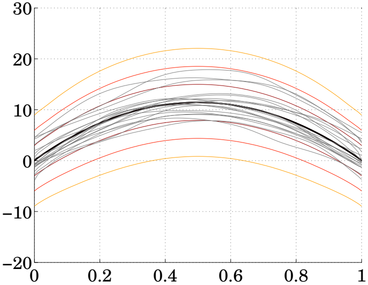
Example 4: 1D elliptic PDE with uncertain coeffs

Taken from Stochastic Galerkin Library (sglib), by Elmar Zander (TU Braunschweig)

$$-\nabla \cdot (\kappa(x,\xi)\nabla u(x,\xi)) = f(x,\xi), \quad x \in [0,1]$$

Measurements are taken at $x_1 = 0.2$, and $x_2 = 0.8$. The means are $\overline{y}(x_1) = 10$, $\overline{y}(x_2) = 5$ and the variances are 0.5 and 1.5 correspondingly.

Example 4: updating of the solution u



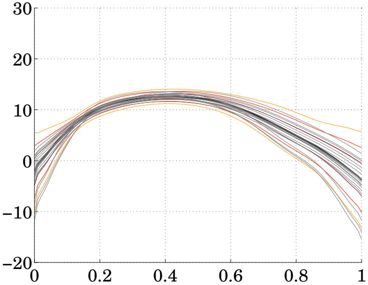


Figure : Original and updated solutions, mean value plus/minus 1,2,3 standard deviations

See more in sglib by Elmar Zander

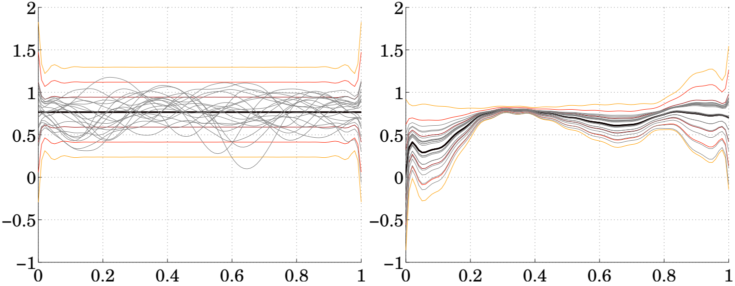


Figure : Original and updated parameter q.

See more in sglib by Elmar Zander

- + Step 1. Introduced a way to derive MMSE φ (as a linear, quadratic, cubic etc approximation, i. e. compute conditional expectation of q, given measurement Y.
- Step 2. Apply φ to identify parameter q
- + All ingredients can be given as gPC.
- + we apply it to solve inverse problems (ODEs and PDEs).
- - Stochastic dimension grows up very fast.

Acknowledgment

I used a Matlab toolbox for stochastic Galerkin methods (sglib)
`https://github.com/ezander/sglib`

Alexander Litvinenko and his research work was supported by the King Abdullah University of Science and Technology (KAUST), SRI-UQ and ECRC centers.

Literature

1. Bojana Rosic, Jan Sykora, Oliver Pajonk, Anna Kucerova and Hermann G. Matthies, Comparison of Numerical Approaches to Bayesian Updating, report on www.wire.tu-bs.de, 2014
2. A. Litvinenko and H. G. Matthies, *Inverse problems and uncertainty quantification* `http://arxiv.org/abs/1312.5048`, 2013
3. L. Giraldi, A. Litvinenko, D. Liu, H. G. Matthies, A. Nouy, *To be or not to be intrusive? The solution of parametric and stochastic equations - the "plain vanilla" Galerkin case*, `http://arxiv.org/abs/1309.1617`, 2013
4. O. Pajonk, B. V. Rosic, A. Litvinenko, and H. G. Matthies, *A Deterministic Filter for Non-Gaussian Bayesian Estimation*, *Physica D: Nonlinear Phenomena*, Vol. 241(7), pp. 775-788, 2012.
5. B. V. Rosic, A. Litvinenko, O. Pajonk and H. G. Matthies, *Sampling Free Linear Bayesian Update of Polynomial Chaos Representations*, J. of Comput. Physics, Vol.

1. *PCE of random coefficients and the solution of stochastic partial differential equations in the Tensor Train format*, S. Dolgov, B. N. Khoromskij, A. Litvinenko, H. G. Matthies, 2015/3/11, arXiv:1503.03210

2. *Efficient analysis of high dimensional data in tensor formats*, M. Espig, W. Hackbusch, A. Litvinenko, H.G. Matthies, E. Zander Sparse Grids and Applications, 31-56, 40, 2013

3. *Application of hierarchical matrices for computing the Karhunen-Loeve expansion*, B.N. Khoromskij, A. Litvinenko, H.G. Matthies, Computing 84 (1-2), 49-67, 31, 2009

4. *Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats*, M. Espig, W. Hackbusch, A. Litvinenko, H.G. Matthies, P. Waehnert, Comp. & Math. with Appl. 67 (4), 818-829, 2012

5. *Numerical Methods for Uncertainty Quantification and Bayesian Update in Aerodynamics*, A. Litvinenko, H. G. Matthies, Book ”Management and Minimisation of Uncertainties and Errors in Numerical Aerodynamics” pp 265-282, 2013