CONTINUOUS SIMULATION OPTIMIZATION WITH MODEL MISMATCH USING GAUSSIAN PROCESS REGRESSION

Alireza Inanlouganji
Giulia Pedrielli
Georgios Fainekos

Sebastian Pokutta

School of Computing, Informatics and
Decision Systems Engineering
Arizona State University
699 S. Mill Ave.
Tempe, AZ, 85281, USA

Stewart School of Industrial & Systems
Engineering
Georgia Institutue of Technology
755 Ferst Drive
NW, Atlanta, GA 30332, USA

ABSTRACT

Multi-fidelity simulation optimization is an emerging area looking at the use of low-fidelity (computationally cheap but inaccurate) models to optimize high-fidelity (expensive and accurate) models. In this context, low-fidelity models exhibit a mismatch to high-fidelity models whose values can be point-wise obtained by querying an expensive simulator. Herein, an efficient multi-fidelity algorithm is proposed for continuous global optimization. The algorithm is made up of an additive model that consolidates low-fidelity and bias (mismatch) predictions. Two sampling criteria with different use of the cumulated high and low-fidelity information are introduced as well as a cheap certificate guiding the decision on whether to sample from the expensive simulator. The performance of proposed algorithms is evaluated using a state of the art stochastic search benchmark algorithm. The results show that the proposed methods can beat the benchmark with improved accuracy, while essentially maintaining the same performance in terms of number of expensive simulations.

1 INTRODUCTION

We investigate the problem of identifying the global minimum of a deterministic black box function $Y_{HF}(\boldsymbol{x})$ whose form is unknown, and its value can only be estimated by querying an oracle at a specific location $\boldsymbol{x}$. More formally, we want to solve: $\boldsymbol{x}^* \in argmin_{x \in \mathbb{X}} Y_{HF}(\boldsymbol{x})$ where $\mathbb{X} \subseteq \mathbb{R}$ is a compact set and $\boldsymbol{x}$ is a d-dimensional decision variable. As an additional "aid", we have access to a low-fidelity (mismatched) version of $Y_{HF}(\boldsymbol{x})$, which we will refer to as the low-fidelity function $Y_{LF}(\boldsymbol{x})$ that has a lower associated computational time to run or optimize. Our objective is then to use the information obtained from sampling $Y_{LF}(\boldsymbol{x})$ to guide the search towards the global optimum of the true function $Y_{HF}(\boldsymbol{x})$. Considering this, we propose an algorithm that learns from this mixed fidelity information. The proposed algorithm is flexible in that no assumption is made about the entity/shape of the mismatch between the low and high-fidelity models. Nonetheless, the choice of mismatched model will clearly influence the performance of the procedure as we will show in the preliminary results section. The proposed algorithm can have various applications. As a motivation to this paper, we have the falsification of Cyber-Physical Systems when a complex non-linear (high-fidelity) and a linearized (low-fidelity) model of such systems are available. This problem was studied in Dokhanchi et al. (2017), and our proposed algorithm extends the approach to handle general model pairs.

The reminder of this paper is organized as follows: Section 1.1 reports the relevant literature and the research contribution is discussed in Section 1.2. The modeling approach and the algorithm are presented in Section 2. Section 3 presents some preliminary results. Finally, Section 4 presents the conclusions and directions for future research.

1.1 Literature Review

To optimize a black-box function, direct search as well as meta-model based approaches have been proposed in the literature (Fu et al. 2015). Direct search methods take decisions related to sampling based on the current point and updating/adapting a sampling rule to establish the next point. The adaptation scheme and the structure of the sampling rule have defined a plethora of different approaches. Examples are improving hit-and-run (Zabinsky et al. 1993; Zabinsky and Smith 1992; Solis and Wets 1981; Brooks 1958), GRASP (Feo and Resende 1995) a greedy adaptive random search sampling technique, and the stochastic ruler method (Yan and Mukai 1992; Alrefaei and Andradòttir 2001).

Meta-model based approaches differ from direct search methods in that they consider all sampled points to make a sampling decision and this is done by constructing a response surface model that emulates the function to optimize where we have not sampled it. In this rich research area, efforts have been dedicated to proposing statistical models to improve the accuracy of the prediction, as well as to the study of effective sampling criteria that are able to handle the exploration/exploitation dilemma (Li et al. 2010).Concerning the model choice, applications of Neural Networks, Kriging models, Regression models and Radial Basis Functions have been proven successful in different application settings (Xu et al. 2015). In this area, Gaussian Processes (GP), which will also be used in this work, have received a lot of attention (Quan et al. 2013; Ankenman et al. 2010; Yin et al. 2011). Nevertheless, most of these models are designed to handle a single fidelity response.

Moving into multi-fidelity optimization research, Xu et al. (2014) proposes MO2TOS (Multi-Fidelity Optimization with Ordinal Transformation and Optimal Sampling) that relies on the concept of Ordinal Transformation (OT). OT is a mapping $\mathbb{X} \to \mathbf{H}$ where $\mathbb{X}$ is d-dimensional discrete space and $\mathbf{H}$ is is a one-dimensional rank space constructed by associating to each point of $\mathbb{X}$ the rank computed according to the corresponding output returned by the low-fidelity model. This mapping, as defined by the authors, can be applied to any finite, countable space $\mathbb{X}$. Once the mapped space is constructed, the solutions are grouped in sub-sets with respect to $\mathbf{H}$ and sampled according to the Optimal Sampling (OS) scheme. The theoretical analysis performed therein shows the properties that the low-fidelity model should have to guarantee an improved performance of the proposed algorithm with respect to the version not using any low-fidelity information. Huang et al. (2015) further extends the previous contribution by using OT not only to transform the solution space, but also to decide on high-fidelity computational budget allocation. Xu et al. (2016a, 2016b) extend the application of OT in sampling. Zhang et al. (2016a) proposes an innovative optimal sampling methodology that maximizes the estimated probability of selecting the best solution. All these approaches handle one low-fidelity model.

In the direction of introducing more fidelities, Zhang et al. (2016b) proposes a novel scheme to weight predictions generated by multiple low-fidelity models creating another equivalent low-fidelity model. More recently, Chen et al. (2017) uses Gaussian Process Regression along with Ordinal Transformation to optimize noisy observations of a black-box function. While the authors still focus on discrete optimization, an approach to handle an arbitrary number of low-fidelity models is proposed.

In the area of continuous optimization, Santner (2013) is the main reference for the modeling of information from several fidelities discussing co-kriging as a possible framework to capture relationships between them when we can "rank" models in terms of their fidelity. Forrester et al. (2007) develops a co-kriging based method where an exchange algorithm is proposed that chooses different sample points at the different fidelity levels. The authors consider a wing optimization problem as a case study to validate and show the applicability of the proposed method. More recently, Ulaganathan et al. (2015), extend the co-kriging formulation with multi-fidelity gradient information. The authors numerically show how

the extra information from the multi-fidelity gradient can help the kriging model achieve better prediction accuracy. Liu et al. (2016) propose to use a multi-fidelity Gaussian Process with Memetic Differential Evolution, while Chen et al. (2015) decompose the high-fidelity response into trend and residual components and use a non-parametric locally weighted regression with smoothing and a kriging model as models for optimization. Osorio and Selvam (2017) propose a method that combines information from evaluating models with different fidelities to optimally control traffic networks. At each iteration, the algorithm decides on the model to evaluate based on an estimate of accuracy loss due to running the low-fidelity model.

In summary, Ordinal Transformation and Co-kriging based methods are the two families of approaches that tackle multi-fidelity simulation-optimization problems. However, Ordinal Transformation cannot handle continuous input variables. This is especially important not only when we have continuous problems, but also when the number of solutions is huge and the computational cost of the low-fidelity model, while low, is not completely negligible. While Co-kriging-based approaches can solve this issue, most of the algorithms require to know the rank of the models in terms of fidelity and several authors highlight the computational burden associated with the estimation of the co-kriging model hyper-parameters (Santner, 2013).

1.2 Contribution

The contribution of this paper is threefold:

1. Our new algorithm, which is markedly different from Ordinal Transformation based approaches, can handle cases where the low-fidelity response is not known across the entire solution space. This is the case in continuous as well as combinatorial optimization.
2. We propose an additive model to capture the relationship between the high and low-fidelity models.
3. Sampling is mainly performed in "low-fidelity": the proposed algorithm generates candidate solutions considering low-fidelity guesses and switches to high-fidelity only if it is required. This becomes particularly important when different optimizers are used to search the low and high-fidelity spaces, respectively.

2 METHODOLOGY

In this section we present our proposed approach. Section 2.1 details the modeling of the relationship between high and low-fidelity data, while Section 2.2 focuses on the search across the solution space.

2.1 Modeling Approach

We assume that the performance metric of interest can be evaluated for a specific configuration by running a very expensive oracle (high-fidelity simulation) or quickly assessed by means of a low precision/low cost oracle (low-fidelity simulator). Suppose that, for a given input vector x, we have $Y_{LF}(\boldsymbol{x})$ and $Y_{HF}(\boldsymbol{x})$ to represent the low and high-fidelity evaluations of the system response at point x. Similar to model (3) discussed in (Son and Choi 2016) where the authors just evaluate the effect of a scale factor and correction on multi-fidelity models, we assume the following relationship between the two:

$$Y_{HF}(\boldsymbol{x}; \theta_{\mathbf{HF}}) = \alpha(\boldsymbol{x}; \theta_\alpha) Y_{LF}(\boldsymbol{M}(\boldsymbol{x}; \theta_{\mathbf{M}}); \theta_{\mathbf{LF}}) + B(\boldsymbol{M}(\boldsymbol{x}; \theta_{\mathbf{M}}); \theta_{\mathbf{B}}), \ \boldsymbol{x} \in \mathbb{X}. \tag{1}$$

At the basis of this work, we assume that Y_{LF} and Y_{HF}, i.e., the low and high-fidelity responses as well as the bias process B representing the distance between them can be modeled as Gaussian processes. In model (1), $\theta_{\mathbf{LF}}$ and $\theta_{\mathbf{B}}$ refer to the vector of the hyper-parameters needed to construct a predictor for the Gaussian Processes Y_{LF} and B, respectively. The multiplier of the low-fidelity model, $\alpha(\boldsymbol{x}; \theta_\alpha)$, can be interpreted as an amplifier of the low-fidelity response and it is generally a function of the location $\boldsymbol{x}$ and it is parametrized through θ_α. Also, $M(\boldsymbol{x}; \theta_{\mathbf{M}})$ represents a general mapping function from the low-fidelity solution space to the high-fidelity solution space. Such a mapping can help in increasing the

similarity between low and high-fidelity models. This mapping function is parametrized through $\theta_{\mathbf{M}}$, which will require development of learning mechanisms to be efficiently estimated. In this first version of the algorithm, we will not account for the mapping function, and we also assume $\alpha = 1$ to obtain the following simpler model structure:

$$Y_{HF}(\boldsymbol{x}; \theta_{\mathbf{HF}}) = Y_{LF}(\boldsymbol{x}; \theta_{\mathbf{LF}}) + B(\boldsymbol{x}; \theta_{\mathbf{B}}), \ \boldsymbol{x} \in \mathbb{X}. \tag{2}$$

In the proposed framework, which will be presented in Section 2.2, the sampling will be performed in low-fidelity, unless the proposed certificate test is not satisfied by the new sample. Therefore, generally much more information on low-fidelity process $Y_{LF}(\boldsymbol{x}; \theta_{\mathbf{LF}})$, will be available than the bias process $B(\boldsymbol{x}; \theta_{\mathbf{B}})$, which is the expensive component in model (2). It is important to understand that the bias term is vital to "correct" predicted low-fidelity responses obtained with abundant information. This correction results in improved prediction accuracy of the high-fidelity response, which, in turn, results in improved efficiency of the search procedure overall.

Section 2.1.1 briefly introduces Gaussian Process estimation, as the basis of the construction of the predictor for model (2). Section 2.1.2 shows the details of applying the Gaussian Process theory to the specifics of model (2).

2.1.1 Gaussian Process Models

A Gaussian Process (GP) is completely characterized by its mean and the covariance function. In this paper, we will adopt the squared exponential covariance function (Santner et al. 2013; Williams and Rasmussen 2006):

$$\mathbf{K}(\mathbf{x_i}, \mathbf{x_j}) = \sigma_f{}^2 \ exp\{\sum_{k=1}^{d} \frac{-|x_i{}^k - x_j{}^k|^2}{l_k{}^2}\}. \tag{3}$$

where $\sigma_f{}^2$ and $l_k, \ k = 1, 2, \ldots d$ are the hyper-parameters of the covariance function (summarized by the vector of hyper-parameters θ) which can be estimated from the training data and $x_i{}^k$ refers to the kth element of the ith input vector, $\boldsymbol{x_i}$. The variance-covariance matrix $\mathbf{K}$ is an $n \times n$ symmetric positive-definite matrix, where n is the number of sampled vectors and the ij^{th} element represents the value of the covariance between the ith and j^{th} input vectors. We refer to the collection of sampled points as the training set $\mathbb{X}_T \subseteq \mathbb{X}$ (such as $|\mathbb{X}_T| = n$) and to the corresponding function values as $\boldsymbol{y}$. Assuming the hyper-parameters θ to be known, then for any new input vector $\boldsymbol{x_0}$, using the theory of GP models, the posterior distribution of the output $\hat{Y}(\boldsymbol{x_0})$ is (Rasmussen and Williams 2006):

$$\hat{Y}(\boldsymbol{x_0})|\mathbb{X}_T; \theta \ \sim \mathbf{N}(\mu_y(\boldsymbol{x_0}), \sigma_y{}^2(\boldsymbol{x_0})). \tag{4}$$

$$\mu_y(\boldsymbol{x_0}) = \mathbf{c^T}(\boldsymbol{x_0}, \mathbb{X}_T)\mathbf{K^{-1}}\boldsymbol{y}. \tag{5}$$

$$\sigma_y{}^2(\boldsymbol{x_0}) = \mathbf{K}(\boldsymbol{x_0}, \boldsymbol{x_0}) - \mathbf{c^T}(\boldsymbol{x_0}, \mathbb{X}_T)\mathbf{K^{-1}}\mathbf{c}(\boldsymbol{x_0}, \mathbb{X}_T). \tag{6}$$

where $\mathbf{c}^T(\boldsymbol{x_0}, \mathbb{X}_T)$ is the vector containing the correlations between $\boldsymbol{x_0}$ and the set of sampled points $\mathbb{X}_T$ calculated using Eq. (3). In next section, the derivation of model (2) using GPs will be discussed.

2.1.2 Multi-fidelity Model Estimation

To derive the model (2) we will have two separate sets of sampled points, $\mathbb{X}_{LF}$ representing the set of points sampled in low-fidelity and $\mathbb{X}_B$, representing the set of points for which both high-fidelity as well as low-fidelity evaluation are available. It should be noted that this additional notation is meant to differentiate between the different sets of points adopted to construct different predictions of the same response. It is apparent that $\mathbb{X}_{LF}$ is the "cheapest" set to construct. Therefore, we will generally have $|\mathbb{X}_B| \ll |\mathbb{X}_{LF}|$.

Given a desired prediction location $\boldsymbol{x_0}$, to estimate $\hat{B}(\boldsymbol{x_0})|\mathbb{X}_B; \theta_{\mathbf{B}}$, which, for simplicity, will be referred to as $\hat{B}(\boldsymbol{x_0})$ from now on, we first obtain the elements of observed bias vector, b using Eq. (7) below.

$$b(\boldsymbol{x}) = y_{HF}(\boldsymbol{x}) - y_{LF}(\boldsymbol{x}), \ \boldsymbol{x} \in \mathbb{X}_B. \tag{7}$$

Using observed bias, b vector as the response, a GP is trained to estimate the bias at any given solution x_0. Using Eq. (4) – (6) we obtain:

$$\hat{B}(x_0) \sim \mathbf{N}(\mu_B(x_0), \sigma_B^2(x_0)). \tag{8}$$

$$\mu_B(x_0) = \mathbf{c}^T(x_0, \mathbb{X}_B)\mathbf{K}^{-1}b. \tag{9}$$

$$\sigma_B^2(x_0) = \mathbf{K}(x_0, x_0) - \mathbf{c}^T(x_0, \mathbb{X}_B)\mathbf{K}^{-1}\mathbf{c}(x_0, \mathbb{X}_B). \tag{10}$$

Concerning the low-fidelity component, we estimate $\hat{Y}_{LF}(x_0)|\mathbb{X}_{LF}; \theta_{LF}$, which for simplicity of notation will be referred to as $\hat{Y}_{LF}(x_0)$ from now on. Using $\mathbb{X}_{LF}$ as input matrix and $y_{LF|\mathbb{X}_{LF}}$ a vector collecting low-fidelity responses for the locations in the set $\mathbb{X}_{LF}$ as the output, another GP model is trained to estimate low-fidelity response at any given solution x_0 as follows:

$$\hat{Y}_{LF}(x_0) \sim \mathbf{N}(\mu_{LF}(x_0), \sigma_{LF}^2(x_0)). \tag{11}$$

$$\mu_{LF}(x_0) = \mathbf{c}^T(x_0, \mathbb{X}_{LF})\mathbf{K}^{-1}y_{LF|\mathbb{X}_{LF}}. \tag{12}$$

$$\sigma_{LF}^2(x_0) = K(x_0, x_0) - \mathbf{c}^T(x_0, \mathbb{X}_{LF})\mathbf{K}^{-1}\mathbf{c}(x_0, \mathbb{X}_{LF}). \tag{13}$$

Finally, we can estimate $\hat{Y}_{HF}(x_0)$ using a liner combination of the two predictions in (8) and (11), obtaining:

$$\hat{Y}_{HF}(x_0) \sim \mathbf{N}(\mu_{HF}(x_0), \sigma_{HF}^2(x_0)). \tag{14}$$

$$\mu_{HF}(x_0) = \mu_{LF}(x_0) + \mu_B(x_0). \tag{15}$$

$$\sigma_{HF}^2(x_0) = \sigma_{LF}^2(x_0) + \sigma_B^2(x_0). \tag{16}$$

We want to remind to the reader that to guarantee adequacy of the additivity of the processes, we should have $\mathbb{X}_{LF} \cap \mathbb{X}_B = \emptyset$, while this is not the case, we have that $|\mathbb{X}_B| \ll |\mathbb{X}_{LF}|$ as already mentioned, this dilutes the dependency effect generated by the sampling policy. This basically implies that the sampling process for the generation of low-fidelity observations and the one for the learning of the low/high-fidelity dependency do not overlap.

In this work, we use a third model to guide the decision whether to sample in high-fidelity or not. Specifically, we derive the predicted low-fidelity response using the same relationship as in Eq. (2), but with the focus of estimating the low-fidelity response, i.e., we look at the relationship in (2) as $Y_{LF}(x; \theta_{LF}) = Y_{HF}(x; \theta_{HF} - B(x; \theta_B)), x \in \mathbb{X}$. While B component is the same as in (8)-(10), $\hat{Y}_{HF}(x; \theta_{LF})$ is estimated only using the high-fidelity evaluations, $y_{HF|\mathbb{X}_B}$. To distinguish this predictor from the one in Eq. (14), we refer to it as $\hat{Y}_{HF}|\mathbb{X}_B(x; \theta_{LF})$ and we refer to this alternative estimation of the low-fidelity model as $\hat{Y}_{LF}|\mathbb{X}_B$ and derive it as follows:

$$\hat{Y}_{LF|\mathbb{X}_B}(x_0) \sim \mathbf{N}(\mu_{LF|\mathbb{X}_B}, \sigma_{LF|\mathbb{X}_B}(x_0)). \tag{17}$$

$$\mu_{LF|\mathbb{X}_B}(x_0) = \mu_{HF|\mathbb{X}_B}(x_0) - \mu_B(x_0). \tag{18}$$

$$\sigma_{LF|\mathbb{X}_B}^2(x_0) = \sigma_{HF|\mathbb{X}_B}^2(x_0) + \sigma_B^2(x_0). \tag{19}$$

We highlight that different from the prediction in (14)-(16), the derivation of the conditional density is not an exact result, because we do not account for the inherent dependency between the processes $\hat{Y}_{HF}$ and $\hat{B}$ generated by the fact that the set of sampled points in high-fidelity $\mathbb{X}_{HF}$ and the set of sampled points in the bias set $\mathbb{X}_B$ are the same. While numerical results are robust to this approximation, this is an aspect requiring further investigation.

2.2 Algorithm

In this section the proposed algorithm will be explained by discussing its components followed by an outline given in Algorithm 1. Section 2.2.1 addresses the sampling part, while the certificate responsible for the choice of whether or not we should sample a point in high-fidelity is dicusses in Section 2.2.2.

2.2.1 Sampling

In this preliminary work, we investigate two versions of the algorithm. The first version uses the low-fidelity evaluation at the best location up to iteration k, i.e., $x^*_{k,LF} \in argmin_{x \in \mathbb{X}_{LF}} Y_{LF}(x)$, where the achieved low-fidelity function value is referred to as $y^b_{k,LF}$ and $\mathbb{X}_{k,LF}$ is the set of locations that have been sampled in low-fidelity up to iteration k. The second approach uses the high-fidelity evaluation at the best location up to iteration k, i.e., $x^*_{k,HF} \in argmin_{x \in \mathbb{X}_{HF}} Y_{HF}(x)$, where the achieved high-fidelity function value is referred to as $y^b_{k,HF}$ and $\mathbb{X}_{k,HF}$ is the set of locations that have been sampled in high-fidelity. We highlight that using low-fidelity information is convenient in sampling since the low-fidelity model may be "easier to optimize" and to sample. In this work, we adopt the Expected Improvement (EI) as sampling criteria (Locatelli 1997, Jones et al. 1998). The EI is a common sampling criterion in surrogate-based optimization and its derivation is discussed in Jones et al. (1998). *Considering the first version of the algorithm*, the improvement at any location,*x*, is derived as $I(x) = max(y^b_{k,LF} - \hat{Y}_{k,LF}(x), 0)$, where $\hat{Y}_{k,LF}(x)$ is the prediction obtained from model (12) at iteration k, when $\mathbb{X}_{k,LF}$ points have been sampled in low-fidelity. The expected improvement will be:

$$x^*_k = argmax_{x \in \mathbb{X}|\mathbb{X}_{k,LF}} E[I(x)] = \int_{-\infty}^{y^b_{LF}} (y^b_{LF} - \hat{Y}_{LF}(x))^+ f(\hat{Y}_{LF}(x)) dy. \quad (20)$$

where $\mathbb{X}_{k,LF}$ is the set of sampled points in low-fidelity, $f\hat{Y}_{LF}(x))$ is the density function of the predicted low-fidelity response which in this case is normal with its mean and variance calculated as functions of candidate solution, *x*, using Eq. (12) and (13). Note that *for the second version of the algorithm*, a similar equation can be derived to calculate EI by simply replacing y^b_{LF} with $y^b_{HF}, \hat{Y}_{LF}(x)$ with $\hat{Y}_{HF}(x)$ and $\mathbb{X}_{k,LF}$ with $\mathbb{X}_{k,B}$.

2.2.2 Low-fidelity Improvement "Certificate"

In addition to choosing the location where sampling should be performed (i.e., a candidate solution), in multi-fidelity frameworks, we need to understand whether simulation in high-fidelity is required. To do so, we devised a statistical certificate of the relationship between the low and high-fidelity models, which is presented in this section. For the candidate point at iteration k, x^*_k, we want to certify the validity of our model in Eq. (2) representing the relationship we constructed between low and high-fidelity models. We achieve this objective with the following statistical test:

$$\begin{cases} H_0 :, & \hat{Y}_{LF}(x^*_k) \sim \mathbf{N}(\mu_{LF|\mathbb{X}_B}(x^*_k), \sigma_{LF|\mathbb{X}_B}(x^*_k)) \\ H_1 :, & \text{otherwise} \end{cases} \rightarrow Q = \frac{y_{LF}(x^*_k) - \mu_{LF|\mathbb{X}_B}(x^*_k)}{\sigma_{LF|\mathbb{X}_B}(x^*_k)} \geq -Z_c. \quad (21)$$

The idea behind this test is to check whether the relationship we have assumed between low-fidelity and high-fidelity models is valid at this location. If the null hypothesis in Eq. (21) cannot be rejected, there is no need for sampling in the high-fidelity space. Otherwise, we sample in high-fidelity to update the low-fidelity predictor using Eq. (11), and the high-fidelity predictor in Eq. (14).

Algorithm 1

		Initialization
Step 0		$k = 0; \mathbb{X}_{B,k} = \emptyset, \mathbb{X}_{LF,k} = \emptyset, y_{LF|\mathbb{X}_B} = \emptyset, y_{LF|\mathbb{X}_{LF}} = \emptyset, y_{HF|\mathbb{X}_B} = \emptyset$. Set the initial number of samples n_{0B}, and n_{0LF}, using a Latin Hypercube Design. Set the maximum number of simulations in low-fidelity CB_{LF} and the maximum number of simulations in high-fidelity CB_{HF}; Set $\mathbb{X}_{B,k} \leftarrow x_{i i=1}^{n_{0,B}}$, update $\mathbb{X}_{LF,k} \leftarrow x_{i i=1}^{n_{0,LF}}$;

	While	$CB_{LF} \geq 0$ & $CB_{HF} \geq 0$
Step 1		**Expensive Update** $\mathbb{X}_B \leftarrow \mathbb{X}_{B,k} \cup \mathbb{X}_B$; $y_{HF|\mathbb{X}_B} \leftarrow \{y_{HF|\mathbb{X}_B} \cup y_{HF|\mathbb{X}_B}(x); x \in \mathbb{X}_{B,k}\}$, $y_{LF|\mathbb{X}_B} \leftarrow y_{LF|\mathbb{X}_B} \cup \{y_{LF|\mathbb{X}_B}(x); x \in \mathbb{X}_{B,k}\}$; Evaluate $b(x) \leftarrow (y_{HF|\mathbb{X}_B}(x) - y_{LF|\mathbb{X}_B}(x))$, $x \in \mathbb{X}_{B,k}$; With $y_{HF|\mathbb{X}_B}, y_{LF|\mathbb{X}_B}, \mathbb{X}_B$ construct: $\hat{B}\ GP(\mu_B, \sigma_B{}^2)$ using Eq. (8)-(10);
Step 2		**Cheap Update** $\mathbb{X}_{LF} \leftarrow \mathbb{X}_{LF,k} \cup \mathbb{X}_{LF}$; $y_{LF|\mathbb{X}_{LF}} \leftarrow y_{LF|\mathbb{X}_{LF}} \cup \{y_{LF|\mathbb{X}_{LF}}(x); x \in \mathbb{X}_{LF,k}\}$; With $y_{LF|\mathbb{X}_{LF}}, \mathbb{X}_{LF}$ construct: $\hat{Y}_{LF} \sim GP(\mu_{LF}, \sigma_{LF}{}^2)$ using Eq. (11)-(13);

		Sampling
		Criteria 1 *Criteria 2*
Step 3.1		Compute $EI(x)$ using Eq. (20), Compute $EI(x)$ using Eq. (20), set:

Criteria 1, set:
$$x_k^* \leftarrow argmax_{x \in \mathbb{X}} EI(x) = \int_{-\infty}^{y_{LF}^b} (y_{LF}^b - \hat{Y}_{LF}(x))^+ f(\hat{Y}_{LF}(x)) dy$$

Criteria 2:
$$x_k^* \leftarrow argmax_{x \in \mathbb{X}} EI(x) = \int_{-\infty}^{y_{HF}^b} (y_{HF}^b - \hat{Y}_{HF}(x))^+ f(\hat{Y}_{HF}(x)) dy$$

Step 3.2		Evaluate $y_{LF}(x_k^*)$; Update $y_{LF}^b \leftarrow \min_{x \in \mathbb{X}_{LF} \cup \mathbb{X}_B} y_{LF|\mathbb{X}_{LF} \cup \mathbb{X}_B}$, $\mathbb{X}_{LF,k+1} \leftarrow x_k^*$, $CB_{LF} \leftarrow CB_{LF} - 1$; **Certificate:** Estimate $\hat{Y}_{LF|\mathbb{X}_B}(x_k^*)$ using Eq. (17)-(19); Use $y_{LF}(x_k^*)$ and $\hat{y}_{LF|\mathbb{X}_B}(x_k^*)$ in Condition (21); **If** (Condition (21) holds TRUE) $\mathbb{X}_{B,k+1} \leftarrow \{\}$ Go to Step 2 **Else** $\mathbb{X}_{B,k+1} \leftarrow x_k^*$; $CB_{HF} \leftarrow CB_{HF} - 1$; Go to Step 1 **End**
	End	

3 NUMERICAL EXPERIMENTS

In this section both versions of proposed algorithm will be evaluated using the Efficient Global Optimization (EGO) method as the benchmark. First, the test function and different low-fidelity models of interest are introduced in Section 3.1. Then, the numerical experiments and results are discussed in Section 3.2. For the test function, as the optimal solution is known, the algorithm stops whenever at least one of the following conditions is met: (1) $\frac{y_{HF}^b - y_{HF}^*}{y_{HF}^*} \leq 0.01$; (2) $Itr_{LF} \geq 500$; (3) $Itr_{HF} \geq 50$. Here Itr_{LF} and Itr_{HF} are the number of samples taken in low and high-fidelity, respectively.

3.1 Test Functions

The function $Y_{HF}(x)$ below will be considered as the high-fidelity model where $x_i \in [0.1, 1]$ (Zabinsky and Bulger 2010).

$$Y_{HF}(\boldsymbol{x}) = -2.5 \prod_{i=1}^{d} sin(\pi x_i) - \prod_{i=1}^{d} sin(5\pi x_i). \tag{22}$$

Where d is the problem dimension. Table 1 reports four different low-fidelity models that we used along with their correlation coefficient characterizing their dependency with high-fidelity model.

Table 1: Different low-fidelity models.

Low-Fidelity Model	Model Form	Correlation
$Y_{LF,1}(\boldsymbol{x})$	$-2\prod_{i=1}^{d} sin(\pi x_i)$	0.87
$Y_{LF,2}(\boldsymbol{x})$	$-0.8\prod_{i=1}^{d} sin(5\pi x_i)$	0.49
$Y_{LF,3}(\boldsymbol{x})$	$2\prod_{i=1}^{d} sin(\pi x_i)$	-0.87
$Y_{LF,4}(\boldsymbol{x})$	$0.8\prod_{i=1}^{d} sin(5\pi x_i)$	-0.49

3.2 Preliminary Results

The proposed algorithm in its two versions along with the benchmark Efficient Global Optimization (EGO, Jones et al. (1998)) are used to optimize the test function introduced above and their performance is compared. EGO only samples in high-fidelity space and works independent of the low-fidelity model of choice to be used in the proposed multi-fidelity algorithm. All algorithms were run for 50 macro-replications and we considered, as performance metric, the relative distance between the reported optimal solution and the real optimal solution of the high-fidelity model $\frac{||\hat{x}_i^* - x^*||}{||x^*||}$, where x^* is the global optimal solution of the high-fidelity model and $\hat{x}_i^*$ is the optimal solution reported by the algorithm at the ith macro-replication. We also report the number of high-fidelity simulations ran before convergence/stoppage of the algorithms over different macro-replications as "High-fidelity Simulations" in both Table 2 and Table 3.

For each macro-replication we ran the proposed algorithms and EGO with the same seed. Upon observing convergence of our algorithm, we also stopped EGO, i.e., EGO could run as many high-fidelity simulations as those needed by the multi-fidelity counter-part to converge.

Table 2, shows the results obtained from the first version of the algorithm that calculates EI over the predicted high-fidelity response. From the results, it can be observed that our multi-fidelity approach outperforms EGO no matter which low-fidelity model is used. Also, the gap between the performance of the first version of the algorithm and EGO is larger when second and fourth low-fidelity models are used. While the reason for this result is highly related to the specific sampling criteria, it reveals that model correlation may not be the only important metric to characterize a low-fidelity model since apparently the nonlinear dependency of the high-frequency component and the original function is helping our search procedure.

Table 2: Statistics on optimality distance for the first version of proposed algorithm.

Low-fidelity Model	Dimension	Proposed Algorithm $\frac{\|\hat{x}_i^* - x^*\|}{\|x^*\|}$		EGO $\frac{\|\hat{x}_i^* - x^*\|}{\|x^*\|}$		High-fidelity Simulations	
		Average	Std Err	Average	Std Err	Average	Std Err
$Y_{LF,1}(\boldsymbol{x})$	3	0.07	0.01	0.17	0.02	2.28	0.17
	4	0.04	0.01	0.16	0.02	2.76	0.23
$Y_{LF,2}(\boldsymbol{x})$	3	0.00	0.00	0.17	0.01	2.38	0.11
	4	0.00	0.00	0.16	0.02	2.30	0.10
$Y_{LF,3}(\boldsymbol{x})$	3	0.09	0.01	0.15	0.01	3.18	0.28
	4	0.05	0.01	0.14	0.01	2.86	0.29
$Y_{LF,4}(\boldsymbol{x})$	3	0.00	0.00	0.17	0.01	2.38	0.11
	4	0.00	0.00	0.16	0.02	2.42	0.16

Table 3: Statistics on optimality distance for the second version of proposed algorithm.

Low-fidelity Model	Dimension	Proposed Algorithm $\frac{\|\hat{x}_i^* - x^*\|}{\|x^*\|}$		EGO $\frac{\|\hat{x}_i^* - x^*\|}{\|x^*\|}$		High-fidelity Simulations	
		Average	Std Err	Average	Std Err	Average	Std Err
$Y_{LF,1}(\boldsymbol{x})$	3	0.00	0.00	0.18	0.01	2.02	0.11
	4	0.00	0.00	0.17	0.02	2.08	0.07
$Y_{LF,2}(\boldsymbol{x})$	3	0.13	0.02	0.10	0.02	15.02	0.97
	4	0.16	0.01	0.03	0.01	31.86	1.21
$Y_{LF,3}(\boldsymbol{x})$	3	0.00	0.00	0.18	0.01	2.02	0.11
	4	0.00	0.00	0.15	0.02	2.60	0.19
$Y_{LF,4}(\boldsymbol{x})$	3	0.12	0.02	0.09	0.02	14.78	1.02
	4	0.17	0.01	0.04	0.01	31.82	1.43

Table 3 shows the results from the second version of the algorithm, which calculates EI over the predicted low-fidelity. Analyzing the results, it can be observed that, while the algorithm still outperforms EGO when the low-fidelity model is highly correlated with the high-fidelity model, it fails when correlation is weak. While optimizing in low-fidelity can have advantages when the low-fidelity model has optimization-amenable characteristics, relying too much on the low-fidelity will slow down the search especially when the low-fidelity model is "bad". It is important to note that, this brings to light the need to have good low-fidelity models when we are optimizing over them.

Considering both Table 2 and Table 3, with good low-fidelity models, the second version of the algorithm outperforms the first version and needs a slightly smaller number of high-fidelity simulations on average. However, with less correlated low-fidelity models, it is the first version of the algorithm that does a better job of guiding the search towards optimal solution to the high-fidelity model. This suggests that a switching mechanism between the two criteria based on local correlation between the models within the algorithm may be effective in more complex cases.

4   CONCLUSIONS & FUTURE RESEARCH

Using low-fidelity information to guide the search towards the optimum of an expensive high-fidelity model in a multi-fidelity optimization framework is an issue that recently has attracted much interest in the Simulation-Optimization research community. In this paper, two versions of an algorithm are proposed to address this issue. Since the proposed framework does not evaluate the low-fidelity model for all possible solutions, it can handle continuous input variables and fairly expensive low-fidelity models. Also, we introduce the concept of Low-Fidelity Improvement Certificate, which allows us to coordinate sampling in low and high-fidelity. Testing with a commonly used test function and different low-fidelity model forms, we observed that with good low-fidelity models the version of the algorithm that samples only relying on low-fidelity information outperforms the other version that samples relying on high-fidelity information. However, this last version is more robust to the quality of the low-fidelity and it always outperforms the benchmark algorithm (EGO). Results show that the low and high-fidelity information can be efficiently used to decrease the number of expensive high-fidelity evaluations needed to obtain the optimal solution, making the problem much less computationally expensive.

For future research, we will work to add space mapping component to the algorithm as a generalization of the Ordinal Transformation proposed for discrete spaces to make it more efficient. Also, we want to test more models and certificates as well as several low-fidelity optimizers. Finally, we will investigate the convergence properties of the proposed methods.

ACKNOWLEDGMENTS

REFERENCES

Alrefaei, M. H., and S. Andradóttir. 2001. "A Modification of the Stochastic Ruler Method for Discrete Stochastic Optimization". *European Journal of Operational Research* 133(1):160-182.

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58(2):371-382.

Brooks, S. H. 1958. "A Discussion of Random Methods for Seeking Maxima". *Operations Research* 6(2):244-251.

Chen, R., J. Xu, S. Zhang, C. H. Chen, and L. H. Lee. 2015. "An Effective Learning Procedure for Multi-fidelity Simulation Optimization with Ordinal Transformation". In *Proceedings of the 11th IEEE International Conference on Automation Science and Engineering (CASE)*, August 24th − 28th, Gothenburg, Sweden, 702-707.

Dokhanchi, A., S. Yaghoubi, B. Hoxha, and G. Fainekos. 2017. "Vacuity Aware Falsification for MTL Request-response Specifications". In *Proceedings of the 13th IEEE International Conference on Automation Science and Engineering (CASE)*, August 20th − 23th, Xi'an, China, 1332-1337.

Feo, T. A., and M. G. Resende. 1995. "Greedy Randomized Adaptive Search Procedures". *Journal of Global Optimization* 6(2):109-133.

Forrester, A. I., A. Sóbester, and A. J. Keane. 2007. "Multi-fidelity Optimization via Surrogate Modelling". *Proceedings of the Royal Society of London a: Mathematical, Physical and Engineering Sciences* 463(2088):3251-3269.

Huang, E., J. Xu, S. Zhang, and C. H. Chen. 2015. "Multi-fidelity Model Integration for Engineering Design". *Procedia Computer Science* 44:336-344.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13(4):455-492.

Li, Y., S. H. Ng, M. Xie, and T. N. Goh. 2010. "A Systematic Comparison of Metamodeling Techniques for Simulation Optimization in Decision Support Systems". *Applied Soft Computing* 10(4):1257-1273.

Liu, B., S. Koziel, and Q. Zhang. 2016. "A Multi-fidelity Surrogate-model-assisted Evolutionary Algorithm for Computationally Expensive Optimization Problems". *Journal of Computational Science* 12:28-37.

Locatelli, M. 1997. "Bayesian Algorithms for One-dimensional Global Optimization". *Journal of Global Optimization* 10(1):57-76.

Min, C. 2017. *Multi-fidelity Optimization with Gaussian Regression on Ordinal Transformation Space*. M.Sc. thesis, Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore.

Osorio, C., and K. K. Selvam. 2017. "Simulation-based Optimization: Achieving Computational Efficiency through the use of Multiple Simulators". *Transportation Science* 51(2):395-411.

Quan, N., J. Yin, S. H. Ng, and L. H. Lee. 2013. "Simulation Optimization via Kriging: a Sequential Search using Expected Improvement with Computing Budget Constraints". *IIE Transactions* 45(7):763-780.

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. Cambridge, Massachusetts: The MIT Press.

Santner, T. J., B. J. Williams, and W. I. Notz. 2013. *The Design and Analysis of Computer Experiments*. NewYork: Springer Science & Business Media.

Solis, F. J., and R. J. B. Wets. 1981. "Minimization by Random Search Techniques". *Mathematics of Operations Research* 6(1):19-30.

Son, S. H., and D. H. Choi. 2016. "The Effects of Scale Factor and Correction on the Multi-fidelity Model". *Journal of Mechanical Science and Technology* 30(5):2075-2081.

Ulaganathan, S., I. Couckuyt, F. Ferranti, E. Laermans, and T. Dhaene. 2015. "Performance Study of Multi-fidelity Gradient Enhanced Kriging". *Structural and Multidisciplinary Optimization* 51(5):1017-1033.

Xu, J., S. Zhang, E. Huang, C. H. Chen, L. H. Lee, and N. Celik. 2014. "An Ordinal Transformation Framework for Multi-fidelity Simulation Optimization". In *Proceeding of the 10th IEEE International Conference on Automation Science and Engineering (CASE),* August $18^{th} - 22^{th}$, Taipei City, Taiwan, 385-390.

Xu, J., E. Huang, C. H. Chen, and L. H. Lee. 2015. "Simulation Optimization: a Review and Exploration in the New era of Cloud Computing and Big Data". *Asia-Pacific Journal of Operational Research* 32(03):1550019.

Xu, J., S. Zhang, E. Huang, C. H. Chen, L. H. Lee, and N. Celik. 2016a. "MO2TOS: Multi-fidelity Optimization with Ordinal Transformation and Optimal Sampling". *Asia-Pacific Journal of Operational Research* 33(03):1650017.

Xu, J., E. Huang, L. Hsieh, L. H. Lee, Q. S. Jia, and C. H. Chen. 2016b. "Simulation Optimization in the era of Industrial 4.0 and the Industrial Internet". *Journal of Simulation* 10(4):310-320.

Yan, D., and H. Mukai. 1992. "Stochastic Discrete Optimization". *SIAM Journal on Control and Optimization* 30(3):594-612.

Yin, J., S. H. Ng, and K. M. Ng. 2011. "Kriging Metamodel with Modified Nugget-effect: the Heteroscedastic Variance Case". *Computers & Industrial Engineering* 61(3):760-777.

Zabinsky, Z. B., and R. L. Smith. 1992. "Pure Adaptive Search in Global Optimization". *Mathematical Programming* 53(1-3):323-338.

Zabinsky, Z. B., R. L. Smith, J. F. McDonald, H. E. Romeijn, and D. E. Kaufman. 1993. "Improving Hit-and-run for Global Optimization". *Journal of Global Optimization* 3(2):171-192.

Zabinsky, Z. B., D. Bulger, and C. Khompatraporn. 2010. "Stopping and Restarting Strategy for Stochastic Sequential Search in Global Optimization". *Journal of Global Optimization* 46(2):273-286.

Zhang, S., J. Xu, E. Huang, and C. H. Chen. 2016a. "A New Optimal Sampling Rule for Multi-fidelity Optimization via Ordinal Transformation". In *Proceedings of the 14th IEEE International Conference on Automation Science and Engineering (CASE),* August $21^{th} - 24^{th}$, Fort Worth, Texas, 670-674.

Zhang, S., J. Xu, E. Huang, C. H. Chen, and S. Gao. 2016b. "Improving Ordinal Transformation through Optimal Combination of Multi-model Predictions". In *Proceedings of IEEE International Conference on Industrial Technology (ICIT)*, March $14^{th} - 17^{th}$, Taipei, Taiwan, 1545-1549.

AUTHOR BIOGRAPHIES

ALIREZA INANLOUGANJI pursues a PhD degree in Industrial Engineering at Arizona State University, USA. He holds MSc. in Industrial Engineering from Sharfi University of Technology. Before joining Arizona State Univeristy, he worked as business analytics professional at different companies including Solico for several years. His current research focus is on Simulation-Optimization, Energy Analytics and Disaster Response. His e-mail address is:ainanlou@asu.edu.

GIULIA PEDRIELLI is an Assistant Professor at the School of Computing, Informatics and Decision Systems Engineering at Arizona State University, USA. Her research focuses on stochastic simulation-optimization in both single and multiple–objectives framework. She is developing her research in meta-model based simulation optimization and learning for simulation and simulation optimization, as well as the development of novel formal languages to embed control in simulation. Her email address is: giulia.pedrielli@asu.edu.

GEORGIOS FAINEKOS is an Associate Professor in the School of Computing, Informatics and Decision Systems Engineering at Arizona State University, USA. Georgios is the director of the Cyber-Physical Systems Lab at ASU. He holds a Ph.D. in Computer and Information Science from the University of Pennsylvania, USA. His research interests include formal specification, modeling, testing, verification, synthesis, control, and monitoring for hybrid dynamical systems and discrete event systems. His email address is: fainekos@asu.edu.

SEBASTIAN POKUTTA is the David M. McKenney Family Associate Professor in the School of Industrial and Systems Engineering, Director of the Interactive Optimization and Learning Lab and an Associate Director of the Center for Machine Learning at Georgia Institute of Technology, USA. His primary research interests are in optimization and machine learning in the context of polyhedral combinatorics. Application areas include -but are not limited to- supply chain management, finance, cyber-physical systems, and predictive analytics. His email is:sebastian.pokutta@isye.gatech.edu.