Journal of Computational Science 11 (2015) 1–11

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Efficient knowledge-based optimization of expensive computational models using adaptive response correction

Slawomir Koziel[a,*], Leifur Leifsson[b]

[a] Faculty of Electronics, Telecommunications & Informatics, Gdansk University of Technology, 80-233 Gdansk, Poland
[b] Department of Aerospace Engineering, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 23 February 2015
Received in revised form 30 July 2015
Accepted 2 August 2015
Available online 7 August 2015

Keywords:
Surrogate modeling
Surrogate-based optimization
Design automation
Adaptive response correction
Microwave engineering
Antenna design
Fluid dynamics modeling
Aerodynamic shape optimization

ABSTRACT

Computer simulation has become an indispensable tool in engineering design as they allow an accurate evaluation of the system performance. This is critical in order to carry out the design process in a reliable manner without costly prototyping and physical measurements. However, high-fidelity computer simulations are computationally expensive. This turns to be a fundamental bottleneck when it comes to design automation using numerical optimization techniques. In particular, direct optimization of simulation models, typically, requires a large number of model evaluations, which may be impractical or even infeasible in a reasonable timeframe. Possibly the most promising approach to alleviate this difficulty is surrogate-based optimization (SBO), where direct optimization of expensive models is replaced by an iterative enhancement and re-optimization of fast surrogate models. While a large variety of surrogate modeling and optimization are available, the methods exploiting the so-called physics-based surrogates seem to be the most efficient ones because the knowledge about the system of interest embedded in the underlying (often simulation-based) low-fidelity model ensures good generalization of the surrogate and a rapid convergence of the SBO algorithm. In this paper, we review a specific technique of this class, that is, the adaptive response correction (ARC). We discuss the formulation of the method, its limitations and generalizations, as well as illustrate its application for solving problems in various areas, including microwave engineering, antenna design, and aerodynamic shape optimization.

1. Introduction

Computational tools have become ubiquitous in contemporary engineering. Initially utilized mostly for verification purposes, nowadays, simulation models a play fundamental role in the design process itself, in particular, design optimization. The latter usually refers to adjustment of geometry and/or material parameters of the system at hand so that given performance specifications are satisfied. Simulations are performed in order to assess the quality of the candidate designs produced during such a procedure, i.e., while seeking for an optimal (in a given sense) set of design parameters. Clearly, automation of the simulation-driven design process by means of numerical optimization algorithms is highly desirable. At the same time, it may be quite challenging. A major bottleneck is the high computational cost of evaluating the simulation models. Despite the progress in the development of computational tools (both hardware- and software-wise), various types of simulations (finite element [1], computational fluid dynamics [2], electromagnetic [3]), especially for complex systems at fine levels of discretization of the structure/system at hand, may be as long as hours, days, or even weeks per design. Consequently, straightforward optimization of such model using conventional algorithms (e.g., gradient-based methods with numerical derivatives [4], pattern search procedures [5], or global optimizers exploiting population-based metaheuristics [6]) is normally impractical because of a large number of simulations required by off-the-shelf methods. Additional problems include numerical noise that is usually present in responses generated by simulation models. The aforementioned issues result in a situation where simulation-driven optimization is often realized as a hands-on process involving considerable interaction with the designer. A typical example is design through repetitive parameter sweeps guided by experience, in which engineering insight is used to decide which parameters of the system should be adjusted and in which order in the search for an improved parameter setup. Although such procedures usually yield acceptable results, they are difficult to automate and cannot lead to truly optimum solutions.

* Corresponding author.
E-mail addresses: koziel@ru.is (S. Koziel), leifur@iastate.edu (L. Leifsson).

http://dx.doi.org/10.1016/j.jocs.2015.08.003
1877-7503/

Various attempts to automate and improve the efficiency of simulation-driven design have been considered and proposed in the literature over the recent years. In the context of numerical optimization, the following approaches (and their combinations) should be mentioned: (i) the use of adjoint sensitivities [7,8], (ii) model order reduction [9] as well as techniques for reducing dimensionality of the design space (e.g., [10,11]), and (iii) surrogate-based optimization (SBO) [12–15]. Adjoint sensitivity allows for evaluating the response gradients with little extra computational effort [16,17], resulting in a considerable speed-up of the gradient-based optimization process, particularly for higher-dimensional design spaces, where gradient estimation using finite differences is computationally prohibitive. Model order reduction techniques aim at reducing the number of internal degrees of freedom (states) of the system of interest (e.g., by using reduced order rational function approximation by means of vector fitting methods [18]), thus, reducing the complexity of the system description. Similarly, methods such as principal component analysis (PCA) [19] allow for exploring correlations between design variables in the search space aiming at reducing the number of independent parameters taken into account in the design process. A good example of PCA efficiency is aerodynamic shape optimization, where, e.g., the coefficients of airfoil geometry parameterization are highly correlated (i.e., their changes have similar effect on the objectives, such as the aerodynamic forces), so that dimensionality of the reduced latent space (spanned by the most important principal components) is significantly lower than the original one, and, therefore, the optimization carried out in the latter is much faster with only a slight degradation of accuracy [10].

Among the methods mentioned in the previous paragraph, SBO is the one we focus on in this work. SBO replaces direct optimization of the expensive simulation model by means of iterative construction and re-optimization of a fast representation of the system, referred to as a surrogate model [20]. The surrogate model can be data-driven (i.e., constructed by approximating data pairs obtained by sampling the search space and acquiring corresponding high-fidelity simulation results [21,22]) or physics-based, i.e., obtained by a suitable correction of an underlying low-fidelity model [23]. The low-fidelity model is typically also simulation-based but with relaxed accuracy (coarse discretization, relaxed convergence criteria, etc. [24]). Physics-based methods are less general due to the fact that low-fidelity models are normally problem dependent. On the other hand, these methods tend to be more efficient because the knowledge embedded in the low-fidelity model ensures a better generalization capability of the surrogate [23]. One of the most popular physics-based SBO techniques is space mapping (SM) [20], originally developed for handling expensive problems in the microwave engineering area [15].

In general, there are several ways of correcting the low-fidelity model in order to create a surrogate. These include transformation of the low-fidelity model parameter space (e.g., input SM popular in electrical engineering [15]), exploitation of certain parameters that are normally fixed in the high-fidelity model but can be adjusted in the low-fidelity one for the sake of reducing misalignment between the two (e.g., implicit SM [25]), adjustment of "global" parameters (such as time and frequency) that allow the scaling of vector-valued responses of the low-fidelity model (e.g., frequency scaling in electrical engineering [15]), and, finally, correction of the low-fidelity model response [20].

Response correction is the simplest, yet, potentially, the most effective approach that allows for aligning the low- and high-fidelity models. In the optimization context, the objective is to ensure at least zero-order consistency between the surrogate and the high-fidelity model [26], i.e., perfect agreement between the model responses at the point at which the model is established (typically, the most recent design produced by the optimization algorithm [12]). First-order consistency [26] is more advantageous, however, it requires high-fidelity model derivatives. Response correction methods, can be divided into two main categories: (i) parametric techniques, where the model correction is formulated using explicit analytical formulas (e.g., output SM [20], AMMO [26], manifold mapping [27]), and (ii) non-parametric ones, where the model correction is defined implicitly by exploring correlations between the low- and high-fidelity models (i.e., through appropriate analysis of the model responses). One of the most promising techniques in this class is shape-preserving response prediction (SPRP) [28]. A comprehensive review of SPRP can be found in [29]. Another one is the adaptive response correction (ARC) [30]. ARC was originally developed and applied to optimization of microwave filters [30]. It is a non-parametric method, where the low-fidelity model correction is implemented by translating the changes (due to the adjustment of, e.g., geometry parameters) of the low-fidelity model response (e.g., scattering parameters vs. frequency [15], or the airfoil pressure distribution vs. chord-line coordinate [43]) into the corresponding changes of the high-fidelity model response without evaluating the latter. Unlike space mapping, ARC does not require extraction of any parameters. Therefore, it is particularly suited to work with relatively expensive low-fidelity models (such as coarse-discretization simulation ones). Similar to SPRP, ARC fully exploits the knowledge contained in the low-fidelity model, but it does not have SPRPs limitations regarding the required response shape similarities of the models of different fidelity [29].

In this paper, we review the formulation of the adaptive response correction method, as well as its applications for solving design optimization problems involving expensive simulation models in various engineering disciplines, including microwave engineering, antenna engineering, and aerodynamics. We, furthermore, discuss the practical issues of ARC and give recommendations for applying ARC in design.

2. Formulation of adaptive response correction method

In this section, we recall the formulation of the simulation-driven design optimization task as a nonlinear minimization problem. We, furthermore, give an outline the fundamentals of surrogate-based optimization, and the construction of surrogate models using the ARC technique.

2.1. Optimization problem formulation

The optimization task is formulated here as a nonlinear minimization problem of the form

$$\boldsymbol{x}^* = arg\min_{x} U(\boldsymbol{f}(\boldsymbol{x})) \quad (1)$$

where $\boldsymbol{f}(\boldsymbol{x}) \in R^m$ denotes the response vector of a high-fidelity simulation model to be optimized; $\boldsymbol{x} \in R^n$ is a vector of deterministic designable variables, e.g., geometry or material parameters to be adjusted. The vector $\boldsymbol{f}(\boldsymbol{x})$ represents relevant deterministic characteristics of the system under design. U is a given scalar merit function that encodes given design specifications. It is formulated so that a better design corresponds to a smaller value of $U(\boldsymbol{f}(\boldsymbol{x}))$. $\boldsymbol{x}^*$ is the optimum design to be determined. Normally, the problem (1) is constrained with possible constraints including lower/upper bounds for the design variables, $\boldsymbol{l} \le \boldsymbol{x} \le \boldsymbol{u}$, as well as nonlinear inequality and equality constraints, $c_{ineq.k}(\boldsymbol{x}) \le 0$, $k = 1, \ldots, n_{ineq}$, $c_{eq.k}(\boldsymbol{x}) = 0$, $k = 1, \ldots, n_{eq}$. We do not consider problems involving uncertainty in the parameters, objectives, and constraints.