Mathematisches Forschungsinstitut Oberwolfach

Report No. 04/2013

DOI: 10.4171/OWR/2013/04

Numerical Methods for PDE Constrained Optimization with Uncertain Data

Organised by
Matthias Heinkenschloss, Houston
Volker Schulz, Trier

27 January – 2 February 2013

Abstract. Optimization problems governed by partial differential equations (PDEs) arise in many applications in the form of optimal control, optimal design, or parameter identification problems. In most applications, parameters in the governing PDEs are not deterministic, but rather have to be modeled as random variables or, more generally, as random fields. It is crucial to capture and quantify the uncertainty in such problems rather than to simply replace the uncertain coefficients with their mean values. However, treating the uncertainty adequately and in a computationally tractable manner poses many mathematical challenges. The numerical solution of optimization problems governed by stochastic PDEs builds on mathematical subareas, which so far have been largely investigated in separate communities: Stochastic Programming, Numerical Solution of Stochastic PDEs, and PDE Constrained Optimization.

The workshop achieved an impulse towards cross-fertilization of those disciplines which also was the subject of several scientific discussions. It is to be expected that future exchange of ideas between these areas will give rise to new insights and powerful new numerical methods.

Mathematics Subject Classification (2010): 35xx, 49xx, 60xx, 65xx, 90xx, 93xx.

Introduction by the Organisers

The workshop *Numerical Methods for PDE Constrained Optimization with Uncertain Data*, organized by Matthias Heinkenschloss (Houston) and Volker Schulz (Trier) was held 27 January – 2 February 2013. One of the main objectives of this meeting was to bring together leading experts from the fields of stochastic

programming, numerical solution of stochastic PDEs, and PDE constrained optimization in order to encourage and foster new approaches by the exchange of state-of-the-art methods and fresh ideas. The achievement of this goal was well reflected by this workshop which was attended by almost fifty active researchers from seven countries including a few students and postdoctoral fellows. A total of thirty presentations was given at the workshop covering a wide spectrum of issues ranging from the analysis of specific theoretical problems to more algorithmic aspects of computational schemes and various applications.

A particular area of active research is the topic of

Optimization methods

This topic was one of the central themes of the workshop addressed in several talks including optimization with probability constraints (Henrion), optimal experimental design (Herzog), preconditioning of full-space SQP methods (Ridzal), finite dimensional stochastic programming (Schultz), robust optimization based on lower order approximations (Ulbrich) and in several application oriented talks.

Another central theme of the workshop was concerned with

Adaptive methods

This topic included the aspects of high order spatial discretizations (Gittelson), adaptive solution of PDE constrained optimization problems (Kouri), low-rank tensor approximations (Litvinenko), accurate discretization schemes (Mohammadi), collocation approaches (Nobile), adaptive quadratur rules (Ritter), and adaptive Smolyak-type approaches (Schillings).

Furthermore, the powerful methodology of

Model reduction and model predictive control

played an important role. Researchers in these areas reported on the usage of Fokker-Planck approaches (Borzi), nonstationery nonlinear model predictive control (Kostina), real-time optimization (Potschka) and approximation of stochastic optimization problems (Römisch).

The important aspect of

Applications

was the subject of ten talks concerned with Pareto-front identification in aerodynamic optimization (Desideri), oil and gas exploration (El-Bakry), groundwater flow (Ernst), optimal experimental design (Herzog), cardiovascular applications (Kunisch), inverse identification (Matthies), radio frequency ablation (Preusser), large scale shape optimization (Schmidt), topology optimization (Stoffel) and thermoelastic shape optimization (Zorn).

Furthermore several talks addressed interesting side aspects not covered within the clusters above, like piecewise deterministic processes (Annunziato), random domains (Harbrecht), and the tractability of multivariate problems (Novak),

It was noticable in several presentations that the idea of bringing the so far seperated disciplines of stochastic programming, numerical solution of stochastic PDEs, and PDE constrained optimization together is timely and very relevant for

applications. However, it has become also clear that this workshop has been just the first step giving an impulse towards cross-fertilization of those fields and that further similar efforts are necessary to unravel the full potential for future joint research.

Workshop: Numerical Methods for PDE Constrained Optimization with Uncertain Data

Table of Contents

Abstracts

Probability Density Function Optimal Control of Piecewise Deterministic Processes

Mario Annunziato

A novel strategy for the optimal control of piecewise deterministic processes (PDP) is illustrated. The PDPs considered here are stochastic processes made of pieces of determisitic motion punctuated of random events that are originated by a Markov process. The control objectives are devised as functionals of the probability density function (PDF) of the process, that follows a desired sequence of configuration. The corresponding optimal control problems are formulated as a sequence of open-loop optimality systems in a nonlinear model predictive control strategy. The governing equation of the time dependent marginal PDFs is a Kolmogorov-Fokker-Planck-type equation, that results to a system of linear hyperbolic PDEs. An upwind discretization scheme, that is positive and conservative preserving, is discussed for this equation and its adjoint. The effectiveness of this new control framework is tested to the problem of optimal control of filtered of dichotomous noise.

This work is part of the talk "A Fokker-Planck Strategy for the Optimal Control of Stochastic Processes and Piecewise Deterministic Processes" that is in collaboration with Prof. Dr. Alfio Borzì (Würzburg Univ.) and it is supported in part by the EU Marie Curie International Training Network Multi-ITN STRIKE Projekt 'Novel Methods in Computational Finance' and in part by the ESF OPTPDE Programme.

References

[1] M. Annunziato, A. Borzì, *A Fokker-Planck control framework for multidimensional stochastic processes*, Journal of Comp. and Appl. Math., **237** (2013) 1, 487–507.

[2] M. Annunziato, A. Borzì, *Optimal control of probability density functions of stochastic processes*, Math. Model. and Analysis, **15** (2010) 4, 393–407.

[3] M. Annunziato, *Analysis of upwind method for piecewise deterministic Markov processes*, Comp. Meth. Appl. Math. **8** (2008) 1, pp. 3–20.

An Optimal Control Strategy for Probability Density Functions of Stochastic Processes and Piecewise Deterministic Processes

Alfio Borzi

(joint work with M. Annunziato, M. Mohammadi)

An efficient framework for the optimal control of probability density functions (PDF) of multidimensional stochastic processes and piecewise deterministic processes is presented. This framework is based on Kolmogorov-Fokker-Planck-type equations that govern the time evolution of the PDF of stochastic processes and

piecewise deterministic processes. In this approach the control objectives are formulated in terms of the PDF to follow a given trajectory and to reach a desired terminal configuration. The corresponding optimal control problems are formulated as a sequence of open-loop optimality systems in a nonlinear model predictive control strategy. Theoretical results concerning the forward and the optimal control problems are provided. In the case of stochastic (Ito) processes, the Fokker-Planck equation is of parabolic type and it is shown that under appropriate assumptions the open-loop bilinear control function is unique. In the case of piecewise deterministic processes (PDP), the Fokker-Planck equation consists of a first-order hyperbolic system. Discretization schemes are discussed that guarantee positivity and conservativeness of the forward solution. The proposed control framework is validated with multidimensional biological and financial models.

References

[1] M. Annunziato and A. Borzì, *Optimal control of probability density functions of stochastic processes*, Mathematical Modelling and Analysis **15** (2010) 393–407.
[2] M. Annunziato and A. Borzì, *A Fokker-Planck control framework for multidimensional stochastic processes*, Journal of Computational and Applied Mathematics **237** (2013) 487–507.
[3] M. Annunziato and A. Borzì, *Fokker-Planck-Based Control of a Two-Level Open Quantum System*, Mathematical Models and Methods in Applied Sciences (M3AS), to appear.
[4] M. Annunziato and A. Borzì, *Optimal Control of Piecewise Deterministic Processes*, submitted.

Multiple-Gradient Descent Algorithm, MGDA, for Pareto-Front Identification - Application to Robust Design

Jean-Antoine Désidéri
(joint work with Jean-Antoine Désidéri, Régis Duvigneau)

We began with a numerical illustration of optimum-shape design in aerodynamics (aircraft wing shape design for minimum drag). If the geometry is optimized at a fixed nominal Mach number, its performance degrades rapidly at off conditions, which illustrates the necessity to perform *robust-design* optimization. This can be realized in practice by Monte-Carlo approaches in which for an assigned (usually Gaussian) distribution of Mach numbers, both mean and variance of drag are minimized, yielding a two-objective optimization formulation. Possibly, due to computationally-demanding flow simulation, one may resort to meta-models of mean and variance in terms of the design parameters. Such meta-models are constructed based on a database of high-fidelity pre-calculations [1]. An alternate route is to consider a multi-point formulation in which the Mach number is discretized, yielding several shape functionals to be minimized. In all cases, algorithm-wise, one is led to consider a multi-objective optimization problem for which we propose a multiple-gradient descent algorithm (MGDA) in which the descent direction is constructed as the minimum-norm element in the convex hull of the gradients [2]. The notion of Pareto-stationarity has been introduced and

proved to be the natural necessary condition to Pareto-optimality. MGDA converges to Pareto-stationary points. Direct variants of MGDA have been proposed using an incomplete well-ordered Gram-Schmidt process, and a scaling by the Hessian matrices [3].

References

[1] M. Martinelli and R. Duvigneau, *On the use of second-order derivative and metamodel-based Monte-Carlo for uncertainty estimation in aerodynamics,* Computers and Fluids, **37-6** (2010).
[2] J.-A. Désidéri, *Multiple-gradient descent algorithm (MGDA) for multiobjective optimization / Algorithme de descente à gradients multiples pour l'optimisation multiobjectif*, Comptes Rendus Mathématique, **350:5-6** (2012), 313-318.
[3] J.-A. Désidéri, *MGDA Variants for Multi-Objective Optimization*, Research Report 8068, INRIA (2012). (http://hal.inria.fr/hal-00732881)

Some Interesting Mathematical Problems in the Oil and Gas Business

Amr El-Bakry

In this talk I will review some important math problems from PDE to optimization and explain the source of uncertainty in these problems.

Travel Time Calculations at the WIPP: UQ for a Groundwater Flow Problem

Oliver G. Ernst
(joint work with K. Andrew Cliffe, Björn Sprungk)

We consider a groundwater flow problem with random transmissivity field arising in a site assessment for the radioactive waste repository Waste Isolation Pilot Plant (WIPP) in Carlsbad, NM (USA). We show how a statistical model may be obtained from measurements of transmissivity and compare three methods for the numerical solution of the resulting boundary value problem with random coefficient: Monte Carlo simulation, sochastic collocation and Gaussian process emulators. The quantity of interest in these calculations is the probability law of the travel time of a particle released in the domain to the domain boundary. We demonstrate that, for rough covariance structures, the number of random parameters one can feasibly accomodate do not allow the CDF of the travel time to be approximated with sufficient accuracy.

References

[1] Andrew Cliffe, Oliver G. Ernst, Björn Sprungk, Elisabeth Ullmann, and K. Gerald van den Boogaart. Probabilistic uncertainty quantification using PDEs with random data: A case study. 2012. in preparation.
[2] Oliver G. Ernst and Björn Sprungk. Stochastic collocation for elliptic PDEs with random data: The lognormal case. 2012. (submitted).
[3] Nicola Stone. *Gaussian Process Emulators for Uncertainty Analysis in Groundwater Flow.* PhD thesis, The University of Nottingham, April 2011.

Merits of High-Order Spatial Discretization for Random PDE

Claude Jeffrey Gittelson

Solutions of a class of random elliptic boundary value problems admit efficient approximations by polynomials on the parameter domain. Each coefficient in such an expansion is a spatially-dependent function, and can be approximated within a hierarchy of finite element spaces. Convergence rates for multilevel approximations using a separate finite element space to approximate each coefficient can be derived based on summability properties of the coefficients in a scale of spatial norms. Sparse tensor product constructions, which impose additional structure on such a multilevel approximation, can always essentially attain the optimal convergence rate, [1].

Surprisingly, simply choosing a single level of refinement for all active coefficients sometimes also reaches the optimal convergence rate of much more flexible multilevel approximations, while avoiding difficulties arising from having to deal with multiple spatial discretizations, [1]. This finding is particularly prevalent for high-order spatial approximations. Thus high-order methods may provide a simpler alternative to sparse tensor product constructions and other multilevel stochastic Galerkin approximations in that both approaches yield improvements in efficiency, but these do not appear to be complimentary, [2].

References

[1] C.J. Gittelson, *Convergence rates of multilevel and sparse tensor approximations for a random elliptic PDE*, submitted.

[2] C.J. Gittelson, *High-order methods as an alternative to sparse tensor products for stochastic Galerkin FEM*, submitted.

Modelling and Simulation of Elliptic PDEs on Random Domains

Helmut Harbrecht

Introduction.

The rapid development of computational resources in recent years allows the accurate numerical solution of large classes of partial differential equation models, *provided* that the problem's input data are given exactly. Often, however, exact input data for numerical simulations in engineering are not known. The practical significance of a highly accurate numerical solution of differential equation models in engineering must thus address how to account for uncertain input data.

If a statistical description of the input data is available, one can mathematically describe the data and the solutions as random fields and aim at the computation of corresponding deterministic statistics of the unknown random solution u. Our particular interest is on elliptic boundary value problems on uncertain domains $D(\omega) \subset \mathbb{R}^n$. Specifically, we like to compute the expectation

$$\mathbb{E}_u(\mathbf{x}) = \int_\Omega u(\mathbf{x}, \omega) \mathrm{d}P(\omega),$$

the two-point correlation

$$\mathrm{Cor}_u(\mathbf{x},\mathbf{y}) = \int_\Omega u(\mathbf{x},\omega)u(\mathbf{y},\omega)\mathrm{d}P(\omega),$$

and from it the variance $\mathbb{V}_u(\mathbf{x}) = \mathrm{Cor}_u(\mathbf{x},\mathbf{x}) - \mathbb{E}_u^2(\mathbf{x})$.

The goal of computation is as follows: *given mean and two-point correlation of the boundary perturbation field, compute, to leading order, the mean and the two-point correlation of the random solution of the boundary value problem.*

Partial differential equations on random domains. For a given random event $\omega \in \Omega$, we shall consider the Dirichlet problem for the Poisson equation on a random domain

$$-\Delta u(\omega) = f \text{ in } D(\omega), \quad u(\omega) = g \text{ on } \partial D(\omega). \tag{1}$$

Here, the random domain $D(\omega) \subset \mathbb{R}^n$ is seen as the random perturbation of a fixed nominal domain $\overline{D} \subset \mathbb{R}^n$, that is

$$\partial D(\omega) = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}(\mathbf{x},\omega) = \mathbf{x} + \varepsilon\kappa(\mathbf{x},\omega)\mathbf{n}(\mathbf{x}),\ \mathbf{x} \in \partial\overline{D}\},$$

where $\mathbf{n}$ denotes the outward normal vector at $\partial\overline{D}$.

For a fixed, small $\varepsilon \in [0,\varepsilon_0]$, one can linearize the problem with respect to ω under additional assumptions on the smoothness of $\kappa(\mathbf{x},\omega)$, necessary to ensure that the perturbed domain is still admissible (e.g. Lipschitz continuous). Namely, for $\mathbf{x} \in K \subset\subset \overline{D}$, we find the first order *shape Taylor expansion*

$$u(\mathbf{x},\omega) = \overline{u}(\mathbf{x}) + \varepsilon\delta u(\mathbf{x})[\kappa(\omega)] + \mathcal{O}(\varepsilon^2), \tag{2}$$

see [7]. Herein, $\overline{u}$ signifies the solution of (1) with respect to the nominal domain $\overline{D}$, i.e.,

$$-\Delta\overline{u} = f \text{ in } \overline{D}, \quad \overline{u} = g \text{ on } \partial\overline{D}, \tag{3}$$

and δu is the related *local shape derivative*

$$\Delta\delta u = 0 \text{ in } \overline{D}, \quad \delta u = \kappa\frac{\partial(g-\overline{u})}{\partial\mathbf{n}} \text{ on } \partial\overline{D}. \tag{4}$$

Under the assumption that $\mathbb{E}_\kappa = 0$, we arrive, in view of the shape Taylor expansion (2), at (cf. [2, 4, 7])

$$\mathbb{E}_u(\mathbf{x}) = \overline{u}(\mathbf{x}) + \mathcal{O}(\varepsilon^2), \quad \mathrm{Cor}_u(\mathbf{x},\mathbf{y}) = \varepsilon^2\mathrm{Cor}_{\delta u}(\mathbf{x},\mathbf{y}) + \mathcal{O}(\varepsilon^3), \tag{5}$$

where $\mathrm{Cor}_{\delta u}$ satisfies the boundary value problem

$$\begin{gathered}(\Delta_\mathbf{x}\otimes\Delta_\mathbf{y})\mathrm{Cor}_{\delta u} = 0 \text{ in } \overline{D}\times\overline{D},\\ \Delta_\mathbf{x}\mathrm{Cor}_{\delta u} = 0 \text{ on } \overline{D}\times\partial\overline{D}, \quad \Delta_\mathbf{y}\mathrm{Cor}_{\delta u} = 0 \text{ on } \partial\overline{D}\times\overline{D},\\ \mathrm{Cor}_{\delta u} = \mathrm{Cor}_\kappa\left[\frac{\partial(g-\overline{u})}{\partial\mathbf{n}}\otimes\frac{\partial(g-\overline{u})}{\partial\mathbf{n}}\right] \text{ on } \partial\overline{D}\times\partial\overline{D}.\end{gathered} \tag{6}$$

Consequently, the expectation of the random solution coincides, with leading order in the size of the random perturbation, with the solution of the Poisson equation (3). The two-point correlation, i.e., the second order statistical moment, is, with

leading order, given by a partial differential equation in tensor product form, which is based on the equation for the local shape derivative (4).

Note that the present linearization approach extents to more general second order diffusion problems [2] and to Neumann and Robin boundary conditions [3]. Random interface problems have been considered in [4]. Moreover, random right hand sides and random coefficients can analogously be treated [6, 8].

An abstract view on the linearization approach. The linearization of a second order elliptic boundary value problem with respect to a given input parameter $\kappa(\omega)$ involves the associated derivative $\delta u(\omega) \in \mathcal{H}(D)$. It is generally given by a boundary value problem

$$\mathcal{A}\delta u(\omega) = f(\omega) \text{ on } D,$$

where $\mathcal{A} : \mathcal{H}(D) \to \mathcal{H}'(D)$ denotes a linear, second order elliptic partial differential operator which is defined on a domain $D \subset \mathbb{R}^n$. Typically one might think of $\mathcal{H}(D)$ being a Sobolev space with dual $\mathcal{H}'(D)$. Moreover, the random input parameter linearly enters the right hand side $f(\omega) \in \mathcal{H}'(D)$ since the mapping $\kappa(\omega) \mapsto \delta u(\omega)$ is linear. The two-point correlation $\mathrm{Cor}_{\delta u} \in \mathcal{H}_{mix}(D \times D) := \mathcal{H}(D) \otimes \mathcal{H}(D)$, which pops up in the asymptotic expansions (5), is thus given by the tensor product problem

$$(\mathcal{A} \otimes \mathcal{A})\mathrm{Cor}_{\delta u} = \mathrm{Cor}_f \text{ on } D \times D. \tag{7}$$

Especially it holds $\mathrm{Cor}_f \in \mathcal{H}'_{mix}(D \times D) = \mathcal{H}'(D) \otimes \mathcal{H}'(D)$.

Combination technique. The starting point for the definition of sparse tensor product spaces for the Sobolev space $\mathcal{H}_{mix}(D \times D)$ are traditional and widely used multilevel hierarchies

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset \mathcal{H}(D),$$

where $\dim(V_j) \sim 2^{jn}$. Then, appropriate complement spaces

$$W_0 := V_0, \qquad W_j := V_j \ominus V_{j-1}, \quad j > 0$$

are chosen to arrive at the multiscale decomposition

$$V_J = W_0 \oplus W_1 \oplus \cdots \oplus W_J.$$

In general, such complement spaces are defined by hierarchical bases like e.g. wavelet or multilevel bases, see [1] and the references therein. The sparse tensor product space $\widehat{V}_J \subset \mathcal{H}_{mix}(D \times D)$ is finally defined in accordance with

$$\widehat{V}_J = \bigoplus_{j+j' \le J} W_j \otimes W_{j'} = \bigoplus_{j=0}^{J} V_j \otimes W_{J-j}.$$

It possesses only $\mathcal{O}(2^{Jn}J)$ degrees of freedom which is much less than the $\mathcal{O}(2^{2Jn})$ degrees of freedom of the full tensor product space $V_J \otimes V_J$. However, the approximation power of the sparse tensor product space and the full tensor product space are essentially (i.e., except for logarithmic factors) identical if extra smoothness is given in terms of Sobolev spaces with dominating mixed derivative [1].

To derive the combination technique, we consider the special complement spaces

$$W_j := (P_j - P_{j-1})\mathcal{H}(D) \subset V_j$$

with $P_j : \mathcal{H}(D) \to V_j$ being the Galerkin projection associated with the operator $\mathcal{A}$. Then, the Galerkin system with respect to the sparse tensor product space $\widehat{V}_J$ decouples due to Galerkin orthogonality. Namely, it holds

$$\big((\mathcal{A} \otimes \mathcal{A})v_{i,i'}, w_{j,j'}\big)_{L^2(D\times D)} = 0 \text{ for all } v_{i,i'} \in W_i \otimes W_{i'},\ w_{j,j'} \in W_j \otimes W_{j'}$$

provided that $i \neq j$ or $i' \neq j'$. As a consequence, the Galerkin solution $\widehat{\mathrm{Cor}}_{\delta u,J}$ to (7) in the sparse tensor product space $\widehat{V}_J$ can be written as

$$\widehat{\mathrm{Cor}}_{\delta u,J} = \sum_{j=0}^{J} \{p_{j,J-j} - p_{j,J-j-1}\} \in \bigoplus_{j=0}^{J} V_j \otimes (V_{J-j} \ominus V_{J-j-1}) = \widehat{V}_J$$

where $p_{j,j'}$ denotes the Galerkin solution of (7) in the full (but small) tensor product space $V_j \otimes V_{j'}$, cf. [3, 6].

Low-rank approximation. A rank-r approximation of a given function $\mathrm{Cor}_f \in L^2(D \times D)$ is defined by

$$\mathrm{Cor}_f(\mathbf{x}, \mathbf{y}) \approx \mathrm{Cor}_{f,r}(\mathbf{x}, \mathbf{y}) := \sum_{\ell=1}^{r} a_\ell(\mathbf{x}) b_\ell(\mathbf{y})$$

with certain functions $a_\ell, b_\ell \in L^2(D)$. Inserting such a low-rank approximation in the tensor product boundary value problem (7) leads to the representation

$$\mathrm{Cor}_{\delta u} = \big(\mathcal{A}^{-1} \otimes \mathcal{A}^{-1}\big)\mathrm{Cor}_f \approx \big(\mathcal{A}^{-1} \otimes \mathcal{A}^{-1}\big)\mathrm{Cor}_{f,r} = \sum_{\ell=1}^{r} \big(\mathcal{A}^{-1} a_\ell\big) \otimes \big(\mathcal{A}^{-1} b_\ell\big),$$

i.e., the tensor product boundary value problem is reduced to $2r$ simple boundary value problems on the domain D. In the case of the second moment analysis in uncertainty quantification, we find the special situation that Cor_f is symmetric and positive semi-definite. Hence, the pivoted Cholesky decomposition can be used to efficiently compute the low-rank approximation to the right hand side, see [4, 5].

References

[1] H.J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica* **13** (2004), 147–269.

[2] H. Harbrecht. A finite element method for elliptic problems with stochastic input data. *Appl. Numer. Math.* **60** (2010), 227–244.

[3] H. Harbrecht. Second moment analysis for Robin boundary value problems on random domains. Preprint 2012-13, Mathematisches Institut, Universität Basel, Switzerland (2012).

[4] H. Harbrecht and J. Li. A fast deterministic method for stochastic elliptic interface problems based on low-rank approximation. Research Report No. 2011-24, Seminar für Angewandte Mathematik, ETH Zürich, Switzerland (2011).

[5] H. Harbrecht, M. Peters, and R. Schneider. On the low-rank approximation by the pivoted Cholesky decomposition. *Appl. Numer. Math.* **62** (2012), 428–440.

[6] H. Harbrecht, M. Peters, and M. Siebenmorgen. Combination technique based k-th moment analysis of elliptic problems with random diffusion. Preprint 2011-02, Mathematisches Institut, Universität Basel, Switzerland (2011).
[7] H. Harbrecht, R. Schneider, and C. Schwab. Sparse second moment analysis for elliptic problems in stochastic domains. *Numer. Math.* **109** (2008), 167–188.
[8] C. Schwab and R. Todor. Sparse finite elements for elliptic problems with stochastic loading. *Numer. Math.* **95** (2003), 707–734.

Optimization Problems with Probabilistic Constraints

René Henrion

A probabilistic constraint is an inequality of the type

$$\mathbb{P}\left(g(z,\xi)\leq 0\right)\geq p, \tag{1}$$

where g is a mapping defining a (random) inequality system and ξ is an s- dimensional random vector defined on some probability space $(\Omega,\mathcal{A},\mathbb{P})$. The probabilistic constraint expresses the requirement that a decision vector z is feasible if and only if the random inequality system $g(z,\xi)\leq 0$ is satisfied at least with probability $p\in[0,1]$. The use of probabilistic constraints is highly relevant for engineering problems involving uncertain data. Among its numerous applications one may find topics like water management, telecommunications, electricity network expansion, mineral blending, chemical engineering etc. For a comprehensive overview on the theory, numerics and applications of probabilistic programming, we refer to, e.g., [8], [9], [10].

From a formal viewpoint, a probabilistic constraint is a conventional constraint $\alpha(z)\geq p$ with $\alpha(z):=\mathbb{P}\left(g(z,\xi)\leq 0\right)$ on the decision vector (because the dependence on ξ vanishes by taking the probability). However, the major difficulty imposed by probabilistic constraints arises from the fact that typically no analytical expression is available for α. All one can hope for, in general, are efficient tools for numerically approximating α. This bears a lot of challenges in the theoretical and algorithmical analysis of such optimization problems, e.g., structure (differentiability, convexity), numerics (computation of gradients) and stability of solutions (perturbations of the underlying probability distribution). The talk addresses some of these aspects along with concrete applications from energy management.

With respect to structural properties, we present some recent results on the Lipschitz continuity of probability functions α as introduced above [5] as well as on the convexity of linear probabilistic constraints with Gaussian coefficient matrix [3, 4]. It is shown that convexity can be derived for sufficiently large probability levels p in (1) which can be explicitely calculated from the given distribution parameters.

The focus of the talk is on gradient formulae for probability functions α which allow to analytically reduce the computation of gradients $\nabla\alpha$ to that of function values $\tilde{\alpha}$ again (possibly with modified distribution parameters). This is possible,

for instance, in case of probabilistic constraints of the form

$$\mathbb{P}(L\xi \leq h(x)) \geq p \quad \text{or} \quad \mathbb{P}(\Xi x \leq h(x)) \geq p$$

where the components of the random vector ξ or of the random coefficient matrix Ξ obey a joint multivariate Gaussian law [1, 6]. The benefit of such reduction is threefold: first, the same efficient codes used to approximate function values can be employed for the computation of gradients; second, proceeding by induction, one may calculate in the same way higher order derivatives; third, thanks to the analytic relation between gradients and function values, there is no risk of increasing the approximation error which is inevitable for function values.

Finally, we touch the aspect of stability of probabilistic optimization problems in the context of empirical (sample average) approximations where the original (typically continuous) probability distribution is approximated by a discrete one (e.g., based on a Monte Carlo sample). Such approach is seductive because it works more or less for arbitrary given distributions and allows to employ techniques from mixed integer programming. While general stability results on probabilistic programming obtained in [2] confirm convergence (actually with Hölder rate with respect to the Kolmogorov distance between probability distributions) of the approximated to the theoretical solution set, simple examples show that the promising exponential term involving the used sample size N does not exclude huge sample sizes needed for a reasonable precision in the approximation of the theoretical solution set even in small dimension [7].

References

[1] W. Van Ackooij, R. Henrion, A. Möller u. R. Zorgati, *On joint probabilistic constraints with Gaussian coefficient matrix*, Operations Research Letters **39** (2011), 99-102.

[2] R. Henrion u. W. Römisch, *Hölder and Lipschitz Stability of Solution Sets in Programs with Probabilistic Constraints*, Mathematical Programming **100** (2004), 589-611.

[3] R. Henrion, *Structural Properties of Linear Probabilistic Constraints*, Optimization **56** (2007), 425-440.

[4] R. Henrion u. C. Strugarek, *Convexity of Chance Constraints with Independent Random Variables*, Computational Optimization and Applications **41** (2008), 263-276.

[5] R. Henrion and W. Römisch, *Lipschitz and differentiability properties of quasi-concave and singular normal distribution functions*, Annals of Operations Research **177** (2010), 115-125.

[6] R. Henrion u. A. Müller, *A gradient formula for linear chance constraints under Gaussian distribution*, Mathematics of Operations Research **37** (2012), 475-488.

[7] R. Henrion, *A critical note on empirical (sample average, Monte Carlo) approximation of solutions to chance constrained programs*, in: D. Hömberg and F. Tröltzsch (eds.): CSMO 2011, IFIP AICT **391**, Springer, Berlin, 2013, pp. 25-37.

[8] A. Prékopa, *Stochastic Programming*, Kluwer, Dordrecht, 1995.

[9] A. Prékopa, *Probabilistic Programming*, Stochastic Programming (A. Ruszczyński and A. Shapiro, eds.), Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.

[10] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming*, MPS-SIAM series on optimization **9**, 2009.

Optimal Experimental Design for Heat Transfer Experiments

Roland Herzog
(joint work with Tommy Etling)

We consider experiments for the determination of the heat transfer coefficient between samples of different materials. The experiment proceeds by heating the two samples to different initial temperatures, then applying the desired contact pressure and observing the temperature evolution by thermographic images. We model the problem as a coupled 1D heat conduction problem for the two temperatures T_1, T_2 in the two probes:

$$\begin{aligned}
\lambda_1 \tfrac{\partial T_1}{\partial n} &= 0 & \lambda_2 \tfrac{\partial T_2}{\partial n} &= 0 && \text{on outer probe boundaries}\\
\varrho_1\, c_{p,1} \tfrac{\partial T_1}{\partial t} &= \lambda_1 \tfrac{\partial^2}{\partial x^2} T_1 & \varrho_2\, c_{p,2} \tfrac{\partial T_2}{\partial t} &= \lambda_2 \tfrac{\partial^2}{\partial x^2} T_2 && \text{within each probe}\\
\lambda_1 \tfrac{\partial T_1}{\partial n} &= \alpha\,(T_2 - T_1) & \lambda_2 \tfrac{\partial T_2}{\partial n} &= \alpha\,(T_1 - T_2) && \text{on inner probe boundaries}\\
T_1(\cdot, 0) &= T_{1,0} & T_2(\cdot, 0) &= T_{2,0} && \text{at initial time.}
\end{aligned}$$

The parameters are the densities ϱ_1, ϱ_2, the heat conduction coefficients λ_1 and λ_2 and the heat transfer coefficient α.

Statistical methods can be used to assess the quality of the parameter estimation. One possibility which is computationally tractable is to calculate confidence regions of the linearized (Gauss-Newton) parameter estimator. Tight confidence regions will be ellipsoidal in shape. Their size is determined by the eigenvalues of the estimator's covariance matrix.

In optimal experimental design, the objective is to choose the experimental conditions so as to maximize the information content of the experiment [2]. In other words, the objective is to reduce the size of the confidence regions. In the heat transfer experiment, experimental conditions available are the initial temperatures of the samples, as well as the choice of spatio-temporal measurements drawn from the wealth of thermographic images. We present numerical experiments for the estimation of the heat transfer coefficient between samples, as well as the thermal conductivity coefficients in each sample. The presentation is based on results obtained in [1].

References

[1] T. Etling, *Optimale Versuchsplanung für die Bestimmung von Parametern bei Wärmeübergangsexperimenten*, Diploma thesis, Chemnitz University of Technology, Germany, 2012,

[2] S. Körkel, *Numerische Methoden für Optimale Versuchsplanungsprobleme bei nichtlinearen DAE-Modellen*, Dissertation, Ruprecht-Karls-Universität Heidelberg, 2002.

Nonlinear Model Predictive Control of Non-Stationary Partial Differential Equations

Ekaterina Kostina
(joint work with Hans Georg Bock, Gregor Kriwet)

Many spatio-temporal processes in the natural and life sciences, and in engineering are described by the mathematical model of non-stationary PDE. It is of high practical relevance as well as a mathematical challenge to use such models for a process optimization subject to numerous important inequality restrictions. However, in the presence of perturbations and model uncertainties the real process will never follow an off-line computed optimal solution. Thus the challenge is to compute feedback controls based on system state and parameter estimation that take these perturbations and uncertainties into account. We present a new optimization method for moving Horizon Estimation (MHE) and Nonlinear Model Prediction Control (NMPC) to meet this challenge. The general principle is to re-estimate states and parameters whenever new information about the process is available, to solve a complete optimal control problem and to apply the first instant of this optimal control as a feedback law. However, the frequency of perturbation information about the system requiring immediate optimal feedback control response is typically orders of magnitude higher than even one single optimization iteration. Therefore we discuss an efficient method for MHE for PDE models which uses the MHE principles for DAE as described in [3, 4]. Then we discuss an innovative multi-level iterations strategy which was first introduce in [1, 2] to make NMPC computations real-time feasible even for PDE constrained optimal control problems. However in the PDE and also the large-scale DAE case, neither does MHE measurement information yield accurate information of the full state and parameter space, nor is this indeed necessary for effective, e.g. stabilizing, NMPC. Hence, a logical further step is the development of a simultaneous MHE and NMPC method "in one step". This requires both the theoretical investigation of joint sensitivity, conditioning and stability analysis of coupled MHE and NMPC and an efficient generalization of the existing algorithmic approaches to combined multi-level real-time iterations. Another open research topic is the so-called "dual" control problem, which simultaneously optimizes the controls for both experimental design for information and for process performance related criteria, thus combining, e.g., the cost function as a performance criterion to be optimized and the contribution of inaccuracies to loss of performance (e.g. the resulting variance of the cost function) to be minimized. As a result, certain degrees of freedom in the controls are exploited to reduce the inaccuracies leading to the most significant performance losses.

References

[1] H. G. Bock, M. Diehl, E. Kostina, and J. P. Schlöder, *Constrained optimal feedback control of systems governed by large differential algebraic equations*, In: L. T. Biegler, O. Ghattas,

M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders (Eds.): Real-Time PDE-Constrained Optimization, SIAM Series: Computational Science & Engineering (2007), 3–24.

[2] H. G. Bock, M. Diehl, P. Kühl, E. Kostina, J. P. Schlöder, and L. Wirsching, *Numerical methods for efficient and fast nonlinear model predictive control*, In: R. Findeisen et al. (Eds.): Assessment and Future Directions, Lecture Notes in Computer Science **358** (2007), 163–179.

[3] M. Diehl, P. Kühl, J. P. Schlöder, and H. G. Bock, *Schnelle Algorithmen für die Zustands- und Parameterschätzung auf bewegten Horizonten*, Automatisierungstechnik **54** (2006), 602–613.

[4] P. Kühl, M. Diehl, T. Kraus, J. P. Schlöder, and H. G. Bock, *A real-time algorithm for moving horizon state and parameter estimation*, Computers and Chemical Engineering **35** (2011), 71–83.

An Approach for the Adaptive Solution of Optimization Problems Governed by Partial Differential Equations with Uncertain Coefficients

Drew P. Kouri

(joint work with Matthias Heinkenschloss, Denis Ridzal, Bart G. van Bloemen Waanders)

The numerical solution of optimization problems governed by partial differential equations (PDEs) with random coefficients is computationally challenging because of the large number of deterministic PDE solves required at each optimization iteration. In this talk, we introduce an efficient algorithm for solving such problems based on a combination of adaptive sparse-grid collocation for the discretization of the PDE in the stochastic space and a trust-region framework for optimization and fidelity management of the stochastic discretization. The overall algorithm adapts the collocation points based on the progress of the optimization algorithm and the impact of the random variables on the solution of the optimization problem. It frequently uses few collocation points initially and increases the number of collocation points only as necessary, thereby keeping the number of deterministic PDE solves low while guaranteeing convergence. Currently an error indicator is used to estimate gradient errors due to adaptive stochastic collocation. The algorithm is applied to three examples, and the numerical results demonstrate a significant reduction in the total number of PDE solves required to obtain an optimal solution when compared with a Newton conjugate gradient algorithm applied to a fixed high-fidelity discretization of the optimization problem.

References

[1] D. P. Kouri, M. Heinkenschloss, D. Ridzal, B. G. van Bloemen Waanders, *A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty*, Tech. Report ANL/MCS-P3035-0912, Department of Mathematics and computer Science, Argonne National Laboratory, 2012.

On Optimal Control of the Bidomain Equations

Karl Kunisch

The bidomain equations describe the electrical activity of the heart. Extracellular current density stimuli are used to act as controls to achieve control objectives as for instance defibrillation of arrhythmias. An open loop formulation is proposed, mathematically justified, and numerically realized for simplified geometries. For closed loop control approximation of the Hamilton-Jacobi-Bellman by model reduction techniques is discussed. The closed loop formulation is robust against system perturbations.

Sampling and Low-Rank Tensor Approximations

Alexander Litvinenko
(joint work with Hermann G. Matthies)

Computing of a response surface (another name is a 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 safe 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 $\mathbf{v}_i \in \mathbb{R}^n$, $i = 1..Z$, 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 $\mathbf{W} = (\mathbf{v}_1, ..., \mathbf{v}_Z) \in \mathbb{R}^{n\times Z}$ and consider the approximation (with accuracy ε)

$$\mathbf{W} \approx \mathbf{W}_k = \mathbf{A}\mathbf{B}^T, \text{ where } \quad \|\mathbf{W} - \mathbf{W}_k\| < \varepsilon,\ k \ll \min\{n, Z\}. \tag{1}$$

To compute factors $\mathbf{A} \in \mathbb{R}^{n\times k}$ and $\mathbf{B} \in \mathbb{R}^{Z\times k}$ we compute the truncated singular value decomposition $\mathbf{W}_k = \mathbf{U}_k\Sigma_k{\mathbf{V}_k}^T$, where $\mathbf{U}_k \in \mathbb{R}^{n\times k}$ contains the first k columns of $\mathbf{U}$, $\mathbf{V}_k \in \mathbb{R}^{Z\times k}$ contains the first k columns of $\mathbf{V}$ and $\Sigma_k \in \mathbb{R}^{k\times k}$ contains the k-biggest singular values of Σ. We denote $\mathbf{A} = \mathbf{U}_k\Sigma_k$ and $\mathbf{B} = \mathbf{V}_k$.

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

$$\overline{\mathbf{v}} = \frac{1}{Z}\sum_{i=1}^{Z}\mathbf{v}_i \approx \overline{\mathbf{v}}_k = \frac{1}{Z}\sum_{i=1}^{Z}\mathbf{A}\cdot\mathbf{b}_i = \mathbf{A}\overline{\mathbf{b}}. \tag{2}$$

The computational complexity is $\mathcal{O}(k(Z+n))$, besides $\mathcal{O}(nZ)$) for usual dense data format. By definition, the covariance matrix is $\mathbf{C} = \frac{1}{Z-1}\mathbf{W}_c\mathbf{W}_c^T$. Now, the covariance matrix can be approximated like

$$\mathbf{C} \approx \frac{1}{Z-1}\mathbf{U}_k\Sigma_k\Sigma_k^T\mathbf{U}_k^T = \frac{1}{Z-1}\mathbf{A}\mathbf{A}^T. \tag{3}$$

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

Lemma: Let $\|\mathbf{W}-\mathbf{W}_k\|_2 \le \varepsilon$, and $\overline{\mathbf{u}}_k$ be a rank-k approximation of the mean $\overline{\mathbf{u}}$. Then a) $\|\overline{\mathbf{u}}-\overline{\mathbf{u}}_k\| \le \frac{\varepsilon}{\sqrt{Z}}$, b) $\|\mathbf{C}-\mathbf{C}_k\| \le \frac{1}{Z-1}\varepsilon^2$.

For the proof see [4, 3].

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

2. Response surface and its low-rank approximation

Let $\mathbf{v}(x,\boldsymbol{\theta})$ be the solution (or a functional of the solution). It can be pressure, density, velocity, lift, drag etc. $\mathbf{v}(x,\boldsymbol{\theta})$ can be approximated in a set of new independent Gaussian random variables (polynomial chaos expansions (PCE) of Wiener [6])

$$\mathbf{v}(x,\boldsymbol{\theta}(\omega)) \approx \sum_{\beta\in\mathcal{J}_{M,p}}\mathbf{v}_\beta(x)H_\beta(\boldsymbol{\theta}) = [...\mathbf{v}_\beta(x)...][...H_\beta(\boldsymbol{\theta})...]^T, \tag{4}$$

where $\boldsymbol{\theta}(\omega) = (\theta_1(\omega),...,\theta_M(\omega))$, $\mathbf{v}_\beta(x)$ are coefficients, $H_\beta(\boldsymbol{\theta})$ the multivariate Hermite polynomials, $\mathcal{J}_{M,p}$ a finite multi-index subset of infinite dimensional multi-index set and p the maximal polynomial order $H_\beta(\boldsymbol{\theta})$. Since Hermite polynomials are orthogonal, the coefficients $\mathbf{v}_\beta(x)$ can be computed by projection:

$$\mathbf{v}_\beta(x) = \frac{1}{\beta!}\int_\Theta H_\beta(\boldsymbol{\theta})\mathbf{v}(x,\boldsymbol{\theta})\,\mathbb{P}(d\boldsymbol{\theta}) \approx \frac{1}{\beta!}\sum_{i=1}^{n_q}H_\beta(\boldsymbol{\theta}_i)\mathbf{v}(x,\boldsymbol{\theta}_i)w_i,$$

where the multidimensional integral over Θ is computed approximately, for example, on a sparse Gauss-Hermite grid where $\boldsymbol{\theta}_i$ are quadrature points, w_i corresponding weights and n_q is the number of quadrature points. Using the rank k

approximation $\mathbf{AB}^T$ of the set of realisations, obtain

$$\mathbf{v}_\beta(x) = \frac{1}{\beta!}[\mathbf{v}(x,\boldsymbol{\theta}_1),...,\mathbf{v}(x,\boldsymbol{\theta}_{n_q})] \cdot [H_\beta(\boldsymbol{\theta}_1)w_1,...,H_\beta(\boldsymbol{\theta}_{n_q})w_{n_q}]^T \approx \mathbf{AB}^T\mathbf{c}_\beta, \tag{5}$$

where vector $\mathbf{c}_\beta := \frac{1}{\beta!}[H_\beta(\boldsymbol{\theta}_1)w_1,...,H_\beta(\boldsymbol{\theta}_{n_q})w_{n_q}]^T$. The matrix of all PCE coefficients will be $[...\mathbf{v}_\beta(x)...] = \mathbf{AB}^T[...\mathbf{c}_\beta...]$, $\beta \in \mathcal{J}_{M,p}$. Put all together, obtain a low-rank representation of RS

$$\mathbf{v}(x,\boldsymbol{\theta}) \approx \mathbf{AB}^T[...\mathbf{c}_\beta...](...,H_\beta(\boldsymbol{\theta}),...)^T. \tag{6}$$

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 $\boldsymbol{\theta}_{n_q+1}$ and evaluate the response surface Eq. 6 in this point. Let $u(x,\boldsymbol{\theta}_{n_q+1})$ be the obtained solution.
(2) Compute residual $\|\mathbf{r}(u(x,\boldsymbol{\theta}_{n_q+1}))\|$. Only if $\|\mathbf{r}\|$ is large solve expensive deterministic problem. It can be also used, e. g., as a start value in an iterative process.
(3) Update $\mathbf{A}$, $\mathbf{B}^T$, $[...\mathbf{c}_\beta...]$ 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 $\boldsymbol{\theta}_{n_q+i}$, $i = 1, 2, ...$ 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

[1] M. Brand. Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra Appl.*, 415(1):20–30, 2006.
[2] M. Espig, W. Hackbusch, A. Litvinenko, H. G. Matthies, and E. Zander. Efficient analysis of high dimensional data in tensor formats. *Sparse Grids and Applications*, editors J. Garcke and M. Griebel, Vol. 88 of *LNCS*, pages 31–56, 2013.
[3] A. Litvinenko and H. G. Matthies and T. A. El-Moselhy, , Sampling and Low-Rank Tensor Approximation of the Response Surface. *Accepted for publication in MCQMC Proceedings, Editors: J. Dick, F. Y. Kuo, G. W. Peters, I. H. Sloan*, 2013.
[4] A. Litvinenko and H. G. Matthies. Uncertainty quantification in numerical aerodynamic via low-rank response surface. *PAMM*, 12(1):781–784, 2012.
[5] W. Nowak and and A. Litvinenko Kriging and spatial design accelerated by orders of magnitude: combining low-rank covariance approximations with FFT-techniques. *Accepted for publication in J. Mathematical Geosciences*, 2013.

[6] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60:897–936, 1938.

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

$$q_a(\cdot) = q_a({}^0H, \ldots, {}^kH, \ldots, {}^nH) = \sum_{k=0}^{n} {}^kH z^{\otimes k}, \tag{1}$$

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 $\mathcal{O}(n^\alpha N^\beta L^\gamma)$, 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

[1] A. Kučerová, J. Sýkora, B. Rosić, H. G. Matthies. Acceleration of uncertainty updating in the description of transport processes in heterogeneous materials, *J. Computational and Applied Mathematics*, 236 (2012) 4862-4872.

[2] H. G. Matthies, A. Litvinenko, O. Pajonk, B. V. Rosić, E. Zander. Parametric and uncertainty computations with tensor product representations. *In: A. Dienstfrey, R. Boisvert (Eds.), Uncertainty Quantification in Scientific Computing*, volume 377 of IFIP Advances in Information and Communication Technology, Springer, Berlin, 2012, pp. 139-150.
[3] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 194 (2005) 1295-1331.
[4] H. G. Matthies and E. Zander. Solving stochastic systems with low-rank tensor compression. *Linear Algebra and its Applications* 436 (2012) 3819-3838.
[5] O. Pajonk, B. V. Rosić, A. Litvinenko, H. G. Matthies. A deterministic filter for non-Gaussian Bayesian estimation—applications to dynamical system estimation with noisy measurements. *Physica D*, 241 (2012) 775-788.
[6] B. V. Rosić, A. Litvinenko, O. Pajonk, H. G. Matthies. Sampling-free linear Bayesian update of polynomial chaos representations. *Journal of Computational Physics*, 231 (2012) 5761-5787.
[7] G. Saad, R. Ghanem. Characterization of reservoir simulation models using a polynomial chaos-based ensemble Kalman filter. *Water Resources Research* 45 (2009) W04417.
[8] E. Zander. Stochastic Galerkin library (sglib). *https://github.com/ezander/sglib*.

Accurate Discretization Schemes for a Class of Fokker-Planck Equations

Masoumeh Mohammadi

For a class of Fokker-Planck equations two different discretization schemes are investigated; considering two cases of bounded and unbounded domains. These equations of parabolic type govern the time evolution of the probability density function of stochastic processes. For the bounded domain, the discretization schemes combine the Chang-Cooper method for spatial discretization with backward first- and second-order finite differencing in time to obtain stable and accurate solutions that satisfy conservation and positivity properties of the probability density function. These properties are theoretically proven and validated by numerical experiments. For the case of unbounded domains, the Hermite spectral discretization method is applied and analyzed. This method is more expensive, but provides us an appropriate means to treat the unbounded domain, and moreover the error decays exponentially as the number of expansion terms in the spatial discretization increases. The results are then generalized to discretize and analyze a Fokker-Planck optimal control system in an unbounded domain.

References

[1] M. Annunziato, A. Borzì, *A Fokker-Planck control framework for multidimensional stochastic processes*, Journal of Computational and Applied Mathematics **237** (2013), 487–507.
[2] J. C. M. Fok, and B. Guo and T. Tang, *Combined Hermite spectral-finite difference method for the Fokker-Planck equation*, Mathematics of computation **240** (2001), 1497–1528.
[3] J. S. Chang and G. Cooper, *A Practical Difference Scheme for Fokker-Planck Equation*, Journal of Computational Physics **6** (1970), 1-16.

Collocation Approaches for Forward Uncertainty Propagation in PDE Models with Random Input Data

Fabio Nobile
(joint work with Raul Tempone)

Mathematical models are widely used in physics and engineering applications as predictive tools. However, in many situations, the input parameters of the model are uncertain due to either a lack of knowledge or an intrinsic variability of the system. Examples are the study of subsurface phenomena, biological tissues, complex materials, whose properties are often heterogeneous, not perfectly characterized and, possibly, changing in time in an uncertain way.

In this work we consider the case in which the uncertainty can be described reasonably well in a probabilistic setting and we focus on the problem of effectively propagating it from the input parameters to the output quantities of interest of the mathematical model. In particular we focus on numerical methods of collocation type that imply solving the problem for a well chosen set of input parameters and make inference on the statistical properties of the output quantities based on the corresponding evaluations.

We consider an abstract problem defined in a domain $D \subset \mathbb{R}^d$ which could involve spatial and temporal variables

$$\text{find } u: \quad \mathcal{L}(\mathbf{y})(u) = \mathcal{F} \qquad \text{in } D \tag{1}$$

where $\mathcal{L}$ is an operator involving partial derivatives of u and $\mathcal{F}$ is the forcing term. Problem (1) has typically to be endowed with suitable boundary or initial conditions on subsets of ∂D.

The operator $\mathcal{L}$ is supposed to depend on a vector of N random parameters $\mathbf{y} = (y_1, \ldots, y_N) \in \mathbb{R}^N$ taking values in a hypercube $\Gamma = \prod_{n=1}^N \Gamma_n \subset \mathbb{R}^N$ and with a known joint probability density function $\rho : \Gamma \to \mathbb{R}_+$. The case $N = \infty$ is also relevant in applications as it allows to tackle cases in which the operator depends on random fields distributed over the domain D (or subsets of it). Only for the sake of exposition, we here assume that the random parameters y_n are statistically independent so that the joint density factorizes as $\rho(\mathbf{y}) = \prod_{n=1}^N \rho_n(y_n)$.

We also assume that problem (1) is well posed for any possible value of the random vector $\mathbf{y} \in \Gamma$, so that there exists a unique solution $u(\mathbf{y})$ in a suitable Hilbert space V and $\|u(\mathbf{y})\|_V \leq C(\mathbf{y})\|\mathcal{F}\|_{V'}$ where the random variable $C(\mathbf{y})$ is L^p integrable with respect to the measure $\rho(\mathbf{y})d\mathbf{y}$ for some $p \geq 2$. Under these assumptions we can think of u as a Hilbert-valued function $u : \Gamma \to V$ which belongs to the Bochner space $L^p_\rho(\Gamma; V)$.

A collocation approach aims at constructing an approximation $\tilde{u}(\mathbf{y})$ of the exact map $u(\mathbf{y})$ starting from point evaluations $u(\mathbf{y}_i)$ for some well chosen points $\mathbf{y}_i \in \Gamma$, $i = 1, \ldots, M$. In particular we look for multivariate polynomial approximations $\tilde{u} \in \mathbb{P}_\Lambda \otimes V$ where $\Lambda \subset \mathbb{N}^N$ is an index set and $\mathbb{P}_\Lambda = \text{span}\{\prod_{n=1}^N y_n^{p_n},\ (p_1, \ldots, p_n) \in \Lambda\}$. The use of global polynomials is sound in many applications because of the high regularity of the map $u(\mathbf{y})$. It is well known, indeed, that for linear elliptic

problems with random diffusion coefficients, the map $\mathbf{y} \mapsto u(\mathbf{y})$ can even be analytic in $\mathbf{y}$ (see e.g. [1, 2]). This is however not the case for a hyperbolic second order equation with random wave speed where only finite regularity has to be expected, in general, as shown in [8] and [9].

Collocation on sparse grids. The first approach we propose is based on the well known Smolyak sparse grid construction (see e.g. [3]) for high dimensional polynomial interpolation. For each random variable y_n, let $i_n > 0$ denote the level of interpolation, $\{y_{n,j}^{i_n}\}_{j=1}^{m(i_n)}$ be a set of good interpolation knots, which could be either Gauss abscissas with respect to the density ρ_n, or nested Gauss abscissas (Gauss-Patterson knots) or Clenshaw-Curtis abscissas for bounded random variables. Here $m(i_n)$ is the number of collocation points used at level i_n and is an increasing function of its argument. We denote by $\mathcal{U}_n^{m(i_n)}$ the 1D polynomial interpolant operator based on the knots $\{y_{n,j}^{i_n}\}_{j=1}^{m(i_n)}$, with $\mathcal{U}_n^{m(0)} = 0$, and by $\Delta^{m(\mathbf{i})} = \bigotimes_{n=1}^{N}(\mathcal{U}_n^{m(i_n)} - \mathcal{U}_n^{m(i_n-1)})$ the hierarchical surplus associated to the multi-index $\mathbf{i} = (i_1, \ldots, i_N) \in \mathbb{N}^N$. Given an index set $\Lambda \subset \mathbb{N}^N$ the sparse grid approximation of u is defined by

$$\tilde{u} = \mathcal{S}_\Lambda u = \sum_{\mathbf{i}\in\Lambda} \Delta^{m(\mathbf{i})} u. \tag{2}$$

Following [3], we propose to optimize the construction following a knapsack approach and select the index set Λ_{opt} corresponding to the most profitable terms in the series. Here, the profit $P(\mathbf{i})$ associated to a hierarchical surplus $\Delta^{m(\mathbf{i})}u$ is defined as $P(\mathbf{i}) = \max_{\mathbf{j}\geq\mathbf{i}} \Delta E(\mathbf{j})/\Delta W(\mathbf{j})$ where $\Delta E(\mathbf{i})$ is the estimated error decrease when adding the term $\Delta^{m(\mathbf{i})}u$ to the approximation in (2) and $\Delta W(\mathbf{i})$ is the corresponding cost (number of evaluation points in $\Delta^{m(\mathbf{i})}u$). In [10] it has been proven that if $\sum_{\mathbf{i}\in\mathbb{N}^N} P(\mathbf{i})^\tau \Delta W(\mathbf{i}) < \infty$ for some $\tau < 1$, then

$$\|u - \mathcal{S}_{\Lambda_{opt}} u\|_{L^2_\rho(\Gamma;V)} \leq C M^{1-\frac{1}{\tau}}$$

where M is the total number of knots included in the sparse grid. This result, applied to a linear elliptic equation with N non-overlapping inclusions of random uniformly distributed conductivity allows us to show that the stochastic collocation method on optimized sparse grids converges with a rate faster than algebraic (see [10] for details). A posterior estimates for the profits based on inexpensive computations have also been proposed in [10], which make this approach practical.

Discrete L^2 projection from random evaluations. An alternative approach to sparse grid approximation consists in taking M independent samples, $\{\mathbf{y}_j\}_{j=1}^M$, distributed with the measure $\rho(\mathbf{y})d\mathbf{y}$ and then computing $\tilde{u}$ by a discrete L^2_ρ projection onto the polynomial space $\mathbb{P}_\Lambda$:

$$\tilde{u} = \operatorname*{argmin}_{v\in\mathbb{P}_\Lambda\otimes V} \frac{1}{M}\sum_{i=1}^{M} \|u(\mathbf{y}_i) - v(\mathbf{y}_i)\|_V^2. \tag{3}$$

The crucial question is how large the sample size M should be with respect to the dimension $|\mathbb{P}_\Lambda|$ of the polynomial space $\mathbb{P}_\Lambda$ to achieve a stable and optimally

convergent approximation. Results in the monovariate case can be found in [5, 6] and their generalization to the multivariate case are given in [4]. In particular, we show that, for N uniformly distributed independent random variables, the condition $M \propto |\mathbb{P}_\Lambda|^2$ guarantees stability of the discrete projection (3). More precisely, the following optimality result is shown in [4]: for any $\gamma > 0$, if $\frac{M}{\log M} \geq C(1+\gamma)|\mathbb{P}_\Lambda|^2$, then, with probability larger than $1 - 2M^{-\gamma}$, we have,

$$\|u - \tilde{u}\|_{L^2_\rho(\Gamma;V)} \leq (1+\sqrt{2}) \inf_{v \in \mathbb{P}_\Lambda \otimes V} \|u - v\|_{L^\infty(\Gamma;V)}.$$

Application of this technique to PDEs with random coefficients can be found in [7]. Still, our numerical experience shows that for a moderate number of random variables, the condition $M \propto |\mathbb{P}_\Lambda|^2$ is over-constraining and a linear relation is often enough for practical purposes, making the discrete projection (3) particularly attractive for moderately high dimensional approximations.

References

[1] I. Babuška, F. Nobile, and R. Tempone, "A stochastic collocation method for elliptic partial differential equations with random input data," *SIAM Review*, vol. 52, pp. 317–355, 2010.

[2] J. Beck, F. Nobile, L. Tamellini, and R. Tempone, "On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods," *Math. Mod. Methods Appl. Sci. (M3AS)*, vol. 22, no. 9, pp. 1250023–1, 1250023–33, 2012.

[3] H.J Bungartz and M. Griebel, "Sparse grids," *Acta Numer.*, 13:147–269, 2004.

[4] A. Chkifa, A. Cohen, G. Migliorati, F. Nobile, R. Tempone. "Discrete least squares polynomial approximation with random evaluations; application to parametric and stochastic elliptic PDEs", in preparation.

[5] A. Cohen, M.A. Davenport and D. Leviatan, "On the stability and accuracy of Least Squares approximations", to appear in JFoCM, 2012.

[6] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone, "Analysis of the discrete L^2 projection on polynomial spaces with random evaluations," MOX-Report 46/2011, Department of Mathematics, Politecnico di Milano, Italy, 2011. submitted.

[7] G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone, "Approximation of quantities of interest in stochastic PDEs by the random discrete L^2 projection on polynomial spaces," *SIAM J. Sci. Comp.*, 2013. accepted for publication.

[8] M. Motamed, F. Nobile, and R. Tempone, "Analysis and computation of the elastic wave equation with random coefficients," MATHICSE Technical Report 32.2012, École Polytechnique Fédérale de Lausanne, 2012. submitted.

[9] M. Motamed, F. Nobile, and R. Tempone, "A stochastic collocation method for the second order wave equation with a discontinuous random speed," *Numer. Math*, vol. 123, no. 3, pp. 493–536, 2013.

[10] L. Tamellini. *Polynomial approximation of PDEs with stochastic coefficients*. PhD thesis, Politecnico di Milano, 2012.

Tractability of Multivariate Problems

Erich Novak

We start with an example: What can we say about the optimal approximation of very smooth functions from the class

$$F_d = \{f : [0,1]^d \to \mathbb{R} \mid \ \|D^\alpha f\|_\infty \leq 1 \text{ for all } \alpha \in \mathbb{N}_0^d\}?$$

We allow all algorithms that use n "pieces of information" $L_1(f), \dots, L_n(f)$, the L_i are linear functionals. It turns out that linear algorithms

$$S_n(f) = \sum_{i=1}^{n} L_i(f)\, g_i$$

are optimal and it is well known that every order of convergence can be achieved for F_d. If

$$e(S_n) = \sup_{f \in F_d} \|f - S_n(f)\|_\infty$$

is the error of S_n, then the complexity of the problem is given by

$$e(n,d) = \inf_{S_n} e(S_n) \qquad \text{or} \qquad n(\varepsilon, d) = \inf\{n \mid e(n,d) \le \varepsilon\}$$

and it is well known that for any d and $r > 0$

$$e(n,d) = \mathcal{O}(n^{-r}) \qquad \text{or} \qquad n(\varepsilon, d) = \mathcal{O}(\varepsilon^{-1/r}).$$

Conventional conclusion: The problem is easy since the order of convergence is excellent. We show that this conclusion is wrong: The order of convergence is not a good measure for the complexity, here we have the curse of dimensionality.

Theorem (Novak and Woźniakowski, 2009): For L_∞-approximation over F_d we have

$$e(n,d) = 1 \quad \text{for all} \quad n \le 2^{\lfloor d/2 \rfloor} - 1$$

or

$$n(\varepsilon, d) \ge 2^{\lfloor d/2 \rfloor} \quad \text{for all} \quad \varepsilon \in (0,1).$$

The proof is short: Take $s = \lfloor d/2 \rfloor$ and consider $f : [0,1]^d \to \mathbb{R}$,

$$f(x) = \sum_{i \in \{0,1\}^s} a_i (x_1 + x_2)^{i_1} (x_3 + x_4)^{i_2} \dots (x_{2s-1} + x_{2s})^{i_s}.$$

The space V_d of such functions has dimension 2^s and

$$\|f\|_\infty = \sup_\alpha \|D^\alpha f\|_\infty \quad \text{for all } f \in V_d.$$

For continuous $N : V_d \to \mathbb{R}^{2^s - 1}$, there is a $f \in V_d$ with $\|f\|_\infty = 1$ such that $N(f) = N(-f)$; this follows from the Borsuk-Ulam Theorem. Hence $S_n(f) = \phi(N(f)) = S_n(-f)$ and $e(S_n) \ge 1$ for $n = 2^s - 1$.

Open problem: Consider the same class

$$F_d = \{ f : [0,1]^d \to \mathbb{R} \mid \|D^\alpha f\|_\infty \le 1 \text{ for all } \alpha \in \mathbb{N}_0^d \}$$

and numerical integration,

$$S_d(f) = \int_{[0,1]^d} f(x)\, \mathrm{d}x.$$

Is this problem tractable? Curse of dimension? For the larger classes

$$F_d^k = \{ f : [0,1]^d \to \mathbb{R} \mid \|D^\alpha f\|_\infty \le 1 \text{ for all } \alpha \in \mathbb{N}_0^d,\ |\alpha| \le k \},$$

we have the curse of dimension, see the recent paper by Hinrichs, Novak, Ullrich, Woźniakowski (2013).

Usually, we cannot obtain tractability even by strong smoothness assumptions, see the L_∞ approximation problem for C^∞ functions. Sometimes: yes, we present recent results about the tractability of the star discrepancy by Heinrich, Novak, Wasilkowski, Woźniakowski (2001) and Hinrichs (2004).

One often can obtain tractability by using additional properties concerning the "structure" of the functions. For example, one often has "partially separable functions". A function $f : [0,1]^d \to \mathbb{R}$ of many variables (d large) may be a sum of functions, that only depend on k variables (k small):

$$f(x_1, x_2, \ldots, x_d) = \sum_\ell g_\ell(x_{i_1}, x_{i_2}, \ldots, x_{i_k}).$$

In optimization such functions are called "partially separable". As a rule: Problems are tractable for such functions (with k fixed and $d \to \infty$), even if the g_ℓ are not very smooth.

More generally, one can study weighted Sobolev spaces. See, for example, Werschulz, Woźniakowski (2009). Unit ball of the space $H_{d,\gamma}$ given by all $f : [0,1]^d \to \mathbb{R}$ with

$$\|f\|^2 = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \int_{[0,1]^d} \left(\frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}} f(x) \right)^2 dx \le 1,$$

where $[d] := \{1, 2, \ldots, d\}$ and $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ are non-negative weights. Results for L_2 approximation for linear (or continuous) information $\Lambda^{\rm all}$ and for function values $\Lambda^{\rm std}$:

- For equal weights $\gamma_{d,\mathfrak{u}} = 1$ the problem is weakly tractable for $\Lambda^{\rm all}$ and not weakly tractable for $\Lambda^{\rm std}$.
- For bounded finite order weights ($\gamma_{d,\mathfrak{u}} = 0$ if $|\mathfrak{u}| > k$) the problem is always polynomially tractable, even for $\Lambda^{\rm std}$.

Various Weights:

- Product weights: $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$. Then
$$H(K_{d,\gamma}) = H(K_{1,\gamma_{d,1}}) \otimes \cdots \otimes H(K_{1,\gamma_{d,d}})$$
and $\gamma_{d,j}$ moderates the influence of x_j
- Finite-order weights:
$\gamma_{d,\mathfrak{u}} = 0$ for all $|\mathfrak{u}| > k$. Then
$$f = \sum_{\mathfrak{u} \subseteq [d], |\mathfrak{u}| \le k} f_{\mathfrak{u}}$$
is a sum of functions depending on at most k variables.

One can model various properties of f by suitable weights. We present various results for integration by Gnewuch, Woźniakowski (2008), Novak, Woźniakowski (2001, 2010), Sloan, Woźniakowski (1998, 2002).

Some problems are not tractable for deterministic algorithms but they are tractable for randomized algorithms. We present several such results.

Solving integral equations with random bits: Compute $u(s)$, integral equation

$$u(x) - \int_{[0,1]^d} k(x,y)u(y)\,\mathrm{d}y = f(x)$$

on $[0,1]^d$ with Lipschitz kernel k, $\|k\|_\infty < \alpha < 1$ and right hand side. Optimal order with MC (Heinrich & Mathé 1993) $e_n \asymp n^{-1/2-1/(2d)}$. With a discretized version of classical MC and results for summation we get the upper bound

$$\text{cost} \le \varepsilon^{-2} + d\,(\log \varepsilon^{-1})^2,$$

only $d\,(\log \varepsilon^{-1})^2$ random bits are needed, see Novak, Pfeiffer (2004).

We present recent results by Daniel Rudolf about Markov chain Monte Carlo: explicit error bounds lead to tractability.

Optimal importance sampling was recently studied by Aicke Hinrichs: $S(f) = \int_{\mathbb{R}^d} f(x)\varrho(x)\,\mathrm{d}x$ for $f \in H$, H a RKHS with

$$\|S\|^2 = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(x,y)\varrho(x)\varrho(y)\,\mathrm{d}x\,\mathrm{d}y < \infty.$$

Randomized error $e(A_n) = \sup_{\|f\|_H \le 1}(E(S(f) - A_n(f))^2)^{1/2}$.

Hinrichs 2010 proved: If $K(x,y) \ge 0$ then with importance sampling

$$e(A_n) \le \left(\frac{\pi}{2}\right)^{1/2} n^{-1/2}\,\|S\|.$$

Hence such problems are strongly polynomially tractable.

Summary: Many problems for functions $f : [0,1]^d \to \mathbb{R}$ are intractable, if considered in the worst case setting for classical function spaces, like $C^k([0,1]^d)$.

There are two major remedies:

- Problems with a structure (weighted spaces)
- Randomized algorithms.

Towards Real-Time Optimization for PDE

Andreas Potschka

(joint work with Hans Georg Bock)

Closed-loop feedback control constitutes a widely used approach to treat uncertainties in time-dependent problems that allow for repeated measurements. We describe an extension of the Real-Time Iteration (RTI) [2] for Nonlinear Model Predictive Control (NMPC) of ordinary differential equations (ODE) to the case with parabolic partial differential equations (PDE). Our work is based on recent advances in the numerical treatment of instationary PDE constrained optimization problems on the basis of a two-grid Newton-Picard decomposition approach [3, 4]. The extension can be analyzed and implemented as an additional level in the Multi-Level Iteration [1] (see also talk by E. Kostina in this workshop).

We assume the Bayesian view of having only uncertain knowledge of the true model parameters and the current system state. However, our knowledge can be

improved through measurements, even though it is usually impossible to measure the complete system state. Via Moving Horizon Estimation (MHE), which determines parameter and state realizations of maximum likelihood and their covariances regarding the most recent measurements on a finite time horizon of the past, we can find the data for an NMPC problem on a finite time horizon of the future. The NMPC result is either a local feedback law or a short control move that we apply to the system and update repeatedly when new measurements become available. The main challenge is to solve the coupled MHE-NMPC problem in real-time and to reduce the feedback delay between measurement arrival and control update to a minimum. Depending on the system under consideration, real-time feasible feedback rates can vary from microseconds (mechanical systems) to minutes (chemical plants). We focus here on the NMPC problem.

We discretize the PDE-NMPC problem first in space and then in time via Direct Multiple Shooting. Through smart grouping of variables into PDE and non-PDE variables $(x_1, x_2) \in \mathbb{R}^{n_1+n_2}, n_1 \gg n_2$, and regrouping of the constraints such that PDE initial value and matching constraints come first, we arrive at a large-scale structured Nonlinear Programming (NLP) problem denoted by

$$\begin{aligned} \min_{(x_1,x_2)\in\mathbb{R}^{n_1+n_2}} \quad & f(x_1,x_2) \\ \text{s.t.} \quad & g_i(x_1,x_2) = 0, \quad i \in \mathcal{E}_1, |\mathcal{E}_1| = n_1, \\ & g_i(x_1,x_2) = 0, \quad i \in \mathcal{E}_2, \\ & g_i(x_1,x_2) \geq 0, \quad i \in \mathcal{I}. \end{aligned}$$

We need to solve a sequence of Quadratic Problems (QP) of the form

$$\begin{aligned} \min \quad & \frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ \text{s.t.} \quad & C_{11}x_1 + C_{12}x_2 = c_1, \quad C_{21}x_1 + C_{22}x_2 = c_2, \quad C_{31}x_1 + C_{32}x_2 \geq c_3. \end{aligned}$$

Our goal is to eliminate all PDE variables x_1 from the QP in a structure exploiting pre- and postprocessing step. The matrix C_{11} has a lot of structure: Due to multiple shooting, it is a large negative unit lower triangular matrix with shooting matrices G^i on the block subdiagonal. Thus, C_{11} is invertible. However, the G^i matrices are large and dense, prohibiting efficient numerical inversion. But if we approximate the constraint and Hessian matrices in a two-grid Newton-Picard fashion, we can still expect fast convergence of the (now inexact) SQP [3] and obtain an efficient formula for the inverse in the following way: We use two spatial discretization grids, denote the prolongation matrix by P, and construct a restriction matrix that satisfies $RP = \mathbf{I}$. Then, we approximate the derivatives $\widetilde{(\,)}$ by their coarse grid counterparts $\hat{(\,)}$, for example $\widetilde{G}^i = P\hat{G}^i R$.

Theorem 1. *With the projectors* $\Pi^{\text{slow}} = \mathbf{I}_{n_{\text{MS}} \times n_{\text{MS}}} \otimes (PR)$ *and* $\Pi^{\text{fast}} = \mathbf{I} - \Pi^{\text{slow}}$, *it holds that* $\widetilde{C}_{11}^{-1}\Pi^{\text{slow}} = (\mathbf{I} \otimes P)\hat{C}_{11}^{-1}(\mathbf{I} \otimes R)$ *and* $\widetilde{C}_{11}^{-1}\Pi^{\text{fast}} = -\Pi^{\text{fast}}$.

In other words, the Newton-Picard approximation of C_{11} acts as a Newton iteration on the slow modes, but as a Picard iteration on the fast modes.

Corollary 2 (Efficient inverse formula)**.** $\widetilde{C}_{11}^{-1} = (\mathbf{I} \otimes P)\left(\hat{C}_{11}^{-1} + \mathbf{I}\right)(\mathbf{I} \otimes R) - \mathbf{I}$.

The approximated QP can then be solved efficiently by virtue of the following theorem, where we have dropped the tildes.

Theorem 3 (Partial nullspace approach)**.** *Let*

$$\begin{aligned} Z &= \begin{pmatrix} -C_{11}^{-1}C_{12} \\ \mathbf{I} \end{pmatrix}, \quad c_1' = C_{11}^{-1}c_1, \qquad b' = B_{21}c_1' + b_2 - C_{12}^{\mathrm{T}}C_{11}^{-\mathrm{T}}(B_{11}c_1' + b_1), \\ B' &= Z^{\mathrm{T}}BZ, \qquad\quad c_2' = c_2 - C_{21}c_1', \quad C_2' = C_{22} - C_{21}C_{11}^{-1}C_{12}, \\ &\qquad\quad\; c_3' = c_3 - C_{31}c_1', \quad C_3' = C_{32} - C_{31}C_{11}^{-1}C_{12}. \end{aligned}$$

Assume $(x_2^*, y_{\mathcal{E}_2}^*, y_{\mathcal{I}}^*) \in \mathbb{R}^{n_2+m_2+m_3}$ *is a primal-dual solution of the reduced QP*

$$(1) \qquad \min_{x_2 \in \mathbb{R}^{n_2}} \quad \frac{1}{2}x_2^{\mathrm{T}}B'x_2 + {b'}^{\mathrm{T}}x_2 \quad \text{s.t.} \quad C_2'x_2 = c_2', \quad C_3'x_2 \geq c_3'.$$

If we choose $x_1^* = C_{11}^{-1}(c_1 - C_{12}x_2^*)$ *and* $y_{\mathcal{E}_1}^* = C_{11}^{-\mathrm{T}}\left((B_{12} - B_{11}C_{11}^{-1}C_{12})x_2^* + B_{11}c_1' + b_1 - C_{21}^{\mathrm{T}}y_{\mathcal{E}_2}^* - C_{31}^{\mathrm{T}}y_{\mathcal{I}}^*\right)$*, then* $(x^*, y^*) := (x_1^*, x_2^*, y_{\mathcal{E}_1}^*, y_{\mathcal{E}_2}^*, y_{\mathcal{I}}^*)$ *is a primal-dual solution of the two-grid Newton-Picard QP.*

If we furthermore choose the two-grid Newton-Picard Hessian approximation $\widetilde{B} = \widetilde{B}^{\mathrm{fast}} + \widetilde{B}^{\mathrm{slow}}$ with

$$\begin{aligned} \widetilde{B}^{\mathrm{fast}} &= \begin{pmatrix} (\mathbf{I} \otimes \Pi^{\mathrm{fast}})^{\mathrm{T}}B_{11}(\mathbf{I} \otimes \Pi^{\mathrm{fast}}) & 0 \\ 0 & 0 \end{pmatrix}, \\ \widetilde{B}^{\mathrm{slow}} &= \begin{pmatrix} (\mathbf{I} \otimes R)^{\mathrm{T}}\hat{B}_{11}(\mathbf{I} \otimes R) & (\mathbf{I} \otimes R)^{\mathrm{T}}\hat{B}_{12} \\ \hat{B}_{21}(\mathbf{I} \otimes R) & \hat{B}_{22} \end{pmatrix}, \end{aligned}$$

then we can see that almost all terms that need to be computed in Theorem 3 can be evaluated completely on the coarse grid due to

$$\widetilde{Z}^{\mathrm{T}}\widetilde{B}^{\mathrm{fast}}\widetilde{Z} = 0, \quad \widetilde{B}' = \widetilde{Z}^{\mathrm{T}}\widetilde{B}\widetilde{Z} = \hat{Z}^{\mathrm{T}}\hat{B}\hat{Z}, \quad \widetilde{B}_{11}\widetilde{C}_{11}^{-1}\widetilde{C}_{12} = (\mathbf{I} \otimes R)^{\mathrm{T}}\hat{B}_{11}\hat{C}_{11}^{-1}\hat{C}_{12}.$$

Only $\widetilde{C}_{11}^{-\mathrm{T}}(\widetilde{B}_{11}c_1' + b_1)$ needs to be evaluated on the fine grid to recover $y_{\mathcal{E}_1}^*$.

In an RTI scheme, it is sufficient to only perform one inexact SQP iteration per NMPC step. Hence, solving two-grid Newton-Picard QP problems on the basis of Theorem 3 enables us to only evaluate one forward solve for c_1, one adjoint solve for b_1, and one matrix-vector-product with the Lagrange-Hessian for $y_{\mathcal{E}_1}$ on the fine grid per NMPC step. This can be carried out within a small multiple of the time of a forward simulation of the system. Hence, high sampling rates appear to be possible also for NMPC of parabolic PDE problems.

Moreover, all fine-grid operations can be postponed to the preparation phase after the updated control was applied to the system. Thus, the feedback delay can be reduced considerably, because after the arrival of the state estimate, we only need to compute a small part of b', c_1', c_2', and c_3' on the coarse grid, solve the reduced QP (1) with efficient warm or hot starts, and immediately apply the first control in x_2 to the system. Thus, we can expect feedback delays which are similar to the case of NMPC with ODE.

References

[1] H.G. Bock, M. Diehl, E.A. Kostina, and J.P. Schlöder. Constrained Optimal Feedback Control for DAE. In L. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes, and B. van Bloemen Waanders, editors, *Real-Time PDE-Constrained Optimization*, chapter 1, pages 3–24. SIAM, 2007.

[2] M. Diehl, H.G. Bock, and J.P. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. *SIAM Journal on Control and Optimization*, 43(5):1714–1736, 2005.

[3] A. Potschka. *A direct method for the numerical solution of optimization problems with time-periodic PDE constraints*. PhD thesis, Universität Heidelberg, 2012.

[4] A. Potschka, M.S. Mommer, J.P. Schlöder, and H.G. Bock. Newton-Picard based preconditioning for linear-quadratic optimization problems with time-periodic parabolic PDE constraints. *SIAM J. Sci. Comp.*, 2012.

Stochastic Collocation for Optimization of Radio Frequency Ablation under Parameter Uncertainty

Tobias Preusser

(joint work with Inga Altrogge, Sabrina Haase, Robert M. Kirby, Tim Kröger, Torben Pätz, Hanne Tiesler, Dongbin Xiu)

Radio frequency (RF) ablation is a minimally invasive alternative to surgical resection for the treatment of tumors and metastases in the human liver. Thereby tumor cells are destroyed by heat, which is produced by a local alternating electric current. The electric current emerges from the electrodes of a needle shaped applicator that is placed into the tumor. Mathematical modeling, simulation and optimization can be used to improve the quality of the treatment by assisting the medical doctor in finding the optimal placement of the RF applicator.

A simple model for RF ablation consists of a system of coupled partial differential and algebraic equations that describe the electric current, the diffusion of heat and the tissue damage. The bio-physical tissue parameters involved in the process (electric and thermal conductivity, density, heat capacity) change non-linearly with the temperature and need to be modeled as well. Moreover these parameters are different for each patient; to a large extend they are not precisely known and thus carry uncertainty.

We model this uncertainty by random fields, which we use in the PDE system and also in optimizing the placement of the probe such that the minimal temperature of the tumor becomes maximal. For the solution of the stochastic PDEs we use the polynomial chaos approach and an adaptive sparse grid collocation method, which is characterized by piecewise multilinear basis functions. Different objective functions involving the fitting of the expectation and stochastic moments are discussed and an SQP method based on stochastic collocation is discussed.

Various examples based on real patient data show the performance of the approaches and underline the practical consequences of our research.

References

[1] H. Tiesler, R. M. Kirby, D. Xiu, T. Preusser, *Stochastic collocation for optimal control problems with stochastic PDE constraints*. SIAM J. Control Optim., **50(5)** (2012), 2659–2682.

[2] I. Altrogge, T. Preusser, T. Kröger, S. Haase, T. Pätz, R. M. Kirby, *Sensitivity Analysis for the optimization of radiofrequency ablation in the presence of material parameter uncertainty*. International Journal for Uncertainty Quantification, **2(3)** (2012), 295–321.

[3] T. Pätz, T. Preusser, *Ambrosio-Tortorelli segmentation of stochastic images*. Proc. ECCV, Lect. Not. Comp. Sci **6315** (2010), 254–267.

Preconditioning of a Full-Space Trust-Region SQP Algorithm for PDE-constrained Optimization

Denis Ridzal

(joint work with Miguel Aguiló and Joseph Young, Sandia National Laboratories)

We develop a new class of preconditioners for augmented (optimality) systems that arise in a full-space trust-region sequential quadratic programming (SQP) algorithm designed for matrix-free optimization. By exploiting the structure of augmented systems and by utilizing SQP as the outer iteration, our approach extends the applicability of recent work on solvers for PDE-constrained optimization [1] to a wide range of optimal design, optimal control and inverse problems.

Introduction. Our optimization algorithm is based on the composite-step trust-region SQP framework of [2]. Let $\mathcal{X}$ and $\mathcal{C}$ be Hilbert spaces and let $f : \mathcal{X} \to \mathbb{R}$ and $c : \mathcal{X} \to \mathcal{C}$ be sufficiently smooth functions. We consider the problem

$$\text{(1a)} \qquad \min \quad f(x)$$
$$\text{(1b)} \qquad \text{s.t.} \quad c(x) = 0.$$

We assume that that the Fréchet derivative $c_x(x)$ of the constraint is surjective. We further assume that the dual spaces satisfy $\mathcal{X}^* = \mathcal{X}$ and $\mathcal{C}^* = \mathcal{C}$. Let $\mathcal{L} : \mathcal{X} \times \mathcal{C} \to \mathbb{R}$,

$$\mathcal{L}(x, \lambda) = f(x) + \langle \lambda, c(x) \rangle_{\mathcal{C}} \,,$$

be the Lagrangian for (1). Let x_k be the k-th SQP iterate and λ_k the Lagrange multiplier estimate at x_k. Let $H_k = H(x_k, \lambda_k)$ be a self-adjoint approximation of the Hessian $\nabla_{xx}\mathcal{L}(x_k, \lambda_k)$ of the Lagrangian. Trust-region SQP methods solve (1) by solving a sequence of convex and nonconvex subproblems derived from

$$\text{(2a)} \qquad \min \quad \tfrac{1}{2} \langle H_k s, s \rangle_{\mathcal{X}} + \langle \nabla_x \mathcal{L}(x_k, \lambda_k), s \rangle_{\mathcal{X}} + \mathcal{L}(x_k, \lambda_k)$$
$$\text{(2b)} \qquad \text{s.t.} \quad c_x(x_k)s + c(x_k) = 0, \ \|s\|_{\mathcal{X}} \leq \Delta_k.$$

To reconcile the linear equality constraint and the trust-region constraint in (2b) we apply a Byrd-Omojokun composite-step approach. The trial step s_k is computed as the sum of a quasi-normal step n_k and a tangential step t_k. The quasi-normal step n_k reduces linear infeasibility by approximately solving

$$\text{(3a)} \qquad \min \quad \|c_x(x_k)n + c(x_k)\|_{\mathcal{C}}^2$$
$$\text{(3b)} \qquad \text{s.t.} \quad \|n\|_{\mathcal{X}} \leq \zeta\Delta_k,$$

where $\zeta \in (0,1)$ is a fixed constant. Once the quasi-normal step n_k is computed, the tangential step t_k is computed as an approximate solution of the subproblem

$$\text{(4a)} \qquad \min \quad \tfrac{1}{2}\langle H_k(t+n_k), t+n_k\rangle_{\mathcal{X}} + \langle \nabla_x \mathcal{L}(x_k,\lambda_k), t+n_k\rangle_{\mathcal{X}} + \mathcal{L}(x_k,\lambda_k)$$
$$\text{(4b)} \qquad \text{s.t.} \quad c_x(x_k)t = 0,\ \|t+n_k\|_{\mathcal{X}} \leq \Delta_k,$$

using a Steihaug-Toint projected conjugate gradient (CG) iteration. The Lagrange multiplier λ_k is estimated by approximately solving

$$\text{(5)} \qquad \min \quad \|\nabla f(x_k) + c_x(x_k)^*\lambda\|_{\mathcal{X}}\,.$$

This outlines a generic composite-step trust-region framework. To make the framework robust and efficient for matrix-free computations, inexact linear system solves arising in the solution of (3), (4) and (5) are incorporated in the convergence analysis, and appropriate linear solver stopping conditions are developed, see [3, 4].

Augmented systems and PDE constraints. Our SQP algorithm solves a sequence of special saddle-point systems, called *augmented systems*. They are used to compute: (a) a Newton direction for the quasi-normal step; (b) the null-space projections in Steihaug-Toint CG; (c) an additional feasibility-restoring null-space projection; and (d) the Lagrange multiplier. The systems are of the form

$$\text{(6)} \qquad \begin{pmatrix} I & c_x(x_k)^* \\ c_x(x_k) & 0 \end{pmatrix} \begin{pmatrix} z \\ y \end{pmatrix} = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix} + \begin{pmatrix} r^1 \\ r^2 \end{pmatrix},$$

where $(b^1\ b^2) \in \mathcal{X}\times\mathcal{C}$ is a given right-hand side vector, and the size of the nonzero residual $(r^1\ r^2) \in \mathcal{X}\times\mathcal{C}$ is controlled via the general condition

$$\|r^1\|_{\mathcal{X}}^2 + \|r^2\|_{\mathcal{C}}^2 \leq \mathcal{R}(\|b^1\|_{\mathcal{X}}, \|b^2\|_{\mathcal{C}}, \|z\|_{\mathcal{X}}, \Delta_k, \xi)\,.$$

Here $\mathcal{R} : [\mathbb{R}^+]^4 \times (0,1) \to \mathbb{R}^+$, and ξ is a prescribed *nominal stopping tolerance*.

In PDE-constrained optimization one often assumes a splitting of optimization variables x into state variables u and control variables g, i.e., $\mathcal{X} = \mathcal{U}\times\mathcal{G}$. The augmented system operators are then written in 3×3 block form

$$A = \begin{pmatrix} I & 0 & c_u(x_k)^* \\ 0 & I & c_g(x_k)^* \\ c_u(x_k) & c_g(x_k) & 0 \end{pmatrix}, \quad \text{or compactly,} \quad A = \begin{pmatrix} I & 0 & c_u^* \\ 0 & I & c_g^* \\ c_u & c_g & 0 \end{pmatrix}.$$

Preconditioning. Assuming the existence of $(c_u(x_k))^{-1}$, and inspired by the solvers developed in [1], we consider two preconditioners for the operator A:

$$P^\star = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & (c_u c_u^* + c_g c_g^*)^{-1} \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & c_u^{-*} c_u^{-1} \end{pmatrix}.$$

We recognize the preconditioner $P^\star$ as an exact Schur-complement preconditioner, in the sense that a $P^\star$-preconditioned Krylov solver for a system given by the operator A converges in at most three iterations, see [6]. However, the inverse of $c_u c_u^* + c_g c_g^*$ is difficult to apply in practice. The approximate preconditioner P is a practical alternative, as it eliminates the control Jacobian product $c_g c_g^*$. As a consequence, the application of P amounts to a 'linearized state solve' followed by an 'adjoint solve', which are readily available in PDE-constrained optimization.

The eigenvalue estimates from [1] extend naturally to the preconditioned operator PA. For the problems studied in [1] the eigenvalues of PA cluster around three values: 1, $(1+\sqrt{5})/2$, and $(1-\sqrt{5})/2$. Owing to the 2×2 identity block in A, note that the *only* requirement for applying P is that $c_u(x_k)$ be invertible!

Numerical results. Augmented systems are the only linear systems solved in our SQP algorithm. They are independent of the Lagrangian $\mathcal{L}$, the objective function f and the second derivatives of the constraint c. While this simplifies preconditioning, the remaining problem information must be handled elsewhere in the SQP loop. This is primarily done through the Steihaug-Toint CG algorithm for subproblem (4), invoked once per SQP iteration. Thus, we not only examine the performance of our preconditioner, but also study the behavior of outer iterations.

We consider four problems with PDE constraints. The details are given in [5]. Problem (A) is an inverse problem in heat transfer (Poisson equation) with full-field measurements. It features a very small regularization term. Problem (B) is an optimal control problem in acoustics (Helmholtz equation). The control is applied in a narrow annular subregion of the computational domain. Problem (C) is a full-field inverse problem in nonlinear elasticity (St. Venant-Kirchhoff model), with a very small regularization term. Problem (D) is an optimal flow-control problem (Navier-Stokes equations), with a boundary control.

We solve augmented systems (6) using GMRES. The nominal stopping tolerance is $\xi = 10^{-2}$. We terminate the SQP algorithm when $\|\nabla_x \mathcal{L}(x_k, \lambda_k)\|_{\mathcal{X}}$ and $\|c(x_k)\|_{\mathcal{C}}$ reach 10^{-9}. The preconditioner P is applied using sparse direct solvers, although iterative solves are possible. We study the mesh dependence of the preconditioner. We report the average number of GMRES iterations. For problem (B) we also increase the wave frequency with the mesh size. Additionally, for all problems we report CG iteration totals for the mid-sized mesh. The results are as follows:

(A) Thermal inversion – CG: 158

Mesh Size	GMRES Iters
64×64	4.7
128×128	4.6
256×256	4.7

(B) Acoustic control – CG: 66

Mesh Size	Freq	GMRES Iters
64×64	112 Hz	9.1
128×128	225 Hz	7.7
256×256	450 Hz	6.4

(C) Elastic inversion – CG: 151

Mesh Size	GMRES Iters
32×32	9.1
64×64	9.0
128×128	9.7

(D) Flow control – CG: 143

Mesh Size	GMRES Iters
352	14.5
1408	17.4
5632	21.4

For problems (A)–(C) the preconditioner's performance is nearly mesh independent. For problem (D) we observe slight mesh dependence. The CG iteration totals are between 66 and 158, while only four to eight SQP iterations are used, see [5]. In [5] we also study a source inversion problem governed by advection-dominated diffusion, and show the preconditioner's robustness to high Péclet numbers.

Conclusion. The augmented system preconditioner is efficient in the context of our matrix-free SQP algorithm for PDE optimization. Numerical examples show mesh independence for various PDE models. The preconditioner is generally applicable, and its performance appears largely independent of the governing PDE.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

[1] T. Rees, H. S. Dollar, A. J. Wathen, *Optimal Solvers for PDE-Constrained Optimization*, SIAM J. Sci. Comput. **32** (2010), 271–298.

[2] J. E. Dennis, M. El-Alem, M. C. Maciel, *A global convergence theory for general trust-region-based algorithms for equality constrained optimization*, SIAM J. Opt. **7** (1997), 177–207.

[3] M. Heinkenschloss, D. Ridzal, *A matrix-free trust-region SQP method for equality constrained optimization*, SIAM J. Opt., To be submitted (2013).

[4] D. Ridzal, M. Aguiló, M. Heinkenschloss, *Numerical study of a matrix-free trust-region SQP method for equality constrained optimization*, SAND2011-9346, Sandia Nat'l Labs (2011).

[5] D. Ridzal, *Preconditioning of a full-space trust-region SQP algorithm for PDE optimization*, Oberwolfach talk, URL `http://www.sandia.gov/~dridzal/docs/owa2013.pdf` (2013).

[6] M. F. Murphy, G. H. Golub, A. J. Wathen, *A note on preconditioning for indefinite linear systems*, SIAM J. Sci. Comput. **21** (2000), 1969–1972.

Quadrature on the Sequence Space

Klaus Ritter
(joint work with M. Gnewuch, F. Hickernell, S. Mayer, T. Müller-Gronbach, B. Niu)

Let ρ be a probability measure on a set D. We study numerical integration of functions $f : D^{\mathbb{N}} \to \mathbb{R}$ w.r.t. the product measure $\mu = \rho^{\otimes\mathbb{N}}$ on the sequence space $D^{\mathbb{N}}$.

The study of quadrature problems for functions of infinitely many variables $x_i \in D$ was initiated in [9], and it has intensively been studied recently, see [1, 5, 6, 8, 12, 13, 14, 16] and the preprints [2, 3, 4]. In the same setting function approximation is studied in [19, 20, 21], linear tensor product problems are studied in [17], and a non-linear problem associated with elliptic PDEs with random coefficients is studied in [11, 10]. Moreover, see [18] for a survey and [7] for some aspects of the analytic foundations of computational problems of this kind.

In the talk we focus on a particular smoothness scale for the integrands f, namely functions with ANOVA components in weighted spaces of bounded mixed smoothness, while the general framework of weighted superpositions of tensor product reproducing kernel Hilbert spaces is only briefly sketched.

We present sharp upper and lower bounds for the minimal errors of randomized algorithms due to [2], which substantially improve the results from [8]. Almost optimal algorithms are based on the stochastic multilevel technique and on randomized QMC methods for certain finite-dimensional subproblems. Tractability results for multivariate problems form an important ingredient in the analysis, see [15] for a comprehensive study. Finally, an integrand from computational finance is used as a numerical example, and to illustrate the need to properly relate smoothness on the path space with smoothness on the sequence space.

Supported by the DFG within SPP 1324 and by $(\mathrm{CM})^2$.

References

[1] J. Baldeaux, *Scrambled polynomial lattice rules for infinite-dimensional integration*, in Monte Carlo and Quasi-Monte Carlo Methods 2012, L. Plaskota and H. Woźniakowski, eds., Springer, Heidelberg, 2012, pp. 255–263.

[2] J. Baldeaux and M. Gnewuch, *Optimal randomized multilevel algorithms for infinite-dimensional integration on function spaces with ANOVA-type decomposition.* arXiv:1209.0882v1 [math.NA], Preprint 2012.

[3] J. Dick and M. Gnewuch, *Infinite-dimensional integration in weighted Hilbert spaces: anchored decompositions, optimal deterministic algorithms, and higher order convergence.* arXiv:1210.4223 [math.NA], Preprint 2012.

[4] M. Gnewuch, *Lower error bounds for randomized multilevel and changing dimension algorithms.* arXiv:1209.1808 [math.NA], Preprint 2012, to appear in: J. Dick, F. Y. Kuo, G. W. Peters, I. H. Sloan (Eds.), Monte Carlo and Quasi-Monte Carlo Methods 2012, Springer.

[5] ———, *Infinite-dimensional integration on weighted Hilbert spaces*, Math. Comp., 81 (2012), pp. 2175–2205.

[6] ———, *Weighted geometric discrepancies and numerical integration on reproducing kernel Hilbert spaces*, J. Complexity, 28 (2012), pp. 2–17.

[7] M. Gnewuch, S. Mayer, and K. Ritter, *On weighted Hilbert spaces and integration of functions of infinitely many variables*. DFG-SPP 1324, Preprint 133, 2012.
[8] F. J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter, *Multi-level Monte Carlo algorithms for infinite-dimensional integration on* $\mathbb{R}^{\mathbb{N}}$, J. Complexity, 26 (2010), pp. 229–254.
[9] F. J. Hickernell and X. Wang, *The error bounds and tractability of quasi-Monte Carlo algorithms in infinite dimensions.*, Math. Comp., 71 (2001), pp. 1641–1661.
[10] F. Y. Kuo, C. Schwab, and I. H. Sloan, *Multi-level quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*. Preprint 2012.
[11] ———, *Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*. Preprint 2011, to appear in SIAM J. Numer. Anal..
[12] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski, *Liberating the dimension*, J. Complexity, 26 (2010), pp. 422–454.
[13] B. Niu and F. J. Hickernell, *Monte Carlo simulation of stochastic integrals when the cost of function evaluations is dimension dependent*, in Monte Carlo and Quasi-Monte Carlo Methods 2008, P. L'Ecuyer and A. B. Owen, eds., Springer, Heidelberg, 2008, pp. 545–560.
[14] B. Niu, F. J. Hickernell, T. Müller-Gronbach, and K. Ritter, *Deterministic multi-level algorithms for infinite-dimensional integration on* $\mathbb{R}^{\mathbb{N}}$, J. Complexity, 27 (2011), pp. 331–351.
[15] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems. Vol. 2: Standard Information for Functionals*, EMS Tracts in Mathematics, European Mathematical Society (EMS), Zürich, 2010.
[16] L. Plaskota and G. W. Wasilkowski, *Tractability of infinite-dimensional integration in the worst case and randomized settings*, J. Complexity, 27 (2011), pp. 505–518.
[17] G. W. Wasilkowski, *On tractability of linear tensor product problems for* ∞*-variate classes of functions*. Preprint 2012.
[18] ———, *Liberating the dimension for function approximation and integration*, in Monte Carlo and Quasi-Monte Carlo Methods 2012, L. Plaskota and H. Woźniakowski, eds., Springer, Heidelberg, 2012, pp. 211–231.
[19] G. W. Wasilkowski, *Liberating the dimension for* L_2*-approximation*, J. Complexity, 28 (2012), pp. 304–319.
[20] G. W. Wasilkowski and H. Woźniakowski, *Liberating the dimension for function approximation*, J. Complexity, 27 (2011), pp. 86–110.
[21] ———, *Liberating the dimension for function approximation: Standard information*, J. Complexity, 27 (2011), pp. 417–440.

Approximation of Stochastic Optimization Problems and Scenario Generation

Werner Römisch

First we recall the main approaches to optimization models under stochastic uncertainty: Chance constrained and two-stage stochastic optimization models (see [6]). Next we review known results on quantitative stability of stochastic optimization problems, i.e., on quantitative continuity properties of optimal values and solution sets with respect to probability distributions belonging to certain spaces equipped with probability metrics [4]. The general results are specialized for linear two-stage stochastic programs. Important specific approximations of the underlying probability distributions are discrete probability measures with finite support. The latter measures reduce integrals to weighted sums and lead, in

addition, to specific constraint structures of the resulting approximate optimization models. For example, in case of linear two-stage models the approximations represent large scale linear programs which contain finitely many realizations or *scenarios* of the underlying random vector.

Next we address the problem of optimal *scenario generation* and discuss a number of methods for generating scenarios based on high-dimensional numerical integration schemes (see [5]). We begin with *empirical* (or Monte Carlo) approximations and argue how available quantitative stability results may be used together with empirical process theory (from asymptotic statistics) to obtain (probabilistic) convergence rates for optimal values and solution sets (see [4, Chapter 4]). Next we report on recent developments in using (i) *optimal quantization methods*, (ii) *Quasi-Monte Carlo methods* [2] and (iii) *sparse grids* for scenario generation (see [5]). We argue that recent results in [3] motivate the use of Quasi-Monte Carlo and sparse grid methods in stochastic optimization if they are combined with effective dimension reduction techniques. In particular, we highlight *randomized lattice rules*, a recently developed class of methods for high-dimensional numerical integration which combine good properties of Monte Carlo and Quasi-Monte Carlo methods [2, 1].

References

[1] J. Dick, F. Y. Kuo and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* (2014) (to appear).
[2] J. Dick and F. Pillichshammer: *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, 2010.
[3] H. Heitsch, H. Leövey and W. Römisch: Are Quasi-Monte Carlo algorithms efficient for two-stage stochastic programs?, Stochastic Programming E-Print Series, 5-2012 (`www.speps.org`) and submitted.
[4] W. Römisch: Stability of stochastic programming problems, in: Stochastic Programming (A. Ruszczynski and A. Shapiro eds.), Handbook, Elsevier, 2003, 483–554.
[5] W. Römisch: Scenario generation, in: Wiley Encyclopedia of Operations Research and Management Science (J.J. Cochran ed.), Wiley 2010.
[6] A. Ruszczyński and A. Shapiro (Eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.

Sparse, Adaptive Smolyak Quadratures for Bayesian Inverse Problems

Claudia Schillings
(joint work with Christoph Schwab)

Based on the parametric deterministic formulation of Bayesian inverse problems with unknown input parameter from infinite dimensional, separable Banach spaces proposed in [10], we develop a practical computational algorithm whose convergence rates are provably higher than those of Monte-Carlo (MC) and Markov-Chain Monte-Carlo methods, in terms of the number of solutions of the forward problem. The focus is on linear, elliptic PDE with unknown diffusion coefficient, however, the derived convergence results are not limited to linear, elliptic PDEs:

analogous results hold for forward maps of a rather wide range of mathematical models.

A basic problem in Bayesian inverse problems consists of determining the unknown diffusion coefficient $u \in X$ from given noisy observation data $\delta = \mathcal{O}(G(u)) + \eta$ (with $\eta \in \mathbb{R}^K$ representing the observation noise, $\mathcal{O} : R \mapsto \mathbb{R}^K$ bounded, linear observation operator and $G : X \to R$ forward response map from some separable Banach space X of unknown parameters into a separable Banach space R of responses) in order to compute the expectation of a quantity of interest. The proposed approach relies on a reformulation of the forward problem with unknown stochastic input data as an infinite dimensional, parametric deterministic problem. Therefore, the unknown diffusion coefficient u is assumed to admit a *parametric representation* of the form

$$u = \bar{a} + \sum_{j \in \mathbb{J}} y_j \psi_j$$

where $y = (y_j)_{j \in \mathbb{J}}$ is an i.i.d sequence of real-valued random variables $y_j \sim \mathcal{U}[-1/2, 1/2]$, ie. the prior is given by $\mu_0(dy) := \bigotimes_{j \in \mathbb{J}} \lambda_1(dy_j)$, $\bar{a},\ \psi_j \in X$ and $\mathbb{J}$ denotes a finite or countably infinite index set, ie. either $\mathbb{J} = \{1, 2, ..., J\}$ or $\mathbb{J} = \mathbb{N}$.

Under appropriate assumptions on the forward and observation model and the prior measure, the posterior distribution on u is absolutely continuous with respect to the prior, see [10]. The density of the posterior with respect to the prior is a Radon-Nikodym derivative that is given by an infinite dimensional version of Bayes rule. Based on this result, we are interested in computing the expectation of a *prediction function* $\phi : X \to S$ for a *Quantity of Interest (QoI)*. The expectation of QoI under the posterior (given data δ) is given by $\mathbb{E}_{\mu^\delta}[\phi(u)] := Z'/Z \in S$ with

$$Z' = \int_{[-1/2,1/2]^{\mathbb{J}}} \Psi(y)\mu_0(dy) \quad Z = \int_{[-1/2,1/2]^{\mathbb{J}}} \Theta(y)\mu_0(dy)\,,$$

$$\Theta(y) = \exp(-\tfrac{1}{2}|\delta - \mathcal{O}(G(u))|^2_\Gamma \tfrac{1}{2}\Big|_{u=\bar{a}+\sum_{j\in\mathbb{J}} y_j\psi_j},\ \Psi(y) = \Theta(y)\phi(u)\Big|_{u=\bar{a}+\sum_{j\in\mathbb{J}} y_j\psi_j}.$$

In [10], joint analyticity of the posterior density as a function of the parameter vector $y \in U$ is proven. In particular, the estimates of the size of domains of analytic continuation which were obtained in [10] allowed to prove rates on so-called *sparse, monotone N-term polynomial chaos approximations* of the posterior density $\Psi(y)$. The resulting approximation rates are independent of the dimension, ie. of the number of active coordinates y_j in the quadrature approximations of Z and Z', and will be the basis for the presented proofs of (dimension independent) convergence rates of the adaptive Smolyak quadrature algorithms. For any finite monotone set $\Lambda \subset \mathcal{F}$, the quadrature operator is defined by

$$\mathcal{Q}_\Lambda = \sum_{\nu \in \Lambda} \Delta_\nu = \sum_{\nu \in \Lambda} \bigotimes_{j \geq 1} \Delta_{\nu_j}$$

with difference operators $\Delta_\nu = \bigotimes_{j\geq 1} \Delta_{\nu_j}$, $\Delta_j = Q^j - Q^{j-1}$ and $(Q^k)_{k\geq 0}$ sequence of univariate quadrature formulas, see [9] for details on the construction of the tensorized multivariate quadrature formulas.

Then, it can be proven (under appropriate assumptions on the univariate quadrature formulas and on the forward model) that sparsity in the unknown coefficient function u, i.e. if $\sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)}^{\sigma} < \infty$ for $0 < \sigma < 1$, implies the existence of two sequences $(\Lambda_N^1)_{N\geq 1}, (\Lambda_N^2)_{N\geq 1}$ of monotone sets $\Lambda_N^{1,2} \subset \mathcal{F}$ such that with $\#\Lambda_N^{1,2} \leq N$

$$|Z - \mathcal{Q}_{\Lambda_N^1}(\Theta)| \leq C_Z N^{-s}\,, \qquad s = \frac{1}{\sigma} - 1\,,$$

and

$$\|Z' - \mathcal{Q}_{\Lambda_N^2}(\Psi)\|_{V^{(m)}} \leq C_{Z'} N^{-s}\,, \qquad s = \frac{1}{\sigma} - 1\,.$$

The construction of the monotone index set $(\Lambda_N^{1,2})_{N\geq 1}$ is based on a greedy-type strategy which attempts to control the global approximation error by locally collecting indices of the current set of reduced neighbors with the largest error contributions. We will present numerical experiments based on the following model parametric elliptic boundary value problem

$$-\mathrm{div}(u\nabla p) = f \quad \text{in } D := [0,1]\,,\ p = 0 \quad \text{in } \partial D\,,$$

with $f(x) = 100 \cdot x$ and diffusion coefficient $u(x,y) = \bar{a} + \sum_{j=1}^{64} y_j \psi_j$, where $\bar{a} = 1$ and $\psi_j = \alpha_j \chi_{D_j}$ with $D_j = [(j-1)\frac{1}{64}, j\frac{1}{64}]$, $y = (y_j)_{j=1,\ldots,64}$ and $\alpha_j = \frac{1.8}{j^\zeta}$, $\zeta = 2,3,4$ in order to numerically verify the theoretical results. Exemplarily, the approximation error of the normalization constant Z for three values of the parameter ζ controlling the sparsity of the unknown input data is shown in Figure 1.

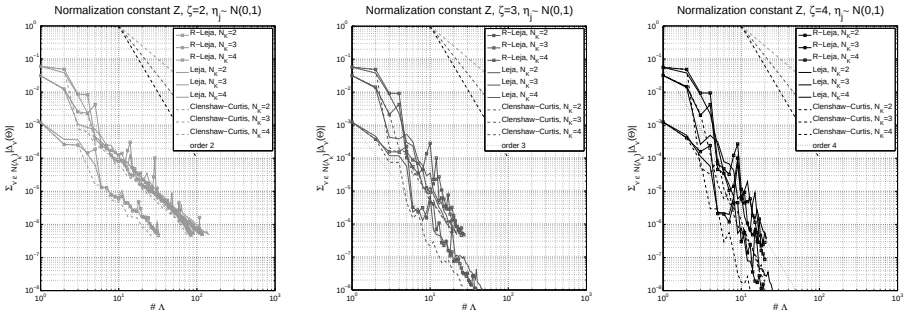

FIGURE 1. Comparison of the error curves of the normalization constant Z with respect to $\#\Lambda_N$ based on the sequences with Clenshaw-Curtis, symmetrized Leja and R-Leja quadrature points with number of observations $K = 2^{N_K} - 1\,,\ N_K = 2,3,4,\ \eta \sim \mathcal{N}(0,1)$ and with $\zeta = 2$ (left), $\zeta = 3$ (middle) and $\zeta = 4$ (right).

Furthermore, we will present numerical results considering a lognormal diffusion coefficient, ie. $\ln(u(x,y)) = \sum_{j=1}^{64} y_j \psi_j$, where $\psi_j = \alpha_j \chi_{D_j}$, indicating the same convergence behavior as in the uniform case, cp. Figure 2.

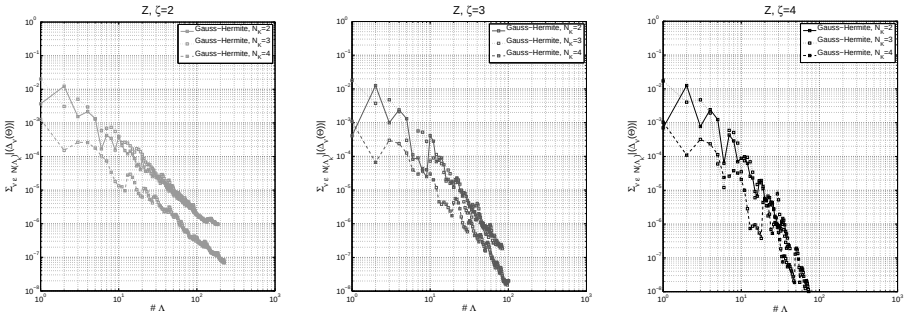

FIGURE 2. Comparison of the error curves of the normalization constant Z with respect to $\#\Lambda_N$ based on the Gauss-Hermite quadrature with number of observations $K = 2^{N_K} - 1$, $N_K = 2, 3, 4$, $\eta \sim \mathcal{N}(0, 1)$ and $\zeta = 2$ (l.), $\zeta = 3$ (m.) and $\zeta = 4$ (r.).

Acknowledgements. This work is supported by Swiss National Science Foundation and by the European Research Council under FP7 Grant AdG247277.

REFERENCES

[1] J.-P. Calvi and M. Phung Van, *On the Lebesgue constant of Leja sequences for the unit disk and its applications to multivariate interpolation* Journal of Approximation Theory **163** (2011), 608–622.

[2] J.-P. Calvi and M. Phung Van, *Lagrange interpolation at real projections of Leja sequences for the unit disk* Proceedings of the American Mathematical Society, (to appear) (2012).

[3] A. Chkifa, *On the Lebesgue constant of Leja sequences for the unit disk*, (submitted) (2012).

[4] A. Chkifa, A. Cohen and C. Schwab, *High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs*, Report 2012-22, SAM, ETH Zürich (2012),

[5] A. Cohen, R. DeVore and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs*, Analysis and Applications **9** (2011), 1–37.

[6] T. Gerstner and M. Griebel, *Dimension-adaptive tensor-product quadrature* Computing **71** (2003), 65–87.

[7] M. Hansen and C. Schwab, *Analytic regularity and best N-term approximation of high dimensional parametric initial value problems*, Report 2011-64, SAM, ETH Zürich (2011).

[8] M. Hansen, C. Schillings and C. Schwab, *Sparse approximation algorithms for high dimensional parametric initial value problems*, Proceedings 5th Conf. High Performance Computing, Hanoi (2012).

[9] C. Schillings and C. Schwab, *Sparse, Adaptive Smolyak Algorithms for Bayesian Inverse Problems*, (submitted) (2012).

[10] C. Schwab and A. M. Stuart, *Sparse Deterministic Approximation of Bayesian Inverse Problems*, *Inverse Problems* **28** (2012), 045003.

Large Scale Shape Optimization for Deterministic Problems

Stephan Schmidt
(joint work with N. Popescu, M. Berggren, V. Schulz, N. Gauger, A. Walther)

Shape optimisation problems are characterised by the fact that the input geometry is the unknown to be found. Thus, using a naive approach, one is forced to compute the sensitivities of the objective function with respect to the input geometry. For problems governed by PDEs, this especially means that the PDE solver has to be differentiated with respect to the input mesh. Because this derivative can be cumbersome and costly to compute, shape calculus will be used to derive an expression of this sensitivity, which exists on the boundary of the domain alone and is an analytic expression that can be evaluated without any knowledge of partial derivatives of the PDE solution procedure with respect to the input mesh. The talk will show some applications in acoustics where is continuous shape variation approach is coupled to dolfin-adjoint and the whole shape optimisation problem is solved using FEniCS. In particular, for the acoustic horn problem, the shape derivative for minimising the sound reflections is given by

$$
\begin{aligned}
dJ[V] &= \int_{\Gamma_{\text{Horn}}} \int_{t_0}^{t_f} \langle V, n\rangle c^2 \left[-\lambda_p \left\langle \frac{\partial u}{\partial n}, n \right\rangle - \nabla_\Gamma \cdot (\lambda_p u) \right] \, dt \, dS \\
&= \int_{\Gamma_{\text{Horn}}} \int_{t_0}^{t_f} \langle V, n\rangle \left(-c^2 \nabla \cdot (\lambda_p u) \right) \, dt \, dS \\
-\dot{\lambda_u} &= -c^2 \nabla \lambda_p \qquad \text{in} \quad \Omega \\
-\dot{\lambda_p} &= -\nabla \cdot \lambda_u \\
\frac{1}{2} \left(\lambda_p - c\langle \lambda_u, n\rangle \right)(t) &= \frac{p(t) - g(t)}{2c} \quad \text{on} \quad \Gamma_{\text{in}}.
\end{aligned}
$$

The talk also serves as a basis for discussing the methodology to solve inverse scattering problems governed by the Maxwell equations given uncertain data and the automatic generation of shape derivatives using UFL. Last but not least, transient problems in CFD are considered with special emphasis on optimal morphing shapes.

References

[1] M. C. Delfour and J.-P. Zolésio. *Shapes and Geometries: Analysis, Differential Calculus, and Optimization.* SIAM Philadelphia, 2001

[2] C. Schillings, S. Schmidt and V. Schulz. Efficient Shape Optimization for Certain and Uncertain Aerodynamic Design. *Computers & Fluids*, 46(1):78-87, 2011.

[3] K. Eppler, S. Schmidt, V. Schulz, and C. Ilic. Preconditioning the pressure tracking in fluid dynamics by shape Hessian information. *Journal of Optimization Theory and Applications*, 141(3):513–531, 2009.

[4] S. Schmidt and V. Schulz. Impulse response approximations of discrete shape Hessians with application in CFD. *SIAM Journal on Control and Optimization*, 48(4):2562–2580, 2009.

[5] S. Schmidt and V. Schulz. Shape derivatives for general objective functions and the incompressible Navier–Stokes equations. *Control and Cybernetics*, 39(3):677-713, 2010

Stochastic Programming in Finite Dimension as Inspiration for PDE-Constrained Optimization under Uncertainty

Rüdiger Schultz
(joint work with Sergio Conti, Benedict Geihe, Martin Rumpf, Harald Held, Martin Pach)

Shape optimization with linearized elasticity and stochastic loading is taken as an example for the transfer of basic modeling principles from finite-dimensional stochastic programming to PDE-constrained optimization under uncertainty. When including stochasticity of data into shape optimization, or PDE-constrained optimization in general, it is not entirely the random vectors representing the stochastic data, but in addition, and with crucial consequences, the availability of information, that matters.

Indeed, there are different modeling approaches, implied by different placement of the constraint that future data information must not be anticipated when making decisions. In stochastic shape optimization, nonanticipativity arises quite naturally. The shape decision has to be made before the random forces are applied to the shape (elastic body), and, of course not vice versa, meaning the shape adjusts to the realization of the random data.

Then the two-stage stochastic programming approach readily translates to shape optimization: decide on the shape , then observe the realization of the stochastic load, compute the displacement by solving the elasticity PDE. With an objective function, such as the compliance, a random quantity now arises – and has to be optimized. The quantity can be seen as stemming from a family of random variables in which "best" members have to be identified

This leads into stochastic decision theory or into answering the question of how to rank random variables. Two generic classes of optimization problems come up this way: risk aversion in the objective by including suitable risk measures, or risk aversion in the constraints by suitable stochastic (partial) orders, also known as stochastic dominance relations.

Again these concepts readily extend to shape optimization, "creating infinite-dimensional cousins" of traditional stochastic programs in finite dimension. These range from risk-neutral expectation models to mean-risk models addressing different attitudes to risk, and finally arrive at models involving stochastic dominance relations.

The talk addresses these modeling alternatives, provides some insight into algorithmic issues aiming at efficient numerical solution of frequently occurring elasticity PDEs., and concludes with reporting illustrative numerical tests. For details, the journal publications [1],[2, 3] can be consulted.

References

[1] S. Conti, H. Held, M. Pach, M. Rumpf, R. Schultz, *Shape optimization under uncertainty - a stochastic programming perspective*, SIAM Journal on Optimization **19** (2008), 1610–1632.
[2] S. Conti, H. Held, M. Pach, M. Rumpf, R. Schultz, *Risk averse shape optimization*, SIAM Journal on Control and Optimization **49** (2011), 927–947.
[3] B. Geihe, M. Lenz, M. Rumpf, R. Schultz, *Risk averse elastic shape optimization with parametrized fine scale geometry*, Mathematical Programming (2012).

Topology Optimization under Aerodynamic Loads

Roland Stoffel, Volker Schulz

Structural optimization is a challenging issue of high practical interest. In this talk, structural optimization is performed within an aerodynamic framework. Therefore a topology optimization method is developed, which is based on the topological derivative. In particular, the topological derivative for linear elasticity is connected with the level-set method, which performed the topological changes implicitly. Efficient implementations as well as classical results and results in practical applications will be presented.

Robust Optimization with PDE Constraints based on Linear and Quadratic Approximations

Stefan Ulbrich
(joint work with Adrian Sichau)

We present a second order approximation for the robust counterpart of PDE-constrained optimization problems with uncertain data. The approach is an extension of first order robust approximations, which have been proposed for uncertain nonlinear optimization problems by [2] and [4], see also [1]. We show how for ellipsoidal uncertainty sets the approximated worst-case functions, which are the essential part of the approximated robust counterpart, can be formulated as trust-region problems that can be solved efficiently. Also, the gradients of the approximated worst-case functions can be computed efficiently by combining a sensitivity and an adjoint approach. However, there might be points where these functions are nondifferentiable. Hence, we introduce an equivalent formulation of the approximated robust counterpart as an MPEC, in which the objective and all constraints are differentiable. Numerical results for the shape optimization in structural mechanics to obtain optimal solutions that are robust with respect to uncertain loadings show the efficiency of the approach. The formulation can further be extended to model the presence of actuators that are capable of applying forces to a structure in order to counteract the effects of uncertainty. We show that also this robust optimization problems with actuators can be solved efficiently.

More precisely, we consider PDE-constrained optimization problems of the form

$$\min_{y\in Y,\ x\in X} \ h_0(y,x;p) \quad \text{s.t.} \quad h_i(y,x;p)\le 0, \quad i\in I,\ C(y,x;p)=0, \tag{1}$$

where $x \in X$ denotes the design variables, $y \in Y$ the state, $p \in \mathbb{R}^{n_p}$ uncertain parameters and $C(y, x; p) = 0$ the PDE constraint. We assume that h_0, $h_i : Y \times X \times \mathbb{R}^{n_p} \to \mathbb{R}, C : Y \times X \times \mathbb{R}^{n_p} \to Z$ are twice continuously differentiable, that $C(y, x; p) = 0$ has a unique solution $y = y(x; p)$ for all relevant x, p and that $\partial_y C \in \mathcal{L}(Y, Z)$ is invertible. The parameter p is uncertain with $p \in \mathcal{U}_p = \{p \in \mathbb{R}^{n_p} : \|p - \bar{p}\|_{B_p} \leq 1\}$, where $\|z\|_{B_p} := \sqrt{z^T B_p z}$ with B_p symmetric positive definite. It would also be possible to consider uncertainty in the design x.

By defining the worst case values of objective function and inequality constraints

$$\phi_i(x) := \max_{y \in Y,\, \Delta p \in \mathbb{R}^{n_p}} h_i(y, \bar{p} + \Delta p) \quad \text{s.t.} \quad C(y, x; \bar{p} + \Delta p) = 0,\ \|\Delta p\|_{B_p} \leq 1 \tag{2}$$

the robust conterpart of (1) is given by

$$\min_{x \in X} \phi_0(x) \quad \text{s.t.} \quad \phi_i(x) \leq 0, \quad i \in I. \tag{3}$$

However, for nonlinear problems this formulation is computationally not tractable. Therefore, it is proposed in [2], see also [4], to approximate the worst case functions (2) by the following approximation based on a linearization around $(\bar{y}, x; \bar{p})$ with $C(\bar{y}, x; \bar{p}) = 0$

$$\begin{aligned} \tilde{\phi}_i(x) := \max_{\Delta y \in Y,\, \Delta p \in \mathbb{R}^{n_p}} & \; h_i(\bar{y}, x; \bar{p}) + \partial_{(y,p)} h_i(\bar{y}, x; \bar{p}) \begin{pmatrix} \Delta y \\ \Delta p \end{pmatrix} \\ \text{s.t.} \quad & C(\bar{y}, x; \bar{p}) + \partial_{(y,p)} C(\bar{y}, x; \bar{p}) \begin{pmatrix} \Delta y \\ \Delta p \end{pmatrix} = 0,\ \|\Delta p\|_{B_p} \leq 1 \end{aligned} \tag{4}$$

and to consider the approximated robust conterpart

$$\min_{x \in X} \tilde{\phi}_0(x) \quad \text{s.t.} \quad \tilde{\phi}_i(x) \leq 0, \quad i \in I. \tag{5}$$

Then (4), (5) can be formulated as an NLP involving second order cone constraints that can be solved efficiently as long as either n_p or $|I|$ is moderate [2].

For problems with large uncertainty sets or strongly nonlinear h_i the approximation (4), (5) can be too inaccurate. Therefore, we propose to use the following approximation of the worst case functions

$$\begin{aligned} \tilde{\phi}_i(x) := \max_{\Delta y \in Y,\, \Delta p \in \mathbb{R}^{n_p}} & \; h_i(\bar{y}, x; \bar{p}) + \partial_{(y,p)} h_i(\bar{y}, x; \bar{p}) \begin{pmatrix} \Delta y \\ \Delta p \end{pmatrix} \\ & + \frac{1}{2} \left\langle \begin{pmatrix} \Delta y \\ \Delta p \end{pmatrix}, \begin{pmatrix} \partial_{yy} h_i & \partial_{yp} h_i \\ \partial_{py} h_i & \partial_{pp} h_i \end{pmatrix} (\bar{y}, x; \bar{p}) \begin{pmatrix} \Delta y \\ \Delta p \end{pmatrix} \right\rangle \\ \text{s.t.} \quad & C(\bar{y}, x; \bar{p}) + \partial_{(y,p)} C(\bar{y}, x; \bar{p}) \begin{pmatrix} \Delta y \\ \Delta p \end{pmatrix} = 0,\ \|\Delta p\|_{B_p} \leq 1. \end{aligned} \tag{6}$$

We show the following, see also [3].

- (6) can be reformulated as a trust region problem and can be characterized equivalently by well known optimality conditions.
- If the hard case does not occur, the derivative of (6) can be computed efficiently by combining sensitivity and adjoint techniques.

- (5), (6) can be formulated as an MPEC that can be solved efficiently as long as n_p and $|I|$ are moderate.

The efficiency of the approach is demonstrated for the shape optimization governed by the linear elasticity equations for uncertain surface loads p. In this case, (5), (6) coincides with the exact robust counterpart (2), (3).

Finally, we consider the case that actuators can counteract the uncertain forces, more precisely, a problem of the form

$$\min_{x\in X} \phi_0^{act}(x) \quad \text{s.t.} \quad c(x) \le 0, \tag{7}$$

where

$$\begin{aligned} \phi_0^{act}(x) := \max_{y\in Y,\, \Delta p\in\mathbb{R}^{n_p}} \min_{\Delta p_a\in\mathbb{R}^{n_p}} & \; h_0(y, \bar{p}+\Delta p+\Delta p_a) \quad \text{s.t. } \|\Delta p_a\|_{B_a} \le 1, \\ & \text{s.t. } C(y, x; \bar{p}+\Delta p+\Delta p_a) = 0,\ \|\Delta p\|_{B_p} \le 1. \end{aligned} \tag{8}$$

We approximate (7), (8) again by using a quadratic approximation of h_0 and a linearization of C at $(\bar{y}, x; \bar{p})$. By assuming that the reduced Hessian of the quadratic model of h_0 is positive definite, which is the case in our application, we derive necessary as well as slightly stronger sufficient conditions for (8) that can be used in an MPEC formulation. For our application it turns out that the sufficient condition is satisfied. Numerical results for the shape optimization of an elastic structure with actuators for uncertain surface loads show the potential of the approach.

References

[1] D. Bertsimas, O. Nohadani, K. M. Teo, *Nonconvex robust optimization for problems with constraints*, INFORMS J. Comput. **22** (2010), 44–58.

[2] M. Diehl, H. G.Bock, E. Kostina, *An approximation technique for robust nonlinear optimization*, Math. Program. **107** (2006), 213–230.

[3] A. Sichau, S. Ulbrich, *A second order approximation technique for robust shape optimization*, Applied Mechanics and Materials **104** (2011), 1–40.

[4] Y. Zhang, *General robust-optimization formulation for nonlinear programming*, J. Optim. Theory Appl. **132** (2007), 111–124.

Thermoelastic Shape Optimization and its Application to a Cast-Iron Cooking Plate

Heinz Zorn

(joint work with Stephan Schmidt, Volker Schulz, Roland Stoffel)

Shape optimization constrained by a single physical discipline is an often discussed and well known problem. Here the optimization of a thermoelastic system, constrained by a coupled system of linear elasticity and the heatequation is considered. The derivation of the shape gradient for this thermoelastic problem is shown and an algorithm for a projected gradient optimization strategy ist introduced. The theory is presented together with its application to a cast-iron cooking plate and the numerical results.

Reporter: Volker Schulz

Participants

Dr. Mario Annunziato
Dipartimento di Matematica
Università degli Studi di Salerno
Via Ponte Don Melillo
84084 Fisciano (SA)
ITALY

Prof. Dr. Alfio Borzi
Institut für Mathematik
Universität Würzburg
Emil-Fischer-Str. 30
97074 Würzburg
GERMANY

Prof. Dr. Jean-Antoine Desideri
INRIA Sophia Antipolis
B.P. 93
2004 Route des Lucioles
06902 Sophia Antipolis Cedex
FRANCE

Prof. Dr. Amr S. El-Bakry
ExxonMobil Production Co.
EMB-2019H
800 Bell St.
Houston, TX 77002
UNITED STATES

Oliver Ernst
Fakultät f. Mathematik u. Informatik
TU Bergakademie Freiberg
Institut f. Num. Math. & Optimierung
09596 Freiberg
GERMANY

Dr. Mike Espig
Max-Planck-Institut für Mathematik
in den Naturwissenschaften
Inselstr. 22 - 26
04103 Leipzig
GERMANY

Prof. Dr. Claude Jeffrey Gittelson
Department of Mathematics
Purdue University
150 N. University Street
West Lafayette IN 47907-2067
UNITED STATES

Jedidiah Gohlke
CAAM Dept. MS 134
Rice University
6100 S. Main Street
Houston, TX 77005-1892
UNITED STATES

Prof. Dr. Helmut Harbrecht
Mathematisches Institut
Universität Basel
Rheinsprung 21
4051 Basel
SWITZERLAND

Prof. Dr. Matthias Heinkenschloss
CAAM Dept. MS 134
Rice University
6100 S. Main Street
Houston, TX 77005-1892
UNITED STATES

Prof. Dr. Rene Henrion
Weierstrass-Institute for Applied
Analysis and Stochastics
Mohrenstr. 39
10117 Berlin
GERMANY

Prof. Dr. Roland Herzog
Fakultät für Mathematik
TU Chemnitz
Reichenhainer Str. 41
09126 Chemnitz
GERMANY

Prof. Dr. Vincent Heuveline
Karlsruher Institut f. Technologie (KIT)
Institut f. Angew. & Num. Mathematik IV
Engineering Mathematics & Computational Lab.
Fritz-Erler-Str. 23
76133 Karlsruhe
GERMANY

Prof. Dr. Michael Hintermüller
Fachbereich Mathematik
Humboldt-Universität Berlin
Unter den Linden 6
10117 Berlin
GERMANY

Prof. Dr. Ekaterina A. Kostina
FB Mathematik & Informatik
Philipps-Universität Marburg
Hans-Meerwein-Strasse (Lahnbg.)
35032 Marburg
GERMANY

Dr. Drew Philip Kouri
Department of Computational and Applied Mathematics
Rice University
P.O. Box 1892
Houston, TX 77251
UNITED STATES

Prof. Dr. Karl Kunisch
Institut für Mathematik und wissenschaftliches Rechnen
Karl-Franzens-Universität Graz
Heinrichstr. 36
8010 Graz
AUSTRIA

Dr. Alexander Litvinenko
Institute of Scientific Computing
Technische Universität Braunschweig
Hans-Sommer-Str. 65
38106 Braunschweig
GERMANY

Prof. Dr. Kurt Marti
Mathematik und Rechneranwendungen
Fakultät LRT
Universität der Bundeswehr München
85577 Neubiberg
GERMANY

Prof. Dr. Hermann G. Matthies
Technische Universität Braunschweig
Institute of Scientific Computing
Hans-Sommer-Str. 65
38106 Braunschweig
GERMANY

Masoumeh Mohammadi
Fakultät für Mathematik
Universität Würzburg
Am Hubland
97074 Würzburg
GERMANY

Prof. Dr. Fabio Nobile
CSQI - MATHICSE
Station 8
École Polytechnique Fédérale de Lausanne
1015 Lausanne
SWITZERLAND

Prof. Dr. Erich Novak
Institut für Stochastik
Friedrich-Schiller-Universität
Ernst-Abbe-Platz 2
07743 Jena
GERMANY

Andreas Potschka
Interdisziplinäres Zentrum für Wissenschaftliches Rechnen
Universität Heidelberg
Im Neuenheimer Feld 368
69120 Heidelberg
GERMANY

Prof. Dr. Tobias Preusser
School of Engineering and Science
Jacobs University Bremen
Postfach 750561
28725 Bremen
GERMANY

Dr. Denis Ridzal
Sandia National Laboratories
Optimization/Uncertainty Quantification
Department (1441)
P.O. Box 5800, MS-1320
Albuquerque NM 87185-1320
UNITED STATES

Prof. Dr. Klaus Ritter
Fachbereich Mathematik
T.U. Kaiserslautern
Erwin-Schrödinger-Straße
67653 Kaiserslautern
GERMANY

Prof. Dr. Werner Römisch
Fachbereich Mathematik
Humboldt Universität Berlin
Unter den Linden 6
10099 Berlin
GERMANY

Prof. Dr. Arnd Rösch
Fakultät für Mathematik
Universität Duisburg-Essen
Campus Duisburg
Forsthausweg 2
47057 Duisburg
GERMANY

Prof. Dr. Ekkehard Sachs
FB IV - Dept. of Mathematics
Universität Trier
54286 Trier
GERMANY

Dr. Claudia Schillings
Seminar für Angewandte Mathematik
ETH-Zentrum
HG G 52.1
Rämistr. 101
8092 Zürich
SWITZERLAND

Dr. Stephan Schmidt
Imperial College London
Department of Aeronautics
South Kensington Campus
London SW7 2AZ
UNITED KINGDOM

Prof. Dr. Rüdiger Schultz
FB Mathematik
Universität Duisburg-Essen
Lotharstr. 65
47057 Duisburg
GERMANY

Prof. Dr. Volker Schulz
Fachbereich IV - Mathematik
Numerik - Optimierung, partielle
Differentialgleichungen
Universität Trier
54286 Trier
GERMANY

Roland Stoffel
Abteilung Mathematik
Fachbereich IV
Universität Trier
54286 Trier
GERMANY

Prof. Dr. Raul F. Tempone
4700 King Abdullah University of
Science
and Technology (KAUST)
Mathematical & Computer Sciences
Thuwal 23955-6900
Jeddah
SAUDI ARABIA

Dr. Hanne Tiesler
Fraunhofer MEVIS
Universitätsallee 29
28359 Bremen
GERMANY

Prof. Dr. Michael Ulbrich
Zentrum Mathematik
TU München
Boltzmannstr. 3
85748 Garching b. München
GERMANY

Prof. Dr. Stefan Ulbrich
Fachbereich Mathematik
Fachgebiet Nichtlineare Optimierung
Technische Universität Darmstadt
Dolviostraße 15
64293 Darmstadt
GERMANY

Prof. Dr. Clayton G. Webster
Computer Science & Mathematics
Division
Oak Ridge National Laboratory
P.O. Box 2008, MS-6164
Oak Ridge, TN 37831
UNITED STATES

Heinz Zorn
Abteilung Mathematik
Fachbereich IV
Universität Trier
54286 Trier
GERMANY