Sequential experimental designs for nonlinear regression metamodels in simulation

M. Isabel Reis dos Santos [a,*], Pedro M. Reis dos Santos [b]

[a] Department of Mathematics, Center for Management Studies, Technical University of Lisbon (IST), Av. Rovisco Pais, 1049-001 Lisboa, Portugal
[b] Department of Computer Science and Engineering, Technical University of Lisbon (IST), Av. Rovisco Pais, 1049-001 Lisboa, Portugal

A B S T R A C T

The construction of a nonlinear regression metamodel for simulation requires experimental designs that better explore the nonlinearities of the system. The proposed sequential procedure focuses on simulation scenarios in sub-regions where the input–output behavior is more interesting. It takes into account, not only the inputs, but also the output (response) value of the metamodel. The resulting experimental design ensures that the scaled response values are evenly spread over a scaled response surface. Although the focus is on nonlinear regression metamodels, the method may be applied to other types of metamodels. This design procedure is illustrated for one-dimensional and two-dimensional inputs, and the results are compared with space filling designs. The use of bootstrapping for metamodel validation stresses the importance of validation in a metamodeling context.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Computer simulation models are one of the most widely used tools to perform sensitivity analysis of a system response. However, the interpretation of large amounts of data produced by the simulation study can become an intimidating task. The representation of the simulation input–output relationships through a simple mathematical function exposes in a straightforward manner the simulation's behavior. Such a representation of the simulation model, or *metamodel* [5], requires fewer computer resources than the computer simulation model itself. Metamodels may be very useful in answering 'what if' questions and can be used for verifying and validating and optimizing the simulation model as well. In the simulation literature, metamodels are also called response surfaces, emulators, etc. A metamodel approximates the simulation system. Due to their simple construction and use, general linear metamodels have been frequently used by simulation analysts and some researchers [8,23]. However, most real-life systems, such as problems involving queuing systems [40], exhibit nonlinear behavior. The use of linear metamodels (e.g. the intensively studied low-order polynomials) are unable to provide an acceptable global fit to curves of arbitrary shape. To overcome this limitation, some researchers have proposed a number of *nonlinear* metamodels that provide a better and more realistic global approximation of the simulation's input–output relationship; for example, nonlinear regression metamodels [32,34], Kriging metamodels [24], and neural nets [3]. Some nonlinear functions may fit the data well and with fewer and more meaningful parameters than linear approximations, thus making the interpretation of the system and its analysis more intuitive. The construction of a metamodel requires the selection of a design, and the determination of the concomitant metamodel type, the estimation of the metamodel parameters, and the validation of the resulting metamodel.

* Corresponding author. Tel.: +351 936 166 769.
  E-mail addresses: maria.santos@tagus.ist.utl.pt (M.I.R. dos Santos), pedro.reis@tagus.ist.utl.pt (P.M.R. dos Santos).

The selection of an efficient set of inputs, or experimental design, at which to run the simulation and compute the output can be an important challenge. Analysts need experimental designs capable of efficiently exploring an intricate simulation model characterized by a complex response surface, possibly containing some nonlinearities. A good experimental design should effectively explore an experimental region of a multi-dimensional input space through a small number of design points. The remaining stages of the metamodel's construction rely heavily on a good selection of the inputs used as training data to predict the output. Some sequential experimental design strategies have been proposed for finding inputs that optimize (instead of answering what-if questions) the output of a simulation model [16].

Several specialized methods for building metamodels have been proposed, including input selection based on Latin hypercube experimental designs [27]. Such one-shot designs, where all design points are selected prior to the execution of the simulation experiments, offer little or no escape for exploring the nonlinearities of the metamodel's response. Indeed, if the model is seriously in doubt, the misspecification of the metamodel itself can lead to completely inappropriate forms of design. Linear regression [11,25] and Kriging [30,42] methods have been proposed and compared [2] with mixed results. Other comparisons based on the relative performances of the metamodeling designs [28,30,39] have been presented in order to aid practitioners decide which designs to use.

The dependency on the true values of the underlying unknown model's parameters is one of the main difficulties associated with optimal experimental designs for nonlinear models [13,31]. These optimal experimental designs require strong *a priori* assumptions on the metamodel's type and on the nature of the response, for example, white noise [22]. As a result, this type of designs may be completely inappropriate when the input–output relation is in doubt.

Frequently, simulation experiments are implemented sequentially, except possibly when parallel computers are used [22]. Therefore, new inputs can be selected based on the information gathered in previous stages of the simulation experiment. A careful selection of inputs may require fewer runs than fixed-sample (one-shot) designs, leading to more efficient experimental designs. This characteristic is most critical when we are dealing with complex systems that require expensive simulations. Such designs are particularly useful when the experimenter may be more interested in obtaining a good prediction over a localized region. Recently, in a Kriging metamodeling context, both Kleijnen and van Beers [24] and Sasena et al. [36] developed sequential designs; see also Alam et al. [1], Brantley and Chen [7], Hendrickx and Dhaene [14] and Keys and Rees [16].

In this article, however, we focus on metamodel techniques for the design of experiments, to be analyzed through a possibly nonlinear regression metamodel. If the metamodel response is nonlinear, then the usual measures of performance of a design depend on the parameters being estimated. A sequential approach is then naturally suggested: one should choose design points so as to maximize a measure of performance evaluated at the estimates obtained from observations made at previous design points.

Santos and Santos [35] proposed a sequential design to improve the overall accuracy of a nonlinear regression metamodel with a single input. The procedure adds new points to the design where there are large gaps in the output. It is noted that this technique needs to be conjugated with initial evenly spaced inputs in order to assess regions where the response is regular. The approach improves the accuracy of the resulting metamodel compared with the results of an evenly spaced design. In this paper, we propose an algorithm to address multiple inputs. This algorithm is based on a wire-frame representation of the response surface using a mesh of triangles. The construction of the wire-frame borrows ideas from 3D computer graphics algorithms, where there is a need to expose the details of 3D objects. The values of each factor are scaled to the interval [0; 1] to account for significantly different ranges of each input; the response is also scaled to the same interval in order to normalize distance calculations. In our method the design points are selected to ensure an approximate uniform response behavior. As a result, it concentrates design points in sub-regions where the input–output behavior is of more interest. Subsequent design points are chosen to minimize the response surface gaps. These gaps correspond to triangles with large areas. The results are compared with an evenly spaced experimental design. Our comparisons are motivated by the fact that relatively little attention has been paid to the interaction between the choice of an experimental design and the resulting nonlinear metamodel accuracy.

This paper is organized as follows. In Section 2, the construction procedure for the general nonlinear multi-dimensional input metamodel is presented, excluding the experimental design component; the construction procedure includes the establishing of asymptotic results for the least squares estimators, the use of a nonparametric test for checking the variance heterogeneity in simulation experiments, and the use of bootstrapping for metamodel validation. Section 3 is dedicated to the proposed experimental design for multi-dimensional input, as well as the unidimensional simplification. In Section 4 two application examples are presented to illustrate unidimensional and two-dimensional input experimental designs. Section 5 summarizes the results and suggests directions for future work.

2. Metamodel construction

The metamodeling construction process should follow a scientific approach if meaningful conclusions are to be drawn from the experimental data [10]. First the simulation practitioner should state the objective of the study (prediction, optimization, ...), choose the controllable input variables (factors), select the response variable (output) and specify the region of interest within a previously determined operating region of the system. To represent the simulation model, an approximating statistical model is postulated for the data generating process. Given a response, Y, and a vector of independent factors $\mathbf{X}$ influencing Y, the relationship between $\mathbf{X}$ and Y is represented by a mathematical transfer function f characterized by a