Computational Statistics and Data Analysis 94 (2016) 250–264

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Multiobjective optimization of expensive-to-evaluate deterministic computer simulator models

Joshua Svenson [a,*], Thomas Santner [b]

[a] JPMorgan Chase & Co., 1111 Polaris Parkway, Columbus, OH, 43240, USA

[b] Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH, 43210, USA

ARTICLE INFO

Article history:
Received 20 June 2014
Received in revised form 6 May 2015
Accepted 22 August 2015
Available online 3 September 2015

Keywords:
Computer experiment
Gaussian process
Kriging
Pareto optimization
Nonseparable GP model
Computer simulator model

ABSTRACT

Many engineering design optimization problems contain multiple objective functions all of which are desired to be minimized, say. This paper proposes a method for identifying the Pareto Front and the Pareto Set of the objective functions when these functions are evaluated by *expensive-to-evaluate* deterministic computer simulators. The method replaces the expensive function evaluations by a rapidly computable approximator based on a Gaussian process (GP) interpolator. It sequentially selects new input sites guided by values of an "improvement function" given the current data. The method introduced in this paper provides two advances in the interpolator/improvement framework. First, it proposes an improvement function based on the "modified maximin fitness function" which is known to identify well-spaced non-dominated outputs when used in multiobjective evolutionary algorithms. Second, it uses a family of GP models that allows for dependence among output function values but which permits zero covariance should the data be consistent with this model. A closed-form expression is derived for the improvement function when there are two objective functions; simulation is used to evaluate it when there are three or more objectives. Examples from the multiobjective optimization literature are presented to show that the proposed procedure can improve substantially previously proposed statistical improvement criteria for the computationally intensive multiobjective optimization setting.

1. Introduction

This paper proposes an algorithm for sequentially designing a sequence of inputs at which to run a set of expensive-to-evaluate functions so as to identify the Pareto Front of function values and the Pareto Set of inputs. Throughout, let $\boldsymbol{y}(\boldsymbol{x}) = (y_1(\boldsymbol{x}), \ldots, y_m(\boldsymbol{x}))$ denote the m functions of interest, d the number of inputs, and $\boldsymbol{x} = (x_1, \ldots, x_d)$ a generic input. The input space for $\boldsymbol{x}$ is denoted by $\mathcal{X} \subset \mathbb{R}^d$ and the function values $y(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{X}$, form the *objective space*.

The goal of this paper is to find the *complement* of the set inputs $\boldsymbol{x} \in \mathcal{X}$ that are dominated by one or more inputs in $\mathcal{X}$. An input $\boldsymbol{x}_1 \in \mathcal{X}$ *weakly dominates* $\boldsymbol{x}_2 \in \mathcal{X}$ ($\boldsymbol{x}_1 \succeq \boldsymbol{x}_2$) if $y_i(\boldsymbol{x}_1) \le y_i(\boldsymbol{x}_2)$ for all $i = 1, \ldots, m$. If at least one inequality is strict, then $\boldsymbol{x}_1$ is said to *dominate* $\boldsymbol{x}_2$ ($\boldsymbol{x}_1 \succ \boldsymbol{x}_2$). Equivalently, an input $\boldsymbol{x}_1$ does *not dominate* $\boldsymbol{x}_2$ ($\boldsymbol{x}_1 \not\succ \boldsymbol{x}_2$) if there exists any i such that $y_i(\boldsymbol{x}_1) > y_i(\boldsymbol{x}_2)$. Geometrically, $\boldsymbol{x}_1 \succ \boldsymbol{x}_2$ if $\boldsymbol{y}(\boldsymbol{x}_1)$ is strictly to the "southwest" of $\boldsymbol{y}(\boldsymbol{x}_2)$.

In an analogous fashion, for $\boldsymbol{y}(\boldsymbol{x}_1)$ and $\boldsymbol{y}(\boldsymbol{x}_2)$ in the objective space, $\boldsymbol{y}(\boldsymbol{x}_1)$ is said to *weakly dominate* $\boldsymbol{y}(\boldsymbol{x}_2)$ ($\boldsymbol{y}(\boldsymbol{x}_1) \succeq \boldsymbol{y}(\boldsymbol{x}_2)$) if $y_i(\boldsymbol{x}_1) \le y_i(\boldsymbol{x}_2)$ for all $i = 1, \ldots, m$. If at least one inequality is strict, then $\boldsymbol{y}(\boldsymbol{x}_1)$ is said to *dominate* $\boldsymbol{y}(\boldsymbol{x}_2)$ ($\boldsymbol{y}(\boldsymbol{x}_1) \succ \boldsymbol{y}(\boldsymbol{x}_2)$).

* Corresponding author.
E-mail addresses: joshua.d.svenson@chase.com (J. Svenson), santner.1@osu.edu (T. Santner).

http://dx.doi.org/10.1016/j.csda.2015.08.011
0167-9473/

An output $\boldsymbol{y}(\boldsymbol{x}_1)$ does *not dominate* output $\boldsymbol{y}(\boldsymbol{x}_2)$ ($\boldsymbol{y}(\boldsymbol{x}_1) \not\prec \boldsymbol{y}(\boldsymbol{x}_2)$) if there exists any i such that $y_i(\boldsymbol{x}_1) > y_i(\boldsymbol{x}_2)$. An input vector $\boldsymbol{x} \in \mathcal{X}$ is *Pareto optimal* if and only if there is no $\boldsymbol{x}' \in \mathcal{X}$ such that $\boldsymbol{x} \prec \boldsymbol{x}'$. (Such $\boldsymbol{x}$ are also referred to as *nondominated inputs*. The image $\boldsymbol{y}(\boldsymbol{x})$ of a nondominated input is sometimes referred to as a *nondominated output*.)

Thus the goal can be restated as that of finding the set of all $\boldsymbol{x} \in \mathcal{X}$ which are not dominated by any other input in $\mathcal{X}$; the set of nondominated inputs is called the *Pareto Set*. The set of $(y_1(\boldsymbol{x}), \ldots, y_m(\boldsymbol{x}))$ corresponding to $\boldsymbol{x}$ in the Pareto Set is termed the *Pareto Front*. This paper proposes an algorithm that uses previous $y(\boldsymbol{x})$ evaluations to determine a sequence of inputs $\boldsymbol{x}$ to identify the Pareto Set and the associated Pareto Front.

In most real-world applications, the Pareto Front is an uncountable set and cannot be found analytically. Therefore this paper, as do virtually all papers that identify Pareto Fronts/Sets, finds a discrete approximation to the Pareto Front. In addition, many current methodologies for approximating Pareto Fronts and Pareto Sets, such as the weighted sum method, the ϵ-constrained method, multiobjective evolutionary algorithms (see Coello et al., 2006), and the MULTIMADS algorithm introduced in Audet et al. (2010) have been designed for applications where many (hundreds, possibly thousands) of function evaluations are feasible. Under these conditions, the algorithms above have proven to be very effective at identifying these two Pareto Sets. However, in multiobjective settings where one is running a detailed deterministic computer simulator that is expensive-to-evaluate, only a few dozen function evaluations may be available. This paper proposes methodology for such cases.

In overview, the proposed methodology builds a rapidly-computable surrogate for $\boldsymbol{y}(\boldsymbol{x})$, which is used to guide the search for nondominated points. The $\boldsymbol{y}(\boldsymbol{x})$ surrogate that the authors employ is an interpolator of the training data based on a *Gaussian process* (GP) stochastic model (see Santner et al., 2003). The search selects the $\boldsymbol{x} \in \mathcal{X}$ which maximizes a heuristically selected "expected improvement" criterion. The methodology proposed in this paper provides two key improvements over other interpolator/expected improvement schemes that have been considered in the literature (Schonlau, 1997; Jones et al., 1998; Keane, 2006; Emmerich et al., 2006; Knowles, 2006), First, it is the only proposed multiobjective expected improvement approach that considers stochastic prediction models which allow for dependence *among* the components of $\boldsymbol{y}(\boldsymbol{x})$; procedures that have used dependence models in other applications can lead to improved procedure performance (Ver Hoef and Cressie, 1993; Williams et al., 2010; Fricker et al., 2013). Second, the improvement criterion is based on the maximin fitness function (see Balling, 2003) from the multiobjective evolutionary algorithm (MOEA) literature. This metric of distance quantifies how much better a given output vector is than the current best estimate of the Pareto front and directs MOEAs towards well-spaced designs that are close to the true Pareto front.

The remainder of this paper is organized as follows. To provide context, Section 2 reviews the expected improvement approach proposed in Schonlau (1997) and Jones et al. (1998) for single-objective functions. Section 3 describes the multivariate Gaussian process model that forms the basis for the proposed objective function emulators. Section 4 introduces the proposed improvement criterion and describes its implementation. Section 5 presents the sequential algorithm used to approximate Pareto Front and Pareto Set. Section 6 presents two examples that contrasts the new method with previous proposals from Keane (2006). Finally, Section 7 contains recommendations as to which methods should be used in practice, compares the proposed approach to the hypervolume-based method of Emmerich et al. (2006), and discusses future research regarding the expected improvement approach to multiobjective optimization.

2. Optimization of a single black-box function

To facilitate understanding the multivariate optimization proposal given in this paper, this section introduces the key ideas for the simpler problem of minimizing a single (expensive-to-evaluate) real-valued function $y(\boldsymbol{x})$ defined on a d-dimensional input space $\mathcal{X}$. The method described is due to Schonlau (1997), Jones et al. (1998) who introduced a methods for minimizing $y(\cdot)$ based on a GP model which they called the "efficient global optimization" (EGO) algorithm. The EGO algorithm uses a probabilistic assessment of $y(\boldsymbol{x})$ given the current data that is provided by the GP model. Specifically, these authors determine the information about the global minimum of $y(\cdot)$ that is in each potential $\boldsymbol{x}$ by the (conditional) expectation of a heuristically selected improvement function.

Suppose that $y(\cdot)$ has been evaluated at each input in the "design" $\mathcal{D}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathcal{X}$. Let $\boldsymbol{y}^n = (y(\boldsymbol{x}_1), \ldots, y(\boldsymbol{x}_n))^T$ denote the corresponding vector of outputs. The deterministic output $y(\boldsymbol{x})$ is regarded as a draw from a stationary GP, $Y(\boldsymbol{x})$, with mean β, variance σ^2, and correlation function

$$Cov\left(Y(\boldsymbol{x}), Y(\boldsymbol{x}^\star)\right) = R\left(\boldsymbol{x} - \boldsymbol{x}^\star \mid \boldsymbol{\theta}\right) = \exp\left\{-\sum_{i=1}^{d} \theta_i \left(x_i - x_i^\star\right)^2\right\}, \tag{1}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. The parameters $(\beta, \sigma^2, \boldsymbol{\theta})$ are unknown and must be estimated to complete specification of the GP model.

The GP provides the basis for interpolation of $y(\cdot)$ and uncertainty assessment of the predicted values. When $\left(\sigma^2, \boldsymbol{\theta}\right)$ is known, the best linear unbiased predictor (BLUP) of $y(\boldsymbol{x}_0)$ is

$$\widehat{y}(\boldsymbol{x}_0) = \widehat{\beta} + \boldsymbol{r}^\top(\boldsymbol{x}_0)\boldsymbol{R}^{-1}\left(\boldsymbol{y}^n - \boldsymbol{1}\widehat{\beta}\right), \tag{2}$$