A modified hypervolume based expected improvement for multi-objective efficient global optimization method

Zheng Li[1] · Xinyu Wang[2] · Shilun Ruan[1] · Zhaojun Li[3] · Changyu Shen[1] · Yan Zeng[1]

Abstract

The hypervolume indicator has been proved as an outstanding metric for the distribution of Pareto points, and the derived hypervolume based expected improvement (HVEI) has received a particular attention in the multi-objective efficient global optimization (EGO) method. However, the high computational cost has become the bottle neck which limits the application of HVEI on many objective optimization. Aiming at this problem, a modified version of HVEI (MHVEI) is proposed in this paper, which is easier to implement, maintains all the desired properties, and has a much lower computational cost. The theoretical study shows that the new criterion can be considered as a weighted integral form of HVEI, and it prefers the new point with a higher uncertainty compared with HVEI. The numerical tests show that the MHVEI performs similar as HVEI on the lower dimensional problem, and the advantage of MHVEI becomes more obvious as the dimension grows.

Keywords Multi-objective · Efficient global optimization · Pareto points · Hypervolume indicator

Nomenclature

$\prec$ Pareto dominance. $\mathbf{a} \prec \mathbf{b} \iff \forall\, 1 \leq i \leq M : a_i \leq b_i \wedge \exists\, 1 \leq i \leq M : a_i < b_i$, where $\mathbf{a}$ and $\mathbf{b}$ denotes two design points.

$\prec\!\prec$ Strong Pareto dominance. $\mathbf{a} \prec\!\prec \mathbf{b} \iff \forall\, 1 \leq i \leq M : a_i < b_i$.

$[\mathbf{l},\mathbf{u}]$ The region of a hyper box which takes $\mathbf{l}$ and $\mathbf{u}$ as the lower and upper limits. If $\mathbf{l} \not\prec \mathbf{u}$, $[\mathbf{l},\mathbf{u}]=\varnothing$.

$V(\mathbf{A})$ The hyper volume of the region $\mathbf{A}$. Note $V(\mathbf{A})=0$ when $\mathbf{A}=\varnothing$

$\mathbf{P}$ Pareto points

$\mathbf{P}_{true}$ True pareto frontier

$\mathbf{r}$ Reference point

$\mathbf{U}$ Local upper bounds

M Number of sub-objectives

d Number of design variables

n Number of points in $\mathbf{P}$

m Number of points in $\mathbf{U}$

✉ Shilun Ruan
   ruansl@dlut.edu.cn

[1] State Key Laboratory of Structural Analysis for Industrial Equipment, Department of Engineering Mechanics, and International Research Center for Computational Mechanics, Dalian University of Technology, Dalian, Liaoning, China

[2] School of Mechanical Engineering, Dalian University of Technology, Dalian, China

[3] School of Information Management, Dezhou University, Dezhou, Shandong, China

1 Introduction

Modern engineering optimization problems usually adopt time-consuming numerical simulations to evaluate the objectives or constraints. For example, a post-buckling analysis of complex aerospace shells usually needs more than two hours (Hao et al. 2012; Hao et al. 2017). Therefore it is desired to limit the number of function evaluations. This requirement has popularized surrogate-based optimization method (Keshtegar et al. 2017; Hao et al. 2015), which uses the cheap mathematical approximation to replace the expensive true objective in the design process. In this area, the Gaussian process models, such as the Kriging surrogate (Krige 1951) and the statistical radial basis functions (Sóbester et al. 2005), have received a great attention, since they not only capture the unknown functions but also provide the probability estimation. With this merit, the efficient global optimization (EGO) methods adopt the statistical infill criteria to select the optima candidate, such as expected improvement (EI) (Jones et al. 1998; Jones 2001)

and generalized EI (GEI) (Schonlau 1998; Sasena et al. 2002), which are able to avoid missing the global optimum to a large extent. Therefore, the EGO methods have been widely used in the modern engineering.

The real-world design problem usually involves more than one sub-objectives which conflict with on another, and this kind of problem is usually mentioned as the multi-objective optimization (MOO) problem. For the case that there is a lack of knowledge about design requirement, it is desired to present the designer multiple diversiform and competitive Pareto designs (also mentioned as Pareto solutions, Pareto frontier, and Pareto points) to decide the most appropriate one. The Pareto designs are defined as the points for which no other point is better in all sub-objectives. The optimization aiming at obtaining the Pareto designs is called Pareto optimization here. In earlier times, there are two main kinds of Pareto optimization approaches.

The first kind of Pareto optimization approach is mentioned as the single objective based approach in this paper, which transforms the MOO problem into different single objective ones and each of them leads to a Pareto point. For example, the weighted sum method (Messac et al. 2000) and the physical programing method (Messac et al. 2004) balance the multiple sub-objectives into one single objective function in a linear or non-linear way. The normal boundary intersection method (Das and Dennis 2006) and normal constraint method (Messac 2003) utilize a group of evenly distributed points on the utopia plane (formed by the optimal solutions of all sub-objectives) to produce different equality or inequality constraints in single objective problems.

The other kind of Pareto optimization approach is the heuristic intelligence algorithms, such as the genetic algorithm (GA) and particle swarming optimization (PSO), etc. The heuristic algorithms have received a great attention in this area, since they are naturally suitable for obtaining Pareto solutions. Some successful algorithms, such as the non-dominated sorting GA (Deb et al. 2002) and strength Pareto evolutionary algorithm (Kim et al. 2004), have been widely used. However, this kind of approaches typically require a massive amount of objective evaluations. Therefore, some researchers used the surrogate model to substitute a part of real simulations to save computational cost. For example, Goel et al. (2007) performed the multi-objective GA on the response surfaces of all the sub-objectives to obtain approximated Pareto frontier. Li et al. (2006) adopted Kriging model to replace a part of true objectives in the GA, and Kriging predicted error was employed to decide whether the prediction or the actual simulation should be performed to evaluate the design points. Jeong and Obayashi (2008) calculated the EI functions as the fitness values in the GA to find non-dominated solutions, and part of them were evaluated to update the Kriging model. Jie et al. (2016) developed the Kriging assisted PSO approach, in which the Kriging prediction took place of true objective

and a normalized summation of EI values was used to determine the most promising designs to be evaluated by actual simulations.

In last two decades, the EGO method has also been extended for the Pareto optimization, and the study mainly focused on the statistical infill criteria for obtaining Pareto solutions. The ParEGO method was developed by Knowles (2006) whose basic idea is using the augmented Tchebycheff functions with different random vectors to balance the multiple sub-objectives into different single objectives. For each single objective, the Kriging surrogate is established, and the EI function is used to obtain new design point. Similar strategy has been adopted in a heuristic algorithm named MOEA/D by Liu et al. (2007) and Zhang et al. (2010).They established the Kriging surrogate for each sub-objective, and further derived the EI function of the balanced single objective based on the multivariate distribution. The balanced single objective could be a weighted sum function (Liu et al. 2007) or the augmented Tchebycheff function (Zhang et al. 2010). The Euclidean distance based EI (EI-ELU) was developed by Keane (2006), in which the expected central point of the probability distribution (provided by Kriging) in the non-dominated region is calculated. Then, the minimum distance between this central point and the obtain Pareto points is maximized. Zhan et al. (2017) defined the EI values rather than all the obtained Pareto points as the EI-Matrix. Then, based on this matrix, three infill criteria which are based on the summation distance, maximum distance and hypervolume were developed respectively.

One kind of statistical infill criterion is developed based on the hypervolume indicator which was firstly proposed by Zitzler and Thiele (1998) and has been proved as an outstanding metric for the quality of Pareto points distribution (Knowles et al. 2003; Emmerich et al. 2005). This indicator is defined as the volume of the dominated region of the Pareto points and bounded by a reference point. With this definition, the statistical infill criterion hypervolume based EI (HVEI) (Emmerich et al. 2011) selects the point with the maximum expected value of hypervolume growth as the new point. Wagner et al. (2010) compared several Pareto optimization infill criteria (before 2010) on a list of necessary and desired properties, and it is found that the HVEI has the best performance in terms of most properties but with a high computational cost. The computational complexity of HVEI mainly comes from the irregular geometry of the integration (non-dominated) region, and it is approximated with Monte Carlo method in earlier times. Emmerich et al. (2011) described a formula for calculating HVEI, in which the integration region is divided into a number of cells, and the integration in HVEI is performed in each one of them. However, the number of cells increases exponentially with the number of objectives. Couckuyt et al. (2014) developed a fast algorithm for HVEI, which allows a much lower number of integration cells, but no computational complexity was provided. As a rough

approximation of HVEI, the lower confidence bound (LCB) (Emmerich et al. 2006) is defined as the hypervolume improvement of a lower confidence point. Since there is no need for the integral, the LCB has a much lower computational cost than HVEI. However, even the runtime of simply calculating the hypervolume still grows exponentially with the number of objectives (Beumea and Emmerich 2007). Zhan et al. (2017) estimated that the infill criterion function could be called for two millions times in the whole optimization process. If one time evaluation of the infill criterion costs one second, the total evaluating time will be twenty three days which might be even longer than the true simulations. Therefore, although the HVEI shows great potential abilities on generating well distributed Pareto points, the high computational complexity limits its application on many objective problems.

In this paper, a modified HVEI (MHVEI) function is developed, which can be considered as a weighted integral form of HVEI. Compared with the original HVEI, this new statistical criterion is easier to implement, maintains all the desired properties, and has a much lower computational cost. This paper is organized as follows. In Section 2, some basic knowledge about the EGO method and HVEI are introduced. In Section 3, the MHVEI is developed and studied theoretically by comparing it with HVEI on a list of desired properties. In Section 4, the performance of MHVEI is investigated with some mathematical test functions.

2 Related works

In this section, the classic EGO method for the single objective problem and the original HVEI function will be introduced.

2.1 Efficient global optimization

2.1.1 Ordinary kriging (OK) surrogate

The OK surrogate is a widely used Gaussian process model. Considering the objective function $y(\mathbf{x})$ is meant to be minimized where $\mathbf{x} \in \mathbb{X}$ denotes the design variables and $\mathbb{X} \subset \mathbb{R}^d$ denotes the design variable space. The OK surrogate assumes that the unknown $y(\mathbf{x})$ meets the normal distribution for any $\mathbf{x} \in \mathbb{X}$

$$y(\mathbf{x}) \sim \mathcal{N}(\beta, \sigma^2) \tag{1}$$

where β and σ^2 are the assumed values of mean and variance. For any two points $\mathbf{x}_i \in \mathbb{X}$ and $\mathbf{x}_j \in \mathbb{X}$,

$$cov[y(\mathbf{x}_i), y(\mathbf{x}_j)] = \sigma^2 r(\mathbf{x}_i, \mathbf{x}_j, \theta) = \sigma^2 exp\left[\sum_{k=1}^d -\theta_k(x_{i,k} - x_{j,k})^2\right] \tag{2}$$

where r denotes a given correlation model function, and $\theta = \{\theta_i | 1 \leq i \leq d\}$ denotes the correlation parameters on

different dimensions. In (2) the Gaussian correlation model is adopted.

The OK surrogate is established based on a group of evaluated design variables $\mathbf{S} = \{\mathbf{S}_i | \mathbf{S}_i \in \mathbb{X}, 1 \leq i \leq N\}$ with their objective values of $\mathbf{Y} = y(\mathbf{S})$. In this paper, $\mathbf{S}$ and $\mathbf{Y}$ will be mentioned as the training data of OK surrogate, and N denotes the number of training data. If the (1) and (2) are considered as the priori hypotheses, and $\mathbf{Y} = y(\mathbf{S})$ is taken as the known event, the conditional probability distribution of $y(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{X}$ will be derived based on the Bayes theory that

$$[y(\mathbf{x})|y(\mathbf{S}) = \mathbf{Y}] | \sim \mathcal{N}\left[\hat{y}(\mathbf{x}), \hat{\sigma}(\mathbf{x})^2\right] \tag{3}$$

where $\hat{y}(\mathbf{x})$ and $\hat{\sigma}(\mathbf{x})^2$ are the predicted mean and variance functions of OK surrogate separately that

$$\hat{y}(\mathbf{x}) = \frac{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{Y}}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}} + \mathbf{r}(\mathbf{x})^T\mathbf{R}^{-1}\left(\mathbf{Y} - \frac{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{Y}}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}}\mathbf{1}\right) \tag{4}$$

$$\hat{\sigma}(\mathbf{x})^2 = \sigma^2\left\{1 + \frac{\left[1 - \mathbf{1}^T\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})\right]^2}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}} - \mathbf{r}(\mathbf{x})^T\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})\right\} \tag{5}$$

where $\mathbf{r}(\mathbf{x}) = \{r(\mathbf{x}, \mathbf{S}_1, \theta), r(\mathbf{x}, \mathbf{S}_2, \theta), \cdots r(\mathbf{x}, \mathbf{S}_n, \theta)\}^T$ denotes the correlation vector between $\mathbf{x}$ and $\mathbf{S}$, $\mathbf{1}$ denotes the vector with n components of 1, and $\mathbf{R} = [r(\mathbf{S}_i, \mathbf{S}_j, \theta)]_{i,j}$ denotes correlation matrix of $\mathbf{S}$. Based on (3), the probability distribution function (PDF) provided by OK surrogate can be written as

$$p[y(\mathbf{x})] = \phi\left[\frac{y(\mathbf{x}) - \hat{y}(\mathbf{x})}{\hat{\sigma}(\mathbf{x})}\right] \tag{6}$$

where ϕ denotes the probability density function of standard normal distribution.

The parameters of OK model including β, σ^2 and θ are usually estimated by maximizing the likelihood function that

$$\max_{\beta, \sigma^2, \theta} p[y(\mathbf{S}) = \mathbf{Y}] = \frac{1}{\sqrt{(2\pi)^n|\mathbf{R}|}} exp\left[\frac{-1}{2\sigma^2}(\mathbf{Y} - \beta\mathbf{1})^T\mathbf{R}^{-1}(\mathbf{Y} - \beta\mathbf{1})\right] \tag{7}$$

More details about the derivation of OK surrogate can be found in the literature (Ginsbourger et al. 2010) which will not be further introduced.

2.1.2 Generalized EI function

During the EGO process, the best obtained objective design is selected from the training data that $Ymin = \min\{\mathbf{Y}\}$. Then, for any unevaluated design $\mathbf{x}$, the improvement rather than $Ymin$ is defined as

$$I(\mathbf{x}) = \max\{0, Ymin - y(\mathbf{x})\} \tag{8}$$

The probability distribution of $I(\mathbf{x})$ can be obtained from (6). Then, the EI method takes the design with the maximal expected value of $I(\mathbf{x})$ as the infilled point

$$EI[y(\mathbf{x}), Ymin] = E[I(\mathbf{x})] = \int_0^\infty Ip(I)dI = u(\mathbf{x})\Phi\left[\frac{u(\mathbf{x})}{\hat{\sigma}(\mathbf{x})}\right] + \sigma(\mathbf{x})\phi\left[\frac{u(\mathbf{x})}{\hat{\sigma}(\mathbf{x})}\right] \tag{9}$$

where $u(\mathbf{x}) = Ymin - \hat{y}(\mathbf{x})$ and Φ denotes the cumulative distribution function of standard normal distribution.

Note that $\hat{\sigma}^2$ is the expected difference between y and $\hat{y}$ that

$$\hat{\sigma}^2 = E\left[\left(y - \hat{y}\right)^2\right] \tag{10}$$

therefore $\hat{\sigma}(\mathbf{x})$ is also mentioned as uncertainty, and used to estimate the accuracy of $\hat{y}(\mathbf{x})$. The EI function prefers to select the point with a higher $\hat{\sigma}$ or a lower $\hat{y}$. The former preference is usually mentioned as global exploration which is meant to improve the accuracy of the surrogate, and the latter one is named as local exploitation which is meant to search the optimum of the surrogate. These two searching preferences can avoid missing the global optimum to a large extent, and they can be further balanced in the GEI function (Schonlau 1998; Sasena et al. 2002) as

$$E[I^g(\mathbf{x})] = \int_0^\infty I^g p(I)dI = \int_0^\infty I^{g-1} Ip(I)I \tag{11}$$

where the positive integer g is used to control the balance. If the I^{g-1} is considered as the integral weight coefficient of the improvement, one can observe that there is a trade-off in choosing between small improvement with large probability (local exploitation) versus large improvement with small probability (global exploration) (Schonlau 1998). As g grows, the balance will incline to global exploration which is more suitable when the unknown objective is poorly approximated (Sasena et al. 2002), otherwise it will incline to local exploitation.

2.2 Hypervolume based expected improvement

Considering there are M objectives $\mathbf{y}(\mathbf{x}) = \{y_1(\mathbf{x}), y_2(\mathbf{x}), \cdots y_M(\mathbf{x})\}$ to be minimized, and all the sub-objectives can be evaluated by one time simulation. In most cases, the dependency between different sub-objective functions is not available, therefore a standard assumption is usually made that all the sub-objectives are independent to one another (Emmerich et al. 2006). One can use the Kriging surrogate to estimate the probability distribution of each sub-objective separately, and the PDF of $\mathbf{y}(\mathbf{x})$ is expressed as

$$p[\mathbf{y}(\mathbf{x})] = \prod_{i=1}^{M}\phi\left[\frac{y_i(\mathbf{x}) - \hat{y}_i(\mathbf{x})}{\hat{\sigma}_i(\mathbf{x})}\right] \tag{12}$$

In each iteration of the EGO process for MOO, the current Pareto points $\mathbf{P} = \{\mathbf{P}_i | 1 \leq i \leq n\}$ are selected from the training data $\mathbf{Y}$. The hypervolume indicator is an outstanding metric for the quality of Pareto points distribution, which is defined as the volume of the region dominated by $\mathbf{P}$ and bounded by a reference point $\mathbf{r} \in \mathbb{R}^M$. It can be mathematically expressed as

$$H(\mathbf{P}) = \int_{\mathbf{a} \prec \mathbf{r} \wedge \exists \mathbf{P}_i \in \mathbf{P}: \mathbf{P}_i \prec \mathbf{a}} d\mathbf{a} \tag{13}$$

The reference point $\mathbf{r}$ should be set with a relative large value to ensure that it is dominated by any true Pareto point. A higher value of hypervolume indicator means a better distribution of Pareto points, therefore the problem of Pareto optimization can be transformed to maximizing $H(\mathbf{P})$. Then, for a given point $\mathbf{y} \in \Omega$, when it is added to the training data, the hypervolume improvement is defined as the growth of $H(\mathbf{P})$ that

$$I_{hv}(\mathbf{y}) = H(\mathbf{P} \cup \mathbf{y}) - H(\mathbf{P}) \tag{14}$$

Note that $I_{hv}(\mathbf{y}) \geq 0$, and $I_{hv}(\mathbf{y}) > 0$ when and only when $\mathbf{y}$ is not strongly dominated by any Pareto point in $\mathbf{P}$. As shown in Fig. 1, $I_{hv}(\mathbf{y})$ is the area of the shaded part. Similar to the classic EI, the HVEI method takes the point with the maximum expected value of $I_{hv}(\mathbf{y})$ as the new point that

$$EI_{hv}(\mathbf{x}) = E[I_{hv}(\mathbf{y}(\mathbf{x}))] = \int_{\mathbf{y} \in \Omega} I_{hv}(\mathbf{y})p(\mathbf{y})d\mathbf{y} \tag{15}$$

where Ω denotes the non-dominated region

$$\Omega = \left\{\mathbf{a} \in \mathbb{R}^M | \mathbf{a} \prec \mathbf{r} \wedge \nexists \mathbf{P}_i \in \mathbf{P}: \mathbf{P}_i \prec \mathbf{a}\right\} \tag{16}$$

The HVEI is monotonically increasing with increasing $\hat{\sigma}_i$ ($i = 1, 2, \cdots M$), therefore it preserves the global exploration of EI method to improve the accuracy of surrogate. And when the accuracy of the surrogate is high (which means $\hat{\sigma}_i$ is low),
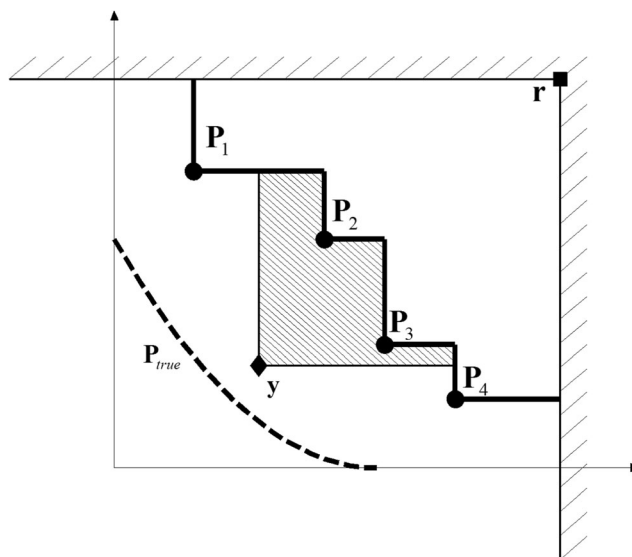

Fig. 1 Reference point, Pareto points and hypervolume improvement

the $EI_{hv}(\mathbf{x})$ will regress to $I_{hv}(\mathbf{y})$ which is directly defined to maximize the hypervolume. Therefore, HVEI is a very effective statistical infill criterion for producing well-distributed Pareto points.

Unfortunately, because of the irregular shape of Ω, there is a lack of direct calculating formula of both $I_{hv}(\mathbf{y})$ and $EI_{hv}(\mathbf{x})$ when $M > 2$ (a formula of HVEI for bi-objective problem is developed in Section 3.2), and the high computational complexity of HVEI (which will be further introduced in Section 3.4) has limited its application on the many objective problems.

3 Modified hypervolume based expected improvement

In this section, a new statistical criterion for Pareto optimization will be developed. In order to make it clear, the related concepts are given first in Section 3.1. In Section 3.2, a modified version of hyper volume improvement is proposed and compared with the original one, and then the MHVEI is derived and the whole optimization process is introduced in Section 3.3. In Section 3.4, the properties of MHVEI will be discussed theoretically.

3.1 Related concepts

- Local upper bounds (LUBs) $\mathbf{U}$: $\mathbf{U} \subset \mathbb{R}^M$ denotes a set of points which cannot be dominated by one another that $\forall \mathbf{U}_i \in \mathbf{U}, \forall \mathbf{U}_j \in \mathbf{U} \Rightarrow \mathbf{U}_i \nprec \mathbf{U}_j$. And it can be used to define the non-dominated region that

$$\Omega = \left\{ \mathbf{a} \in \mathbb{R}^M \middle| \exists \mathbf{U}_i \in \mathbf{U} : \mathbf{a} \prec\prec \mathbf{U}_i \right\} \quad (17)$$

- Grid cells: Using the j_{th} component of each $\mathbf{P}_i$ to divide the j_{th} dimension of $\mathbb{R}^M$ into $n+1$ intervals, then all the intervals on different dimensions will form $(n+1)^M$ hyper boxes. All the hyper boxes whose lower bounds are inside Ω are defined as the grid cells. In this paper, $\mathbf{l}_i = (l_{i,1}, l_{i,2}, \cdots l_{i,M})$ and $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \cdots u_{i,M})$ are used to denote the lower and upper bounds of the i_{th} grid cell, where $l_{i,j}$ and $u_{i,j}$ denote the bounds corresponding to the j_{th} sub-objective function.

More details of the LUBs can be found in the work of Klamroth et al. (2015) which will not be further introduced. The grid cells have been adopted by Emmerich et al. (2011) to calculate HVEI. Figure 2 shows the relationship between $\mathbf{P}$ and $\mathbf{U}$, and the grid cells are denoted by the dotted lines. Note, every LUB is the upper limit of a grid cell.

3.2 Modified hypervolume improvement function

In EGO method, the improvement function should be defined based on the design purpose. For example, maximizing $I(\mathbf{x})$ is

the same to minimizing the single objective function, and maximizing I_{hv} is the same to maximizing the hypervolume. In other words, the improvement function for the MOO problem should be able to produce well distributed Pareto points, when it is used as the single objective function balancing the multiple sub-objectives (like the methods weighted sum and physical programming, see Section 1). Three desired properties of the improvement function for Pareto optimization are listed as follows:

Porperty 1.   The improvement function should be monotonically decreasing with increasing $y_j(j = 1, 2, \cdots M)$.

Porperty 2.   When $\mathbf{P} \subset \mathbf{P}_{true}$, the improvement function should prefer the point which mostly augments the current Pareto frontier.

Porperty 3.   A simple and integrable formula of the improvement function is desired.

Among these properties, Property 1 can ensure the maximum point of improvement function will always locate on $\mathbf{P}_{true}$, Property 2 relates to the quality of the Pareto points distribution, and Property 3 is meant to ensure that the direct calculating formula of the statistical infill criterion exists.

Because $\max I_{hv}(\mathbf{y}) \Longleftrightarrow \max H(\mathbf{P} \cup \mathbf{y})$, the I_{hv} satisfies the first two properties perfectly but fails on the last one. By adopting the concept of the LUBs, the $I_{hv}(\mathbf{y})$ can be written in the form of

$$I_{hv}(\mathbf{y}) = \sum_{i=1}^{m} V([\mathbf{y}, \mathbf{U}_i]) - V_{overlap}(\mathbf{y}) \quad (18)$$

where the first term has considered the volume of the whole region dominated by $\mathbf{y}$, and the $V_{overlap}$ denotes the volume of the overlapped part in the first item. Note, for a given $\mathbf{U}_i \in \mathbf{U}$, if $\mathbf{y} \nprec \mathbf{U}_i$, $[\mathbf{y}, \mathbf{U}_i] = \varnothing$ and $V([\mathbf{y}, \mathbf{U}_i]) = 0$. As shown in Fig. 2, the area with different shaded lines denotes the region of different $[\mathbf{y}, \mathbf{U}_i]$ in the first term, and the overlapped shaded area relates to the
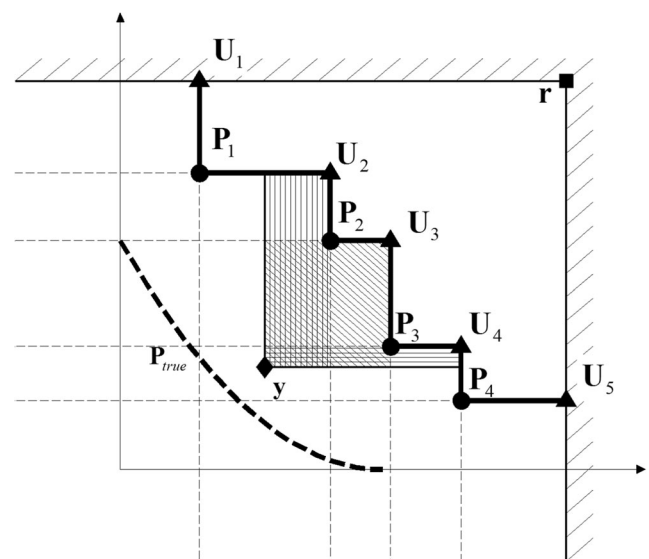

Fig. 2 LUBs, grid cells and modified hypervolume improvement

value of $V_{overlap}$.

For the bi-objective problem, every Pareto point dominates two LUBs, then $V_{overlap}(\mathbf{y})$ can be calculated as

$$V_{overlap}(\mathbf{y}) = \sum_{i=1}^{n} V([\mathbf{y}, \mathbf{P}_i]) \tag{19}$$

Substituting (18) and (19) into (15), the formula of HVEI can be written as

$$EI_{hv}(\mathbf{x}) = \sum_{i=1}^{m}\left[\prod_{j=1}^{M} EI\left(y_j, U_{i,j}\right)\right] - \sum_{i=1}^{n}\left[\prod_{j=1}^{M} EI\left(y_j, P_{i,j}\right)\right] \tag{20}$$

Unfortunately, (19) dose not hold true when $M > 2$, therefore the HVEI still lacks a calculating formula. However, if we defined the first term of (18) as an improvement function, it will preserve all the three properties above. This new function is named as the modified version of hypervolume improvement which is given as

$$I_{hv}^{mod}(\mathbf{y}) = \sum_{i=1}^{m} V([\mathbf{y}, \mathbf{U}_i]) \tag{21}$$

and the three properties of I_{hv}^{mod} will be discussed in the following.

Obviously, I_{hv}^{mod} satisfies Property 1 and Property 3 based on its definition, see (21). In order to discuss Property 2 of I_{hv}^{mod}, the difference between I_{hv} and I_{hv}^{mod} will be first analyzed by adopting the concept of grid cells. In the i_{th} grid cell, the part dominated by $\mathbf{y}$ can be expressed as

$$[\mathbf{y}, \mathbf{u}_i] \cap [\mathbf{l}_i, \mathbf{u}_i] = [\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i] \tag{22}$$

where $\max\{\mathbf{y}, \mathbf{l}_i\} = (\max\{y_1, l_{i, 1}\}, \max\{y_2, l_{i, 2}\}, \cdots \max\{y_M, l_{i, M}\})$. Then $I_{hv}(\mathbf{y})$ can be expressed as the summation of dominated region in all the grid cells

$$I_{hv}(\mathbf{y}) = \sum_i V\left([\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i]\right) \tag{23}$$

The intersection of $[\mathbf{y}, \mathbf{U}_j]$ and the i_{th} grid cell can be written as

$$[\mathbf{y}, \mathbf{U}_j] \cap [\mathbf{l}_i, \mathbf{u}_i] = \begin{cases} [\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i], & \text{if } \mathbf{l}_i \prec\prec \mathbf{U}_j \\ \varnothing & , \text{else} \end{cases} \tag{24}$$

then similar to (23), $I_{hv}^{mod}(\mathbf{y})$ can be rewritten as

$$\begin{aligned} I_{hv}^{mod}(\mathbf{y}) &= \sum_{j=1}^{m} V\left([\mathbf{y}, \mathbf{U}_j]\right) \\ &= \sum_{j=1}^{m} \sum_{i}^{\mathbf{l}_i \prec\prec \mathbf{U}_j} V\left([\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i]\right) \\ &= \sum_i g_i V\left([\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i]\right) \end{aligned} \tag{25}$$

where g_i denotes the number of LUBs strongly dominated by $\mathbf{l}_i$. By comparing (23) and (25), we can get that when $\mathbf{y}$ dominates more LUBs, the difference between $I_{hv}(\mathbf{y})$ and $I_{hv}^{mod}(\mathbf{y})$ will become more obvious. In the contrary, if $\mathbf{y}$ dominates only one LUB, $I_{hv}(\mathbf{y}) = I_{hv}^{mod}(\mathbf{y})$. It implies that $I_{hv}(\mathbf{y})$ and $I_{hv}^{mod}(\mathbf{y})$ will have a more similar landscape when $\mathbf{y}$ is closer to the LUBs.

When $\mathbf{P} \subset \mathbf{P}_{true}$, all the LUBs will be close to $\mathbf{P}_{true}$, then $I_{hv}^{mod}(\mathbf{y})$ will have a similar ability of augmenting Pareto frontier as $I_{hv}(\mathbf{y})$ because of the similar landscape. Note, when $\mathbf{P} \subset \mathbf{P}_{true}$ and $M = 2$, any point $\mathbf{y} \in \mathbf{P}_{true}$ can strongly dominate only one (at most) LUB, therefore $I_{hv}(\mathbf{y}) = I_{hv}^{mod}(\mathbf{y})$. It means that I_{hv}^{mod} will work the same as I_{hv} on augmenting the Pareto frontier for the bi-objective case.

Generally speaking, I_{hv}^{mod} shows better than I_{hv} on Property 3. And based on the discussion above, I_{hv}^{mod} can also produce well distributed Pareto points (Property 1 and Property 2). The quality of Pareto points obtained by I_{hv}^{mod} and I_{hv} will be compared with numerical tests in Section 4.1.

3.3 Optimization with MHVEI

Based on (17) and (21), $I_{hv}^{mod}(\mathbf{y}) \geq 0$, and $I_{hv}^{mod}(\mathbf{y}) > 0$ when and only when $\mathbf{y} \in \Omega$. Then similar to the HVEI, the expected value of $I_{hv}^{mod}(\mathbf{y})$ is defined as the MHVEI that

$$EI_{hv}^{mod}(\mathbf{x}) = E\left[I_{hv}^{mod}(\mathbf{y}(\mathbf{x}))\right] = \sum_{i=1}^{m} E[V([\mathbf{y}, \mathbf{U}_i])] = \sum_{i=1}^{m}\left[\prod_{j=1}^{M} EI\left(y_j, U_{i,j}\right)\right] \tag{26}$$

where $EI(y_j, U_{i, j})$ is the EI function, see (9).

The optimization with the MHVEI follows a typical process of EGO method as shown in Fig. 3. At the very beginning, the initial training data is usually generated with Latin Hypercube method
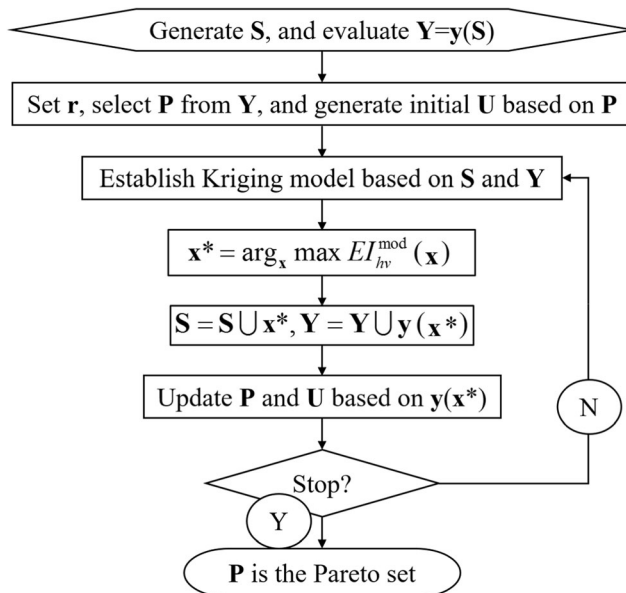

Fig. 3 Optimization process with MHVEI

(Mckay et al. 1979). Before the optimization, the initial LUBs are obtained based on initial **P** and **r**, and they will be updated with new design $\mathbf{y}(\mathbf{x}^*)$ in each iteration. An efficient algorithms for generating and updating the LUBs has been proposed by Przybylski et al. (2010), and further improved by Klamroth et al. (2015). In their tests, updating the LUBs with one new point takes no more than 10^{-5} second when $m = 10000$ and $M = 6$, therefore the computational cost can be ignored. The Matlab code of Przybylski's approach can be found in the Appendix 1.

3.4 Properties of MHVEI

In the following, the performance of MHVEI will be discussed on several desired properties of Pareto infill criterion. All these properties are summarized from the work of Wagner et al. (2010)

- Global exploration and local exploitation

The statistical infill criterion should be monotonically decreasing with increasing $\hat{y}_i$, and monotonically increasing with

increasing $\hat{\sigma}_i$ ($i = 1, 2, \cdots M$). Obviously this property has been preserved in MHVEI, since it is monotonically increasing with increasing $EI(y_j, U_{i,j})$.

In addition, the balance between global exploration and local exploration is the main difference between HVEI and MHVEI. If we substituting (23) and (25) into (15) and (26) separately, we will get

$$EI_{hv}(\mathbf{x}) = \int_{\mathbf{y} \in \Omega} \left[\sum_i V\left([\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i] \right) \right] p(\mathbf{y}) d\mathbf{y}$$

$$= \sum_i \int_{\mathbf{y} \in \Omega} V([\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i]) p(\mathbf{y}) d\mathbf{y} \qquad (27)$$

$$EI_{hv}^{mod}(\mathbf{x}) = \int_{\mathbf{y} \in \Omega} \left[\sum_i g_i V\left([\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i] \right) \right] p(\mathbf{y}) d\mathbf{y}$$

$$= \sum_i g_i \int_{\mathbf{y} \in \Omega} V([\max\{\mathbf{y}, \mathbf{l}_i\}, \mathbf{u}_i]) p(\mathbf{y}) d\mathbf{y} \qquad (28)$$

If g_i is considered as the weight coefficient, EI_{hv}^{mod} will be the weighted integral form of EI_{hv}, and EI_{hv}^{mod} will prefer the region which dominates more LUBs comparing with EI_{hv}. Usually, if $\mathbf{y}$ dominates more LUBs, it will have a higher $I_{hv}(\mathbf{y})$ but with a smaller probability of $p(\mathbf{y})$. Then, based on


Fig. 4 Fitness landscape on bi-objective case
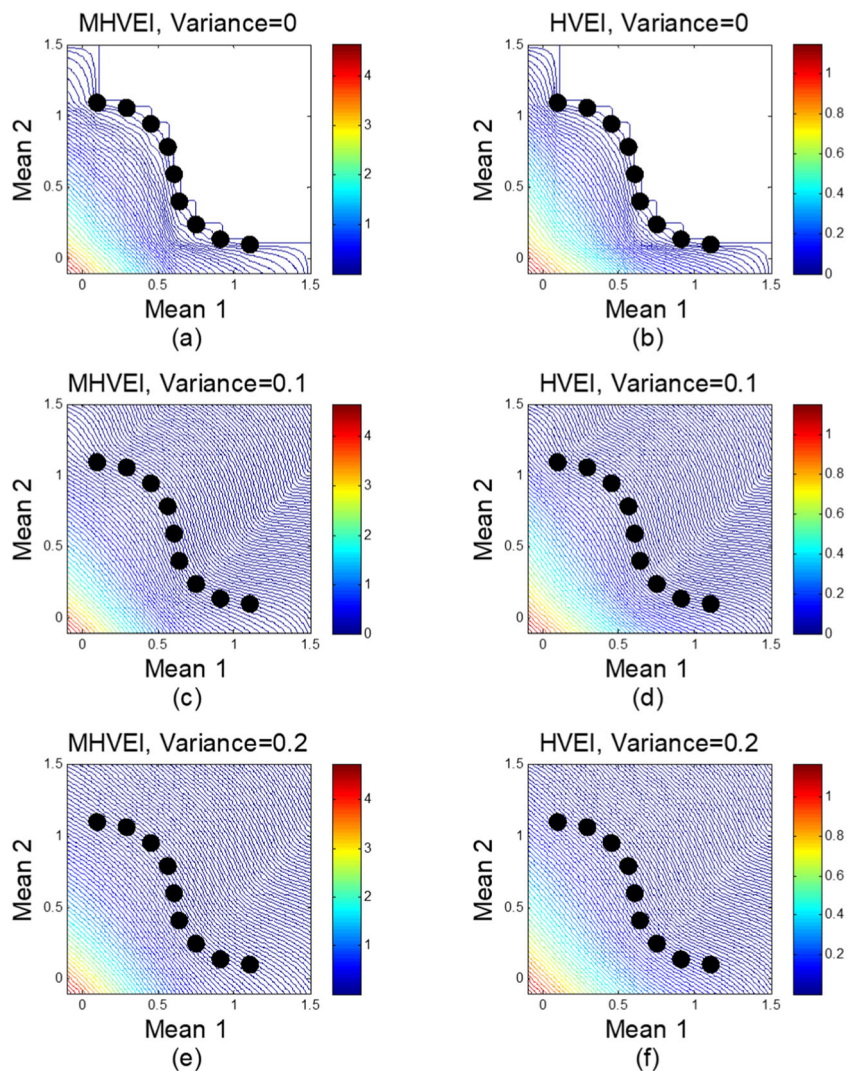
Table 1 Comparison on computational complexity

	$M = 2$	$M = 3$	$M > 3$
HVEI	$O(n)$	$O[n\log(n)]$	$O(M * n^{2\lfloor M/2 \rfloor})$
MHVEI	$O(n)$	$O(n)$	$O(M * n^{\lfloor M/2 \rfloor})$

the discussion above, we can conclude that comparing with HVEI, MHVEI prefers the design point with a higher improvement $I_{hv}(\mathbf{y})$ but a smaller probability of $p(\mathbf{y})$. This difference is similar to the one between GEI functions (Schonlau 1998) with different orders, see Section 2.1. Therefore, we can also conclude that the balance (between the global exploration and local exploitation) in MHVEI more inclines to the global exploration than HVEI.

• Improving the Pareto set distribution when $\hat{\sigma}$ is small

The HVEI shows outstanding advantage on this property, because it is defined to directly maximize the hypervolume. When $\hat{\sigma}$ is small, the statistical infill criterion will regress to the improvement function. As discussed in Section 3.2, the I_{hv}^{mod} has a very similar ability of producing well distributed Pareto points as I_{hv}, therefore MHVEI can show a similar performance as HVEI on improving the Pareto set distribution. In order to demonstrate this, the approach of Wagner et al. (2010) is adopted. Note, both MHVEI and HVEI are functions with respect to $\hat{\mathbf{y}}$ and $\hat{\sigma}$. The ability of statistical infill criterion on improving the distribution depends on the fitness landscapes with respect to varied $\hat{\mathbf{y}}$. Therefore the functions of both criteria
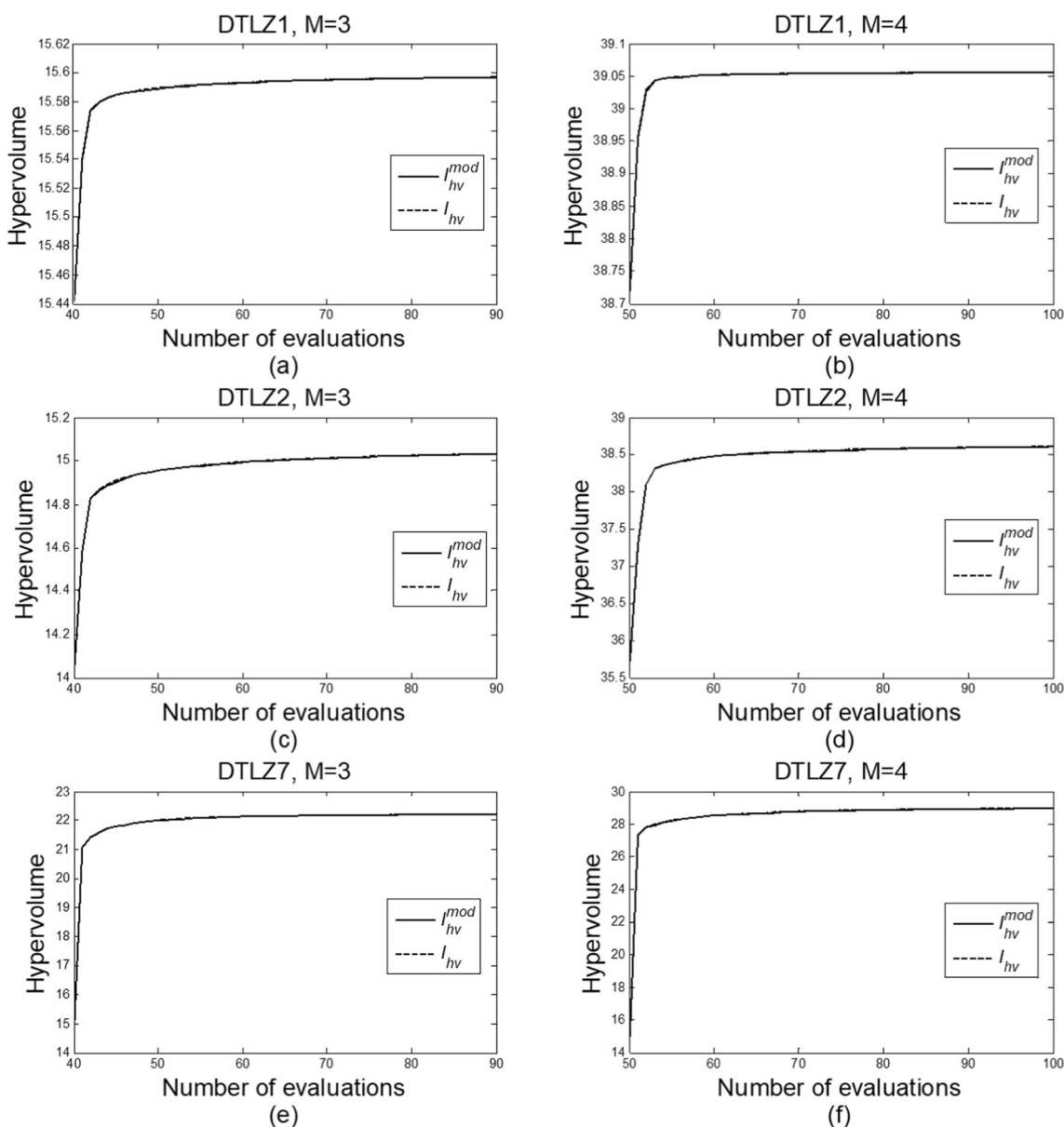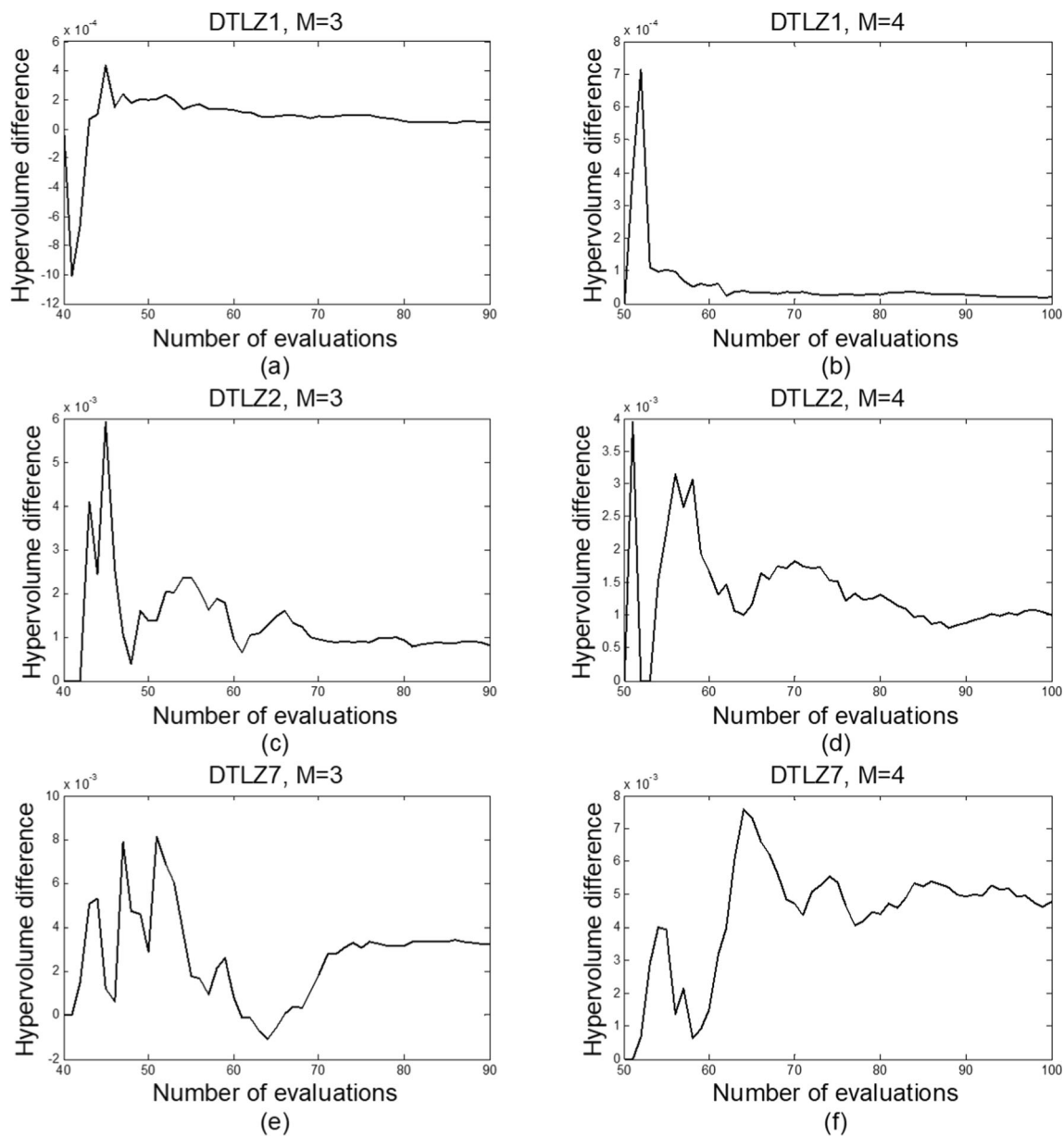
Fig. 5 Comparison of improvement functions

Fig. 6 Hypervolume difference

with respect to varied $\hat{y}$ and constant $\hat{\sigma}$ ($\hat{\sigma} = 0, 0.1$ and 0.2) on bi-objective case are compared in Fig. 4. For each $\hat{\sigma}$, a group of Pareto points with both convex and concave shapes are generated. It can be observed that, although the two criteria refer to very different scale, they have very similar fitness landscapes.

- Easy to search

The infill criterion should be continuous, differentiable and the plateaus of its fitness landscape should be avoided, then the direct optimizer (especially the gradient based methods) can easily reach the local optima of the infill criterion. Clearly the MHVEI is continuous and differentiable since it has an analytical expression. According to Wagner et al. (2010), HVEI

shows strong gradients to their local optima by checking its fitness landscapes, therefore the comparison in Fig. 4 also implies that the MHVEI preserves this property as well as HVEI.

- Easy to implement and efficient to calculate

The MHVEI is easy to implement, since the direct calculating formula has been given in (26). By observing (26), the computing amount of MHVEI is proportional to $M * m$. Clearly, $m = n + 1$ for the bi-objective case, and m is tightly upper bounded by $O(n^{\lfloor M/2 \rfloor})$ when $M > 2$ (Klamroth et al. 2015). Therefore it can be got that the MHVEI has a computational complexity of $O(M * n^{\lfloor M/2 \rfloor})$ in the worst case.
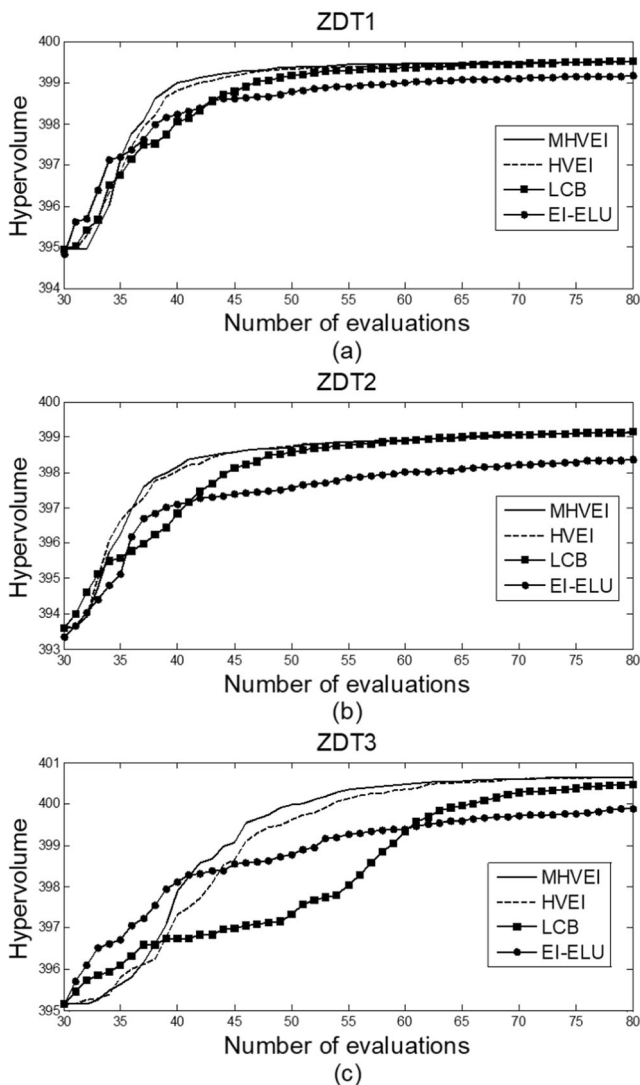
Fig. 7 Comparison on ZDT functions

In Kriging model, predicting $\hat{\sigma}_i$ involves a matrix-vector multiply operation, and the computational complexity is $O(N^2)$ where N denotes the number of training data. Since $\mathbf{P} \subset \mathbf{Y}$, $n \leq N$. Therefore MHVEI has the same computational complexity as the Kriging model when $M \leq 5$ (the computational complexity is $O(n^2)$ when $M = 5$).

The computational complexity of MHVEI are compared with the fastest algorithms (as far as we know) of HVEI in Table 1, where the algorithm for bi-objective case has been developed in Section 3.2, see (20). The algorithms of HVEI for the problem when $M = 3$ and $M > 3$ were developed by Yang et al. (2017) and Couckuyt et al. (2014) separately. Note, Couckuyt et al. (2014) have not provided the computational complexity in their work, but it can be estimated as $O(M * q^2)$ based on their calculating formula, where q denotes the number of multiple disjoint cells covering the whole region of Ω. Based on (17), it can be deduced that m is the smallest number of cells which can cover the whole Ω, therefore $q \geq$

m, and the computational complexity will be $O(M * n^{2\lfloor M/2 \rfloor})$ in the best case.

Generally speaking, the Table 1 shows that MHVEI has an obvious advantage on the computational complexity comparing with HVEI when M is large, but it is still more time-consuming than the other infill criteria which is not developed based on the hypervolume indicator.

4 Numerical tests

The performance of MHVEI will be investigated with mathematical test functions. The hypervolume indicator is used to measure the obtained Pareto solutions, see (13). All the programs used in this paper are coded with Matlab. The Kriging model is established with DACE tool box (Nielsen et al. 2002). The new design point is obtained by a hybrid strategy of PSO algorithm and gradient based method. To be specific, the PSO is used first to obtain a solution, and then this solution and all the current Pareto designs are used as the initial guesses in the "fmincon" function provided by Matlab. The HSO algorithm (While et al. 2005) is adopted to calculate the hypervolume indicator.

Six test functions from previous literatures are used in this paper. For the bi-objective case, the test functions of ZDT1, ZDT2 and ZDT3 are adopted, which have different shape of Pareto frontier, (convex, concave and discontinue). For the higher M, the test functions of DTLZ1, DTLZ2 and DTLZ7 (Deb et al. 2001) are adopted, whose Pareto frontier are linear, concave and discontinue. The mathematical expression of the test functions can be found in Appendix 2. Note, all the test functions with the same d have the same design variable space of $\mathbb{X} = \{\mathbf{x}|0 \leq x_i \leq 1, i = 1, 2, \cdots d\}$. In order to average out the randomness caused by the initial samples, fifty groups of Latin Hypercube Samples are generated for each $\mathbb{X}$, and the number of initial samples always equals $10 * d$. For all the test functions, it is set that $d = M + 1$.

4.1 Comparison between improvement functions

The definitions of improvement function are important for the Pareto statistical infill criteria, and three desire properties have been proposed in Section 3.2. In order to further investigate the first two properties (generating well distributed Pareto points) of $I_{hv}^{mod}(\mathbf{y})$, $I_{hv}^{mod}(\mathbf{y})$ and $I_{hv}(\mathbf{y})$ will be compared on the DTLZ test functions ($M = 3$ and 4) in Section 4.1. To be specific, both the improvement functions are directly used to obtain the infill design point that

$$\max_{\mathbf{x}} I_{hv}^{mod}[\mathbf{y}(\mathbf{x})] \tag{29}$$

and

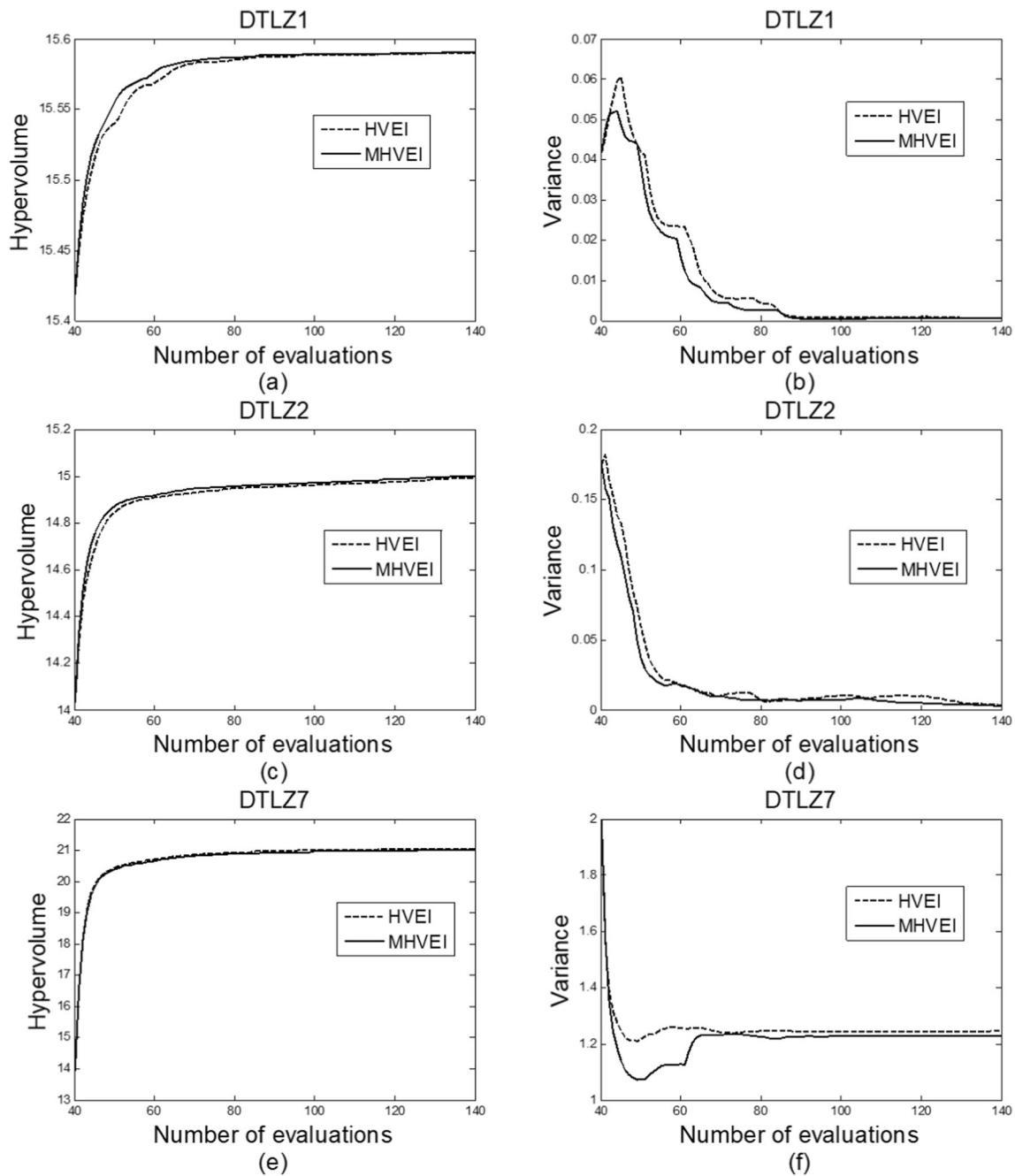$$\max_{\mathbf{x}} I_{hv}[\mathbf{y}(\mathbf{x})] \tag{30}$$

Fig. 8 Comparison on DTLZ functions with $M = 3$ and $d = 4$

where $\mathbf{y}(\mathbf{x})$ denotes the true objective functions. Note, the (29) and (30) can be understood as the MHVEI function and HVEI function with setting $\hat{\mathbf{y}} = \mathbf{y}$ and $\hat{\sigma} = 0$. The obtained values of hypervolume are compared in Fig. 5, and the Fig. 6 shows the difference (the hypervolume obtained by I_{hv} minus the one obtained by I_{hv}^{mod}).

The results show that the I_{hv} is able to achieve a higher hypervolume than I_{hv}^{mod}. This is understandable, since

maximizing I_{hv} is exactly the same to maximizing the hypervolume. However, the two curves shown in Fig. 5 basically overlap with each other, and the advantage of I_{hv} shown in Fig. 6 can be almost ignored. This comparison proves that I_{hv}^{mod} can produce Pareto points distributed as well as I_{hv}, and it also implies that the MHVEI works similar as HVEI in the absence of uncertainty.
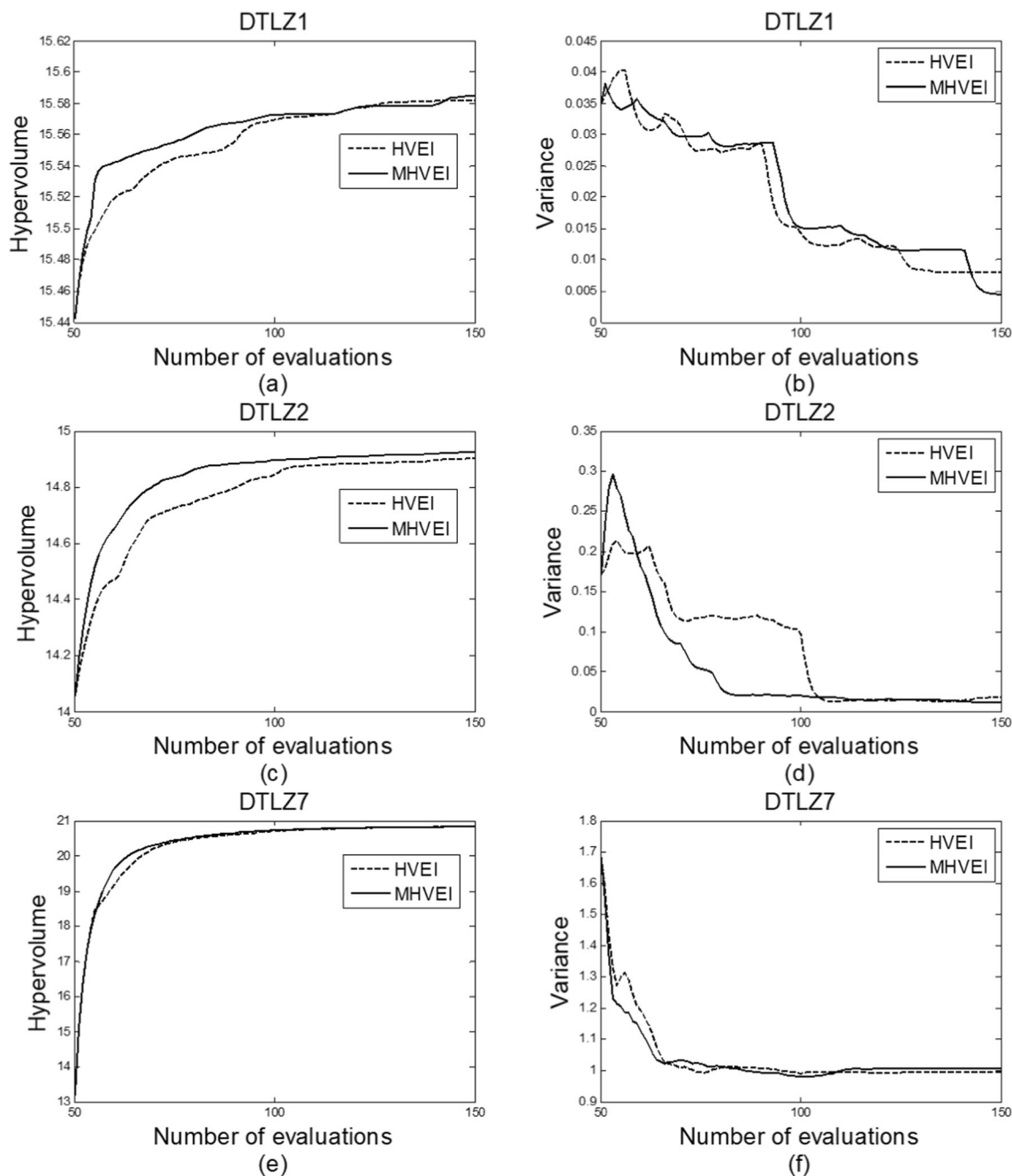
Fig. 9 Comparison on DTLZ functions with $M = 3$ and $d = 5$

In addition, I_{hv} lacks a calculating formula when $M > 3$, and the computational complexity also grows exponentially with M. Therefore the I_{hv}^{mod} can be used to replace the I_{hv} in the heuristic intelligence algorithms such as the SMS-EMOA (Beumea and Emmerich 2007), when the computational cost becomes a matter of concern.

4.2 Comparison with other infill criteria

In this section, the MHVEI is compared with three existing infill criteria including HVEI, EI-ELU and LCB. The EI-ELU is developed by Keane (2006) and defined as

$$EI_{elu}(\mathbf{x}) = P[I_{hv}(\mathbf{y}) > 0]\min\{\|\mathbf{y}^* - \mathbf{P}_i\| | 1 \leq i \leq n\} \qquad (31)$$
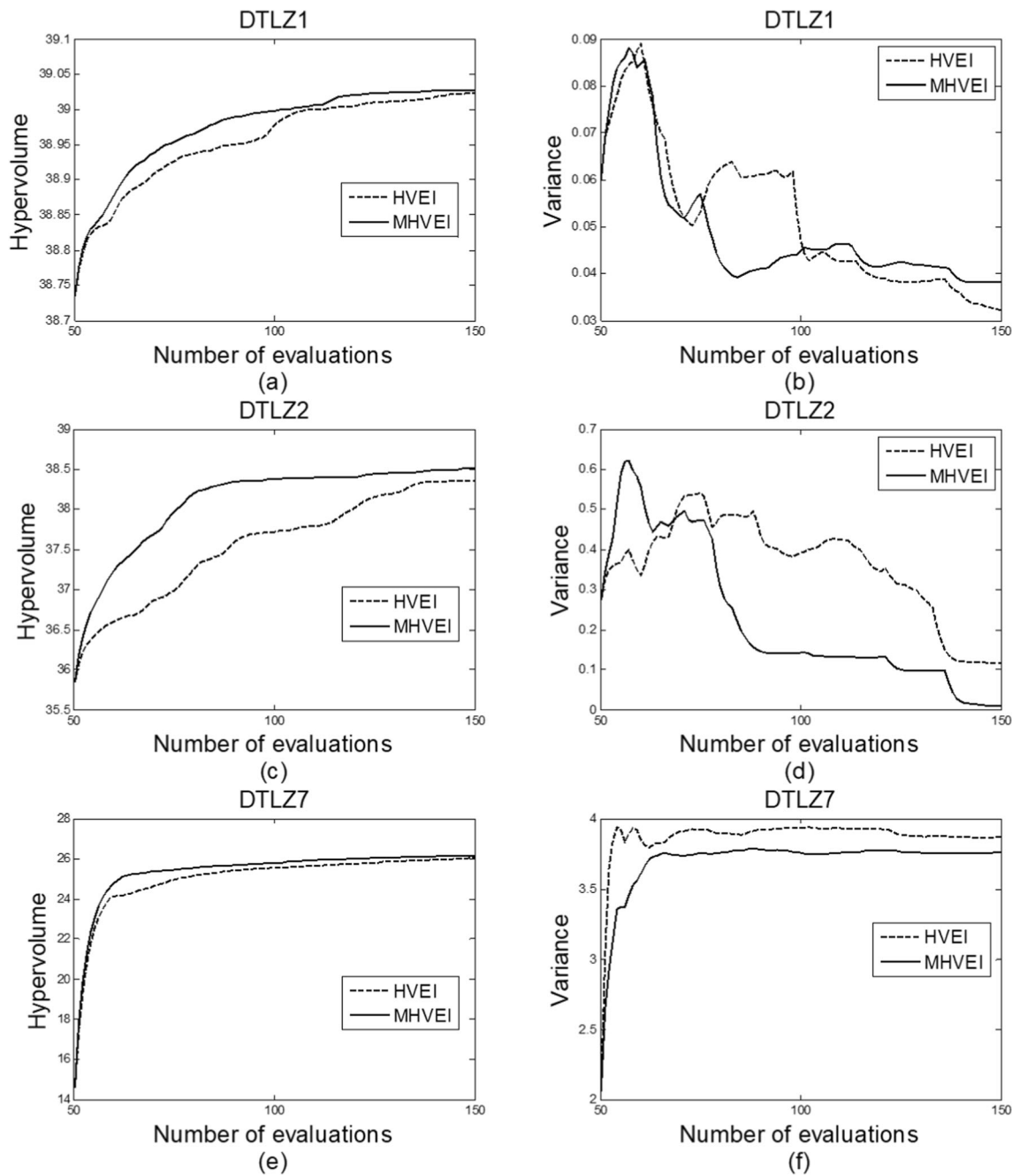
Fig. 10 Comparison on DTLZ functions with $M = 4$ and $d = 5$

where $P[I_{hv}(\mathbf{y}) > 0]$ denotes multiple objective probability improvement, and $\mathbf{y}^*$ is the expected centroid point of $\mathbf{y}$ in Ω. The LCB can be understood as the hypervolume improvement of a lower confidence point that $I_{hv}(\hat{\mathbf{y}} - w\hat{\sigma})$ where w is the gain factor. In this work, w is set with the value suggested by Emmerich et al. (2006). All these criteria can be easy to implement with a low computational cost on the bi-objective problem, but the computational complexity of both EI-ELU and LCB also grows exponentially with M. With the consideration of computational cost,

the MHVEI will be compared with all the three criteria on the ZDT functions ($M = 2$), and it will be compared with only HVEI on the DTLZ functions ($M = 3, 4$).

4.2.1 Comparison on ZDT functions

The averaged hypervolume obtained by MHVEI, HVEI, EI-ELU and LCB are compared in Fig. 7, and the results can be discussed with the desired properties mentioned in Section 3.3.
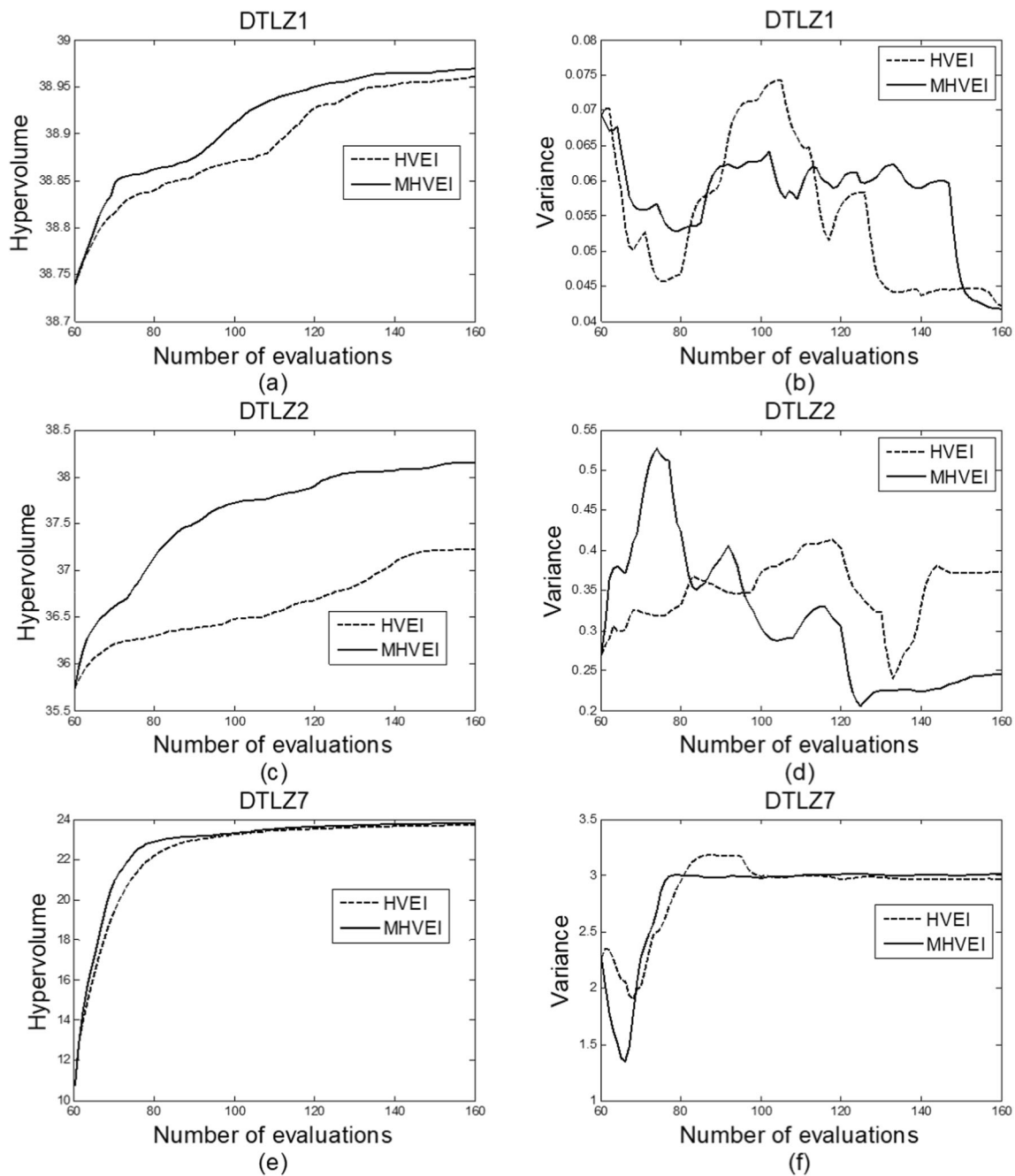
Fig. 11 Comparison on DTLZ functions with $M = 4$ and $d = 6$

The EI-ELU performs worst among all the four criteria in most cases, which can be attributed to that the EI-ELU is not monotonically decreasing with increasing $\hat{y}_i$. Comparing with HVEI and MHVEI, the LCB can achieve a similar hypervolume at the end of the optimization, but the hypervolume grows slower during the process. Note, the LCB balances the global exploration and local exploitation roughly in a linear way in the purpose of reducing the computational cost (the LCB can be faster calculate than the HVEI).

However the two searching preferences have a strong effect on both the convergence rate and global searching ability of the statistical infill criterion.

The HVEI and MHVEI perform similar on the ZDT functions, and they will be further compared below.

4.2.2 Comparison on DTLZ functions

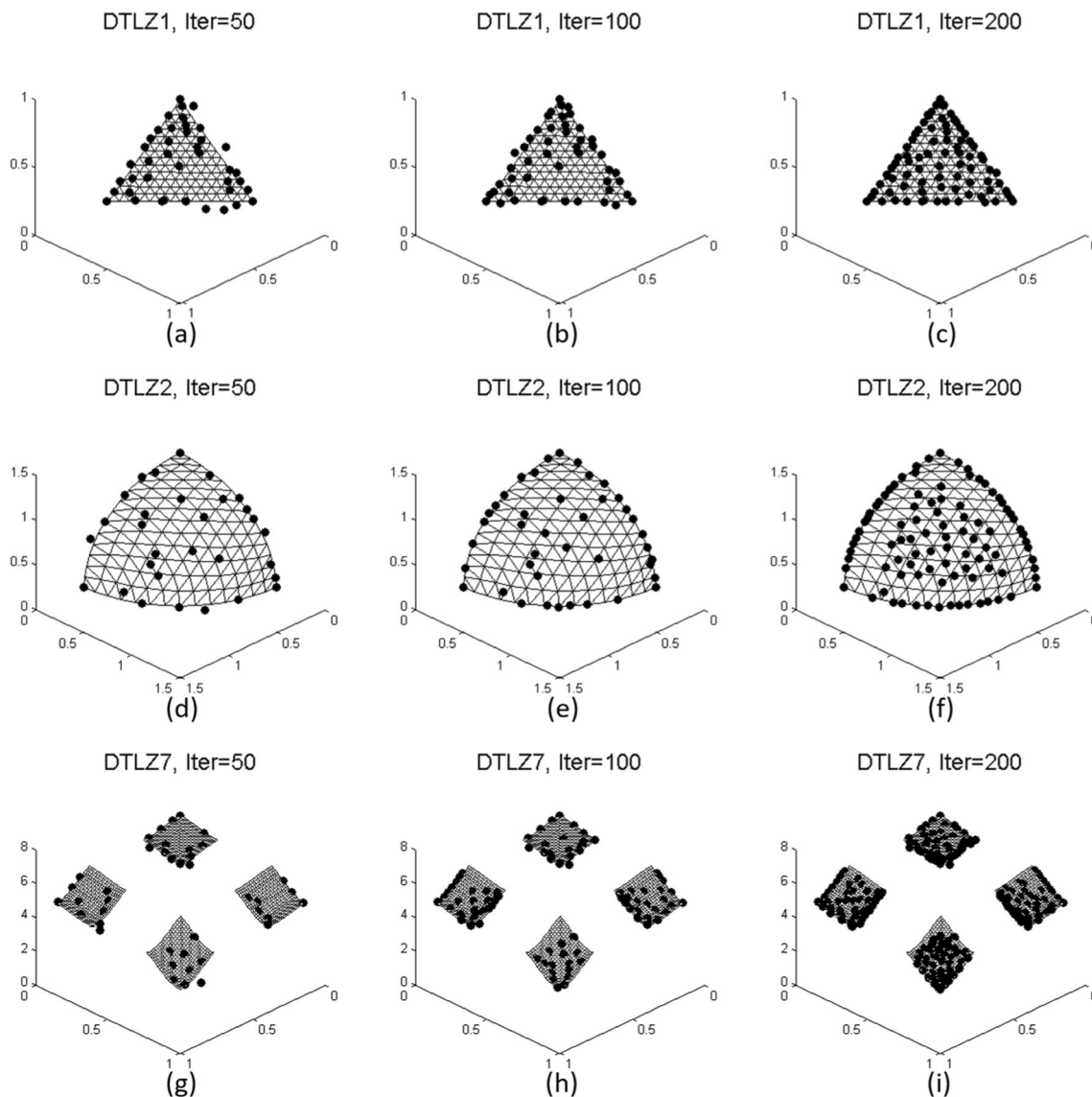Four groups of DTLZ functions with different number of sub-objectives and design variables are adopted. The

Fig. 12 Pareto points obtained by MHVEI in the first 200 iterations

averaged hypervolume and the relative variance are shown and compared in Figs. 8, 9, 10 and 11. Generally speaking, MHVEI performs no worse than HVEI in all the cases, and the advantage of MHVEI is quite obvious on the DTLZ1 and DTLZ2 with $d > 4$. The reason can be explained by the difference between the two infill criteria. When d is larger, the unknown sub-objectives will be more difficult to approximate by the Kriging model. As demonstrated in Section 3.4, comparing with HVEI, MHVEI prefers the point with higher $\hat{\sigma}$, and it is more helpful to improve the accuracy of the surrogate.

4.3 Investigation on Higher Number of Objectives

Figure 12 shows the Pareto points obtained by MHVEI during the optimization process with the first (of fifty)

group of initial samples. It can be observed that the Pareto points keep exploring $\mathbf{P}_{true}$ as the optimization continues. However, one can image that when M is larger, much more Pareto points will be needed to cover the higher dimensional space (the dimension of $\mathbf{P}_{true}$ is $M-1$ in our tests).

In this section, the performance of MHVEI on many objective problem ($M \geq 3$) will be investigated. In order to eliminate the effects of dimension on the hypervolume indicator, the normalized hypervolume defined as $H(\mathbf{P})/H(\mathbf{P}_{true})$ is used as the metric to measure Pareto points distribution. Only the test functions of DTLZ1 and DTLZ2 are adopted here, because the $H(\mathbf{P}_{true})$ can be calculated analytically with arbitrary value of M. The optimization process in the first 200 iterations with $M = 3$, $4, \cdots 7$ and $d = M + 1$ are compared.
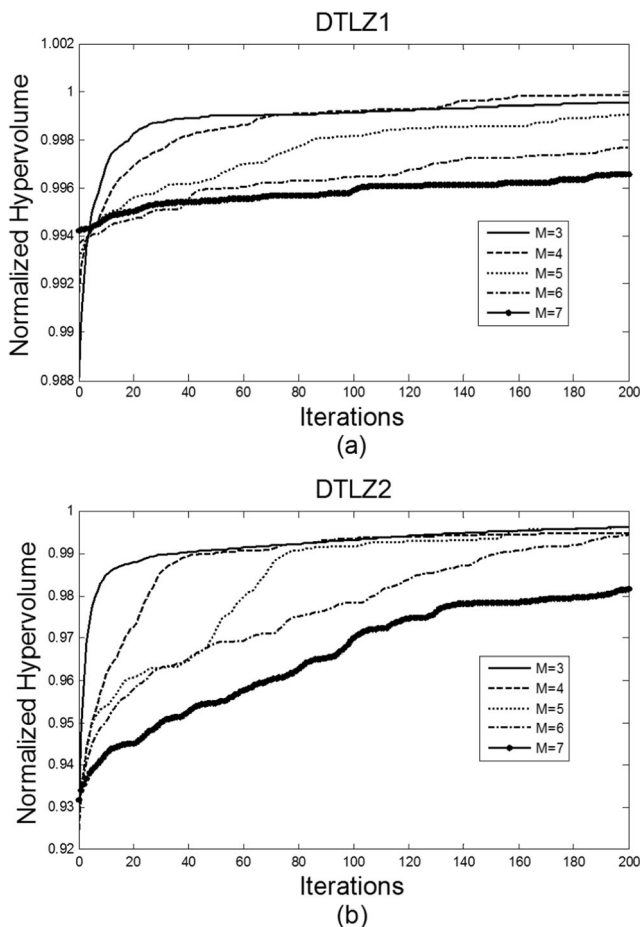
Fig. 13 Normalized hypervolume obtained by MHVEI

Note that $H(\mathbf{P})/H(\mathbf{P}_{true}) \leq 1$, and $H(\mathbf{P})/H(\mathbf{P}_{true}) = 1$ when and only when $\mathbf{P} = \mathbf{P}_{true}$. Figure 13 shows that the algorithm works well at the end of optimization when M is small ($M \leq 4$ for DTLZ1, and $M \leq 6$ for DTLZ2), but the hypervolume grows slower as M increases. It means that for the higher dimensional (both the design variable space and objective space) problem, more iterations will be needed to improve the distribution of the Pareto points. However, the computational cost might be unacceptable in the real application because of the time consuming evaluation. We believe that the parallel EGO methods (Haftka et al. 2016), which generate multiple new points in one iteration, are more desired for the Pareto optimization rather than the single objective problem.

5 Conclusion

The HVEI function has been proved as an outstanding statistical infill criterion for the multi-objective EGO method. However, the high computational cost has become the bottle neck to limit its application on the many objective optimization. Aiming at this problem, a new statistical infill criterion named MHVEI has been developed. The theoretical study shows that the MHVEI maintains all the desired properties of HVEI, but has a much lower computational complexity. The difference between HVEI and MHVEI is also studied theoretically, and it shows that compared with HVEI, the MHVEI more inclines to the global exploration. The new criterion is compared with several existing ones on Mathematical functions. The results show that the MHVEI and HVEI perform better than the other criteria on bi-objective problem, and MHVEI performs no worse than HVEI on lower dimensional (design variable space) problems but better on the higher dimensional ones.

The results in Section 4.1 show that maximizing the I_{hv}^{mod} can produce as well distributed Pareto points as directly maximizing the hypervolume. Therefore we believe the I_{hv}^{mod} can be used in the heuristic intelligence algorithms for MOO, such as the SMS-EMOA, when the computational cost becomes a matter of concern. In addition, the MHVEI is a statistical infill criterion which is based on the probability estimated by the Kriging surrogate in this work. In theory, this approach can be extended with other Gaussian process models, such as the stochastic radial basis functions.

Some problems have also been observed in our method. Because of the curse of dimension, when the number of design variables is higher, the surrogate will need more training data to properly approximate the objectives. And even if the accuracy of the surrogate is high enough, it still will need more points to explore the higher dimensional true Pareto frontier. Therefore, we believe that the multiple points infill criteria for the parallel computation is desired in the future. However, one question needs to be considered first that should the multiple points explore the multi modal of the infill criterion (like the approach for the single objective problem), or should they try to augment the obtained Pareto solutions on different region? In addition, although the MHVEI has a much lower computational complexity than the existing algorithms of HVEI, it is still more time-consuming than the other criteria which are not developed directly to improve the hypervolume. Therefore more efficient algorithms are still expected.

Appendix 1: Code for obtaining and updating LUBs

Code for updating LUBs

Matlab code for updating LUBs with a new Pareto point

```
1.  function U=UpdateLUBs(U,p)
2.  %INPUT:
3.  %U: the LUBs to be updated
4.  %p: the new Pareto point
5.  %OUTPUT:
6.  %U: the updated LUBs
7.  [m,M]=size(U);%n: number of LUBs, M: number of sub-objectives
8.  D=max(repmat(p,m,1)-U,[],2);
9.  SDU=U(D<0,:);% Current LUBs strongly dominated by p
10. WDU=U(D==0,:);% Current LUBs dominated by p but not strongly
11. % Delete the LUBs strongly dominated by p
12. U(D<0,:)=[];
13. % Take the projection points of p on the boundary of
14. % each [-inf, SDU(i,:)] as the candidates of new LUBs
15. [n,~]=size(SDU);
16. C=repmat(SDU,M,1);
17. for i=1:M
18.    C(n*(i-1)+1:n*i,i)=repmat(p(i),n,1);
19. end
20. % Update the LUBs
21. U=[U;-ParetoSet(-[C;WDU])];
```

The algorithm first delete the invalid old LUBs on line 12, and then append the new LUBs which are selected from a group of candidates on line 22. The invalid LUBs are the points strongly dominated by p and denoted by SDU in the code, see line 9. Part candidates of the new LUBs are the projection points of p on the boundary of the region bounded by each SDU point, these candidates are denoted by C in the code, see line 15–19. The other part candidates are the old LUBs which are dominated by p but not strongly, and these LUBs are denoted by WDU in the code, see line 10. The new LUBs are the upper bounds of these candidates, which can be understood as the Pareto points selected from the candidates when all the sub-objective are meant to be maximized. The "ParetoSet" on line 21 denotes the function of selecting the Pareto points. For example, the in the EGO process, the Pareto points **P** are selected from the training data **Y**, then **P**=ParetoSet(**Y**). The code of "ParetoSet" will not be further introduced.

The code for generating LUBs

Matlab code for generating LUBs with a group of Pareto points and a reference point

```
1.  function U=GenerateLUBs(P,r)
2.  % INPUT:
3.  % P: A group of Pareto points
4.  % r: Reference point
5.  %OUTPUT:
6.  %U: LUBs
7.  U=r;% LUB with no Pareto point
8.  % Update the LUBs with each Pareto point
9.  [n,~]=size(P);
10. for i=1:n
11.    U=UpdateLUBs(U,P(i,:));
12. end
```

When there is no Pareto point, the reference point will be the only LUB, see line7. Then the whole LUBs can be obtained by updating with the Pareto point one by one, see line 9–12.

Appendix 2: Test functions

ZDT functions

$$ZDT1: \quad \begin{aligned} & \min y_1 = x_1 \\ & \min y_2 = g\left(1-\sqrt{y_1/g}\right) \\ & s.t. 0 \le x_i \le 1, i = 1, 2, \cdots d \end{aligned} \quad (32)$$

$$ZDT2: \quad \begin{aligned} & \min y_1 = x_1 \\ & \min y_2 = g\left[1-(y_1/g)^2\right] \\ & s.t. 0 \le x_i \le 1, i = 1, 2, \cdots d \end{aligned} \quad (33)$$

$$ZDT3: \quad \begin{aligned} & \min y_1 = x_1 \\ & \min y_2 = g\left[1-\sqrt{y_1/g}-(y_1/g)\sin(10\pi y_1)\right] \\ & s.t. 0 \le x_i \le 1, i = 1, 2, \cdots d \end{aligned} \quad (34)$$

where $g = 1 + 9\sum_{i=1}^{d} x_i/(d-1)$ in all the three ZDT functions.

DTLZ functions

$$\text{DTLZ1}: \begin{array}{l} \min y_1 = 0.5 g_M \prod_{i=1}^{M-1} x_i \\ \min y_2 = 0.5 g_M (1-x_{M-1}) \prod_{i=1}^{M-2} x_i \\ \vdots \\ \min y_{M-1} = 0.5 g_M x_1 (1-x_2) \\ \min y_M = 0.5 g_M (1-x_2) \\ s.t. 0 \le x_i \le 1, i = 1, 2, \cdots d \end{array} \quad (35)$$

$$\text{DTLZ2}: \begin{array}{l} \min y_1 = g_M \prod_{i=1}^{M-1} \cos(0.5\pi x_i) \\ \min y_2 = g_M \sin(0.5\pi x_{M-1}) \prod_{i=1}^{M-2} x_i \cos(0.5\pi x_i) \\ \min y_3 = g_M \sin(0.5\pi x_{M-2}) \prod_{i=1}^{M-3} x_i \cos(0.5\pi x_i) \\ \vdots \\ \min y_{M-1} = g_M \sin(0.5\pi x_2) \cos(0.5\pi x_1) \\ \min y_{M-1} = g_M \sin(0.5\pi x_1) \\ s.t. 0 \le x_i \le 1, i = 1, 2, \cdots d \end{array}$$

$$(36)$$

where $g_M = 1 + \sum_{j=M}^{d} (x_j - 0.5)^2$ in DTLZ1 and DTLZ2 functions. Note that the original DTLZ1 has an extremely large number of local Pareto frontier, which cannot be approximated by the surrogate, therefore, the DTLZ1 function has been modified in (35)

$$\text{DTLZ7}: \begin{array}{l} \min y_1 = x_1 \\ \min y_2 = x_1 \\ \vdots \\ \min y_{M-1} = x_{M-1} \\ \min y_M = g_M \left[M - \sum_{j=1}^{M-1} y_1 (1 + \sin(3\pi y_i)/g_M) \right] \\ s.t. 0 \le x_i \le 1, i = 1, 2, \cdots d \end{array}$$

$$(37)$$

where $g_M = 2 + 9 \left(\sum_{j=M}^{d} x_j \right) / (d - M + 1)$ in DTLZ7 function.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Beumea N, Emmerich M (2007) SMS-EMOA: Multiobjective selection based on dominated hypervolume. Eur J Oper Res 181(3):1653–1669

Couckuyt I, Deschrijver D, Dhaene T (2014) Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. J Glob Optim 60(3):575–594

Das I, Dennis JE (2006) Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. SIAM J Optim 8(3):631–657

Deb K, Thiele L, Laumanns M, Zitzler E (2001) Scalable test problems for evolutionary multi-objective optimization. Evolutionary Multiobjective Optimization: Theoretical Advances and Applications Chapter 6: 105-145.

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput 6(2):182–197

Emmerich MTM, Beume N, Naujoks B (2005) An EMO Algorithm Using the Hypervolume Measure as Selection Criterion. In: International Conference on Evolutionary Multi-Criterion Optimization, Guanajuato, Mexico.

Emmerich MTM, Giannakoglou KC, Naujoks B (2006) Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels. IEEE Trans Evol Comput 10(4):421–439

Emmerich MTM, Deutz AH, Klinkenberg JW (2011) Hypervolume-based expected improvement: Monotonicity properties and exact computation. In: In 2011 IEEE Congress of Evolutionary Computation, New Orleans, LA, USA.

Ginsbourger D, Le Riche R, Carraro L (2010) Kriging is well-suited to parallelize optimization. Springer Berlin Heidelberg

Goel T, Vaidyanathan R, Haftka RT, Wei S, Queipo NV, Tucker K (2007) Response surface approximation of Pareto optimal front in multi-objective optimization. Comput Methods Appl Mech Eng 196(4): 879–893

Haftka RT, Villanueva D, Chaudhuri A (2016) Parallel surrogate-assisted global optimization with expensive functions – a survey. Struct Multidisc Optim 54: 3-13.

Hao P, Wang B, Li G (2012) Surrogate-Based Optimum Design for 871 Stiffened Shells with Adaptive Sampling. AIAA J 50(11):2389–2407

Hao P, Wang B, Li G, Meng Z, Wang L (2015) Hybrid Framework for Reliability-Based Design Optimization of Imperfect Stiffened Shells. AIAA J 53(10):2878–2889

Hao, P., Wang, B., Tian, K., Li, G., Sun, Y., & Zhou, C. (2017) Fast procedure for Non-uniform optimum design of stiffened shells under buckling constraint. Structural and Multidisciplinary Optimization 55(4):1503-1516

Jeong S, Obayashi S(2005) Efficient global optimization (EGO) for multi-objective problem and data mining, In: IEEE Congress on Evolutionary Computation, Edinburgh, Scotland.

Jie H, Wu Y, Zhao J, Ding J and Liangliang (2016) An efficient multi-objective PSO algorithm assisted by Kriging metamodel for expensive black-box problems. J Glob Optim 67: 1–25

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

Keane AJ (2006) Statistical Improvement Criteria for Use in Multiobjective Design Optimization. AIAA J 44(4):879–891

Keshtegar B, Hao P, Wang Y, Li Y (2017) Optimum design of aircraft panels based on adaptive dynamic harmony search. Thin-Walled Structures 118: 37–45.

Kim M, Hiroyasu T, Miki M, Watanabe S (2004) SPEA2+: Improving the Performance of the Strength Pareto Evolutionary Algorithm 2. Lect Notes Comput Sci 3242(4):742–751

Klamroth K, Lacour R, Vanderpooten D (2015) On the representation of the search region in multi-objective optimization. Eur J Oper Res 245(3):767–778

Knowles J (2006) ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Trans Evol Comput 10(1):50–66

Knowles JD, Corne DW, Fleischer M(2003) Bounded archiving using the lebesgue measure, In: IEEE Congress on Evolutionary Computation, Canberra, Australia.

Krige DG (1951) A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis University of the Witwatersrand

Li M, Li G, Azarm S (2006) A Kriging Metamodel Assisted Multi-Objective Genetic Algorithm for Design Optimization. J Mech Des 130(3):405–414

Liu W, Zhang Q, Tsang E, Liu C and Virginas B (2007) On the Performance of Metamodel Assisted MOEA/D, In: Advances in Computation and Intelligence, Second International Symposium, ISICA 2007, Wuhan, Proceedings, pp 547–557

Mckay MD, Beckman RJ, Conover WJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2):239–245

Messac A (2003) The normalized normal constraint method for generating the Pareto frontier. Struct Multidiscip Optim 25(2):86–98

Messac A, Puemi-Sukam C, Melachrinoudis E (2000) Aggregate Objective Functions and Pareto Frontiers: Required Relationships and Practical Implications. Optim Eng 1(2):171–188

Messac A, Dessel SV, Mullur AA, Maria A (2004) Optimization of large-scale rigidified inflatable structures for housing using physical programming. Struct Multidiscip Optim 26(1–2):139–151

Lophaven, SN, Søndergaard J, Nielsen HB (2002). DACE A Matlab Kriging Toolbox. IMM Informatiocs and Mathematical Modelling. Retrieved from http://imedea.uib-csic.es/master/cambioglobal/Modulo_V_cod101615/Lab/lab_maps/krigging/DACE-krigingsoft/dace/dace.pd

Przybylski A, Gandibleux X, Ehrgott M (2010) A two phase method for multi-objective integer programming and its application to the assignment problem with three objectives. Discret Optim 7(3):149–165

Sasena MJ, Papalambros P, Goovaerts P (2002) Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization. Eng Optim 34(3):263–278

Schonlau M (1998) Computer experiments and global optimization. University of Waterloo, Waterloo, Canada

Sóbester A, Leary SJ, Keane AJ (2005) On the Design of Optimization Strategies Based on Global Response Surface Approximation Models. J Glob Optim 33(1):31–59

Wagner T, EmmerichM, PonweiserW(2010) On expected-improvement criteria for model-based multi-objective optimization. In: International Conference on Parallel Problem Solving From Nature, Krakow, Poland.

While R, Bradstreet L, Barone L and Hingston P (2005) Heuristics for Optimising the Calculation of Hypervolume for Multi-objective Optimisation Problems 3(2): 2225–2232

Yang K, EmmerichM, Deutz A and Fonseca CM(2017) Computing 3-D expected hypervolume improvement and related integrals in asymptotically optimal time. In: International Conference on Evolutionary Multi-Criterion Optimization, Munster, Germany.

Zhan D, Cheng Y, Liu J (2017) Expected Improvement Matrix-based Infill Criteria for Expensive Multiobjective Optimization. IEEE Trans Evol Comput PP(99):1

Zhang Q, Liu W, Tsang E, Virginas B (2010) Expensive Multiobjective Optimization by MOEA/D With Gaussian Process Model. IEEE Trans Evol Comput 14(3):456–474

Zitzler E, Thiele L (1998) Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study. In: 5th International Conference on Parallel Problem Solving from Nature, Amsterdam, Netherlands.