Journal of Computational Physics 308 (2016) 124–152

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Variational Bayesian strategies for high-dimensional, stochastic design problems ☆

P.S. Koutsourelakis [1]

Professur für Kontinuumsmechanik, Technische Universität München, Boltzmannstrasse 15, 85747 Garching (b. München), Germany

ARTICLE INFO

Article history:
Received 24 July 2015
Received in revised form 12 November 2015
Accepted 14 December 2015
Available online 17 December 2015

Keywords:
Uncertainty quantification
Variational Bayes
Optimization
Dimensionality reduction
Dictionary learning

ABSTRACT

This paper is concerned with a lesser-studied problem in the context of model-based, uncertainty quantification (UQ), that of optimization/design/control under uncertainty. The solution of such problems is hindered not only by the usual difficulties encountered in UQ tasks (e.g. the high computational cost of each forward simulation, the large number of random variables) but also by the need to solve a nonlinear optimization problem involving large numbers of design variables and potentially constraints. We propose a framework that is suitable for a class of such problems and is based on the idea of recasting them as probabilistic inference tasks. To that end, we propose a Variational Bayesian (VB) formulation and an iterative VB–Expectation-Maximization scheme that is capable of identifying a local maximum as well as a low-dimensional set of directions in the design space, along which, the objective exhibits the largest sensitivity. We demonstrate the validity of the proposed approach in the context of two numerical examples involving thousands of random and design variables. In all cases considered the cost of the computations in terms of calls to the forward model was of the order of 100 or less. The accuracy of the approximations provided is assessed by information-theoretic metrics.

1. Introduction – motivation

With the increased computational capabilities afforded by the utilization of peta-scale computing resources throughout engineering and the physical sciences, the issue of confidence in simulation results has come at the center of current research. The objective of obtaining an average computational representation of a physical process is being replaced by the new paradigm of *predictive simulations* where the analysis delivers a quantification of uncertainty due to stochasticity in parameters, data and models. Decisions that are based on high-fidelity computational simulations due to their potential economic or societal impact cannot be accepted without quantitative information on the confidence in the computed result.

The field of model-based, uncertainty quantification has seen marked advances in recent years. Naturally, the majority of the efforts have been directed towards *forward* uncertainty propagation i.e. the computation of output statistics given input uncertainties. While several important challenges still remain unanswered, the ultimate objective of the analysis of physical processes and engineering systems is to enable their control and optimization with respect to design objectives.

☆ This paper is based on the homonymous talk given during the international symposium on “Big Data and Predictive Computational Modeling” that took place in 18–21 May 2015 at TUM-IAS, Munich, Germany.

E-mail address: p.s.koutsourelakis@tum.de.
URL: http://www.contmech.mw.tum.de.
[1] Tel.: +49 89 289 16690.

http://dx.doi.org/10.1016/j.jcp.2015.12.031
0021-9991/

Problems of optimization in the presence of uncertainty have attracted much less attention. On one hand, this is because they encompass all the difficulties encountered in uncertainty propagation. Amongst these, the most important stems from the complexity of the forward problem and the increased computational expense associated which each call to the forward solver. It is generally the number of such forward solves that determines the overall computational cost. Secondly, the high-dimensionality of the vector of random variables. Especially in cases where spatiotemporal discretizations of random processes and fields are necessary, one must frequently deal with thousands of random variables. Furthermore, in stochastic optimization problems, there is the additional need to solve a demanding, nonlinear optimization problem which might itself involve thousands of design variables as well as equality/inequality constraints.

Significant advances have been achieved in deterministic optimization and control of complex systems particularly with the development of adjoint-based techniques [1–3] as well as by making use of reduced-order modeling strategies [4,5]. Nevertheless their direct application in the stochastic counterparts of these problems would be infeasible or impractical as the integration with respect to uncertainties poses an insurmountable task.

While decision-making under uncertainty was pioneered in the 1950s [6], applications to large-scale physical models are scarce due to the inherent computational difficulties. Advances in stochastic/robust control and optimization [7–9] or reliability-based design optimization [10–12] are generally applicable to small systems or rely on specific system structure. Techniques using surrogate models and response surfaces [13] or generalized Polynomial Chaos expansions [14] might fail to provide good approximations to the response quantities of interest in the optimization if the number of uncertainties is large, irreducible or non-Gaussian. Furthermore, there is a difficulty in quantifying the error introduced due to the discrepancy between the surrogate and reference model. A critical problem in that respect is the ability to deal with noisy evaluations of the objective functions, its gradient and higher-order derivatives.

The stochastic optimization framework advocated in the present paper is motivated by the following desiderata:

- The ability to seamlessly utilize deterministic simulators and deterministic optimization components such as a first and second order parametric derivatives of model outputs.
- The ability to deal with high-dimensional vectors of random and design variables.
- Least possible number of forward solutions for the same accuracy level.
- The ability to quantify the sensitivity of the expected value/gain/utility to variations in the design variables in the vicinity of the optimum and to provide information on the design features that lead to the largest decay in the expected utility.
- The ability to utilize approximate, reduced-order models or surrogates in order to expedite the solution process.

The objective functions considered in this paper can be written in a general form as:

$$V(\boldsymbol{z}) = \int U(\boldsymbol{\theta}, \boldsymbol{z})\, p_\theta(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{1}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{d_\theta}$ denotes the vector of random variables with a probability density function $p_\theta(\boldsymbol{\theta})$ and $\boldsymbol{z} \in \mathbb{R}^{d_z}$ denotes the vector of design variables. The function $U(\boldsymbol{\theta}, \boldsymbol{z})$ depends on the *output* of the mathematical model and in turn, implicitly depends on random and design variables. Each evaluation of $U(\boldsymbol{\theta}, \boldsymbol{z})$ implies a forward model solution which is assumed *expensive* as in most challenging applications. Naturally the optimization problem can be augmented with constraints with regards to the design variables as it will be demonstrated in the stochastic topology optimization problem that will be considered in the last section. We adopt the term *gain function* (opposite of a loss function) for $U(\boldsymbol{\theta}, \boldsymbol{z})$ and expected gain for $V(\boldsymbol{z})$ and, without loss of generality, pose the corresponding problem as one of *maximization* [15].

The formulation above is quite general and can be readily adapted to cases of practical interest. For example if $U(\boldsymbol{\theta}, \boldsymbol{z}) = \mathbf{1}_{\mathcal{A}}(\boldsymbol{\theta}, \boldsymbol{z})$ is the indicator function of an event $\mathcal{A}$ of interest (e.g. non-failure, or non-exceedance of a response threshold) then maximizing $V(\boldsymbol{z})$ in Equation (1) is equivalent to the maximization of the probability associated with the event $\mathcal{A}$ (similarly one can minimize the probability of event $\mathcal{A}$ by employing the indicator function of the complementary even $\mathcal{A}^c$ in place of U in Equation (1)). The case that would be of principal concern in this paper involves gain functions of the following form[2]:

$$U(\boldsymbol{\theta}, \boldsymbol{z}) = \exp\{-\frac{1}{2} \| \boldsymbol{Q}^{1/2}(\boldsymbol{u}_{target} - \boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{z})) \|^2\}, \tag{2}$$

where $\boldsymbol{u}(\boldsymbol{\theta}, \boldsymbol{z}) \in \mathbb{R}^n$ denotes an output vector of interest (i.e. displacements, velocities, temperature etc.), $\boldsymbol{u}_{target} \in \mathbb{R}^n$ a target/desired response and $\boldsymbol{Q}$ a positive definite matrix of choice. In the examples considered in this paper, $\boldsymbol{Q} = \tau_Q \boldsymbol{I}_n$ where τ_Q^{-1} expresses the allowed variability of $\boldsymbol{u}$ from $\boldsymbol{u}_{target}$. One can readily introduce a diagonal, but anisotropic $\boldsymbol{Q}$ implying that certain response components are more less/important than others. Maximizing the corresponding expected gain, implies finding $\boldsymbol{z}$ for which the response quantities of interest are, *on average*, as close (in the norm defined by $\boldsymbol{Q}$) to

[2] As it will become apparent in the subsequent derivations, the exponent in Equation (2) is used in order to simplify the presentation and several other options to the same effect are possible.