Structural and Multidisciplinary Optimization (2023) 66:30
https://doi.org/10.1007/s00158-022-03484-7

RESEARCH PAPER

A proportional expected improvement criterion-based multi-fidelity sequential optimization method

Hanyan Huang[1] · Zecong Liu[1] · Hongyu Zheng[1] · Xiaoyu Xu[1] · Yanhui Duan[1]

Received: 28 June 2022 / Revised: 7 October 2022 / Accepted: 10 November 2022 / Published online: 17 January 2023

Abstract

Multi-fidelity surrogate models fusing data from different fidelity systems can significantly reduce the computational cost while ensuring the model accuracy. The focus of this paper is on the sequential design for multi-fidelity models for expensive black-box problem. A Co-kriging-based multi-fidelity sequential optimization method named proportional expected improvement (PEI) is proposed with the objection to be more efficient for global optimization and to be more reasonable to evaluate the costs and benefits of candidate points from different levels of fidelity. The PEI method is an extension of expected improvement (EI) and uses an integrated criterion to determine both location and fidelity level of the subsequent. In the integrated criterion, a proportional factor which is adaptively adjusted according to the sample density is added in EI to adjust the tendency between exploration and exploitation during the search process. Meanwhile, Kullback–Leibler divergence is used to measure the credibility of a point from system with different fidelities, and the cost and constraint of different fidelities are also considered. The effectiveness and advantage of the proposed method were demonstrated by seven analytical functions and then applied to the aerodynamic shape optimization of NACA0012 airfoil. Experiments show that the proportional factor makes the proposed algorithm better search for the global optimum, and the KL divergence can describe the relationship between high and low fidelity more significantly.

Keywords Co-Kriging · Multi-fidelity · Expected improvement · Sequential sampling · Kullback–Leibler divergence

1 Introduction

Bayesian optimization is a powerful global optimization method for solving expensive black-box problems (Greenhill et al. 2020; Shahriari et al. 2015), especially when the objective function is non-convex and expensive to explore. Compared with other black-box optimization algorithms, such as genetic algorithm (GA) (Holland 1975) and particle swarm optimization algorithm (PSO) (Kennedy and Eberhart 1995), Bayesian optimization is a surrogate model-based method, which enables to search the global optimum with much less number of expensive evaluation points. The surrogate model and the acquisition function are the two cores of Bayesian optimization. The surrogate model providing a basic model hypothesis of the system is constructed and updated iteratively. The most popular surrogate model by far is the Gaussian Process (GP) model(Santner et al. 2018), thus Bayesian optimization with GP model is also often referred to as sequential Kriging optimization (SKO) (Huang et al. 2006a) and efficient global optimization (EGO). The acquisition function guides the selection rule of the next sampling point. Under a GP model, the commonly used acquisition functions include expected improvement method (EI) (Jones et al. 1998), probability improvement method (PI) (Jones 2001), upper confidence bound method (UCB) (Srinivas et al. 2018), and knowledge gradient method (Frazier et al. 2008). Among them, EI is the most popular method due to its closed-form acquisition function. The core idea of EI is that if a certain point can bring the greatest expected improvement to the current best point, it will be selected as the next sample point. However, EI is very greedy, it tends to over exploit the fitted GP model, and it is very easy to fall into the local optimum (Qin et al. 2017; Bull 2011). Some

Responsible Editor: Lei Wang.

✉ Hanyan Huang
huanghy99@mail.sysu.edu.cn

[1] School of Systems Science and Engineering, Sun Yat-sen University, No.132, Outer Ring East, Guangzhou Higher Education Mega Center, GuangZhou 510006, Guangdong, People's Republic of China

recent efforts are developed to remedy the greediness of EI, including Snoek et al. (2012), Chen et al. (2017), and so on. But such methods diminish a key advantage of EI: efficient queries via a closed-form criterion. Chen et al. (2020) proposed a hierarchical EI (HEI) criterion by modifying the prior distribution and increasing the variance, the method corrects the greediness of EI while preserving a closed-form acquisition function.

Yet, in many practical instances, the evaluation of a real system of interest is too expensive, and one may consider drawing data from surrogate experiment systems with lower cost. For example, the computer simulations can be used to approximate physical experiments; the numerical simulation might involve simplifying the mathematical model of the physical reality, and it can be run at different levels of complexity. We call these systems "low-fidelity systems" (LFMs) and the real systems the "highest-fidelity systems" (HFMs).

Multi-fidelity models (MFMs) combine both HFMs and LFMs in order to achieve the desired accuracy at lower cost, and it attracted much attention from uncertainty quantification or optimization. MFMs involve generally construction of surrogate models by using variable-fidelity data. Fernández-Godino et al. (2016) reviewed a large variety of MFM implementations, and classified it as multi-fidelity surrogate models (MFSMs) and multi-fidelity hierarchical models (MFHMs). The main concept of multi-fidelity surrogate models (MFSMs) is to use an algebraic surrogate model to correct the LFMs using HFMs. Four main correction methods are space mapping, multiplicative correction, additive correction, and comprehensive correction. The key idea of space mapping is the generation of an appropriate transformation of the vector of fine model parameters $\mathbf{x}_{\mathrm{HF}}$ to the vector of coarse model parameters $\mathbf{x}_{\mathrm{LF}}$, that is $\mathbf{x}_{\mathrm{LF}} = F(\mathbf{x}_{\mathrm{HF}})$, and this technique allows the vectors $\mathbf{x}_{\mathrm{HF}}$ and $\mathbf{x}_{\mathrm{LF}}$ to have different dimensions. The latter three methods can be uniformly described as $\hat{y}_{\mathrm{HF}}(\mathbf{x}) = \rho(\mathbf{x}) \cdot y_{\mathrm{LF}}(\mathbf{x}) + \delta(\mathbf{x})$, the differences among those methods are the treatment of $\rho(\mathbf{x})$ and $\delta(\mathbf{x})$. When multiple fidelity levels are involved, Loïc et al. (2020) reviewed the fusion frameworks based on Gaussian Process (GP) and classified them into four types: linear model of coregionalization (LMC), classical auto-regressive (AR1), nonlinear auto-regressive multi-fidelity Gaussian process (NARFGP), and multi-fidelity Deep Gaussian Process (MF-DGP). AR1 model is widely used in engineering design field, and the first Co-Kriging model developed by Kennedy and O'Hagan (2000) for multi-fidelity computer experiments is just based on this framework. The Co-Kriging method assumed that the covariance structure of the observed data has Markov properties and the relationship between adjacent level of fidelity has an autoregression structure. It provides a closed form of prediction uncertainty in addition to its predictive ability. Most development in the literature were within Kennedy and O'Hagan's framework, including Higdon et al. (2004), Reese et al. (2004), Qian and Wu (2008), Le Gratiet and Garnier (2014), and Le Gratiet (2013). Besides, Han and Stefan (2012) proposed the hierarchical Kriging, in the hierarchical Kriging model, the low-fidelity Kriging model is directly used as the trend of the multi-fidelity model to avoid nested sample points. Tuo et al. (2013) proposed a class of non-stationary Gaussian Process models to link the output from different fidelity levels.

The sequential sampling strategy of multiple fidelity data has also received great interest for the purpose of constructing a sufficiently global accurate metamodel (Xiong et al. 2009; Jin et al. 2002; Le Gratiet and Cannamela 2015) or for optimization. This article focuses on the latter. Currently, the multi-fidelity Bayesian optimization still mainly focused on the expansion of EI criterion, PI criterion, and UCB criterion. In the early days, Xiong et al. (2008) applied the confidence boundary as the acquisition function of the bi-fidelity model in sequence sampling. Xf et al. (2020) proposed a variable-fidelity probability of improvement method. Among the extended methods of EI, Huang et al. (2006b) proposed an augmented EI criterion based on Co-Kriging and called it MFSKO. The method uses the correlation between different fidelities to measure the credibility, and then integrates the correlation and the evaluation cost with the EI acquisition function to determine the location and the fidelity level. As this criterion is based on EI, thus it is unavoidable from the greedy nature of EI and the samples generated may cluster within a certain area and reduce the efficiency of global optimization. He et al. (2017) proposed the EQIE criterion by dividing the EI acquisition function by a cost factor to make the cost within the consideration range of the acquisition function. Kim et al. (2018) used EI acquisition function for the hierarchical kriging model. However, those methods do not take into account the accuracy of the fidelity and the cost issue between different fidelities, thus the lower-fidelity samples are preferred since it is computationally cheaper. Liu et al. (2018) introduced a sample density function into the acquisition function proposed by Huang et al. and proposed an enhanced Co-Kriging based sequential optimization method to reduce the computational cost. Zhang et al. (2018) proposed a multi-fidelity global optimization method based on the hierarchical Kriging model to extend the MFEI. Shu et al. (2021) used the hierarchical Kriging method and proposed the expectation to further improve acquisition criteria.

The focus of this paper is on the sequential design for multi-fidelity models. Since compared with other methods, Co-Kriging can better quantify the relationship between different fidelities, and it is better at dealing with problems with random errors, so as to select new sampling points. We select the similar MFSMs of Huang et al. (2006b). The objective of this article is to develop a sequential sampling strategy such that (1) the method should be more efficient

for global optimization, as most of the current multi-fidelity sequential optimization methods in literatures tend to the local optimum, the reason is that those method is based on classical EI or PI, and the surrogate model is not globally accurate. In this paper, we will firstly develop the revised expected improvement method to strike the correct balance between accurate global surrogate model and high probability of improvement. The proposed method may adaptively adjust the right tendency between exploration and exploitation during the search process; (2) the method should adaptively add sampling points from different levels of fidelity, thus, an effective criterion is required to evaluate the costs and credibility of candidate points from different levels of fidelity. In this paper, the Kullback–Leibler divergence is used to measure the credibility of a point from system with different fidelities.

The paper is organized as follows: in Sect. 2, the optimization problem and the adopted Co-Kriging method are described, after that the prediction uncertainty of the metamodel is discussed. In Sect. 3, the proposed sequential sampling strategy, especially the PEI criterion is presented and then the convergence property of PEI based method is discussed. In Sect. 4, seven analytical functions are used to compare the proposed method with MFSKO, and then the proposed method is applied to the aerodynamic shape optimization of NACA0012 airfoil. Concluding remarks are given in Sect. 5.

2 Background

2.1 The optimization problem

Suppose there are currently m systems with different fidelities (including the real system) to obtain the estimate of the same real system. We consider the systems as black boxes. Denote the input vector as $\mathbf{x}$, and the output of those systems in increasing order of fidelity by $f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_m(\mathbf{x})$. The goal is to optimize the response of the real system within the feasible region $\chi \subset R^d$ and under some constraints, i.e.,

$$\begin{aligned} &\min_{\mathbf{x}\in\chi} f_m(\mathbf{x}) \\ &s.t.\ g_i(\mathbf{x}) \le 0,\ i = 1, \cdots, N_{\mathrm{C}} \end{aligned} \tag{1}$$

where $g_i(\mathbf{x})$ denotes the constraint function, whose closed form maybe known or unknown, N_{C} is the number of constraints. Each system is associated with a cost-per-evaluation denoted by $C_1, C_2, \ldots, C_m$ respectively. And there is an assumption that even a lower-fidelity evaluation is expensive and it is cheaper than a higher-fidelity evaluation, i.e., $C_1 < C_2 < \cdots < C_m$. Thus it is necessary to generate a surrogate model (for example a Kriging model) in order to quickly determine the next search region or for other purposes.

2.2 Co-Kriging for multiple fidelity systems

Kennedy and O'Hagan (2000) proposed a Co-Kriging model based on autoregressive assumption. In this paper, we adopt the simple version method of Huang et al. (2006b), that is

$$f_l(\mathbf{x}) = f_{l-1}(\mathbf{x}) + \delta_l(\mathbf{x}), l = 2, 3 \ldots m \tag{2}$$

where $\delta_l(\mathbf{x}), l = 1, 2, 3 \ldots m$ is independent of $f_1(\mathbf{x})$, $f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})$, and for convenience let $f_1(\mathbf{x}) = \delta_1(\mathbf{x})$. $\delta_l(\mathbf{x})$ is thought of as the "discrepancy" between a lower-fidelity system and the next higher-fidelity system. The reason is that in engineering practice, the results of equivalent experiments will be translated into condition of the real experiment firstly, and then the "discrepancy" between those systems is analyzed and quantified. Quantization of uncertain discrepancy is important and difficult, here we use GP or Kriging model to metamodel the discrepancy. As $\delta_l(\mathbf{x})$ is usually small in scale as compared to $f_l(\mathbf{x})$. In Kriging metamodeling, $\delta_l(\mathbf{x})$ is described as

$$\delta_l(\mathbf{x}) = b_l(\mathbf{x})^T \beta_l + Z_l(\mathbf{x}) + \varepsilon_l, l = 2, 3 \ldots m \tag{3}$$

where $b_l(\mathbf{x})^T$ and β_l are the basis function and its coefficient, $Z_l(x)$, is a stationary Gaussian Process and ε_l is used to describe random error with variance $\sigma^2_{\varepsilon,l}$. In engineering practice, any physics system will have random error, and a computer experiment has no random error. Generally, a comprehensive test will start from computer simulation with lower fidelity and then maybe simulation with higher fidelity, and as followed as physical equivalent test and real test.

The basis functions are often selected as polynomials. In order to simplify the calculation, the constant term is generally selected. $Z_l(\mathbf{x})$ is often assumed to be a stationary Gaussian Process with zero-mean, and generally taken a Gaussian kernel function. Then the covariance between two point $\mathbf{x} = (x_1, x_2 \cdots x_d)$ and $\mathbf{x}' = (x'_1, x'_2 \cdots x'_d)$ in system l is described as

$$\mathrm{cov}[\delta_l(\mathbf{x}), \delta_l(\mathbf{x}')] = \sigma^2_{Z,l} \exp\left[-\sum_{j=1}^{d} \theta_{l,j}(x_j - x'_j)^2\right] \tag{4}$$

where $\theta_{l,j}$ is a "roughness" parameter in the kernel function when a higher value implies lower spatial correlation in dimension j, and $\sigma^2_{Z,l}$ denotes the variance of the stochastic process. Based on equation (4), the covariance between a point $\mathbf{x}$ from system l and another point $\mathbf{x}'$ from system l' can be derived as

$$\operatorname{cov}[f_l(\mathbf{x}), f_{l'}(\mathbf{x}')] = \sum_{i=1}^{\min(l,l')} \operatorname{cov}[\delta_i(\mathbf{x}), \delta_i(\mathbf{x}')] \tag{5}$$

Assume we got observations $Y_1, Y_2, \cdots, Y_n$ from n points $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ located in system indexes $l_1, l_2, \cdots, l_n$, respectively. Let $\mathbf{y}^T = [Y_1, Y_2, \cdots, Y_n]$, and $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_m]^T$, $\mathbf{h}_l(\mathbf{x}) = [b_1(\mathbf{x})^T, b_2(\mathbf{x})^T, \cdots, b_l(\mathbf{x})^T, 0, \cdots, 0]^T$. The best linear prediction (BLP) of $\boldsymbol{\beta}$ and $f_m(\mathbf{x})$ can be given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T\mathbf{V}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{V}^{-1}\mathbf{y} \tag{6}$$

$$\hat{f}_m(\mathbf{x}) = \mathbf{h}_m(\mathbf{x})^T\hat{\boldsymbol{\beta}} + \mathbf{t}_m(\mathbf{x})^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}}) \tag{7}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{l_1}(\mathbf{x})^T \\ \mathbf{h}_{l_2}(\mathbf{x})^T \\ \vdots \\ \mathbf{h}_{l_n}(\mathbf{x})^T \end{bmatrix}$$

$$\mathbf{V} = [\operatorname{cov}(Y_i, Y_j)]_{1\le i,j\le n} = [\operatorname{cov}(f_{l_i}(\mathbf{x}_i), f_{l_j}(\mathbf{x}_j))]_{1\le i,j\le n} + [\sigma^2_{\varepsilon,l}\eta_{ij}]_{1\le i,j\le n}$$

$$\mathbf{t}_l(\mathbf{x}) = \left[\operatorname{cov}(f_{l_1}(\mathbf{x}_1), f_l(\mathbf{x})), \cdots \operatorname{cov}(f_{l_n}(x_n), f_l(\mathbf{x}))\right]^T$$

Let $f_l^p(\mathbf{x})$ be the posterior distribution of the system response of input $\mathbf{x}$ from the real system, thus, the posterior mean of $f_l^p(\mathbf{x})$ is equal to the BLP predictor in (7). The posterior covariance between different points from different fidelities is

$$\operatorname{cov}\left[f_l^p(\mathbf{x}), f_{l'}^p(\mathbf{x}')\right] = \operatorname{cov}\left[f_l(\mathbf{x}), f_{l'}(\mathbf{x}')\right] - \left[\mathbf{h}_l(\mathbf{x})^T, \mathbf{t}_l(\mathbf{x})^T\right] \begin{bmatrix} \mathbf{0} & \mathbf{H}^T \\ \mathbf{H} & \mathbf{V} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{h}_{l'}(\mathbf{x}') \\ \mathbf{t}_{l'}(\mathbf{x}') \end{bmatrix} \tag{8}$$

when $l = l'$ and $\mathbf{x} = \mathbf{x}'$, then (8) is an estimate of the variance of a certain estimated point.

2.3 Estimation of the hyper-parameters

As to estimate the hyper-parameters $\theta_{l,j}$ and $\sigma^2_{Z,l}$, in order to reduce the amount of calculation, it is acceptable to estimate the hyper-parameters for each individual system, and once the hyper-parameters of the low-fidelity system are estimated, we assume they are fixed and the data from systems above are ignored.

For instances, when there are only two fidelities, and the basis function is a constant, then $h_l(\mathbf{x}) = [1, 0]$ or $h_l(\mathbf{x}) = [1, 1]$. Let $A_t(x_i, x_j) = \exp\{-b_t(x_i - x_j)^T(x_i - x_j)\}$, and $A_1(\mathbf{D}_1)$ denote the covariance matrix of the lower-fidelity sampling point, then $\theta_{l,j}$ is referred to as b_1 and $\sigma^2_{Z,l}$ is referred to as σ_1^2, b_1 and σ_1^2 can be estimated by minimizing the following negative log likelihood function

$$\log|\mathbf{A}_1(\mathbf{D}_1)| + n_1\log\sigma_1^2 + (\mathbf{y}_1 - \hat{\boldsymbol{\beta}}_1\mathbf{1}_{n_1})^T\left\{\sigma_1^2\mathbf{A}_1(\mathbf{D}_1)\right\}^{-1}(\mathbf{y}_1 - \hat{\boldsymbol{\beta}}_1\mathbf{1}_{n_1}) \tag{9}$$

For higher-fidelity system with $A_2(\mathbf{D}_2)$ as the covariance matrix, we define $\mathbf{d}_2 = y_2 - f_1(\mathbf{D}_2)$, where $f_1(\mathbf{D}_2)$ denotes the output from $f_1(\cdot)$ at points in $\mathbf{D}_2$. Then b_2, σ_2^2 are estimated by minimizing the following function

$$\log|\mathbf{A}_2(\mathbf{D}_2)| + n_2\log\sigma_2^2 + (\mathbf{d}_2 - \hat{\boldsymbol{\beta}}_2\mathbf{1}_{n_2})^T\left\{\sigma_2^2\mathbf{A}_2(\mathbf{D}_2)\right\}^{-1}(\mathbf{d}_2 - \hat{\boldsymbol{\beta}}_2\mathbf{1}_{n_2}) \tag{10}$$

For more than two levels of fidelity, the hyper-parameters are estimated sequentially, and as the input space of the latter layer is smaller, it is much easier to solve.

2.4 Discussion about the prediction uncertainty

Let $s_m(\mathbf{x})$ be the standard deviation of the estimation, which can be calculated by the following formula

$$s_m(\mathbf{x}) = \sqrt{\operatorname{cov}[f_l^p(\mathbf{x}), f_l^p(\mathbf{x})]} \tag{11}$$

Based on equation (8) and (11), the posterior mean and variance of $f_l^p(\mathbf{x})$ are obtained. As we all know, the posterior variance is also used as a measure of prediction uncertainty, the smaller the variance, the narrower the confidence interval. A domain with high certainty means less opportunity to be explored in the EI sampling criterion. However, as the hyper-parameters of the Gaussian Process are estimated from limited numbers of observations by maximum likelihood estimation, when the observations do not carry enough information about f, the estimation of hyper-parameters will lead to very disappointing results. Thus, as the variance of the $f_l^p(\mathbf{x})$ is estimated by limited numbers of observations and we have enough reasons to make a proportional revise of the variance and make

$$s_m^P(\mathbf{x}) = s_m(\mathbf{x}) \cdot \lambda, \lambda > 0 \tag{12}$$

That is to say, if $\lambda > 1$, the confidence interval of the estimated value in the actual problem should be larger than the confidence interval estimated by the surrogate model.

3 The proposed approach

In this part, we explain in detail the multi-fidelity sequential optimization method we proposed.

3.1 Multi-fidelity sequential optimization framework

The framework of the proposed multi-fidelity sequential optimization employing Co-Kriging in conjunction with the proposed PEI criterion is illustrated in Fig. 1.

The step-by-step description of the proposed method is as follows:

Step 1 Initial design. Compared with the traditional method, the proposed method has a greater error tolerance rate in experimental design. In this paper, Latin hypercube design(Michael 1987) is adopted in low or lower fidelity, the high-fidelity points are not needed to be the subset of low fidelity. But the optimal Latin hypercube design method is chosen in high or higher fidelity to ensure a uniformly distributed test point design combination.

Nevertheless, it is often difficult to choose a suitable initial design in practice, especially the sample size for different fidelities, the initial design of multi-fidelity is still an area worthy of further research.

Step 2 Construct the Co-Kriging model for multiple fidelity systems.

Here, we select the method illustrated in Sect. 2.2 to construct the Kriging meta-models for multiple fidelity systems.

Step 3 Sequential sampling strategy.

In the following paper, we will propose a new sequential sampling criterion called PEI($\mathbf{x}, l$), and the location and fidelity level of the next evaluation are selected by maximizing PEI($\mathbf{x}, l$), that is

$$\left(\mathbf{x}_{n+1}, l_{n+1}\right) = \arg\max_{\mathbf{x},l} \mathrm{PEI}(\mathbf{x}, l) \tag{13}$$

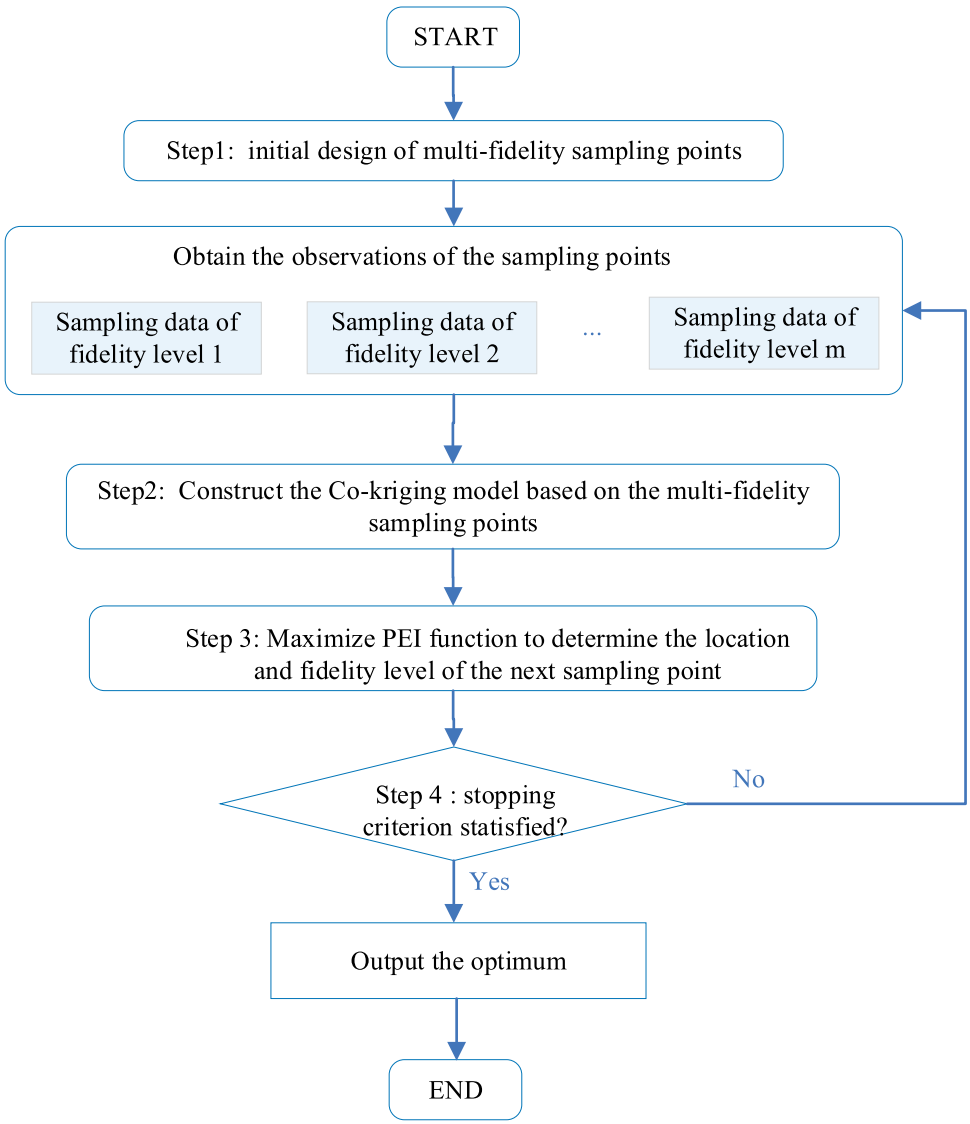


Fig. 1 Sequential optimization framework of the proposed method

Step 4 Stopping criterion.

In the setting of the stopping criterion, we believe that the stopping criterion needs to be adapted to the magnitude of the data.

$$\Delta_s = \max(Y_1, Y_2, \cdots, Y_n) - \min(Y_1, Y_2, \cdots, Y_n) \tag{14}$$

The optimization process stops when

$$\max_{\mathbf{x},l} \mathrm{PEI}(\mathbf{x}, l) < \alpha \cdot \Delta_s \tag{15}$$

In addition, it is required that the point that meets the stopping criterion must be on the highest fidelity. If the point that meets the stopping criterion is not the point on the highest fidelity, the highest-fidelity point closest to the stopping criterion is taken as the optimal value. Generally, we use (15) as the stop criterion. However, in Sect. 4.2, in order to compare the efficiency of the proposed method and MFSKO, we also use iteration numbers as the stopping criterion.

3.2 The proposed sequential sampling strategy

The standard EGO method employs an expected improvement (EI) function to select a new sample point in single-fidelity optimization. Huang et al. (2006b) adopted an augmented EI function to select the location and fidelity level of the next evaluation. However, both methods are based on classical EI function, and are unavoidable to the greedy drawback of EI. And in fact, we found that the samples generated by augmented EI function tend to stuck in local optima and fail to converge to a global optimum. To address this, we propose a new acquisition function called Proportional EI (PEI), which encourages exploration by increasing model uncertainty.

Proportional EI (PEI) acquisition function is defined as follows to determine the location and the fidelity level of the next sample point.

$$\mathrm{PEI}(\mathbf{x}, l) = EI_{\mathrm{P}}(\mathbf{x}) \cdot \alpha_1(\mathbf{x}, l) \cdot \alpha_2(l) \cdot \alpha_3(\mathbf{x}, l) \cdot \alpha_4(\mathbf{x}, l) \tag{16}$$

As shown in equation (16), PEI is an acquisition function composed of five items: revised expected improvement, Kullback–Leibler divergence-based model difference, ratio of cost, random errors, and constraint handling item. The next five subsections will discuss these five items, respectively. Compared with Huang's (2006b) augmented EI function, the PEI method is improved from three espects. Firstly, in order to avoid the greediness of EI, revised expected improvement function $EI_{\mathrm{P}}(\mathbf{x})$ is proposed, where the proportional factor is used to balance exploration and exploitation. Secondly, to measure differences in posterior distribution between system m and system l, Kullback–Leibler divergence-based model difference $\alpha_1(\mathbf{x}, l)$ is used to replace the correlation. Besides, we add a new term $\alpha_4(\mathbf{x}, l)$ to handle constraints, so that the PEI method can be applied to optimization problem with unknown constraint scenarios. As for $\alpha_2(l)$ and $\alpha_3(\mathbf{x}, l)$, they are the same as described in augmented EI function.

3.2.1 The revised expected improvement

In (16), what we concern is the expected improvement of the highest fidelity and the expectation can be calculated as follows:

$$\begin{aligned} EI_{\mathrm{P}}(\mathbf{x}) = {} & \left(\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})\right)\Phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m^P(\mathbf{x})}\right) \\ & + s_m^P(\mathbf{x})\phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m^P(\mathbf{x})}\right) \end{aligned} \tag{17}$$

where ϕ and Φ are the standard normal probability density and cumulative distribution function, respectively. $\mathbf{x}^*$ is the current optimal solution, and $\mathbf{x}^* = \arg\max_{x \in \{\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_n\}}[u(\mathbf{x})]$, and $u(\mathbf{x})$ is called utility function; in our method, we make $u(\mathbf{x}) = -\hat{f}_m(\mathbf{x}) - s_m(\mathbf{x})$.

In (17), according to (12), the measure of uncertainty changed. In Appendix A, we will prove that for a given index λ, there exists a corresponding γ called proportional factor satisfying

$$\begin{aligned} EI_{\mathrm{P}}(\mathbf{x}) = {} & \underbrace{\left(\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})\right)\Phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})}\right)}_{\exp\,\mathit{loitation}} \\ & + \underbrace{\gamma s_m(\mathbf{x})\phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})}\right)}_{\exp\,\mathit{loration}} \end{aligned} \tag{18}$$

And the monotonous trend of γ is consistent to λ. In the following section, we will further discuss the influence of the value of γ to the searching process and will adaptively adjust the value of γ by the sample density.

3.2.2 Kullback–Leibler divergence-based model difference

In (16), $\alpha_1(\mathbf{x}, l)$ is the measure of the credibility or information contribution of a point from system with different fidelities, so as to control the number of low-fidelity samples. In many relevant literatures, it was computed by the correlation between the posterior estimate of point $\mathbf{x}$ from system l to the posterior estimate of point $\mathbf{x}$ from system m, see Huang et al. (2006b), that is

$$\alpha_1(\mathbf{x}, l) = corr\left[f_l^p(\mathbf{x}), f_m^p(\mathbf{x})\right]$$

But in our test, we found that as the point is very close to the observed point, the denominator is close to 0, and the values of points far away from the samples are nearly the same. Our idea is to use the posterior distribution difference between system l and system m located at $\mathbf{x}$ to measure the credibility or the information contribution. That is if the posterior distribution of system l is the same as system m at location $\mathbf{x}$, then the contributions should be the same. The larger the posterior distribution difference, the lower the contribution of the point $(\mathbf{x}, l)$. There are many ways to measure differences in distribution, such as the Kullback–Leibler (KL) divergence (Gultekin and Paisley 2017), Jensen-Shannon, Wasserstein distance, and so on. Here, we use KL divergence to quantify the posterior distribution differences. Given the probability density function p and q, as the KL divergence is asymmetric, the symmetrized KL divergence is defined as

$$D_{\mathrm{KL}}(p, q) = \frac{1}{2} D_{\mathrm{KL}}(p|q) + \frac{1}{2} D_{\mathrm{KL}}(q|p) \tag{19}$$

where $D_{\mathrm{KL}}(p|q) = \int_{-\infty}^{+\infty} p(\mathbf{x}) \cdot \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}$. Specializing to normal distribution $p \sim N\left(\hat{f}_m(\mathbf{x}), s_m(\mathbf{x})\right)$ and $q \sim N\left(\hat{f}_l(\mathbf{x}), s_l(\mathbf{x})\right)$, then the KL divergence is given by

$$D_{\mathrm{KL}}(p, q) = \ln \frac{s_m(\mathbf{x})}{s_l(\mathbf{x})} + \frac{\left(\hat{f}_m(\mathbf{x}) - \hat{f}_l(\mathbf{x})\right)^2 + s_l^2(\mathbf{x})}{2 s_m^2(\mathbf{x})} - \frac{1}{2} \tag{20}$$

The smaller the KL divergence is, the closer the distribution is. KL divergence is equal to 0 if and only if they are identically distributed. Then we define

$$\alpha_1(\mathbf{x}, l) = \frac{1}{1 + \eta\left(D_{\mathrm{KL}}(p, q)\right)^{\kappa}} \tag{21}$$

in which η and κ are parameters that are used to control the change rate. Obviously, as $D_{\mathrm{KL}}(p, q) \geq 0$, $0 < \alpha_1(\mathbf{x}, l) \leq 1$. If the distribution of system l is exactly the same as system m at location $\mathbf{x}$, then it is regarded as equal to the real system. If $l = m$, for any $\mathbf{x}$, $\alpha_1(\mathbf{x}, m) = 1$, it stands that the point from the highest-fidelity system owns the highest credibility. Generally speaking, the credibility of the surrogate system will gradually decrease with the increase of the distribution difference, which indicates that the greater the difference between the distribution of the surrogate system and the distribution of the real system, the smaller the chance the point located to that surrogate system, this is intuitive. Specially, if a point $(\mathbf{x}, l)$ or $(\mathbf{x}, m)$ has been evaluated, $s_l(\mathbf{x}) = 0$, or $s_m(\mathbf{x}) = 0$, we directly define $\alpha_1(\mathbf{x}, l) = 0$, that is to say that the point $(\mathbf{x}, l)$ won't be select again. That is consistent to the common sense, if observation is done at a certain location in any higher fidelity, adding new lower-fidelity observation at the same location contributes very few to the improvement of PEI. Thus, for any $\mathbf{x}$, $\alpha_1(\mathbf{x}, l) \in [0, 1]$. The values of η and κ together with $\alpha_2(l)$ will decide the level index of the next point, and we should decide their values according to actual situation.

3.2.3 Ratio of cost

In (16), $\alpha_2(l)$ is used to control the cost of evaluations. If the cost of evaluation from system l is represented by C_l, then take $\alpha_2(l) = \frac{C_m}{C_l}$. If $l = m$, $\alpha_2(l) = 1$. If $l \neq m$, $\alpha_2(l) > 1$, thus the inclusion of $\alpha_2(l)$ tends to select a point from the cheapest system.

If there are no random errors and do not consider the constraints, $\alpha_1(\mathbf{x}, l)$ and $\alpha_2(l)$ together decide the fidelity level. If at all locations $\mathbf{x}$, we have $\alpha_1(\mathbf{x}, l) * \alpha_2(l) < 1 = \alpha_1(\mathbf{x}, m) * \alpha_2(m)$, then all points are selected in the highest fidelity. If at all locations $\mathbf{x}$, we have $\alpha_1(\mathbf{x}, l) * \alpha_2(l) > 1 = \alpha_1(\mathbf{x}, m) * \alpha_2(m)$, then no point will be selected from the highest fidelity. That is not what we want, as the cost factor $\alpha_2(l)$ is often known to us. For the case where the algorithm time is complex and unknown, we need to run the computer code with different fidelities to get an approximate estimate of the cost. In (16), we should set suitable values of η and κ to make $\forall \mathbf{x}, l$ satisfying $\alpha_1(\mathbf{x}, l) * \alpha_2(l) \in [i, 1] \cup [1, j]$.

3.2.4 Consideration of random errors

In (16), the existence of $\alpha_3(\mathbf{x}, l)$ is to adjust $EI_{\mathrm{P}}(\mathbf{x})$ if the output of system l contains random errors, as more replicates are added, the posterior standard deviation will reduce. Like the work of Huang et al. (2006b), the following penalization function is introduced to limit replications.

$$\alpha_3(\mathbf{x}, l) = 1 - \frac{\sigma_{\varepsilon, l}}{\sqrt{s_l^2(\mathbf{x}) + \sigma_{\varepsilon, l}^2}} \tag{22}$$

And if the variance of the random error is 0, then $\alpha_3(\mathbf{x}, l) = 1$.

3.2.5 Constraint handling

The constraints in engineering optimization can be divided into two categories: known constraints and unknown constraints. The former is easy, a penalty function approach is used to handle the known constraints. The later is much more complex, and after constructing a Kriging model, the prediction of $g(\mathbf{x})$ obeys a normal distribution $g(\mathbf{x}) \sim N\left(\hat{g}(\mathbf{x}), s_g^2(\mathbf{x})\right)$, then the probability of satisfying the constraint is

$$p\left[g(\mathbf{x}) \le 0\right] = \Phi\left(\frac{-\hat{g}(\mathbf{x})}{s_g^2(\mathbf{x})}\right) \tag{23}$$

Thus, in (16), $\alpha_4(\mathbf{x}, l)$ can be defined as

$$\begin{aligned}\alpha_4(\mathbf{x}, l) &= p\left[g(\mathbf{x}) \le 0\right] \\ &= \begin{cases} 1, \text{ if known } g(\mathbf{x}) \le 0 \\ 0, \text{ if known } g(\mathbf{x}) > 0 \\ \Phi\left(\frac{-\hat{g}(\mathbf{x})}{s_g^2(\mathbf{x})}\right), \text{ if unknown } g(\mathbf{x}) \end{cases}\end{aligned} \tag{24}$$

3.3 Further discussion about the revised expected improvement criterion

In this part, we will further discuss the value of γ to the optimum and to the iterative numbers of PEI$(\mathbf{x}, l)$.

3.3.1 The impact of γ to the optimum

In (18), we can see that EI is a tradeoff between exploitation (optimization of the predictor) and exploration (seeking areas of maximum uncertainty), while PEI introduces a coefficient γ to adjust the trend to exploitation or exploration. In particular, we will record $EI_{\mathrm{P}}(\mathbf{x})$ as $EI_{\mathrm{P}}(\mathbf{x}, \gamma)$ in the following section. Clearly, if $\gamma = 1$, PEI is just the standard EI.

If $\gamma > 1$, like the work of Johns et al. (1998), one can get the partial derivative of $EI_{\mathrm{P}}(\mathbf{x}, \gamma)$ with respect to s_m, here recorded $I(\mathbf{x}) = \hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})$, as ϕ and Φ are the standard normal probability density and cumulative distribution function, the result is

$$\begin{aligned}\frac{\partial EI_{\mathrm{P}}(\mathbf{x}, \gamma)}{\partial s_m(\mathbf{x})} &= -\left(\frac{I(\mathbf{x})}{s_m(x)}\right)^2 \phi\left(\frac{I(\mathbf{x})}{s_m(x)}\right) + \gamma\phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &\quad - \gamma\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right)\phi'\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &= -\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right)^2 \phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) + \gamma\phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &\quad + \gamma\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right)^2 \phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &= \left[\gamma + (\gamma - 1)\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right)^2\right] \cdot \phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right)\end{aligned} \tag{25}$$

If $\gamma > 1$, then $\frac{\partial EI_{\mathrm{P}}(\mathbf{x},\gamma)}{\partial s_m(\mathbf{x})} > \phi(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}) = \frac{\partial EI(\mathbf{x})}{\partial s_m(\mathbf{x})}$. That is to say as γ increases, under the same uncertainty, the improvement of $EI_{\mathrm{P}}(\mathbf{x}, \gamma)$ is larger, $\gamma > 1$ will yield to explore more, and can help the multi-fidelity acquisition function jump out of the local optimum and tend to choose more "exploratory" points, which helps preventing falling into a local optimum while $\gamma < 1$ will yield to exploit more.

3.3.2 The impact of γ to the numbers of iterations

Obviously, any $\gamma > 1$ weakens the influence of $\left[\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})\right]$ on $PEI(\mathbf{x})$, while $I(\mathbf{x}) = \hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x}) > 0$ means the improvement of the objective function, itself is positively related to the stopping rule, we prove that when $\gamma > 1$, the iterative steps will increase. One can get several terms that cancel and then the partial derivative of $I(\mathbf{x})$ to (18) is

$$\begin{aligned}\frac{\partial EI_{\mathrm{P}}(\mathbf{x}, \gamma)}{\partial I(\mathbf{x})} &= \Phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) + \frac{I(\mathbf{x})}{s_m(\mathbf{x})}\phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) + \gamma\phi'\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &= \Phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) + (1 - \gamma)\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &= \frac{\partial EI(\mathbf{x})}{\partial I(\mathbf{x})} + (1 - \gamma)\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right)\end{aligned} \tag{26}$$

For $I(\mathbf{x}) = \hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x}) > 0$, then $\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\phi\left(\frac{I(\mathbf{x})}{s_m(\mathbf{x})}\right) > 0$. Thus, if $\gamma > 1$, we have $\frac{\partial EI_{\mathrm{P}}(\mathbf{x},\gamma)}{\partial I(\mathbf{x})} < \frac{\partial EI(\mathbf{x})}{\partial I(\mathbf{x})}$, and as γ increases, $\frac{\partial EI_{\mathrm{P}}(\mathbf{x},\gamma)}{\partial I(\mathbf{x})}$ decreases.

As $\frac{\partial EI(\mathbf{x})}{\partial I(\mathbf{x})} \cong \frac{\Delta EI(\mathbf{x})}{\Delta I(\mathbf{x})}$, thus, for the same $\Delta I(x)$, $\Delta EI_{\mathrm{P}}(\mathbf{x}, \gamma) < \Delta EI(\mathbf{x})$, and if $\gamma_1 > \gamma_2$, we have $\Delta EI_{\mathrm{P}}(\mathbf{x}, \gamma_1) < \Delta EI_{\mathrm{P}}(\mathbf{x}, \gamma_2)$. $EI_{\mathrm{P}}(\mathbf{x}, \gamma)$ with a lager γ improve less can lead to the same improve of $I(\mathbf{x})$. Assume the real optimum is $\mathbf{x}^{**}$, that is to say gap between $\hat{f}_m(\mathbf{x}^*)$ and $\hat{f}_m(\mathbf{x}^{**})$ is given. Then as the iterative optimization goes on, according to the iterative stopping criteria, $EI_{\mathrm{P}}(\mathbf{x}, \gamma)$ will be slow to stop, and the sampling points will cluster.

That is consistent to our experiments, and it was found through experiments that when γ is large enough, the improvement of the results will be indistinctive and the points cluster seriously, and when $\gamma < 1$, the iterative will stop fast but easy to fall into the local optimum. We suggest select the value of γ between [0.1, 10].

3.3.3 Adaptive adjust of γ by the sample density

A larger γ yield to sample at the location with high uncertainty and contribute well to the modeling accuracy, thus at the initial stage of sampling, which is appropriate to choose a larger value of γ, it will avoid early stop of iteration because of low modeling accuracy. However, at the end of the iteration, a larger γ may lead to the clustering of samples and large iteration numbers. As the computational cost is very expensive but the accuracy of the results improved very little, it is not cost-effective. In order to solve that problem, we suggest that the proportional factor γ should be adaptively adjusted during the iteration.

The adaptive criteria can be varied. This paper proposes an adaptive criterion based on the sample density, with "exploitation-exploration-exploitation" as the search mode. We select three levels $\gamma_1 < \gamma_2 < \gamma_3$. They, respectively, indicate active

exploitation, balanced exploration and exploitation, and active exploration. γ will be valued by the following criteria:

(1) When PEI($\mathbf{x}, l$) increase in successive iterations keeps getting smaller, it means that the current exploitation is sufficient. Then we should switch from γ_1 to γ_2 in order to increase the exploratory.

(2) When max PEI($\mathbf{x}, l$) $< k\alpha\Delta_s$, that means the expected improvement may be very small, we should convert from γ_2 to γ_3, and turn to active exploration. In this paper, we adopt $k = 10,\ \gamma_1 = 0.5,\ \gamma_2 = 1,\ \gamma_3 = 2$.

(3) The expected improvement in successive iterations has been increasing, which means that we have discovered new peaks through increasing exploration. At this time, in order to avoid sampling too densely, we should convert from γ_3 to γ_1 in order to hurry up the iteration process. In this paper, when the number of iterations is 1.5 times of the initial design number, we let $\gamma_1 = 0.5$.

Above all, our method can be summarized in Algorithm 1.

Algorithm 1 PEI algorithm

Input: design space Ω, test function $f_1, f_2, ..., f_m$;
Output: y_{min};

```
Latin Hypercube Sample in Ω to obtain sample points D_1, D_2, ..., D_m;
calculate response y_l = f_l(D_l);
calculate stop criterion by (15);
γ=1;
while stop criterion is not satisfied do
    estimate hyperparameters by (9) and (10);
    update cokriging by (2) - (8);
    x* = arg max_{x∈{x_1,x_2,...x_n}} [u(x)];
    calculate PEI with different l;
    (x_new, l)=argmaxPEI(x,l);
    y_new = f_l(x_new);
    D_l = D_l ∪ x_new;
    y_l = y_l ∪ x_new;
    choose γ = 0.5, 1 or 2;
end while
y_min = min(y_m)
return y_min.
```

4 Test cases of analytical functions and application

In this section, the effectiveness of the proposed multi-fidelity sequential optimization method will be verified by test functions and an engineering application example. Obviously, the difference between the optimization result and the real optimum is an important index to measure the quality of the sequential optimization design method. On the other hand, as compared with the complex and time-consuming finite element simulation, the computational amount of sequential optimization algorithm is relatively small, so the evaluation cost (the number of high- and low-fidelity test samples) is also considered here.

4.1 An illustrative example

In this subsection, the proposed multi-fidelity sequential optimization method is illustrated with a one-dimensional method. Firstly, we use a couple of test functions created by (Sasena et al. 2002):

$$f_h(x) = -\sin(x) - \exp(x/100) + 10, (0 < x < 10)$$
$$f_l(x) = -\sin(x) - \exp(x/100) + 10.3 + 0.03 \times (x-3)^2 \quad (0 < x < 10)$$

As Sasena function is also used by Huang et al. (2006b) to explain MFSKO method, here we will draw a comparison between PEI method and MFSKO method.

Here, the initial design of the test is six points located at $x = \{0, 2, 4, 6, 8, 10\}$ in low fidelity, and two points at $x = \{3.5, 6.5\}$ in high fidelity. Fig. 2 shows the optimization result of the proposed method by letting $\gamma = 1$, which is similar to Huang's MFSKO method. We can see that the search scope of MFSKO is only around the left local optimal peak and stop searching after four times iterations. The global optimum of the right peak is not found. Fig. 3 shows the searching result of the proposed method by letting $\gamma = 2$, the method jumps out of the local optimal, and successfully finds the global optimum. A series of examples listed below will further demonstrate that the proposed PEI method with adaptive value of γ is indeed helpful to explore the global optimum.

The reason for this is that the EI function itself has a tendency to fall into a local optimum, that is the greediness of the EI. So a single-fidelity optimization example is given. We choose the high-fidelity Sasena function above as test function. The initial design is $x = \{1, 4, 6, 9\}$. EI function and Revised EI with γ=2 are adopted as acquisition functions. As shown in Fig. 4, the EI function itself has a tendency to fall into the local optimum, and this shortcoming can be avoided by Revised EI.

To help understanding each item in equation (16), Fig. 5 displays the breakdown of PEI just after the initial design. $EI_p(x)$ in equation (16) incorporates the prediction mean and the prediction uncertainty, while the proportion between exploitation and exploration is adjusted during the search process, this term is independent of the fidelity level, we can see that the left domain of the searching space is preferred.

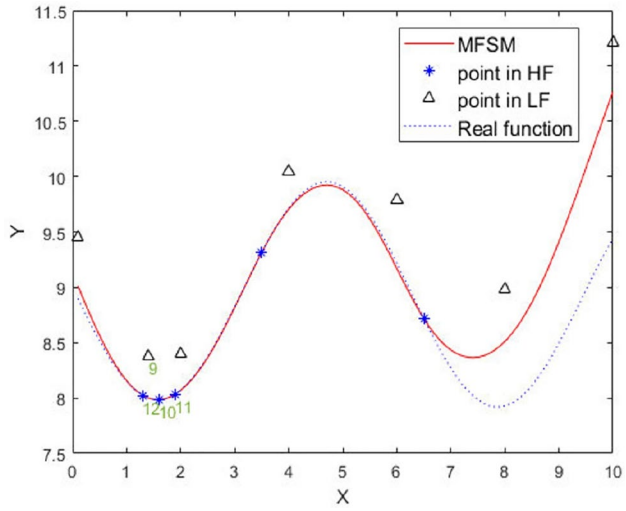


Fig. 2 Search pattern for Sasena function when $\gamma = 1$

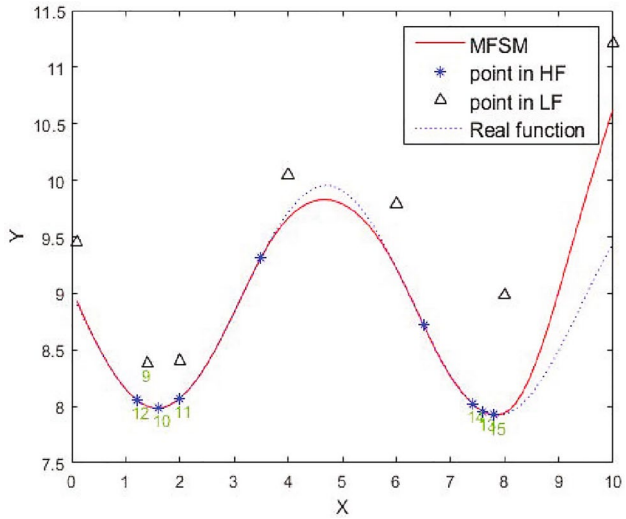
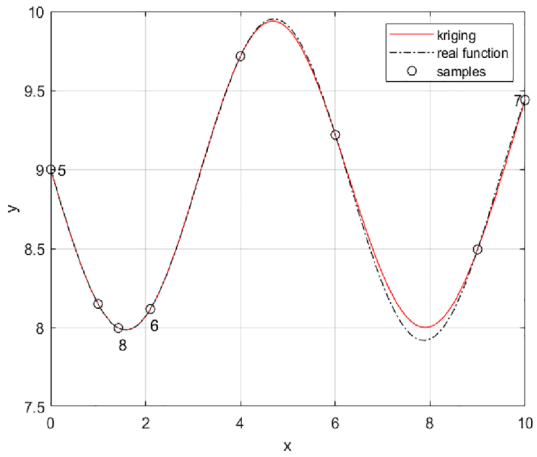
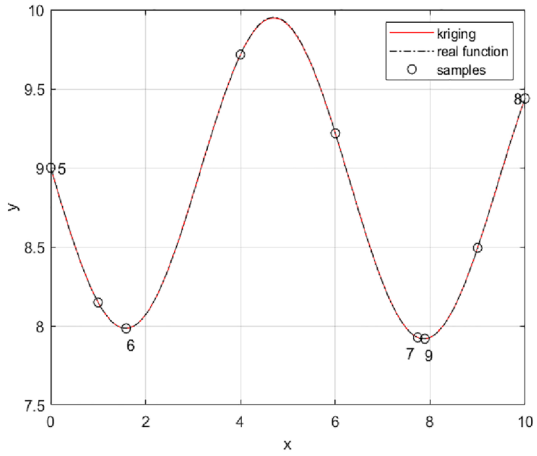


Fig. 3 Search pattern for Sasena function when $\gamma = 2$



(a) EI function for single-fidelity Sasena function



(b) PEI function for single-fidelity Sasena function when $\gamma = 2$

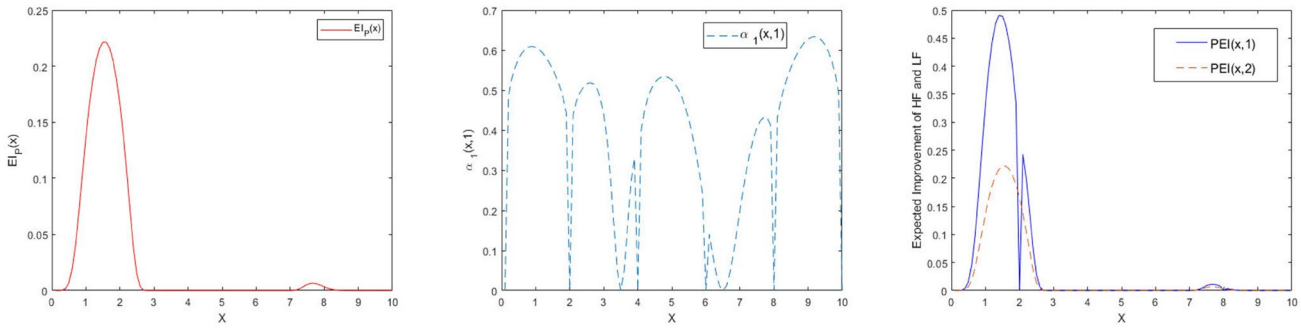
Fig. 4 contrastive test to explain the greediness of EI

Here, the value of $\alpha_1(x, 1)$ is based on the distribution difference of different fidelities, and it measures the credibility of expected improvement when a lower-fidelity evaluation is added. Here, η and κ in equation (21) are set as 0.5 and 1, respectively. As the random error and the constraint are not considered, $\alpha_3(x, l) \cdot \alpha_4(x, l) = 1$, based on the costs of evaluation, let $\alpha_2(1) = 4$ and $\alpha_2(2) = 1$. The PEI values of different fidelities are shown in Fig. 5.

Fig. 6 shows the value of $\alpha_1(x, 1)$ from the 9th to the 14th iteration; according to the figure, the values are very small near the tested points, and after six iterations, the values in the whole design space of lower fidelity are too small to be selected as the next sample points, which is consistent with the expectation of Bayesian optimization that the sample points in the later stage of optimization are all in high fidelity.

4.2 Other numerical tests

In this subsection, a series of test functions from literature will be used to further investigate the properties of the proposed PEI method. Seven tests are conducted as listed in Table 1. The first five test functions are unconstrained, the sixth one is constrained, and the seventh one is a three-fidelity function (Xiao et al. 2002).



(a) $EI_p(x)$ in (16) **(b)** $\alpha_1(x,1)$ in (16) **(c)** PEI value

Fig. 5 The PEI function of initial design

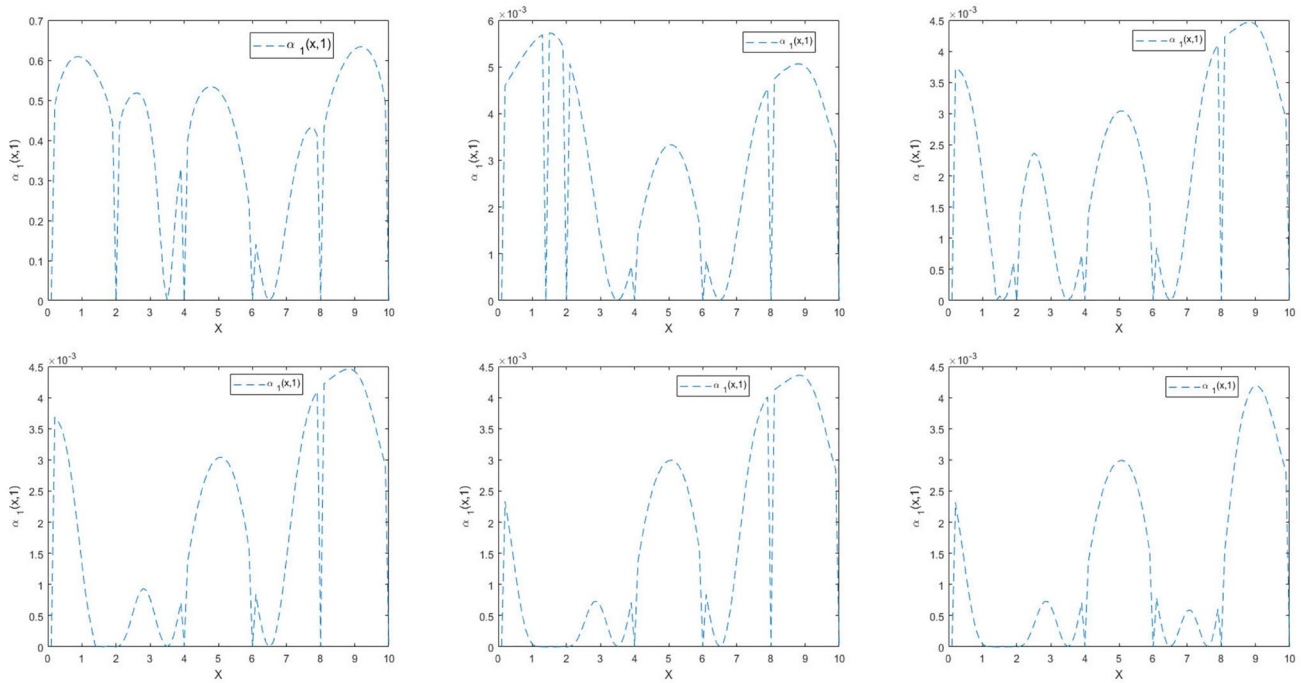


Fig. 6 The values of $\alpha_1(x, l)$ during the 9th–14th iteration

4.2.1 Contrast of optimization result

In this subsection, we focus on the optimization results of the test functions. In all of the test, Latin hypercube is used to generate the initial design.

Specially, the search pattern of Case 4 is shown in Fig. 7, from which we can see after 1 time of high fidelity and 3 times of low-fidelity search, the peak where the optimal value located is successfully searched. Within that peak, after 5 times of high fidelity and 8 times of low-fidelity search, the iteration stops and the optima is searched.

Case 6 is a multi-fidelity optimization problem with unknown constraint. Fig. 8 displays the contour map of MFSM after optimization and the sample points. The points without labels are the initial design, including 9 high-fidelity points and 15 low-fidelity points. After 3 times of low-fidelity observations and 5 times of high-fidelity observations, the search stops and got optimal value as (0.75, 1.75) and optimum $f_h(x_{best}) = 3.2344$. If there is no constraint, the best advantage should be (0.1, 0.1), so the constraint comes into play.

Case 7 is a three-fidelity optimal problem, where l_1 represents first-low fidelity, l_2 represents second-low fidelity, and h represents the high fidelity. We set C_m/C_{l_2}=2 and C_m/C_{l_1}=4. The initial design is $x = \{0.1, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1\}$ at first-low fidelity, $x = \{0.2, 0.4, 0.6, 0.8\}$ at second-low

Table 1 Test functions

source	Description of functions
Case 1	$f_l(x) = -\sin(x) - \exp(x/100) + 10 + k \cdot \left(0.3 + 0.03 \times (x-3)^2\right)$ $f_h(x) = -\sin(x) - \exp(x/100) + 10$ $x \in [0, 10]$
Case 2	$f_l = \frac{\sin(20\times x)/(1+x)+3\times x^2\times\cos(5x)+10\times(x-0.5)^2-0.6}{2} + \frac{\sin(15\pi\times(x+0.1))}{5}$ $f_h = \frac{\sin(20\times x)/(1+x)+3\times x^2\times\cos(5x)+10\times(x-0.5)^2-0.6}{2}$ $x \in [0, 1]$
Case 3	$f_l = 0.5 \times (6x-2)^2 \times \sin(12x-4) + 10 \times (x-0.5) + 5$ $f_h = (6x-2)^2 \times \sin(12x-4)$ $x \in [0, 1]$
Case 4	$f_l = -(x_2+47) \times \sin(\sqrt{\lvert\frac{x_1}{2} + x_2 + 47\rvert}) - x_1 \times \sin(\sqrt{\lvert x_1 - x_2 - 47\rvert}) + 20 * \sin[4\pi(\frac{x_1}{150} + \frac{x_2}{150})]$ $f_h = -(x_2+47) \times \sin(\sqrt{\lvert\frac{x_1}{2} + x_2 + 47\rvert}) - x_1 \times \sin(\sqrt{\lvert x_1 - x_2 - 47\rvert})$ $x_1, x_2 \in [-75, 75]$
Case 5	$f_h = -\sum_{i=1}^{4} c_i \exp[-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2]$ $f_l = -\sum_{i=1}^{4} c_i \exp[-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2] + MA3(x).$ $MA3(x) = 0.585 - 0.324x_1 - 0.379x_2 - 0.431x_3 - 0.208x_1x_2 + 0.326x_1x_3 + 0.193x_2x_3$ $+0.225x_1^2 + 0.263x_2^2 + 0.274x_3^2$ $A = \begin{pmatrix} 3 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3 & 10 & 30 \\ 0.1 & 10 & 35 \end{pmatrix}, C = \begin{pmatrix} 1 \\ 1.2 \\ 3 \\ 3.2 \end{pmatrix}, P = \begin{pmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.03815 & 0.5743 & 0.8828 \end{pmatrix}$
Case 6	$f_h(x) = 4{x_1}^2 + x_2^3 + x_1x_2$ $g_h(x) = 1/x_1 + 1/x_2 - 2$ $f_l(x) = 4(x_1+0.1)^2 + (x_2-0.1)^3 + x_1x_2 + 0.1$ $g_l(x) = 1/x_1 + 1/(x_2+0.1) - 2 - 0.001$ $x_1, x_2 \in [0.1, 10]$
Case 7	$f_{l_1} = 0.5 \times (6x-2)^2 \times \sin(12x-4) + 10 \times (x-0.5) + 5$ $f_{l_2} = 0.4 \times (6x-2)^2 \times \sin(12x-4) - x - 1$ $f_h = (6x-2)^2 \times \sin(12x-4)$ $x \in [0, 1]$

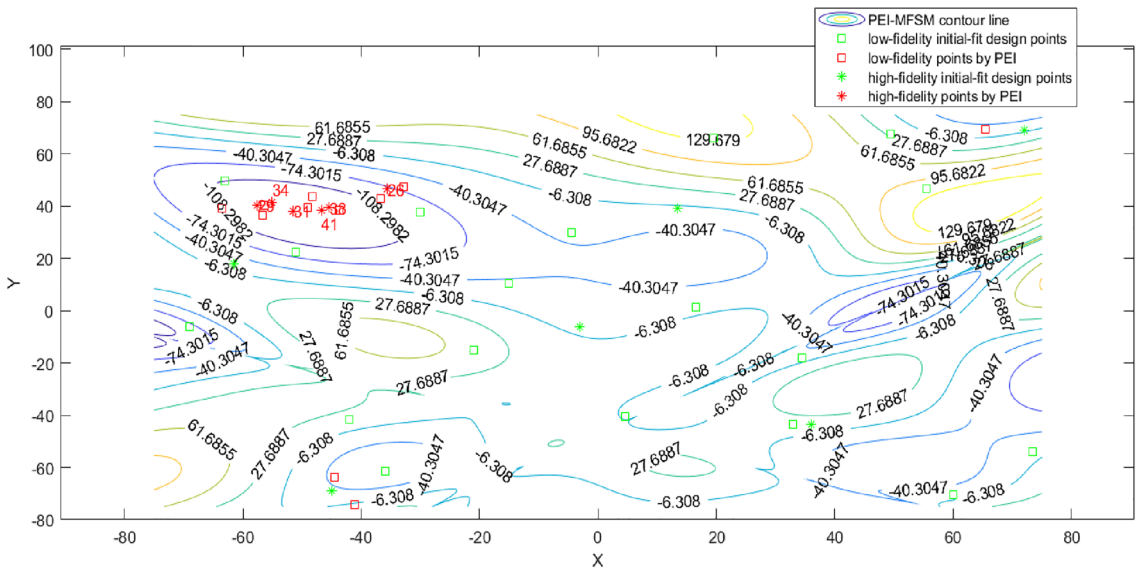


Fig. 7 The search process of Case 4

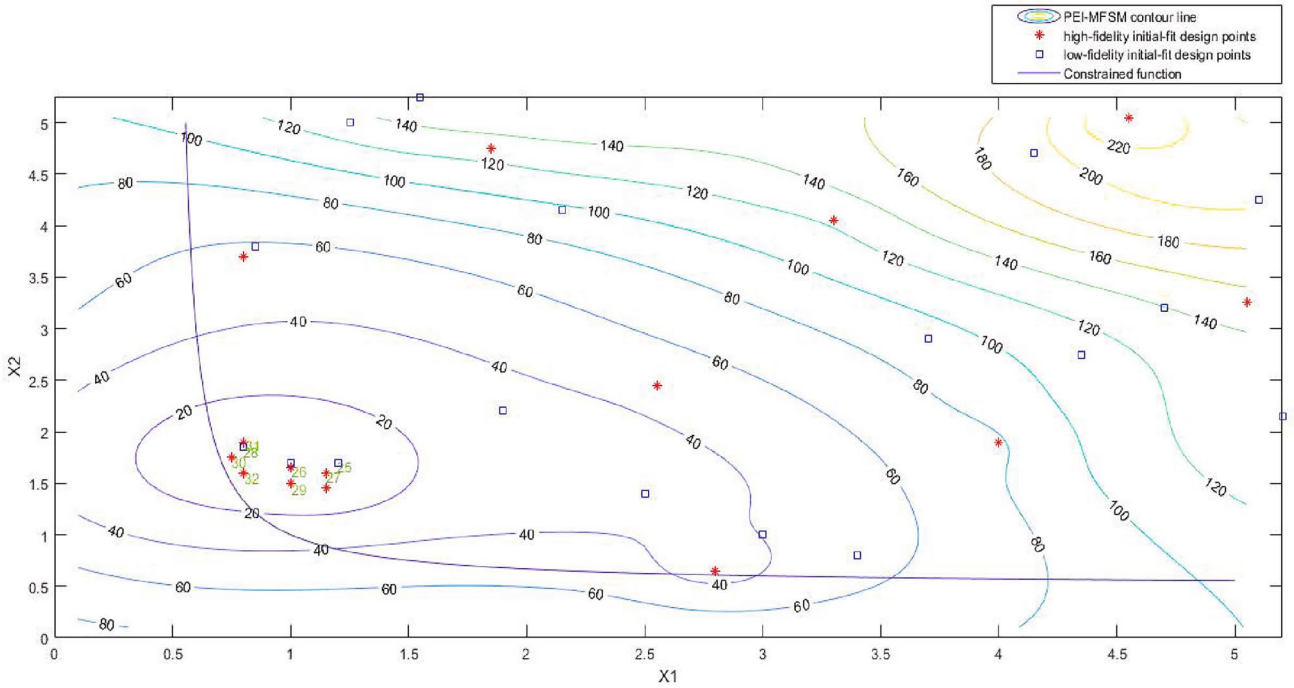


Fig. 8 The search process of Case 6

fidelity, and $x = \{0.35, 0.65\}$ at high fidelity. As shown in Fig. 10, it is obvious that the Co-kriging model is closer to the real function compared with that in Fig. 9 and the PEI method finds the optimal value -6.02 after 13 iterations.

Table 2 shows the results of PEI and MFSKO for the former five test functions. From Table 2, we can see that PEI can jump out of the local optimum. When both methods search into the peak of the real optimum, the results of PEI are still closer to the real optima than the results of MFSKO. When the observations especially the high-fidelity observations are inadequate or the observations are worse space-filling, there is insufficient information to metamodel the real system, and the expected improvement deduced by the constructed MFSM is sometimes underestimated and hence causes premature stopping. Toward the introduction of γ and selecting a large value of it, the probability of exploration will increase, then PEI can jump out of local optimum and have the chance to get a better point.

Besides, to verify the efficiency and accuracy of the algorithm, we use the same number of iterations as the stopping criterion. As shown in Table 3, PEI can find the optima more accurate. Even when the accuracy is the same, PEI saves the cost, such as Case 3 and Case 4, because the different choices of two algorithms for high or low fidelity lead to different cost. For example, in Case 3, the total number of searches for both algorithms is 24,

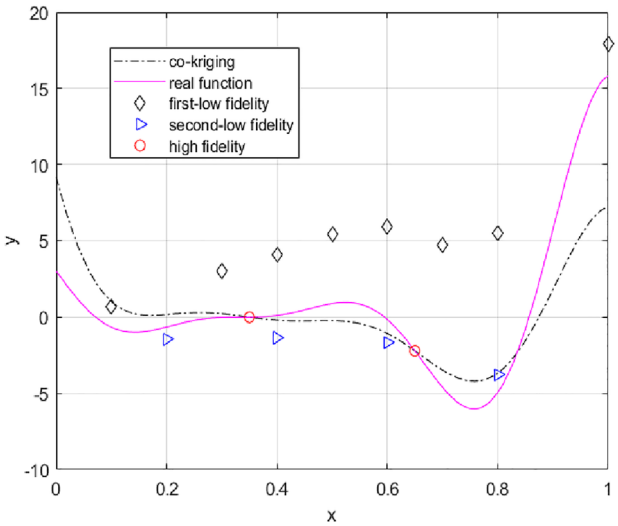


Fig. 9 Initial design of Case 7

but the high-fidelity number of PEI is 9, which is less than that of MFSKO.

4.2.2 Contrast of test cost

In this part, the performances of PEI method and MFSKO method are compared when the cost ratio changes are compared. Fig. 12 shows the cost change process in the search

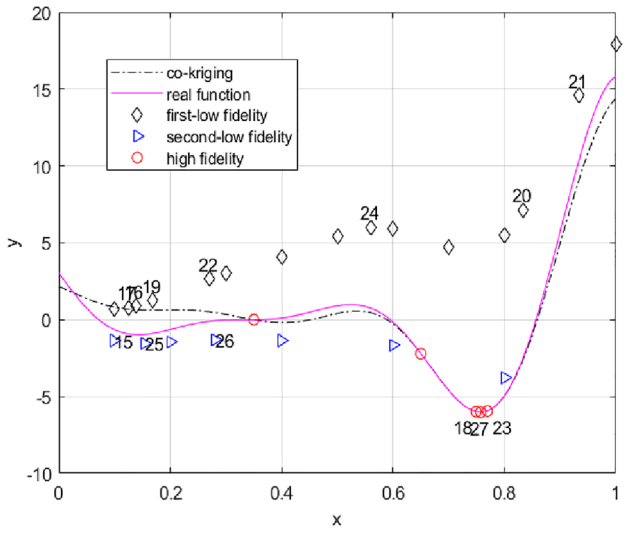


Fig. 10 The search process of Case 7 by PEI method

Table 2 The optimization results of the former five test functions

	real optima	In the optimal peak?		Gap to real optima	
		MFSKO	PEI	MFSKO (%)	PEI (%)
Case 1	7.9189	Yes	Yes	0.03	0.03
Case 2	− 1.0391	Yes	Yes	1.27	0.19
Case 3	− 6.0167	Yes	Yes	0.19	0
Case 4	− 126.3537	Yes	Yes	1.50	0.40
Case 5	− 3.8627	No	Yes	3.72	1.46

Table 3 The optimization results and total cost of the former five test functions when the stop criterion is iteration number

	initial	Gap to real optima		Total cost	
		MFSKO (%)	PEI (%)	MFSKO	PEI
Case 1	6+2	0.03	0	15+9	11+13
Case 2	6+2	0.14	0	13+11	14+10
Case 3	6+2	0	0	14+10	15+9
Case 4	18+6	0	0	30+42	44+28
Case 5	25+9	0.04	0	60+42	37+65

(1) The second column of the table shows the number of initial designs, for example, 6+2 represents 6 low-fidelity samples and 2 high-fidelity samples

(2) cost-per-evaluation of real system is 1

(3) The stop criterion is set as two times for initial design

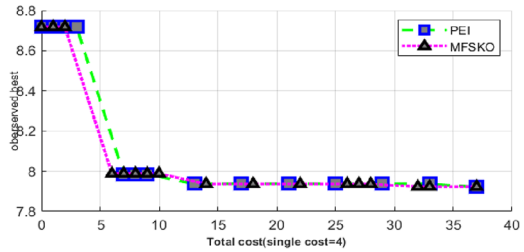
process of former five test functions, where the abscissa is the cost currently spent, and the ordinate is the optimal value observed so far. By comparison, we found that both the search process of PEI method and MFSKO method will change when the cost ratio changes. And as the cost ratio increases from 4:1 to 20:1, the cost increase of each step of both methods becomes slower, that means more observations on low-fidelity systems are carried out. Fig. 11 also reflects the higher optimization accuracy of PEI method than MFSKO method. The reason for this is that the KL divergence is more significant than correlation in describing the relationship between high and low fidelity, which is not affected by the cost ratio.

Table 4 shows the total cost of the MFSKO and PEI as cost ratio changes and the details for high- and low-fidelity sample can refer to Table B1 in Appendix B. Due to the increase of exploratory nature, one may consider the total cost of PEI criterion may increase compared with the MFSKO method; however, from Table 4, we can see it is not always true, especially when the convergence accuracy is the same, the cost of PEI is smaller than that of MFSKO. We attribute this to the adjustment of proportional factor γ during the sampling process. In this paper, we adopted a simple adjustment method, and we believe that a better adaptive criterion of γ will lead to the improvement of PEI method to achieve high accuracy and considerable cost.

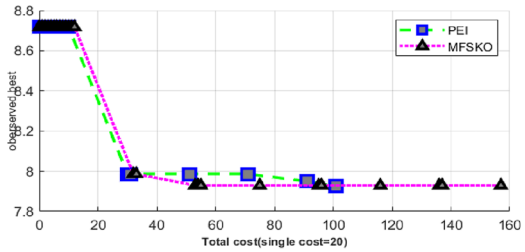
4.2.3 Discussion of $\alpha_1(\mathbf{x}, 1)$

This section focuses on discussing the value of $\alpha_1(\mathbf{x}, 1)$ in multi-fidelity sequential criteria. We want to know as the difference between low fidelity and high fidelity increases, what will happen to the value of $\alpha_1(\mathbf{x}, 1)$. Here, we use Case 1 as test function and take the value of k as 1, 3, and 5, and the value curves of $\alpha_1(\mathbf{x}, 1)$ are shown in Fig. 12. According to Fig. 12, as the value of k increases, i.e., the difference between low fidelity and high fidelity increases, the general trend of the curve will not change, but the overall value will always be smaller, that means on the same condition, the probability of point from lower fidelity will be selected reduces. This is consistent with the intuition.

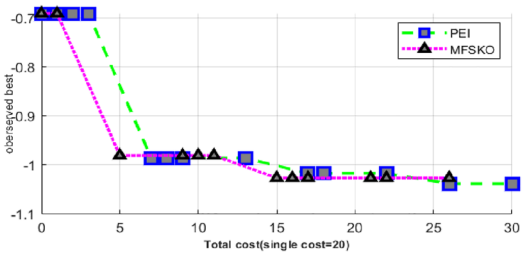
$\alpha_1(\mathbf{x}, 1)$ measures the credibility of fidelity level and KL divergence method is used in this paper. From Fig. 12 and Fig. 6, we can see that the value curve of $\alpha_1(\mathbf{x}, 1)$ is basically continuous and valued as 0 at the test points. $\alpha_1(\mathbf{x}, 1)$ measures the credibility of fidelity level. That is similar to using the correlation, but we think it is more explanatory as it is based on the distribution difference. Furthermore, the method is more flexible as we can change the value of η, κ



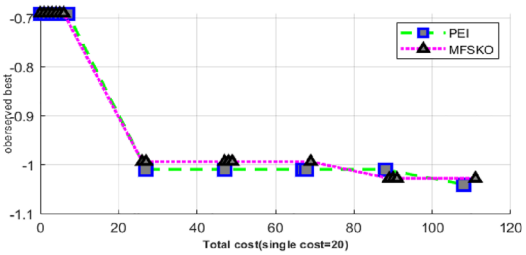
(a) Case 1 with cost ratio 4:1



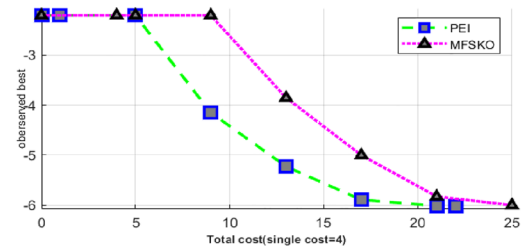
(b) Case 1 with cost ratio 20:1



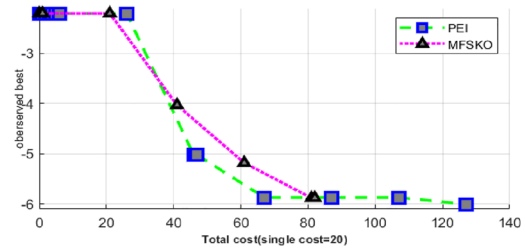
(c) Case 2 with cost ratio 4:1



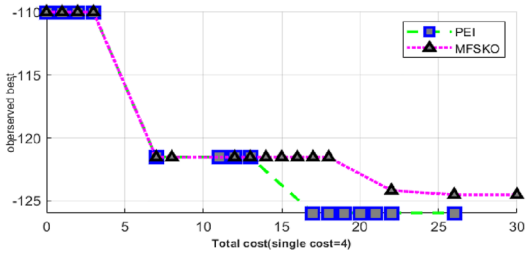
(d) Case 2 with cost ratio 20:1



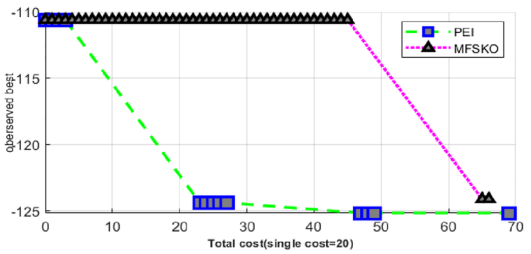
(e) Case 3 with cost ratio 4:1



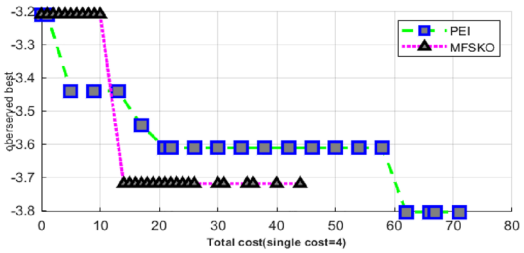
(f) Case 3 with cost ratio 20:1



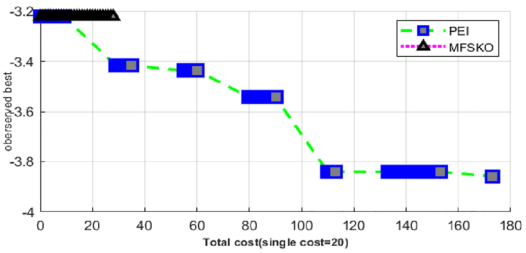
(g) Case 4 with cost ratio 4:1



(h) Case 4 with cost ratio 20:1



(i) Case 5 with cost ratio 4:1



(j) Case 5 with cost ratio 20:1

Fig. 11 Computational cost changes as iterations proceed of PEI and MFSKO

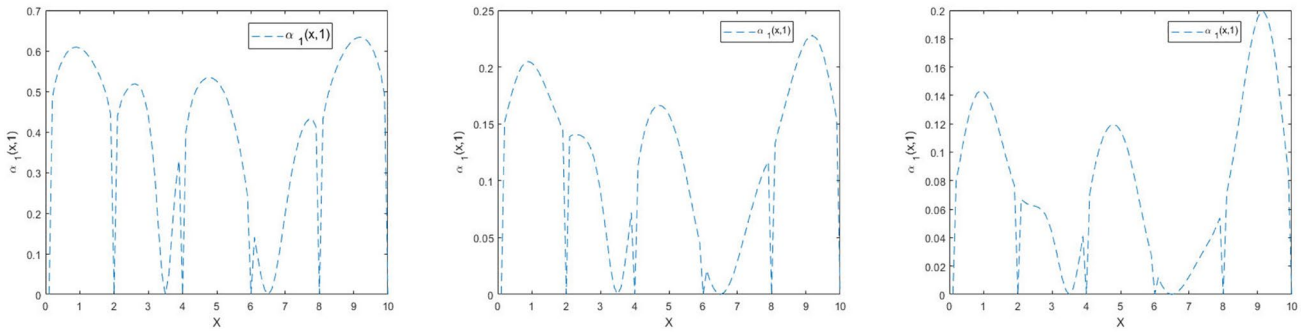


Fig. 12 the value of $\alpha_1(x, 1)$ as k changes in Case 1

Table 4 The total cost of the former five test functions

	$C_m/C_l = 4$		$C_m/C_l = 20$	
	MFSKO	PEI	MFSKO	PEI
Case 1	12.75	12.75	10.15	7.85
Case 2	10	11	7.85	7.7
Case 3	9.75	9.0	6.4	8.65
Case 4	19	17	11.15	11.35
Case 5	26.25	33	11.2	18.9

(1) Intial design is the same as Table 3

(2) cost-per-evaluation of real system is 1

together with the cost ratio to meet the preference of fidelity level. For instanse, the parameters are set as $\eta = 2,\ \kappa = 0.5$ in case 2 and $\eta = 2,\ \kappa = 1$ in case 3.

4.3 Application to air drag minimization of the NACA 0012 airfoil

Aerodynamic shape optimization design of airfoil is one of the key problems in aircraft design. With the development of CFD Technology and optimization algorithms, the optimization design of aerodynamic shape has made great progress. In particular, the high-efficiency computer has shortened the cost of optimization design. However, a high-accuracy simulation still takes a long time. So, two major problems in aerodynamic shape optimization design should be considered: on the one hand, how to design an optimization algorithm that can quickly and accurately find the global optimal solution; on the other hand, in many practical instances, the evaluation of a real system of interest is too expensive, and one may consider drawing data from surrogate experiment systems with lower cost. Here, we will apply the proposed PEI method to optimize the NACA0012 airfoil.

The airfoil profile needs to be parameterized before optimizing. An excellent method of parameterized characterization of airfoil is to use as few design variables as possible to represent the design space with given constraints, so as to effectively reduce the computational cost. At present, the commonly used methods for the parametric characterization of airfoil include linear function perturbation method, characteristic parameter description method, and Class-Shape Function Transformation (CST) method (Ivanov et al. 2017), which is adopted in this paper.

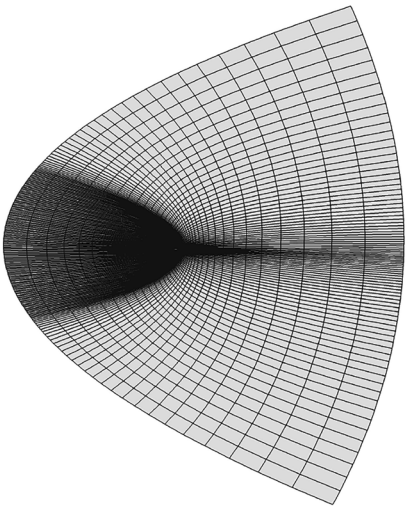
The mathematical expression of CST parameterization method of NACA0012 airfoil is

$$y = (x^{0.5} \times (1 - x)) \cdot S(x) + x \cdot z_{te} \tag{27}$$

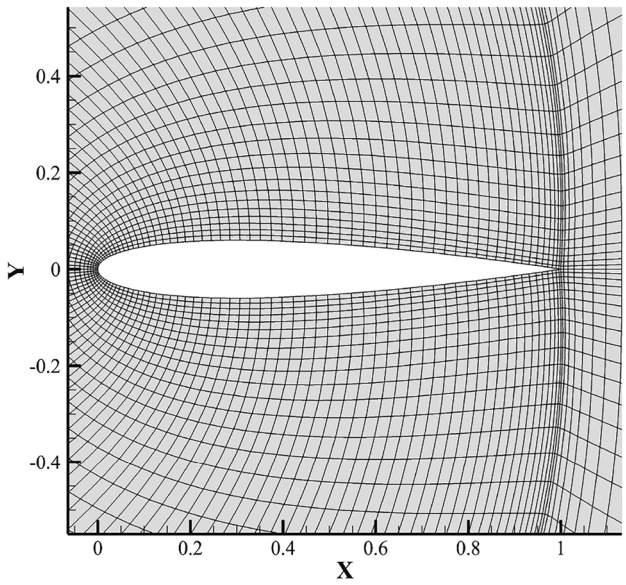
in which x represents the dimensionless coordinate value in the chord direction of the airfoil, y represents the dimensionless coordinate value in the thickness direction of the airfoil, z_{te} is the half thickness of the trailing edge, and $S(x)$ represents the shape function, which is usually expressed by the N-order Bernstein polynomial given by

$$S(x) = \sum_{i=0}^{N} A_i \cdot S_i(x) \tag{28}$$

where $S_i(x) = C_N^i x^i (1-x)^{N-i} = \frac{N}{i!(N-i)!} x^i (1-x)^{N-i}$. For n-order Bernstein polynomials, the upper and lower surfaces of NACA0012 airfoil have a total of $N + 1$ design variables, which are $A_0, A_1, \cdots, A_N$. However, since the design variable positions on the upper and lower surfaces of the leading edge position are the same, there are $2N + 1$ design variables that need to be optimized. In this example, the Bernstein polynomial order N is selected as 5, so the airfoil optimization design problem involves 11 design variables. The three most important design variables were screened out through sensitivity analysis, and were denoted as a_1, a_2, and a_3. The other eight design variables were fixed to optimize the airfoil. The value ranges of the three design variables are $a_1 \in [0.153849, 0.188038]$, $a_2 \in [0.139910, 0.171001]$, and $a_3 \in [0.143035, 0.1748200]$.



(a) Overall grid division



(b) Grid division of airfoil

Fig. 13 Computational grid of NACA0012 airfoil

The work aims at minimizing the air drag of the NACA0012 airfoil at a free-stream Mach number $M_\alpha = 0.8$ and an angle of attack $\alpha = 0°$, Reynolds number $\mathrm{Re} = 6.5 \times 10^6$. When the Mach number is 0.8, the compressibility effect of the gas makes strong positive shock waves form on the airfoil surface, and then shock resistance is generated. This drag can be significantly reduced by changing the geometry of the airfoil, so airfoil drag reduction design examples at transonic speeds are often used to verify the algorithm. In our work, CFD is employed for aerodynamic analysis and a C-type grid is generated to divide the computing domain as shown in Fig. 13. Two kinds of grids with different sparsity, respectively, are 321×65 and 193×33 which are used to simulate and obtain high- and low-fidelity data for the NACA0012 airfoil. And the simulation time of high fidelity and low fidelity is treated as the cost, which is nearly 4:1. The CFD simulation is built by us, for details refer to (Duan et al. 2012, 2019).

We obtain 25 low-fidelity and 9 high-fidelity initial samples by using the Latin hypercube design. Table 5 shows the last 9 samples before stopping, and the right column of the table gives the airfoil drag value. It can be seen that the proposed PEI method still has a tendency to search low fidelity first and then high fidelity even near the stopping point. The high-fidelity point searched in step 42 is taken as the optimization result.

A total of 31 high-fidelity and 45 low-fidelity experiments were performed in the whole calculation, including the initial design. However, a total of 36 high-fidelity and 50 low-fidelity experiments were conducted by the MFSKO method. At the same time, we also adopted the particle swarm optimization algorithm(PSO) to optimize the design. As an engineering example, the real optimal value is unknown in advance, so the PSO algorithm is employed to approximate the real optimal value. 30 particle points were selected on the real system, and after 38 iterations, the optimal value 2.444136E-03 was obtained. As shown in Table 6, PEI and MFSKO are also compared.

The airfoil curves after optimization by the three method are shown in Fig. 14, we can see that compared to the baseline airfoil, the thickness and curvature are both reduced,

Table 5 The result of the last 8 samples before iteration stop

Iteration	Best solution	Fidelity level	maximal PEI	Response
35	(0.159029, 0.171001, 0.164546)	H	0.032	1.322173E-02
36	(0.153849, 0.143051, 0.147530)	L	0.030	3.869752E-03
37	(0.179059, 0.162208, 0.143035)	H	0.027	3.869752E-03
38	(0.163518, 0.159067, 0.143035)	H	0.026	7.297047E-03
39	(0.180440, 0.139910, 0.159409)	H	0.022	5.547299E-03
40	(0.153849, 0.139910, 0.143035)	L	0.0014	3.110536E-03
41	(0.156957, 0.139910, 0.143035)	L	0.0019	3.307508E-03
42	(0.154194, 0.139910, 0.143035)	H	3.5625E-07	2.458502E-03

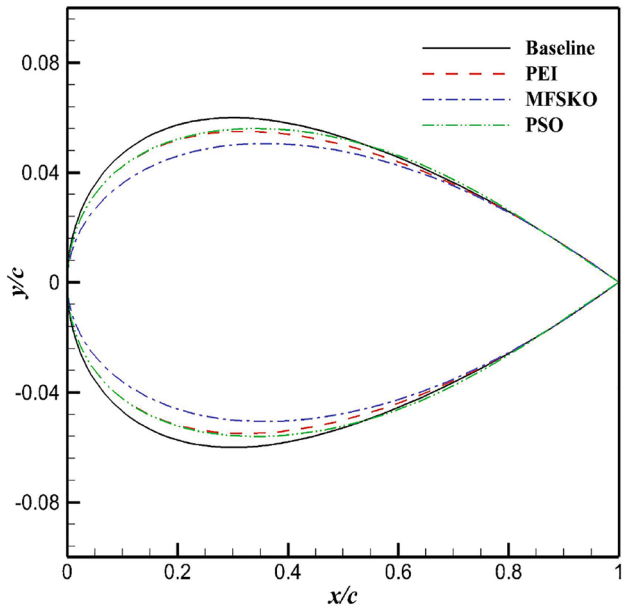


Fig. 14 Optimized front and rear airfoils

Table 6 Comparison of airfoil optimization results with different methods

method	HF samples	LF samples	optimum
PEI	31	45	2.458502E-03
MFSKO	36	50	2.581709E-03

and the maximum thickness of the airfoil moves to the right, which can reduce the drag. In addition, Fig. 15 shows the static pressure contours of the baseline and optimized airfoil. The results show that the pressure recovery of the trailing edge of the airfoil is delayed, the pressure distribution becomes flat, and the shock wave intensity is weakened.

5 Summary and future work

In this paper, a Co-Kriging-based multi-fidelity sequential optimization method is proposed for expensive black-box problem, and the objective of the paper is to develop a more efficient sequential sampling strategy, named proportional expected improvement (PEI). The PEI criterion is an extension of EI criterion and used an integrated criterion to determine both location and fidelity level of the subsequent search.

In the integrated criterion, a proportional factor γ which is adaptively adjusted according to the sample density is added to adjust the tendency between exploration and exploitation during the search process. Meanwhile, Kullback–Leibler divergence is used to measure the credibility

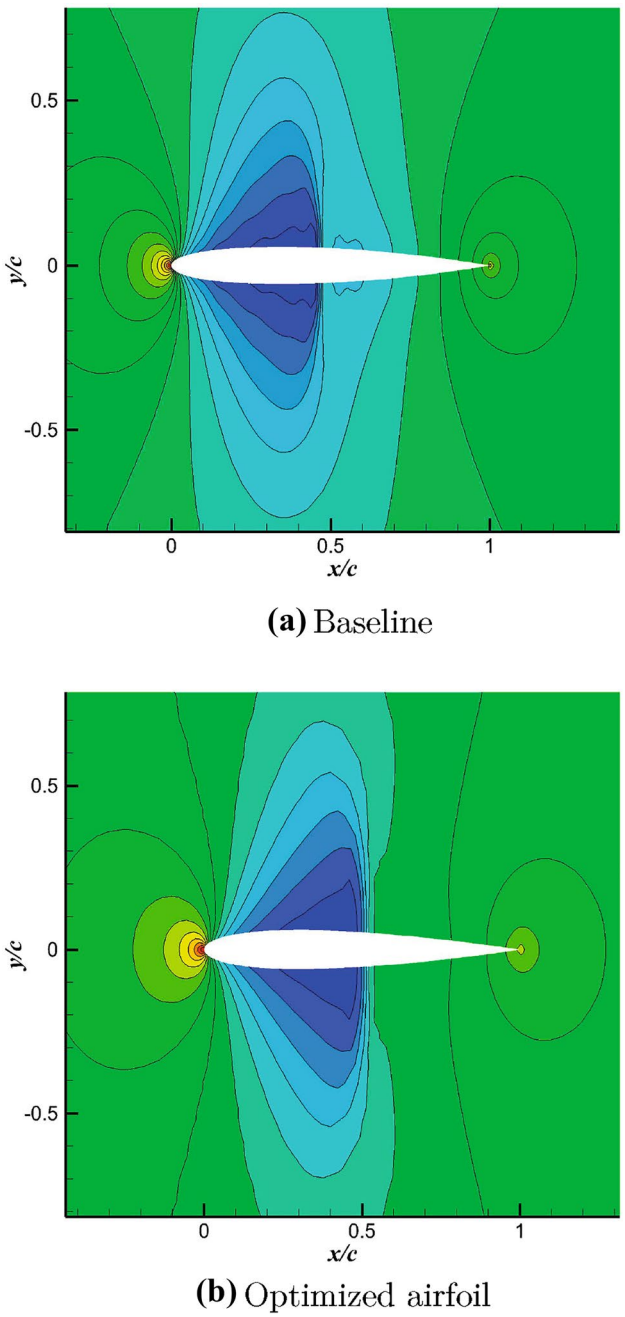


Fig. 15 Static pressure contours of NACA0012 airfoil

or information contribution of a point from system with different fidelities. The PEI method was then validated by application to seven analytical functions, and we found that the introduction of factor γ helps to improve the greedy nature of EI criterion and helps to search for the global optima. However, theoretical analysis shows that a large value of γ may lead to the increase of total cost and sample points may cluster. To solve this problem, we suggest adaptively adjusting γ by the sample density, and it will make a tradeoff between the accuracy of optimum

and the numbers of iterations. The contrast test for five analytical functions also shows that adaptively adjusting the value of γ can achieve high accuracy and considerable cost. Besides, when the accuracy of two methods is the same, PEI can save the total cost, which means that PEI is a more efficient method.

Moreover, a Kullback–Leibler divergence-based item is used to measure the credibility of point from different fidelities, and it decides the fidelity level of the next observation along with the cost item. The effectiveness and advantage of the proposed method were compared with MFSKO method by seven analytical functions, and also demonstrated for aerodynamic shape optimization of NACA0012 airfoil.

The research in this paper is based on the existing Co-Kriging method, and the currently widely used multi-fidelity surrogate model is mostly based on autoregressive architecture. This kind of treatment is relatively easy to handle, but if the autoregression hypothesis does not meet the actual situation, it will result in incorrect results. As the surrogate model may be more important to the optimization results than the sequential criterion, thus how to describe the fidelity difference and the consequently modeling problem is a creative potential. In this paper, Case 7 is an three-fidelity optimal problem, and the computational cost of building a Co-Kriging model increases as the sample size increases. So, the more advanced metamodel is worth studying. In the PEI sequential criterion proposed in this paper, an adaptive proportional factor is introduced to improve greedy and to achieve better optimization effect, which is also applicable to single-fidelity case, but how to adjust the proportional factor adaptively to achieve high accuracy and considerable cost remains to be further studied. And we will also focus on applying the PEI method to higher dimensional optimization, it will be challenging as the dimension is larger than 50. On the other hand, through in this paper, we can handle some kind of unknown constraints, it is still an essential direction.

Appendix A

Theorem 1 (Equivalence theorem about exploration and exploitation) *The PEI function defined by* (17) *is equivalent to the following form*:

$$EI_{\mathrm{P}}(\mathbf{x}) = \underbrace{(\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x}))\Phi(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})})}_{\exp\,\mathrm{loi}tation} + \underbrace{\gamma s_m(\mathbf{x})\phi(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})})}_{\exp\,loration}$$

and $\mathrm{sgn}[(\lambda-1)(\gamma-1)] \geq 0$, *in which* γ *is the proportional factor of exploration and exploitation*, *and* λ *is the scaling of variance*.

Proof Similar to the work of Johns et al. (1998), $EI_{\mathrm{P}}(\mathbf{x})$ and $EI(\mathbf{x})$ are treated as the function of $\hat{f}_m(\mathbf{x})$ and $s_m(\mathbf{x})$, that is to record $EI_{\mathrm{P}}(\mathbf{x}) = EI_{\mathrm{P}}(\hat{f}_m(\mathbf{x}), s_m(\mathbf{x}))$. Obviously, EI function is the special case of PEI, that is

$$\begin{aligned} EI_{\mathrm{P}}(\mathbf{x}) &= EI_{\mathrm{P}}(\hat{f}_m(\mathbf{x}), s_m(\mathbf{x})) = EI(\hat{f}_m(\mathbf{x}), \lambda s_m(\mathbf{x})) \\ &= EI\big(\hat{f}_m(\mathbf{x}), s_m(\mathbf{x}) + (\lambda-1)s_m(\mathbf{x})\big) \end{aligned}$$

We get the partial derivative of prediction mean and prediction variance $s_m(\mathbf{x})$, and we use the binary Lagrange mean value theorem

$$f(x_0, y_0 + k) = f(x_0, y_0) + k \cdot \nabla f_y(x_0, y_0 + \theta k)$$

In which $0 < \theta < 1$, ∇EI_{s_m} is the partial derivative of EI to $s_m(\mathbf{x})$. To further simplify, as ϕ and Φ are the standard normal probability density and cumulative distribution function, based on the partial derivative relationship, we have

$$\begin{aligned} &EI_{\mathrm{P}}(\hat{f}_m(\mathbf{x}), s_m(\mathbf{x})) \\ &= (\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x}))\Phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(x)}\right) \\ &\quad + s_m(\mathbf{x})\phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &\quad + \phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x}) + \theta(\lambda-1)s_m(\mathbf{x})}\right) \cdot (\lambda-1)s_m(\mathbf{x}) \\ &= (\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x}))\Phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})}\right) \\ &\quad + \hat{\gamma}(\mathbf{x})s_m(\mathbf{x})\phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})}\right) \end{aligned} \tag{A1}$$

In which

$$\hat{\gamma}(\mathbf{x}) = 1 + \frac{\phi(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x}) + \theta(\lambda-1)s_m(x)})}{\phi(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{\lambda s_m(\mathbf{x})})}(\lambda-1)$$

As $\frac{\phi(\frac{\hat{f}_m(\mathbf{x}^*)-\hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})+\theta(\lambda-1)s_m(\mathbf{x})})}{\phi(\frac{\hat{f}_m(\mathbf{x}^*)-\hat{f}_m(\mathbf{x})}{\lambda s_m(\mathbf{x})})} > 0$, $\frac{\hat{\gamma}(\mathbf{x})-1}{\lambda(\mathbf{x})-1} > 0$, thus we have $\mathrm{sgn}((\lambda(\mathbf{x})-1)(\hat{\gamma}(\mathbf{x})-1)) > 0$. If $\lambda = 1, \hat{\gamma} = 1$, PEI is just the standard EI function, $\mathrm{sgn}((\lambda(\mathbf{x})-1)(\hat{\gamma}(\mathbf{x})-1)) = 0$. To sum up, for any $\lambda(\mathbf{x})$, there exist a $\hat{\gamma}(\mathbf{x})$ to meet (A1) and $\mathrm{sgn}((\lambda(\mathbf{x})-1)(\hat{\gamma}(\mathbf{x})-1)) \geq 0$

Thus, for any point $\mathbf{x}$, after get its $\hat{f}_m(\mathbf{x})$ and $s_m(\mathbf{x})$, we can surely get a $\lambda(\mathbf{x})$ and after scaling of $s_m(\mathbf{x})$ to make $\hat{\gamma}(\mathbf{x})$ the same everywhere, we record it as γ and have

$$EI_{\mathrm{P}}(\mathbf{x}) = (\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x}))\Phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})}\right) + \gamma s_m(\mathbf{x})\phi\left(\frac{\hat{f}_m(\mathbf{x}^*) - \hat{f}_m(\mathbf{x})}{s_m(\mathbf{x})}\right)$$

□

Appendix B

See Table 7.

Table 7 samples number of the former five test functions

	$C_m/C_l = 4$		$C_m/C_l = 20$	
	MFSKO	PEI	MFSKO	PEI
Case 1	15+9	11+10	23+8	17+7
Case 2	12+7	12+8	17+7	14+7
Case 3	7+8	8+7	8+6	13+8
Case 4	28+12	28+10	63+8	27+10
Case 5	49+14	28+26	44+9	78+15

Note: this table is the supplement of Table 4 indicating the number of samples of high and low fidelity. For example, 15+9 represents 15 low-fidelity samples and 9 high-fidelity samples

Acknowledgements This work is supported by the Fundamental Research Funds for the Central Universities of China (No. 2021qntd03) and National Natural Science Foundation of China (No. 12201656).

Funding This research received no external funding.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

Replication of results The MATLAB codes used to generate results are available upon request.

References

Bull AD (2011) Convergence rates of efficient global optimization algorithms. J Mach Learn Res 12:2879–2904

Chen RB, Wang W, Wu C-FJ (2017) Sequential designs based on Bayesian uncertainty quantification in sparse representation surrogate modeling. Technometrics 59(2):139–152

Chen Z, Mak S, Wu C-FJ (2019) A hierarchical expected improvement method for Bayesian optimization. arXiv:1911.07285

Duan Y, Cai J, Li Y (2012) Gappy proper orthogonal decomposition-based two-step optimization for airfoil design. AIAA J 50(4):968–971

Duan Y, Wu W, Zhang P, Tong F, Fan Z, Zhou G, Luo J (2019) Performance improvement of optimization solutions by POD-based data mining. Chin J Aeronaut 32(4):826–838

Fernández-Godino MG, Park C, Kim N-H, Haftka RT (2016) Review of multi-fidelity models. arXiv:1609.07196

Frazier PI, Savas WD, Dayanik S (2008) A Knowledge-Gradient Policy for sequential information collection. SIAM J Control Optim 47(5):2410–2439

Greenhill S, Rana S, Gupta S, Vellanki P, Venkatesh S (2020) Bayesian optimization for adaptive experimental design: a review. IEEE Access, pp 13937–13948

Gultekin S, Paisley J (2017) Nonlinear Kalman filtering with divergence minimization. IEEE Trans Signal Process 65(23):6319–6331

Han Z-H, Stefan G (2012) Hierarchical kriging model for variable-fidelity surrogate modeling. AIAA J 50(9):1885–1896

He X, Tuo R, Wu CFJ (2017) Optimization of multi-fidelity computer experiments via the EQIE criterion. Technometrics 59(1):58–68

Holland JH (1975) Adaptation in natural and artificial systems. The University of Michigan Press, Michigan

Higdon D, Kennedy M, Cavendish JC, Cafeo JA, Ryne RD (2004) Combining field data and computer simulation for calibration and prediction. SIAM J Sci Comput 26:448–466

Huang D, Allen TT, Notz WI, Zeng N (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models. J Glob Optim 34(3):441–466

Huang D, Allen TT, Notz WI, Miller RA (2006) Sequential kriging optimization using multiple-fidelity evaluations. Struct Multidisc Optim 32(5):369–382

Ivanov TD, Simonović AM, Svorcan JS, Peković OM (2017) VAWT optimization using genetic algorithm and CST airfoil parameterization. FME Trans 45(1):26–31

Jones DR (2001) A taxonomy of global optimization methods based on response surfaces. J Glob Optim 21(4):345–383

Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. J Glob Optim 13(4):455–492

Jin R, Chen W, Sudjianto A (2002) On sequential sampling for global metamodeling in engineering design. ASME international design engineering technical conferences and computers and information in engineering conference, Montreal, Canada

Kennedy J, Eberhart R (1995) Particle swarm optimization. Proc IEEE Int Conf Neural Netw 4:1942–1948

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

Kim v, Lee S, Yee K, Rhee D (2018) High-to-low initial sample ratio of hierarchical kriging for film hole array optimization. J Propul Power 34(1):108–115

Le Gratiet L (2013) Bayesian analysis of hierarchical multi-fidelity codes. SIAM/ASA J Uncertain Quant 1:1244–269

Le Gratiet L, Garnier J (2014) Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. Int J Uncertain Quant 4(5):365–386

Le Gratiet L, Cannamela C (2015) Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. Technometrics 57(3):418–427

Loïc B, Balesdent M, Hebbal A (2020) Overview of Gaussian process based multi-fidelity techniques with variable relationship between fidelities, application to aerospace systems. Aerospace Sci Technol 1:107106339

Liu Y, Chen S, Wang F, Xiong F (2018) Sequential optimization using multi-level cokriging and extended expected improvement criterion. Struct Multidisc Optim 58(3):1155–1173

Michael S (1987) Large sample properties of simulations using Latin hypercube sampling. Technometrics 29(2):143–151

Qin C, Klabjan D, Russo D (2017) Improving the expected improvement algorithm. In: 31st conference on neural information processing systems, Long Beach, CA, USA

Qian PZG, Wu CFJ (2008) Bayesian hierarchical modeling for integrating low-accuracy and high accuracy experiments. Technometrics 50:192–204

Reese CS, Wilson AG, Hamada M, Martz HF, Ryan KJ (2004) Integrated analysis of computer and physical experiments. Technometrics 46:153–164

Ruan X, Jiang P, Zhou Q, Hu J, Shu L (2020) Variable-fidelity probability of improvement method for efficient global optimization of expensive black-box problems. Struct Multidisc Optim 62:3021–3052

Sasena MJ, Papalambros P, Goovaerts P (2002) Exploration of metamodeling sampling criteria for constrained global optimization. Eng Optim 34(3):263–278

Santner TJ, Williams BJ, Notz W, Williams BJ (2018) The design and analysis of computer experiments. Springer, Berlin

Srinivas N, Krause A, Kakade S, Seeger M (2010) Gaussian process optimization in the bandit setting: No regret and experimental design. The 27th international conference on international conference on machine learning, Israel

Shahriari B, Swersky K, Wang Z, Adams RP, Freitas ND (2015) Taking the human out of the loop: a review of Bayesian optimization. Proc IEEE 104:148–175

Shu L, Jiang P, Wang Y (2021) A multi-fidelity Bayesian optimization approach based on the expected further improvement. Struct Multidisc Optim 63(4):1709–1719

Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian optimization of machine learning algorithms. In: Advances in neural information processing systems, pp 2951–2959

Tuo R, Wu CJ, Yu D (2013) Surrogate modeling of computer experiments with different mesh densities. Technometrics 56(3):372–380

Xiao M, Zhang G, Breitkopf P, Villon P, Zhang W (2018) Extended Co-Kriging interpolation method based on multi-fidelity data. Appl Math Comput 323:120–131

Xiong Y, Chen W, Tsui KL (2008) A new variable-fidelity optimization framework based on model fusion and objective-oriented sequential sampling. J Mech Des 130(11):699–708

Xiong F, Xiong Y, Chen W, Yang S (2009) Optimizing Latin hypercube design for sequential sampling of computer experiments. Eng Optim 41(8):793–810

Yu Z, Han Z-H, Zhang K (2018) Variable-fidelity expected improvement method for efficient global optimization of expensive functions. Struct Multidisc Optim 58(4):1431–1451