Applied Mathematics and Computation 261 (2015) 8–27

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A search grid for parameter optimization as a byproduct of model sensitivity analysis

Jan Verwaeren*, Pieter Van der Weeën, Bernard De Baets

KERMIT, Department of Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Coupure links 653, Ghent B-9000, Belgium

ARTICLE INFO

Keywords:
Grid search
Parameter estimation
Sensitivity analysis
Sphere packing

ABSTRACT

Inverse problem solving, i.e. the retrieval of optimal values of model parameters from experimental data, remains a bottleneck for modelers. Therefore, a large variety of (heuristic) optimization algorithms has been developed to deal with the inverse problem. However, in some cases, the use of a grid search may be more appropriate or simply more practical. In this paper an approach is presented to improve the selection of the grid points to be evaluated and which does not depend on the knowledge or availability of the underlying model equations. It is suggested that using the information acquired through a sensitivity analysis can lead to better grid search results. Using the sensitivity analysis information, a Gauss–Newton-like matrix is constructed and the eigenvalues and eigenvectors of this matrix are employed to transform naive search grids into better thought-out ones. After a theoretical analysis of the approach, some computational experiments are performed using a simple linear model, as well as more complex nonlinear models.

1. Introduction

Simulations with computer models offer a great time reduction when compared to laboratory experiments. However, the performance of these models strongly depends on the knowledge that is incorporated into the model as well as a correct parametrization of the model [1]. Solving the inverse problem, i.e. finding good values of model parameters using experimental data, remains a bottleneck when modeling (a)biological processes [2,3]. Often, an inverse problem is considered to be a (continuous) mathematical optimization problem, where one needs to find values for optimization variables (here, the model parameters) that minimize an objective function (here, the cost function or the misfit of the model on experimental data). To solve these optimization problems, a myriad of optimization algorithms is available. Some of the more popular optimization algorithms used for continuous optimization problems are quasi-Newton methods such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [4], sequential quadratic programming [5], genetic algorithms [6], (simplex) simulated annealing [7,8], stochastic steepest descent [9], and particle swarm optimization [10]. Depending on the properties of the optimization problem, some algorithms will be more successful in solving the problem than others. Important properties are: smoothness of the cost function, number of parameters (i.e. the dimensionality of the problem), the presence of local optima, constraints on feasible values for some parameters, and the time it takes to evaluate the cost function (typically a model run, or a forward solve of the model).

In this paper, we focus on a simple, yet often used principle known as grid search. A grid search evaluates the cost function at a predefined set of locations (points) in the parameter space. Typically, this set of points forms a regularly spaced grid in the

* Corresponding author. Tel.: +32 92645931; fax: +32 92646220.
E-mail address: jan.verwaeren@ugent.be (J. Verwaeren).

http://dx.doi.org/10.1016/j.amc.2015.03.064
0096-3003/

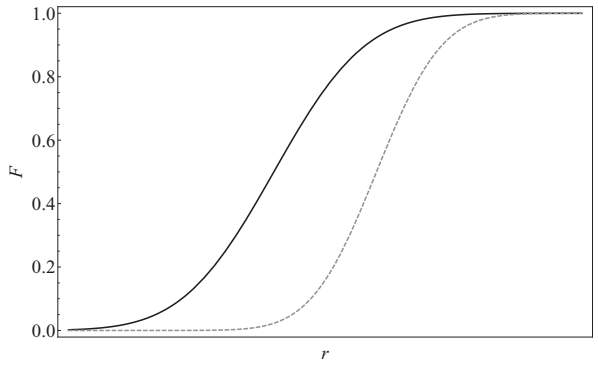

Fig. 1. An example of two CDFs $\mathcal{F}_{\mathbf{x}}$ (black) and $\mathcal{F}_{\mathbf{y}}$ (gray, dashed) of the random variables $F_{\mathbf{x}}$ and $F_{\mathbf{y}}$ for which $F_{\mathbf{x}} \lhd F_{\mathbf{y}}$.

parameter space (hence the name grid search). Subsequently, the point with the smallest cost is selected and returned as the *best* point. Even though grid search is often criticized as being inferior to the methods referred to above, it is still used rather frequently in settings that: (1) count only a small number of parameters, and (2) require highly time consuming model runs (or forward solves). Such settings appear quite frequently, for instance in chemical engineering [11,12], modeling using computational fluid dynamics [13] and tuning of hyper-parameters in machine learning [14]. In those settings, grid search can be tempting as it is simple to implement and is non iterative (i.e. future simulations do not depend on past simulations). The latter property is very tractable when we have multiple computing nodes at our disposal such that all required model evaluations can be performed in parallel.

The main challenge with grid search is to choose the search grid. Choosing a grid essentially amounts to selecting a number of points in the parameter space at which to evaluate the cost function. The naive approach is to take the same number of parameter values equidistantly according to each of the dimensions of the parameter space, and consider all combinations of these values. In this paper, we show that the information obtained from a sensitivity analysis of the model [15] can be exploited to construct an improved search grid. The improved search grid is motivated theoretically using known results on the analysis of cost functions of (linearized) inverse problems [16]. To quantify the improvement that can be obtained by using an optimized search grid, we follow a probabilistic approach. In that setting, we show that the probability distribution function of the cost associated with the *best* grid point can be improved (i.e. high costs become less likely) in terms of stochastic dominance.

The remainder of this paper is organized as follows. In Section 2, we present our main criterion for judging the quality of a search grid. Moreover, we propose a strategy that leads to search grids that are optimal with respect to this criterion. In Section 3, we provide some background on sensitivity analysis and how the measures obtained through this analysis can be employed to improve the selection of a subset of the parameter space for evaluation. The proposed approach is exemplified using a simple linear model as well as two nonlinear models in Section 4 after which the conclusions follow in Section 5.

2. Comparing search grids

2.1. The quality of a search grid

As a starting point, consider an objective function $f:\mathbb{R}^q \to \mathbb{R}$. For a point $\mathbf{x} \in \mathbb{R}^q$, we call $f(\mathbf{x})$ the objective function value at the point $\mathbf{x}$. The problem of finding a minimizer of f is called a minimization problem. Parameter estimation often leads to minimization problems. In those problems, it is natural to compare the *quality* or fitness of two points (or parameter vectors) $\mathbf{x}, \mathbf{y} \in \mathbb{R}^q$ by means of their objective function values. Indeed, if $f(\mathbf{x}) < f(\mathbf{y})$, we naturally say that $\mathbf{x}$ is better than $\mathbf{y}$. As we illustrate hereafter, this idea can easily be extended to judge the quality of search grids.

A search grid is a (mostly finite) countable subset of $\mathbb{R}^q$. Now, let $G = \{\mathbf{x}_i\}_{i=1}^N$ be a search grid of N points. When using a grid search to minimize an objective function, the general strategy is to compute the objective function value at each point in the search grid and to retain only the point with the lowest objective function value. Therefore, we define the objective function value (or quality) $f(G)$ of a search grid G as follows:

$$f(G) = \min_{\mathbf{x} \in G} f(\mathbf{x}). \tag{1}$$

Consequently, given two search grids G_1 and G_2, we say that G_1 is a better search grid than G_2 if $f(G_1) < f(G_2)$.

This reasoning is now extended to settings in which the objective function value at a given point $\mathbf{x}$ is a random variable. We denote this random variable as $F_{\mathbf{x}}$; in statistical literature, $F_{\mathbf{x}}$ is called a random field [17]. The cumulative distribution function (CDF) of $F_{\mathbf{x}}$ is defined as $\mathcal{F}_{\mathbf{x}}(r) = \mathbf{Pr}(F_{\mathbf{x}} \leq r)$. To describe the quality of a point $\mathbf{x}$, the stochastic nature of the objective function has to be taken into account. For example, the quality of $\mathbf{x}$ can be measured through the expected value of $F_{\mathbf{x}}$. However, in this paper, we will use a more stringent condition, known as *stochastic dominance* [18]. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^q$, then we say that $F_{\mathbf{x}}$ is better than $F_{\mathbf{y}}$ if $F_{\mathbf{y}}$ stochastically dominates $F_{\mathbf{x}}$, which is denoted as $F_{\mathbf{x}} \lhd F_{\mathbf{y}}$. Explicitly, $F_{\mathbf{x}} \lhd F_{\mathbf{y}}$ expresses that (see also Fig. 1):