A multi-fidelity Bayesian optimization approach based on the expected further improvement

Leshi Shu[1,2], Ping Jiang[1], Yan Wang[2,*]

[1]The State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science & Technology, Wuhan 430074, PR China
[2]Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

*Corresponding author:
Prof. Yan Wang
Email: yan-wang@gatech.edu
Mail: 813 Ferst Drive NW, Atlanta, GA 30332-0405, USA.
Telephone: +1(404)894-4714
Fax: +1(404)894-9342

Abstract: Sampling efficiency is important for simulation-based design optimization. While Bayesian optimization (BO) has been successfully applied in engineering problems, the cost associated with large-scale simulations has not been fully addressed. Extending the standard BO approaches to multi-fidelity optimization can utilize the information of low-fidelity models to further reduce the optimization cost. In this work, a multi-fidelity Bayesian optimization approach is proposed, in which hierarchical Kriging is used for constructing the multi-fidelity metamodel. The proposed approach quantifies the effect of HF and LF samples in multi-fidelity optimization based on a new concept of expected further improvement. A novel acquisition function is proposed to determine both the location and fidelity level of the next sample simultaneously, with the consideration of balance between the value of information provided by the new sample and the associated sampling cost. The proposed approach is compared with some state-of-the-art methods for multi-fidelity global optimization with numerical examples and an engineering case. The results show that the proposed approach can obtain global optimal solutions with reduced computational costs.

1. Introduction

Bayesian optimization (BO) is a metamodel based global optimization approach, where the search process is assisted by constructing and updating a metamodel iteratively, and the sequential sampling is guided by an acquisition function to incorporate uncertainty (Ghoreishi and Allaire 2019; Tran et al. 2019b). The construction of metamodels helps improve the search efficiency, while the sequential sampling guided by the acquisition function reduces the overall number of samples. The sequential sampling strategy is particularly helpful when high-cost simulations or physical experiments are involved. Different definitions of acquisition functions have been developed to balance between exploration and exploitation, such as expected improvement (EI), probability of improvement, and lower confidence bound. BO with the EI acquisition function is also called efficient global optimization (EGO) by some researchers.

Similar to other metamodel based global optimization methods (Queipo et al. 2005; Wang and Shan 2007), the computational challenge for BO to solve large-scale problems still exist. Because the number of samples to cover the search space grows exponentially as the dimension of the space increases. Multi-fidelity (MF) surrogate modeling is one approach to reduce the cost by combining sample points predicted by high-fidelity (HF) and low-fidelity (LF) models to construct the surrogates, since running LF models is less costly (Huang et al. 2006; Jones 2001; Shu et al. 2019a; Xiong et al. 2008; Zhou et al. 2017). The existing MF metamodels can be categorized as three types. The first type is the scaling function based MF metamodeling, which tunes the LF model according to the HF model responses (Chang et al. 1993; Zhou et al. 2015). The second type is space-mapping MF metamodels, in which a transformation operator is applied to map the LF design space to the HF space and the optimal sample point in the HF space can be estimated (Bakr et al. 2001; Bandler et al. 1994; Koziel et al. 2006). The third type is MF Kriging models, such as co-Kriging model (Kennedy and O'Hagan 2000) and hierarchical Kriging model (Han and Görtz 2012). Co-kriging models are constructed with the information of covariance between the LF and HF samples. However, they are constructed based on the nested HF sample points, which adds limitations in their applications. In hierarchical Kriging models, the LF Kriging model is directly used as the trend of the MF metamodel, without the requirement of nested sample points (Han and Görtz 2012; Zhang et al. 2018). Hierarchical Kriging allows designers to choose sample points more freely in the optimization process. Because of its flexibility in sampling, hierarchical Kriging received much more attentions in engineering design optimization (Courrier et al. 2016; Palar and Shimoyama 2017; Zhang et al. 2015). MF metamodels for multi-objective optimization (Shu et al. 2019b; Zhou et al. 2016), incorporating gradient information (Song et al. 2017; Ulaganathan et al. 2015), and adaptive hybrid scaling method (Gano et al. 2005) have also been developed.

The acquisition guided sequential sampling has been applied in MF metamodel based design optimization. For instance, Xiong et al. (Xiong et al. 2008) applied the lower confidence bound in sequential sampling to construct MF metamodels. Kim et al. (Kim et al. 2017) used the EI acquisition function for hierarchical Kriging model. However, these methods merely adopt the high-fidelity simulation data to update the MF metamodels. The acquisition functions were only applied to determine the locations of new samples, while

the different costs associated with LF and HF samplings are not considered. To solve this problem, Huang et al. (Huang et al. 2006) developed an augmented EI acquisition function for co-kriging, in which EI is augmented by the correlation of predictions between different fidelity models and a ratio of sampling costs so that both the sample location and the fidelity level can be determined by maximizing the acquisition. Liu et al. (Liu et al. 2018) improved the augmented EI criterion with the consideration of the sample cluster issue to reduce the computational cost of the co-kriging. Ghoreishi et al. (Ghoreishi et al. 2018) proposed to identify the next best fidelity information source and the best location in the input space via a value-gradient policy. Then they considered more information sources with different fidelity levels and explicitly account for the computational cost associated with individual sources (Ghoreishi et al. 2019). Zhang et al. (Zhang et al. 2018) proposed a multi-fidelity global optimization approach based on hierarchical Kriging model, in which an MFEI acquisition function is extended from EI with different uncertainty levels corresponding to the samples of low and high fidelities. Tran et al. (Tran et al. 2020) proposed to combine the overall posterior variance reduction and computational cost ratio to select the fidelity level.

In this paper, a new MF Bayesian optimization (MFBO) approach with the hierarchical Kriging model is developed. A MF acquisition function based on a new concept of expected further improvement is proposed, which enables the simultaneous selections of both location and fidelity level for the next sample. The different costs of HF and LF samples as well as the extra information of HF samples are considered altogether. A constrained MF acquisition function for unknown constraints is also introduced. The proposed MFBO approach is compared with the standard EGO method and the MFEI method (Zhang et al. 2018) using five numerical examples and one engineering case.

The remainder of this paper is organized as follows. In Section 2, the hierarchical Kriging model and standard EGO method are reviewed. In Section 3, the proposed MFBO approach and the new acquisition function are described in details. Five numerical examples and one engineering case study with the comparisons of results are presented in Section 4, followed by concluding remarks in Section 5.

2. Background

2.1 Hierarchical Kriging

Hierarchical Kriging is a MF metamodeling method, in which the LF Kriging model is taken to predict the overall trend whereas the HF samples are used to correct the LF model. The metamodel can be expressed as

$$Y(\boldsymbol{x}) = \beta_0 \hat{y}_l(\boldsymbol{x}) + Z(\boldsymbol{x}) \tag{1}$$

where $\hat{y}_l(\boldsymbol{x})$ is the predicted mean of the LF Kriging model, which is constructed based on LF sample points, β_0 is a scaling factor, and $Z(\boldsymbol{x})$ is a stationary random process with zero mean and a covariance of

$$Cov[Z(\boldsymbol{x}), Z(\boldsymbol{x}')] = \sigma^2 R(\boldsymbol{x}, \boldsymbol{x}') \tag{2}$$

where σ^2 is the process variance. $R(\boldsymbol{x}, \boldsymbol{x}')$ is the spatial correlation function which only depends on the distance between two design sites, $\boldsymbol{x}$ and $\boldsymbol{x}'$. Given HF sample points $X_h = \{\boldsymbol{x}_{h,1}, \boldsymbol{x}_{h,2}, ..., \boldsymbol{x}_{h,n}\}$ and their responses $\boldsymbol{f}_h(X_h) = \{f(\boldsymbol{x}_{h,1}), f(\boldsymbol{x}_{h,2}), ..., f(\boldsymbol{x}_{h,n})\}$, the predicted mean and variance of the hierarchical Kriging model at an unobserved point can be calculated as

$$\hat{y}(\boldsymbol{x}) = \beta_0 \hat{y}_l(\boldsymbol{x}) + \boldsymbol{r}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}[\boldsymbol{f}_h(X_h) - \beta_0 \mathbf{F}] \tag{3}$$

and

$$\mathrm{MSE}\{\hat{y}(\boldsymbol{x})\} = \beta_0 \hat{y}_l(\boldsymbol{x}) + \boldsymbol{r}^{\mathrm{T}}(\boldsymbol{x})\boldsymbol{R}^{-1}[\boldsymbol{f}_h(X_h) - \beta_0 \mathbf{F}] \tag{4}$$

respectively, where $\boldsymbol{r}(\boldsymbol{x})$ is the correlation vector with elements $r_i(\boldsymbol{x}) = R(\boldsymbol{x}, \boldsymbol{x}_i), \boldsymbol{x}_i \in X_h$. $\boldsymbol{R}$ is the correlation matrix with elements $\boldsymbol{R}(i, j) = R(\boldsymbol{x}_i, \boldsymbol{x}_j), \boldsymbol{x}_i, \boldsymbol{x}_j \in X_h$. $\mathbf{F}$ is the vector of predictions by the LF Kriging model at the locations of HF samples. The Gaussian correlation function

$$R(\boldsymbol{x}, \boldsymbol{x}') = \prod_{k=1}^{m} R_k(\theta, x_k - x_k') = \prod_{k=1}^{m} \exp(-\theta_k \left| x - x' \right|^2) \tag{5}$$

is used in this paper. The hyper-parameters of hierarchical Kriging can be trained by maximizing the likelihood function

$$L(\beta_0, \sigma^2, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n |\mathbf{R}|}} \exp(-\frac{1}{2} \frac{(\boldsymbol{f}_h(X_h) - \mathbf{F}\beta_0)^T \mathbf{R}^{-1}(\boldsymbol{f}_h(X_h) - \mathbf{F}\beta_0)}{\sigma^2}) \tag{6}$$

More details of hierarchical Kriging can be found in (Han and Görtz 2012).

2.2 Efficient global optimization approach

The EGO is a Bayesian optimization method where the EI acquisition function is used. The standard EGO method was originally proposed by Jones et al. (Jones et al. 1998) for expensive black-box problems. For Kriging and hierarchical Kriging models, the prediction at an unsampled point $\boldsymbol{x}$ can be regarded as a random variable and obeys a normal distribution $Y(\boldsymbol{x}) \sim N(\hat{y}(\boldsymbol{x}), \sigma^2(\boldsymbol{x}))$, where $\hat{y}(\boldsymbol{x})$ and $\sigma^2(\boldsymbol{x})$ are the predicted mean and variance. The improvement at $\boldsymbol{x}$ for a minimization problem is

$$I(\boldsymbol{x}) = \max(f_{\min} - Y(\boldsymbol{x}), 0) \tag{7}$$

where $f_{\min}$ is the best solution in the current sample set. The expected improvement is

$$EI(\boldsymbol{x}) = E[\max(f_{\min} - Y(\boldsymbol{x}), 0)] \tag{8}$$

By expressing the right-hand side of Eq. (8) as an integral, one can obtain the EI in the closed form as

$$EI(\boldsymbol{x}) = (f_{\min} - \hat{y}(\boldsymbol{x}))\Phi(\frac{f_{\min} - \hat{y}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}) + \sigma(\boldsymbol{x})\phi(\frac{f_{\min} - \hat{y}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}) \tag{9}$$

where $\phi(\bullet)$ and $\Phi(\bullet)$ are the probability density function and cumulative distribution function of the standard normal distribution, respectively. The EGO method helps obtain the next sample point by maximizing the EI function expressed in Eq. (9). Then the new sample point is used to update the metamodel. The iteration continues until the algorithm converges. More details of EGO can be found in (Jones et al. 1998).

3. The Proposed MFBO Approach

The standard EGO method provides a way to select a new sample point in single-fidelity optimization. However, in MF optimization, the decision of choosing the next sample at either HF or LF level needs to be made to update the MF metamodel. In the proposed MFBO, a new acquisition function is developed to support the sequential sampling strategy for selecting sample points of different fidelity levels adaptively in MF optimization.

3.1 Acquisition function based on the expected further improvement

In general, HF sample points are more expensive to obtain but can provide more precise information, whereas LF sample points are less costly but less reliable. In MF

optimization, the sample points in both HF and LF levels are chosen to update the MF metamodel. Here, a new acquisition function is defined so that the choice of fidelity level incorporates the considerations of both cost and benefit in LF and HF samples.

With the sequential sampling, the maximum EI gradually decreases as the BO algorithm converges to the optimal solution. If a sample point $\boldsymbol{x}^*$ is selected to update the metamodel, the EI will decrease from $EI(\boldsymbol{x}^*)$ calculated from Eq. (9) to zero. Therefore, the reduction of the EI value can be equivalently used to guide the sequential sampling. In the proposed MFBO, the reduction of EI value incorporates the different effects of LF and HF samples. If a HF sample is chosen at $\boldsymbol{x}^*$, the HF effect on reducing the EI is

$$\Delta EI_{\mathrm{h}}(\boldsymbol{x}^*) = EI(\boldsymbol{x}^*) - 0 = EI(\boldsymbol{x}^*) \tag{10}$$

where $EI(\cdot)$ is the EI function of hierarchical Kriging based on the existing HF samples and LF samples. If a LF sample at this location $\boldsymbol{x}_l^*$ is chosen instead, the LF effect on reducing the EI, which is named further improvement given that $\boldsymbol{x}_l^*$ is taken, is calculated as

$$\Delta EI_l(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*)) = EI(\boldsymbol{x}^*) - EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*)) \tag{11}$$

where $Y_l(\bullet)$ is the LF metamodel with predicted mean $\hat{y}_l(\bullet)$ and variance $\sigma_l^2(\boldsymbol{x}^*)$, $EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))$ represents the EI if calculated by HF sample $\boldsymbol{x}^*$ given that a LF sample $\boldsymbol{x}_l^*$ was taken instead at the same location of $\boldsymbol{x}^*$ with the predicted LF response $Y_l(\boldsymbol{x}_l^*)$. Note that $Y_l(\boldsymbol{x}_l^*)$ is a random variable which follows a Gaussian distribution. Therefore, both the conditional expected value $EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))$ and the conditional further improvement in Eq.(11) vary according to $Y_l(\boldsymbol{x}_l^*)$. The overall expected value of $EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))$ is obtained as

$$\begin{aligned} E[EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))] &= \int P[Y_l(\boldsymbol{x}_l^*)] EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*)) dY_l(\boldsymbol{x}_l^*) \\ &= \int \phi(\frac{Y_l(\boldsymbol{x}_l^*) - \hat{y}_l(\boldsymbol{x}^*)}{\sigma_l(\boldsymbol{x}^*)}) EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*)) dY_l(\boldsymbol{x}_l^*) \end{aligned} \tag{12}$$

Note that $EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))$ in Eq.(12) is calculated based on the HF metamodel with HF sample $\boldsymbol{x}^*$. Thus the expected further improvement is

$$E[\Delta EI_l(\boldsymbol{x}^*)] = E[EI(\boldsymbol{x}^*) - EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))]$$

$$= EI(\boldsymbol{x}^*) - E[EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))] \tag{13}$$

Considering the different costs of HF and LF samples and assuming that the cost ratio of a HF sample to a LF sample is T, we define a new acquisition function as

$$a(\boldsymbol{x}, fidelity) = \begin{cases} \dfrac{1}{T}\Delta EI_h(\boldsymbol{x}), \text{ if } fidelity = 2 \\ E[\Delta EI_l(\boldsymbol{x})], \text{ if } fidelity = 1 \end{cases} \tag{14}$$

$$= EI(\boldsymbol{x}) + (fidelity - 1)\frac{1-T}{T}EI(\boldsymbol{x}) + (fidelity - 2)E[EI(\boldsymbol{x}^* | Y_l(\boldsymbol{x}_l^*))]$$

to decide both the sample location and the fidelity level, where *fidelity* equals 1 for LF level and *fidelity* equals 2 for HF level. The location and fidelity level can be obtained by maximizing the acquisition function.

The proposed approach can be further extended to problems with multiple fidelity levels. The expected value of $EI(\boldsymbol{x}^* | Y_j(\boldsymbol{x}_j^*))$ at the *j*-th fidelity level can be calculated similarly as in Eq.(12), whereas $EI(\boldsymbol{x}^* | Y_j(\boldsymbol{x}_j^*))$ itself is calculated based on metamodel $Y_{j+1}(\boldsymbol{x}) = \beta_j \hat{y}_j(\boldsymbol{x}) + Z_j(\boldsymbol{x})$ that is similar to Eq.(1). The acquisition function in Eq.(14) can be adjusted to include all fidelity levels with different cost ratios accordingly.

Because of the multiple integral in Eq. (12), direct calculation of the acquisition function in Eq. (14) can be computationally expensive. An alternative approach can be taken here to search the maximum of the acquisition function. From Eq. (11), it is seen that $\Delta EI_l(\boldsymbol{x}^*)$ tends to be large when $EI(\boldsymbol{x}^*)$ is large. $\Delta EI_l(\boldsymbol{x}^*)$ has a similar trend as $EI(\boldsymbol{x}^*)$. Thus the location of the new LF sample point tends to be selected near the location where a large EI is obtained. Therefore, the EI function for HF sampling can be used in the searching of maximum, which approximates the true location of the maximum expected further improvement. From the sample location, the acquisition function in Eq. (14) can be evaluated based on the surrogates at both fidelity levels, and the fidelity level which leads to a larger acquisition value is selected. The worst-case scenario of this heuristic searching approach is that its searching efficiency is the same as the standard EGO.

3.2 Constrained acquisition function

In general, the constraints in engineering optimization can be divided into two categories: known constraints and unknown constraints. Known constraints can be evaluated easily and analytically without running a simulation. In contrast, unknown constraints are much more complex and usually related to design performance. Whether they are satisfied or not can only be determined after running a simulation. In this work, the penalty function approach (Coello 2000; Shu et al. 2017) is used to handle known constraints where the objective function is penalized. Researchers have proposed different approaches to hand unknown constraints such as constrained EI (Schonlau et al. 1998) and surrogates of constraints (Gardner et al. 2014; Gelbart et al. 2014; Tran et al. 2019a). Here, unknown constraints are incorporated in the new acquisition function.

For an unknown constraint $g(\boldsymbol{x}) \leq 0$, we define an indication function $F(\boldsymbol{x})$ as

$$F(\boldsymbol{x}) = \begin{cases} = 1, \ if \ g(\boldsymbol{x}) \leq 0 \\ = 0, \ if \ g(\boldsymbol{x}) > 0 \end{cases} \tag{15}$$

Since $g(\boldsymbol{x})$ cannot be evaluated without running a simulation, we can also construct a hierarchical Kriging model and assume that the prediction of $g(\boldsymbol{x})$ obeys a normal distribution $g(\boldsymbol{x}) \sim N(\hat{g}(\boldsymbol{x}), \sigma_g^2(\boldsymbol{x}))$. Then the constrained acquisition function is defined as

$$\begin{aligned} a_C(\boldsymbol{x}, fidelity) &= E[F(\boldsymbol{x})a(\boldsymbol{x}, fidelity)] \\ &= E[F(\boldsymbol{x})]E[a(\boldsymbol{x}, fidelity)] + \mathrm{cov}[F(\boldsymbol{x}), a(\boldsymbol{x}, fidelity)] \\ &= E[F(\boldsymbol{x})]a(\boldsymbol{x}, fidelity) + \mathrm{cov}[F(\boldsymbol{x}), a(\boldsymbol{x}, fidelity)] \end{aligned} \tag{16}$$

In general, the correlation between $F(\boldsymbol{x})$ and $a(\boldsymbol{x}, fidelity)$ can be ignored. Thus $\mathrm{cov}[F(\boldsymbol{x}), a(\boldsymbol{x}, fidelity)] = 0$. According to the definition of $F(\boldsymbol{x})$, $E[F(\boldsymbol{x})]$ can be calculated as

$$E[F(\boldsymbol{x})] = P[g(x) \leq 0] = \Phi(\frac{-\hat{g}(\boldsymbol{x})}{\sigma_g(\boldsymbol{x})}) \tag{17}$$

Then Eq. (16) can be expressed as

$$a_C(\boldsymbol{x}, fidelity) = \Phi(\frac{-\hat{g}(\boldsymbol{x})}{\sigma_g(\boldsymbol{x})})a(\boldsymbol{x}, fidelity) \tag{18}$$

4. Examples and Results

In this section, five numerical examples and one engineering case study are used to demonstrate the applicability and performance of the proposed approach. The formulations of the five numerical examples and the respective optimal solutions (Cai et al. 2016; Zhang et al. 2018; Zhou et al. 2016) are listed in Table 1. $f_h(\boldsymbol{x})$ and $f_l(\boldsymbol{x})$ represent the HF and LF models, respectively. $g_h(\boldsymbol{x})$ and $g_l(\boldsymbol{x})$ represent the HF and LF constraints, respectively. $\boldsymbol{x}_{best}$ is the optimal solution and $f_h(\boldsymbol{x}_{best})$ is the corresponding response.

The proposed approach is compared with the standard EGO (Jones et al. 1998) and the MFEI method (Zhang et al. 2018). Note that the cost difference between HF and LF samples is not considered in the MFEI method, in contrast to our approach. The computational cost is calculated as

$$\text{cost} = n_h + \frac{n_l}{T} \tag{19}$$

where n_h and n_l are the numbers of HF and LF samples, respectively. T is the cost ratio.

4.1 The one-dimensional example

The first numerical example in Table 1 is used for the illustration of the proposed approach as well as a detailed comparison between different approaches. Here we assume that the cost of a HF sample point is 4 times of a LF sample point (T=4).

The initial hierarchical Kriging model is constructed based on six LF sample points $S_l = \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ and three HF sample points $S_h = \{0.0, 0.5, 1.0\}$. The initial samples are uniformly distributed in the design space. The initial sample points, the constructed HF and LF models, the initial hierarchical Kriging model, and the EI function are shown in Figure 1(a). The maximum value of EI function is at $x_1 = 0.9093$. At this location, $\frac{1}{T}\Delta EI_h(x_1) = 1.8598$ and $E[\Delta EI_l(x_1)] = 6.7459$. Hence, a LF sample point is added at this location.

The updated hierarchical Kriging model and EI function are shown in Figure 1(b). Similarly, a LF sample point is added at $x_2 = 0.8232$ in the second iteration, where $\frac{1}{T}\Delta EI_h(x_2) = 1.7698$ and $E[\Delta EI_l(x_2)] = 5.7174$. The updated hierarchical Kriging model and
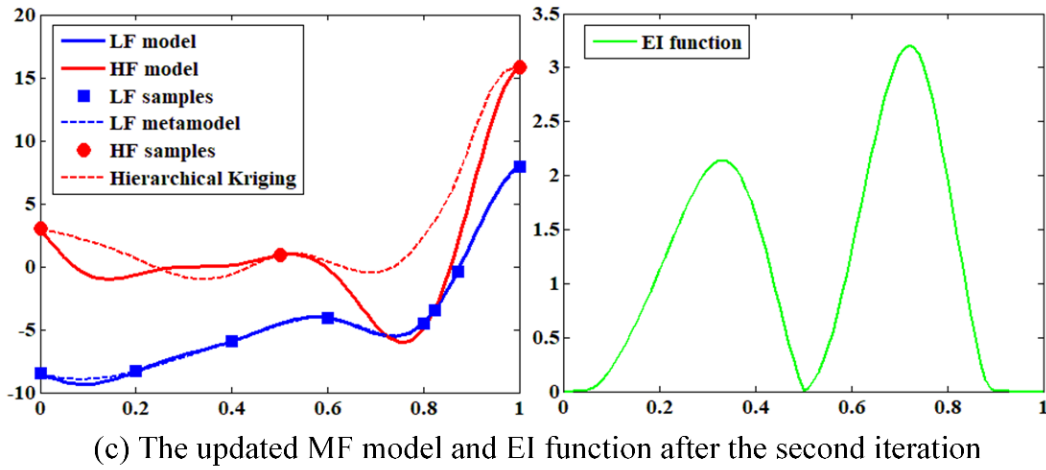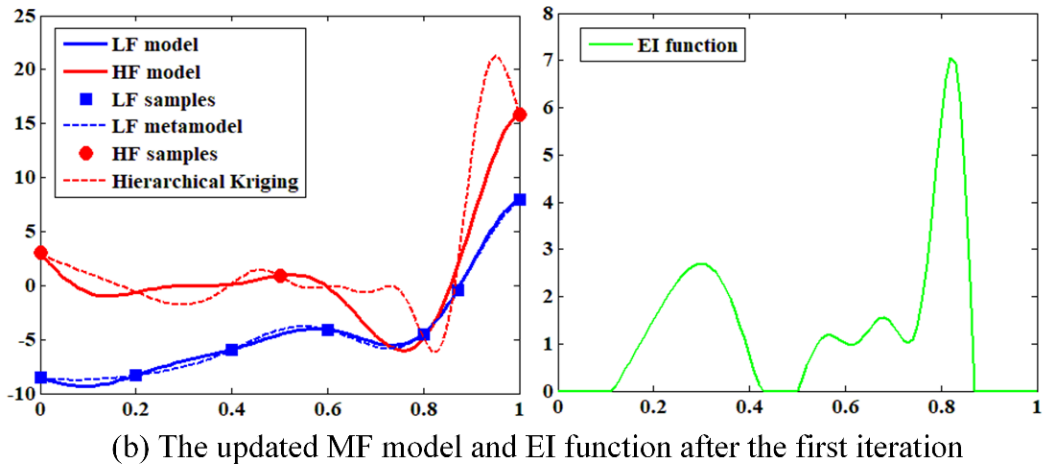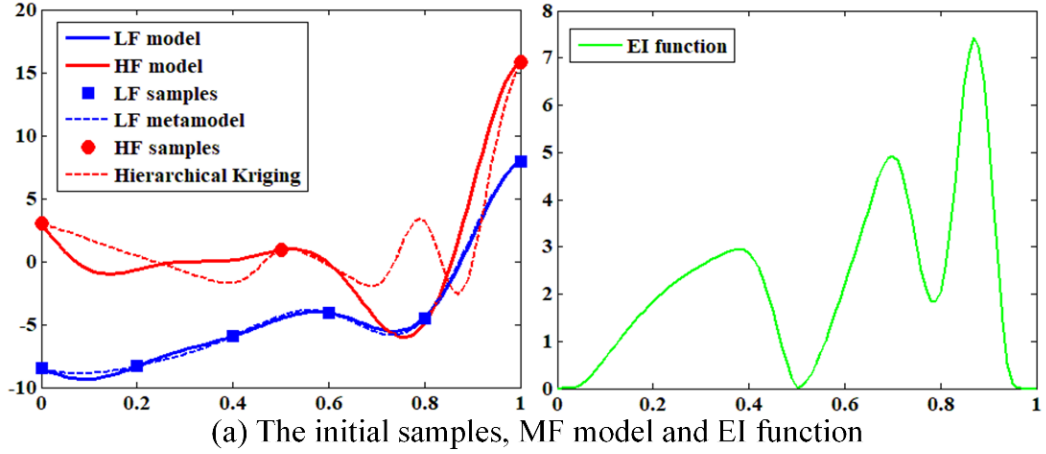
EI function are shown in Figure 1(c). The maximum value of EI function in this iteration is at $x_3 = 0.7211$. At this location, $\frac{1}{T}\Delta EI_{\mathrm{h}}(x_3) = 0.8018$ and $\mathrm{E}[\Delta EI_l(x_3)] = -1.8252$. Hence, a HF sample point is added in the third iteration.

The searching process of the proposed approach in the first numerical example is listed in Table 2. The proposed approach requires three LF samples and three HF samples to find the optimal solution. The termination criterion for this numerical example is set as

$$\left| f_{\min} - f_h(x_{best}) \right| \le \varepsilon \tag{20}$$

where $f_{\min}$ is the best observed objective function, and ε is set to be 0.01.

(a) The initial samples, MF model and EI function

(b) The updated MF model and EI function after the first iteration

(c) The updated MF model and EI function after the second iteration

Figure 1. The hierarchical Kriging model and EI function of the one-dimensional function

Table 1. Formulations and solutions for five numerical examples.

Case	Formulation	Optimal solution
Case 1	$f_h(x) = (6x-2)^2 \sin(12x-4)$ $f_l(x) = 0.5 f_h(x) + 10(x-0.5) - 5$ $x \in [0,1]$	$x_{best} = 0.7573$ $f_h(x_{best}) = -6.0207$
Case 2	$f_h(\boldsymbol{x}) = 4x_1^2 + x_2^3 + x_1 x_2$ $g_h(\boldsymbol{x}) = 1/x_1 + 1/x_2 - 2$ $f_l(\boldsymbol{x}) = 4(x_1+0.1)^2 + (x_2-0.1)^3 + x_1 x_2 + 0.1$ $g_l(\boldsymbol{x}) = 1/x_1 + 1/(x_2+0.1) - 2 - 0.001$ $x_1, x_2 \in [0.1, 10]$	$\boldsymbol{x}_{best} = [0.8846, 1.1500]$ $f_h(\boldsymbol{x}_{best}) = 5.6684$
Case 3	$f_h(\boldsymbol{x}) = 4x_1^2 - 2.1x_1^4 + x_1 x_2 - 4x_2^2 + 4x_2^4$ $f_l(\boldsymbol{x}) = 4(x_1+0.1)^2 + (x_2-0.1)^3 + x_1 x_2 + 0.1$ $x_1, x_2 \in [-2, 2]$	$\boldsymbol{x}_{best} = [-0.0898, 0.7127]$ $f_h(\boldsymbol{x}_{best}) = -1.0316$
Case 4	$f_h(\boldsymbol{x}) = -\sum_{i=1}^{4} C_i \exp[-\sum_{j=1}^{3} A_{ij}(x_j - P_{ij})]$ $f_l(\boldsymbol{x}) = f_h(\boldsymbol{x}) + 7.6 \text{MA3}(\boldsymbol{x})$ $where \ A = \begin{pmatrix} 3 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3 & 10 & 30 \\ 0.1 & 10 & 35 \end{pmatrix}, C = \begin{pmatrix} 1 \\ 1.2 \\ 3 \\ 3.2 \end{pmatrix},$ $P = \begin{pmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.03815 & 0.5743 & 0.8828 \end{pmatrix}$ $\text{MA3}(\boldsymbol{x}) = 0.585 - 0.324x_1 - 0.379x_2 - 0.431x_3 - 0.208x_1 x_2$ $+ 0.326x_1 x_3 + 0.193x_2 x_3 + 0.225x_1^2 + 0.263x_2^2 + 0.274x_3^2$ $0 \le x_i \le 1, \ for \ i = 1, 2, 3;$	$\boldsymbol{x}_{best} = [0.114, 0.556, 0.852]$ $f_h(\boldsymbol{x}_{best}) = 3.8627$
Case 5	$f_h(\boldsymbol{x}) = \sum_{i=1}^{10} \exp(x_i)(A(i) + x_i - \ln(\sum_{k=1}^{10} \exp(x_k)))$ $f_l(\boldsymbol{x}) = \sum_{i=1}^{10} \exp(x_i)(B(i) + x_i - \ln(\sum_{k=1}^{10} \exp(x_k)))$ $A = [-6.089, -17.164, -34.054, -5.914, -24.721,$ $-14.986, -24.100, -10.708, -26.662, -22.179]$ $B = [-5, -10, -30, -5, -25, -15, -20, -10, -25, -20]$ $x_i \in [-5, 5], i = 1, 2, \cdots, 10$	$\boldsymbol{x}_{best} = [1,1,1,1,1,1,1,1,1,1]$ $f_h(\boldsymbol{x}_{best}) = 0$

Table 2. The searching process of the proposed approach

No.	Location	Fidelity level	The best observed objective function
1	0.8712	1	0.9093
2	0.8232	1	0.9093
3	0.7211	2	-5.4033
4	0.7283	2	-5.6147
5	0.2953	1	-5.6147
6	0.7565	2	-6.0204
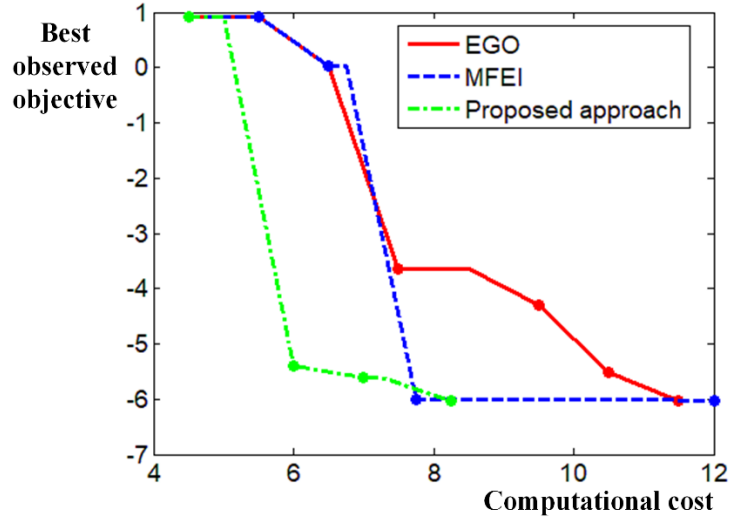
Figure 2. The convergence curves of the three approaches for the one-dimensional example

The convergences of the three approaches for the first numerical example are compared in Figure 2. The numbers of total HF and LF sample points, including the initial samples, and the computational costs of the three approaches are listed in Table 3. Note that the initial samples also need to be included in estimating the overall costs. For the standard EGO which itself is for single-fidelity optimization, the same numbers of initial LF and HF samples are recorded for comparison. After the initial model is constructed, the sample points added in the following iterations are counted as HF samples. It is seen from Table 3 that the proposed approach requires the least computational cost to find the optimal solution for the first example. The convergence criterion in Eq.(20) is applied in all three approaches.

Table 3. The number of sample points and computational cost in different approaches for Case 1

	Number of HF sample points	Number of LF sample points	Computational cost
MFEI	9	9	11.25
Standard EGO	10	6	11.5
Proposed approach	6	9	8.25

4.2 Numerical Examples for Cases 2 to 4

For numerical examples of Cases 2 to 4, Latin hypercube sampling (LHS) (Park 1994; Wang 2003) is used to generate the initial HF and LF sample sets. The sizes of the initial HF and LF sample sets are set to be 3 times and 6 times of the dimensions of the problems respectively. To account for the influence of randomness, each of these cases is solved 30 times with each of the three approaches. The results for the average numbers of LF and HF sample points are compared in Table 4. The same convergence criterion in Eq. (20) is applied for all cases. To illustrate the effect of the cost ratio on the proposed approach, two different cost ratios (T=4 and T=10) are tested for the proposed approach. Based on the acquisition functions in Eqs. (14) and (18), the proposed approach tends to select more LF sample points if the cost of LF sampling is lower (i.e., a higher cost ratio).

For Case 2, the EGO and proposed approach require almost the same number of LF and HF sample points, while the MFEI method samples much more LF sample points than the other two approaches. However, this does not reduce the number of HF samples required in the MFEI method. For Case 3 and Case 4, the MFEI method and the proposed approach require fewer HF sample points than the EGO by supplementing with LF sample points. One key difference between the MFEI method and the proposed approach is that the cost ratio is not considered in the MFEI method. The proposed approach tends to sample more LF sample points and fewer HF sample points as the cost ratio increases.

The computational costs of different approaches for T=4 and T=10 according to Eq. (18) are listed in Table 5. For Case 2, the EGO and the proposed approach have the similar computational costs, which are lower than that of the MFEI method. For Case 3 and Case 4, the MFEI method is more efficient than EGO, and the proposed approach has the lowest cost among the three approaches.

Table 4. The average numbers of sample points in different approaches for Cases 2, 3 and 4

	Case 2		Case 3		Case 4	
	Number of LF samples	Number of HF samples	Number of LF samples	Number of HF samples	Number of LF samples	Number of HF samples
Standard EGO	12.00	44.86	12.00	25.40	18.00	19.10
MFEI	38.50	44.76	19.63	23.97	27.00	15.90
Proposed approach (T=4)	12.26	45.77	21.63	20.47	24.20	15.77
Proposed approach (T=10)	12.63	44.50	26.43	19.93	29.00	14.87

Table 5. Computational costs of different approaches for Cases 2, 3, and 4

	T=4			T=10		
	Case 2	Case 3	Case 4	Case 2	Case 3	Case 4
Standard EGO	48.67	29.77	23.60	46.87	27.97	20.90
MFEI	54.39	28.88	22.65	48.61	25.93	18.60
Proposed approach	48.84	25.88	21.82	45.76	22.58	17.77

4.3 Case 5: A high-dimensional example

The fifth numerical example is used to test the ability of the different approaches to solve high-dimensional optimization problems. LHS is applied to generate the 40 initial HF samples and 100 initial LF samples. In this example, we set the maximum number of iterations to 200 to observe the convergence process of different optimization approaches.

The convergences of the objective values along with the computational costs for the three different approaches are compared in Figure 3. The best observed objectives, the numbers of LF and HF sample points, and the corresponding computational costs (T=4) for the three approaches after convergence are listed in Table 6.

Table 6. Comparing results of the three approaches

	Best observed objective	Number of LF samples	Number of HF samples	Computational cost
Standard EGO	8.36	100	210	235
MFEI	5.57	204	49	100
Proposed approach	4.85	160	100	140

From Figure 3 and Table 6, it is seen that the Standard EGO and the proposed approach can obtain a better optimal solution than the MFEI approach. Compared to the EGO, the proposed approach has a lower computational cost to converge to the optimal solution. After 200 iterations, the cost of MFEI is the least, since the MFEI approach added the most LF samples and the fewest HF samples. The overreliance on LF samples led to the missing out on the opportunities to reach a better solution.
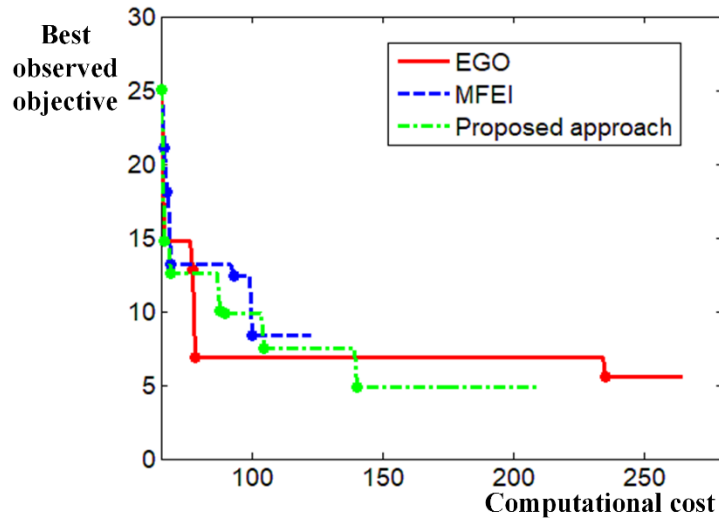

Figure 3. The convergence curves of the three approaches for the high-dimensional example

4.4 Engineering case study: Impedance optimization of the long base

As an engineering case study, the proposed approach is applied to optimize the long base of a ship. The simulation model of the problem consists of a cylindrical shell and a long base, which is shown in Figure 4. The optimization objective is to maximize the minimum impedance of the pedestal while keeping the weight below 3.4 tons. The mechanical impedance of a vibrating system is the complex ratio of a harmonic excitation

to its response. In this example, the impedance is the origin impedance, which is the complex ratio of a harmonic excitation to its response at the same location. To calculate the impedance, two unit harmonic forces are loaded in the Y-axis direction at point A and point B of the long base in Figure 4. The frequency of the unit harmonic forces ranges from 0 to 350 Hz. The displacements at the ends of the cylindrical shell and the part of the base connected to the bulkhead are fixed to zeros. The six design variables shown in Figure 4 are listed in Table 7. Other fixed parameters related to materials and geometry are shown in Table 8.
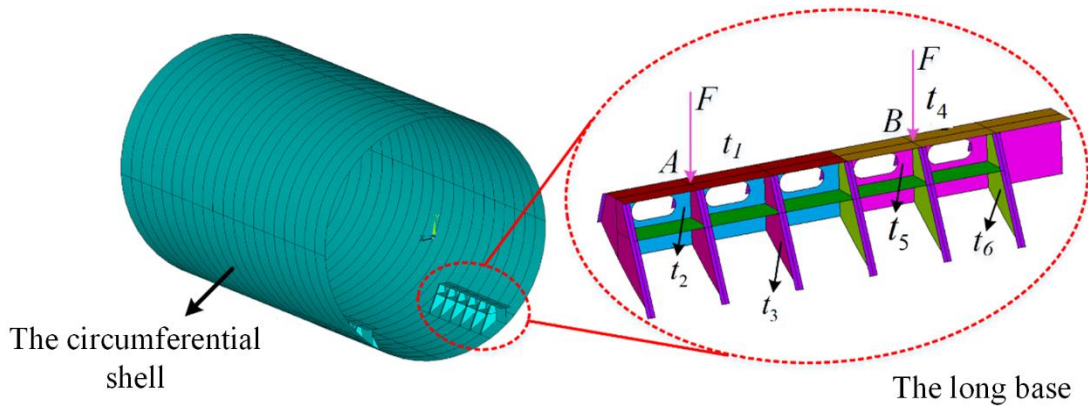
**Figure 4.** The geometric model of the cylindrical shell and the long base

For the HF model, the step size of frequency calculation is chosen to be 2.5 Hz. For the LF model, the step size of calculation is 10 Hz. The computational cost of the HF model is 4 times of the LF model (T=4). The convergence of the objective values along with the computational costs for the three approaches are plotted in Figure 5. The best observed objectives, the numbers of LF and HF sample points, and the corresponding computational costs for the three approaches after convergence are listed in Table 9. For further comparison, the simulation resolution is further reduced to the step size of 25Hz and applied as the LF model (T=10). The results are also listed in Table 9.

Table 7. The ranges of the design variables

Design variables		Range
	Thickness of the panel t_1	40-90 mm
The first half of the base	Thickness of the web t_2	12-40 mm
	Thickness of the rib t_3	10-60 mm
	Thickness of the panel t_4	40-90 mm
The second half of the base	Thickness of the web t_5	12-40 mm
	Thickness of the rib t_6	10-60 mm

Table 8. The fixed parameters related to materials and geometries

Parameters	Values
Young's modulus	2.09×10^5 MPa
Density	7850 kg/m^3
Poisson's ratio	0.3
Length of the shell	12000 mm
Radius of the shell	3300 mm
Spacing of ribs	600 mm
Radius of the hole in the web	75 mm
Width of the hole in the web	210 mm
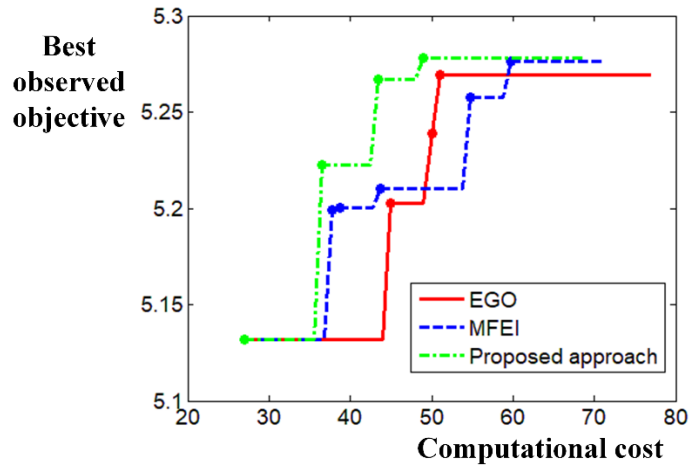
Figure 5. The convergence curves of the three approaches

Table 9. Comparing results of the three approaches

	Best observed objective	number of LF samples	number of HF samples	computational cost
Standard EGO	5.2696	36	40	51
MFEI	5.2782	43	49	59.75
Proposed approach (*T*=4)	5.2763	40	37	47.75
Proposed approach (*T*=10)	5.2774	46	29	40.5

From Figure 5 and Table 9, it is seen that the MFEI method and the proposed approach can obtain a better optimal solution than the standard EGO. Compared to the EGO and MFEI, the proposed approach has a lower computational cost to converge to the optimal solution. The proposed approach can rely more on LF sample points when the LF model is cheaper. This indicates that the proposed approach can adjust the sampling process adaptively according to the cost of obtaining extra information.

5. Concluding Remarks

In this paper, a MFBO approach for global optimization is proposed based on the hierarchical Kriging model and a new acquisition function. In the new acquisition function, the value of LF sample points is quantified as the expected further improvement and the cost ratio between HF and LF sampling is considered. Both the location and fidelity level of the next sample point are determined simultaneously by maximizing the acquisition function. For constrained problems, the acquisition function can be further generalized with the surrogates of constraints. The proposed approach has been demonstrated with five numerical problems and one engineering design case. Compared to single-fidelity BO and an existing multi-fidelity BO method, the new approach incorporates the sampling cost differences in the sequential process and shows a higher level of efficiency.

The major limitation of the proposed acquisition function is the cost of direct computation. In this paper, a heuristic approach is taken in searching the maximum of acquisition based on the EI of HF model. The search efficiency is usually better than or at least not worse than the standard EGO. In future work, efficient computational methods for the new acquisition function with the expected further improvement will be investigated. Numerical integration methods such as quadrature and importance sampling can be helpful.

In the proposed acquisition function, the cost ratio of HF to LF samples plays a major role. In all examples of this paper, the ratios were assumed to be known a priori. When the costs of HF and LF simulations are not previously known in simulation-based design optimization, an initial ratio can be estimated. During the sequential sampling process, the cost ratio can be updated on the fly once the simulations are run and actual costs become available. Thus the acquisition function can be adjusted adaptively. Nevertheless, the overall sampling cost proposed in Eq.(19) to evaluate the performance of MFBO approaches requires further study for its fairness in comparisons.

Scalability has been a major issue for Kriging based metamodeling. The number of samples increases exponentially as the dimension of the searching space increases. Approaches such as batch parallelization (Tran et al. 2019a, 2019b) and sparse Gaussian process (McIntire et al. 2016; Zhang et al. 2019) have been applied in Bayesian optimization to alleviate the dimensionality challenge of Kriging. Yet much work of Bayesian optimization for high-dimensional problems remains.

Acknowledgements

Notes

The authors declare no conflict of interest.

Replication of Results

Data or implementation code are available upon request.

References

Bakr MH, Bandler JW, Madsen K, Søndergaard J (2001) An introduction to the space mapping technique. Optimization and Engineering 2:369-384

Bandler JW, Biernacki RM, Chen SH, Grobelny PA, Hemmers RH (1994) Space mapping technique for electromagnetic optimization. IEEE Transactions on Microwave Theory and Techniques 42:2536-2544

Cai X, Qiu H, Gao L, Yang P, Shao X (2016) An enhanced RBF-HDMR integrated with an adaptive sampling method for approximating high dimensional problems in engineering design. Structural and Multidisciplinary Optimization 53:1209-1229

Chang KJ, Haftka RT, Giles GL, Kao I-J (1993) Sensitivity-based scaling for approximating structural response. Journal of Aircraft 30:283-288

Coello CAC (2000) Use of a self-adaptive penalty approach for engineering optimization problems. Computers in Industry 41:113-127

Courrier N, Boucard P-A, Soulier B (2016) Variable-fidelity modeling of structural analysis of assemblies. Journal of Global Optimization 64:577-613

Gano SE, Renaud JE, Sanders B (2005) Hybrid variable fidelity optimization by using a kriging-based scaling function. AIAA Journal 43:2422-2433

Gardner JR, Kusner MJ, Xu ZE, Weinberger KQ, Cunningham JP (2014) Bayesian optimization with inequality constraints. In: ICML. pp 937-945

Gelbart MA, Snoek J, Adams RP (2014) Bayesian optimization with unknown constraints. arXiv preprint arXiv:1403.5607.

Ghoreishi SF, Allaire D (2019) Multi-information source constrained Bayesian optimization. Structural and Multidisciplinary Optimization 59:977-991

Ghoreishi SF, Molkeri A, Arróyave R, Allaire D, Srivastava A (2019) Efficient use of multiple information sources in material design. Acta Materialia 180:260-271

Ghoreishi SF, Molkeri A, Srivastava A, Arroyave R, Allaire D (2018) Multi-information source fusion and optimization to realize ICME: Application to dual-phase materials. Journal of Mechanical Design 140:111409

Han Z-H, Görtz S (2012) Hierarchical kriging model for variable-fidelity surrogate modeling. AIAA journal 50:1885-1896

Huang D, Allen TT, Notz WI, Miller RA (2006) Sequential kriging optimization using multiple-fidelity evaluations. Structural and Multidisciplinary Optimization 32:369-382

Jones DR (2001) A taxonomy of global optimization methods based on response surfaces. Journal of global optimization 21:345-383

Jones DR, Schonlau M, Welch WJ (1998) Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization 13:455-492

Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. Biometrika 87:1-13

Kim Y, Lee S, Yee K, Rhee D-H (2017) High-to-low initial sample ratio of hierarchical kriging for film hole array optimization. Journal of Propulsion and Power 34:108-115

Koziel S, Bandler JW, Madsen K (2006) A space-mapping framework for engineering optimization—Theory and implementation IEEE Transactions on Microwave. Theory and Techniques 54:3721-3730

Liu Y, Chen S, Wang F, Xiong F (2018) Sequential optimization using multi-level cokriging and extended expected improvement criterion. Structural and Multidisciplinary Optimization 58:1155-1173

McIntire M, Ratner D, Ermon S. (2016). Sparse Gaussian processes for Bayesian optimization. Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence, pp. 517-526.

Palar PS, Shimoyama K (2017) Multi-fidelity uncertainty analysis in CFD using hierarchical kriging. In: 35th AIAA Applied Aerodynamics Conference, p. 3261

Park J-S (1994) Optimal Latin-hypercube designs for computer experiments. Journal of Statistical Planning and Inference 39:95-111

Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK (2005) Surrogate-based analysis and optimization. Progress in aerospace sciences 41:1-28

Schonlau M, Welch WJ, Jones DR (1998) Global versus local search in constrained optimization of computer models. Lecture Notes-Monograph Series 11-25

Shu L, Jiang P, Song X, Zhou Q (2019a) Novel approach for selecting low-fidelity scale factor in multifidelity metamodeling. AIAA Journal 57:5320-5330

Shu L, Jiang P, Wan L, Zhou Q, Shao X, Zhang Y (2017) Metamodel-based design optimization employing a novel sequential sampling strategy. Engineering Computations 34:2547-2564

Shu L, Jiang P, Zhou Q, Xie T (2019b) An online variable-fidelity optimization approach for multi-objective design optimization. Structural and Multidisciplinary Optimization 60:1059-1077

Song C, Song W, Yang X (2017) Gradient-enhanced hierarchical kriging model for aerodynamic design optimization. Journal of Aerospace Engineering 30:04017072

Tran A, Sun J, Furlan JM, Pagalthivarthi KV, Visintainer RJ, Wang Y (2019a) pBO-2GP-3B: A batch parallel known/unknown constrained Bayesian optimization with feasibility classification and its applications in computational fluid dynamics. Computer Methods in Applied Mechanics and Engineering 347:827-852

Tran A, Tran M, Wang Y (2019b) Constrained mixed-integer Gaussian mixture Bayesian optimization and its applications in designing fractal and auxetic metamaterials. Structural and Multidisciplinary Optimization 59:2131-2154

Tran A, Wildey T, McCann S (2020) sMF-BO-2CoGP: A sequential multi-fidelity constrained Bayesian optimization framework for design applications. Journal of Computing and Information Science in Engineering 20:031007

Ulaganathan S, Couckuyt I, Ferranti F, Laermans E, Dhaene T (2015) Performance study of multi-fidelity gradient enhanced kriging. Structural and Multidisciplinary Optimization 51:1017-1033

Wang GG (2003) Adaptive response surface method using inherited Latin hypercube design points. Journal of Mechanical Design 125:210-220

Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. Journal of Mechanical design 129:370-380

Xiong Y, Chen W, Tsui K-L (2008) A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling. Journal of Mechanical Design 130:111401

Zhang J, Yao X, Liu M, Wang Y. (2019) A Bayesian discrete optimization algorithm for permutation problems. Proceedings of 2019 IEEE Symposium Series on Computational Intelligence (SSCI 2019), pp.871-881.

Zhang Y, Han Z-H, Liu J, Song W-P Efficient variable-fidelity optimization applied to benchmark transonic airfoil design. In: 7th Asia-Pac int Symp Aerosp Technol, Cairns, Australia, 2015. pp 25-27

Zhang Y, Han Z-H, Zhang K-S (2018) Variable-fidelity expected improvement method for efficient global optimization of expensive functions. Structural and Multidisciplinary Optimization 58:1431-1451

Zhou Q, Shao X, Jiang P, Gao Z, Wang C, Shu L (2016) An active learning metamodeling approach by sequentially exploiting difference information from variable-fidelity models. Advanced Engineering Informatics 30:283-297

Zhou Q, Shao X, Jiang P, Zhou H, Shu L (2015) An adaptive global variable fidelity metamodeling strategy using a support vector regression based scaling function. Simulation Modelling Practice and Theory 59:18-35

Zhou Q, Wang Y, Choi S-K, Jiang P, Shao X, Hu J (2017) A sequential multi-fidelity metamodeling approach for data regression. Knowledge-Based Systems 134:199-212