2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

Kriging Model Based Many-Objective Optimization with Efficient Calculation of Expected Hypervolume Improvement

Chang Luo, Koji Shimoyama, and Shigeru Obayashi
Institute of Fluid Science
Tohoku University
Sendai 980-8577, Japan
Email: {luo, shimoyama, obayashi}@edge.ifs.tohoku.ac.jp

Abstract—**The many-objective optimization performance of using expected hypervolume improvement (EHVI) as the updating criterion of the Kriging surrogate model is investigated, and compared with those of using expected improvement (EI) and estimation (EST) updating criteria in this paper. An exact algorithm to calculate hypervolume is used for the problems with less than six objectives. On the other hand, in order to improve the efficiency of hypervolume calculation, an approximate algorithm to calculate hypervolume based on Monte Carlo sampling is adopted for the problems with more objectives. Numerical experiments are conducted in 3 to 12-objective DTLZ1, DTLZ2, DTLZ3 and DTLZ4 problems. The results show that, in DTLZ3 problem, EHVI always obtains better convergence and diversity performances than EI and EST for any number of objectives. In DTLZ2 and DTLZ4 problems, the advantage of EHVI is shown gradually as the number of objectives increases. The present results suggest that EHVI will be a highly competitive updating criterion for the many-objective optimization with the Kriging model.**

I. INTRODUCTION

In recent years, solving many-objective optimization problems that involve four or more objectives [1-5] has been one of the main research fields in the evolutionary multi-objective optimization (EMO). However, the applications of evolutionary algorithm based optimization methods to practical engineering designs need high computational cost due to a large number of expensive performance analyses such as three-dimensional computational fluid dynamics for complex geometries. A common strategy to reduce the computational effort is to use surrogate models. A number of surrogate models such as response surface model (RSM) [6], radial basis function (RBF) [7], Kriging model [8], and neural network (NN) [9] have been applied to practical engineering designs. Among these surrogate models, the Kriging model can estimate the deviation between the response model and sample points, and automatically adapt to the sample points. Additionally, the Kriging model has a characteristic that an assumption of order of the approximate function is not needed, so it is superior to general RSM. In this study, the Kriging model is adopted as the surrogate model.

With regard to the updating criterion of the Kriging model, Jones et al. [10] suggested that the expected improvement (EI) of an original objective function was maximized to determine the location of a new additional sample point, and designated the efficient global optimization algorithm (EGO). Jeong et al. [11] extended EGO for multi-objective problems, in which the EIs in terms of all objective functions were maximized and some of the non-dominated solutions in the EIs space were selected as the additional sample points. Emmerich et al. [12] proposed the expected hypervolume improvement (EHVI) as the updating criterion of the Gaussian random field metamodels, and evaluated EHVI using the Monte Carlo integration method. Shimoyama et al. [13] compared the optimization performance of the EHVI, EI and estimation (EST) updating criteria in the multi-objective optimization. The results indicated EHVI kept a good balance between accurate and wide search for non-dominated solutions. Subsequently, EHVI updating criterion was applied to the many-objective optimizations of the 2 to 8-objective DTLZ1 and DTLZ2 problems [14].

This paper continues to investigate the many-objective optimization performance of using EHVI as the updating criterion in other test problems with more objectives, for comparison with EI and EST. In order to improve the efficiency of many-objective optimization, faster algorithms to calculate hypervolume are adopted.

II. BACKGROUND

A many-objective problem with M objectives is defined as follows:

$$\begin{aligned} &\text{minimize} \quad \boldsymbol{F}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_M(\boldsymbol{x})) \\ &\text{subject to} \quad \boldsymbol{x} \in X \end{aligned} \tag{1}$$

where $\boldsymbol{x}$ is a vector of n decision variables, $\boldsymbol{F}(\boldsymbol{x})$ is a vector of M objective functions, $f_i(\boldsymbol{x})$ denotes the i-th objective to be minimized, and X is the feasible region delimited by the problem's constraints.

A. Kriging Model

The Kriging model has its original applications in mining and geostatistical fields referring to spatially and temporally correlated data [15]. The Kriging model is a combination of global model and localized departures as follows:

$$f(x) = \mu + Z(x) \tag{2}$$

where $f(x)$ denotes an unknown function of interest, and μ denotes a known global approximation model. $Z(x)$ is a realization of a stochastic process with mean zero and variance σ^2, and the covariance matrix of $Z(x)$ is given by

$$Cov\left[\mathrm{Z}(x^i), \mathrm{Z}(x^j)\right] = \sigma^2 \boldsymbol{R} \qquad \boldsymbol{R} = \left[R(x^i, x^j)\right] \tag{3}$$

In Eq. (3), $\boldsymbol{R}$ is an $(n_S \times n_S)$ correlation matrix symmetric with ones along the diagonal, $R(x^i, x^j)$ is the correlation function between any two points x^i and x^j among n_s sample points. The correlation function needs to be selected by users. This paper employs the following Gaussian correlation function:

$$R(x^i, x^j) = \exp\left[-\sum_{k=1}^{n} \theta_k \left|x_k^i - x_k^j\right|^2\right] \tag{4}$$

where n denotes the number of design variables. θ_k is the weight of the distance along the k-th design variable.

In the Kriging model, the values of μ, σ^2 and $\boldsymbol{\theta} = [\theta_1, \ \theta_2, \cdots, \ \theta_n]$ are determined by maximizing the likelihood function. First, $\boldsymbol{\theta}$ is obtained by maximizing the concentrated log-likelihood function as

$$\max_{\theta_k > 0} \left(-\frac{n_s}{2} \ln\left(\sigma^2\right) - \frac{1}{2} \ln\left(\left|\boldsymbol{R}\right|\right) \right), \ k = 1, 2, \cdots, n \tag{5}$$

μ and σ^2 that maximize the likelihood function are represented in closed form as

$$\mu = \frac{\mathbf{1}^T \boldsymbol{R}^{-1} \boldsymbol{f}}{\mathbf{1}^T \boldsymbol{R}^{-1} \mathbf{1}} \tag{6}$$

$$\sigma^2 = \frac{(\boldsymbol{f} - \mathbf{1}\mu)^T \boldsymbol{R}^{-1} (\boldsymbol{f} - \mathbf{1}\mu)}{n_s} \tag{7}$$

where $\boldsymbol{f} = [f(x^1), f(x^2), \cdots, f(x^{n_s})]^T$, and $\mathbf{1}$ is an n_s-dimensional unit vector. After $\boldsymbol{\theta}$ is obtained, μ and σ^2 are obtained by Eqs. (6) and (7), respectively. Thereinto, $\boldsymbol{R}$ is calculated by Eq. (4). Now, Eq. (2) can be written to be the Kriging model predictor as

$$\hat{f}(x) = \mu + \boldsymbol{r}^T \boldsymbol{R}^{-1} (\boldsymbol{f} - \mathbf{1}\mu) \tag{8}$$

where $\boldsymbol{r}$ is an n_s-dimensional vector whose i-th element is $Cov[\mathrm{Z}(x), \mathrm{Z}(x^i)]$.

The accuracy of the predicted value $\hat{f}(x)$ depends greatly on the distances between predicted point x and the sample points. The mean squared error $\hat{s}^2(x)$ for a predicted point x using the Kriging model predictor is defined by

$$\hat{s}^2(x) = \sigma^2 \left[\mathbf{1} - \boldsymbol{r}^T \boldsymbol{R}^{-1} \boldsymbol{r} + \frac{\left(\mathbf{1} - \mathbf{1}^T \boldsymbol{R}^{-1} \boldsymbol{r}\right)^2}{\mathbf{1}^T \boldsymbol{R}^{-1} \mathbf{1}} \right] \tag{9}$$

B. Expected Hypervolume Improvement

EHVI is based on the theory of the hypervolume indicator [16], and now we first introduce the hypervolume indicator. Hypervolume indicator is a popular metric which is used for comparing performance between different multi-objective optimizers. The hypervolume of a set of solutions S measures the size of the portion of objective space that is dominated by the set S collectively. In the field of EMO, the hypervolume indicator is the only unary indicator that is known to be strictly monotonic with regard to Pareto dominance. This characteristic is of high interest and relevance for the problems with a large number of objectives.

Hypervolume calculation requires high computational effort. Several algorithms have been proposed for calculating hypervolume exactly. Wu et al. [17] proposed the inclusion-exclusion algorithm (IEA) for hypervolume calculation, and the complexity of this algorithm is $O(M2^N)$ for N solutions and M objectives. Fleischer [18] introduced the algorithm based on the Lebesgue measure, and its complexity is $O(N^M)$. While et al. [19] suggested a fast hypervolume by slicing objective (HSO) algorithm, and the complexity is $O(N^{M-1})$. Based on HSO, Fonseca et al. [20] proposed an improved dimension-sweep algorithm for calculating hypervolume. The proposed algorithm achieved $O(N^{M-2}\log N)$ complexity in the worst case. The fastest algorithm yet known for exact hypervolume calculation is the Waking Fish Group (WFG) algorithm, proposed by While et al. [21].

Because of the high computation cost especially in the problems with more than five objectives, some approximate algorithms to calculate hypervolume have also been developed in recent years. Bader et al. [1] proposed the approximate algorithm based on Monte Carlo sampling. Bringmann et al. [22] presented a fast approximation algorithm, and proved that it calculates a solution with contribution at most $(1+\varepsilon)$ times the minimal contribution with probability at least $(1-\delta)$ for arbitrarily given $\varepsilon, \delta > 0$. Ishibuchi et al. [23] proposed the approximation algorithm using achievement scalarizing functions with uniformly distributed weight vectors.

EHVI is the expected value of hypervolume improvement in the Kriging model. The hypervolume improvement $HVI[f_1(x), f_2(x), \cdots f_M(x)]$ is defined as the difference of hypervolume between the current sample set and the next sample set, as illustrated in Fig. 1, and its expected value $EHVI[f_1(x), f_2(x), \cdots f_M(x)]$ is expressed as

$$EHVI\left[f_1(x), f_2(x), \cdots, f_M(x)\right] = \int_{-\infty}^{f_{1ref}} \int_{-\infty}^{f_{2ref}} \cdots \int_{-\infty}^{f_{Mref}} HVI\left[f_1(x), f_2(x), \cdots, f_M(x)\right] \times \phi_1(F_1)\phi_2(F_2)\cdots\phi_M(F_M)\, dF_1 dF_2 \cdots dF_M \quad (10)$$

where F_i denotes the Gaussian random variable $N[\hat{f}_i(x), \hat{s}_i^2(x)]$. $\phi_i(F_i)$ is the probability density function, and f_{iref} is the reference value used for calculating hypervolume. The maximization of EHVI is considered as the updating criterion to determine the location of an additional sample point.

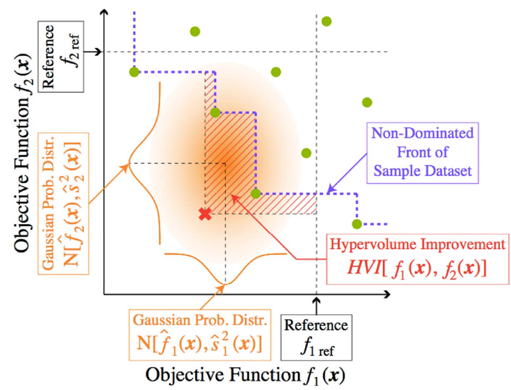


Fig. 1. EHVI updating criterion of the Kriging model [14]

In the present optimizations, objective function values are normalized between 0 and 1 for all given sample points. When the number of objectives is small (≤ 5), hypervolume is calculated using the fastest exact algorithm WFG [21], and the reference values are set to be 1.1 for all objective functions. When the number of objectives is large (>5), hypervolume is estimated based on the approximate algorithm as described in [1], and the number of samples used in this approximate algorithm is set to be 300 for 8 and 10-objective problems, and 500 for 12-objective problems.

III. NUMERICAL EXPERIMENTS

A. Problem Definition

Three to twelve-objective DTLZ1, DTLZ2, DTLZ3 and DTLZ4 problems [24] are considered in the present numerical experiments. All these problems are non-constrained, and the corresponding Pareto-optimal fronts lie in [0, 0.5] for DTLZ1 or [0, 1] for other DTLZ problems. Table 1 lists these test problems and their key properties.

Table 1. The test problems and key properties
(M is the number of objectives, and n is the number of variables)

Problems	M	n	Properties
DTLZ1	3, 5, 8, 10, 12	M+4	Multimodal, separable
DTLZ2	3, 5, 8, 10, 12	M+9	Concave, separable
DTLZ3	3, 5, 8, 10, 12	M+9	Multimodal, concave, separable
DTLZ4	3, 5, 8, 10, 12	M+9	Concave, separable

B. Computational Procedure

In order to investigate the performance of EHVI, two other updating criteria EI and EST [14] are adopted for comparing with EHVI. Figure 2 gives the flowchart of many-objective optimization with the Kriging model.

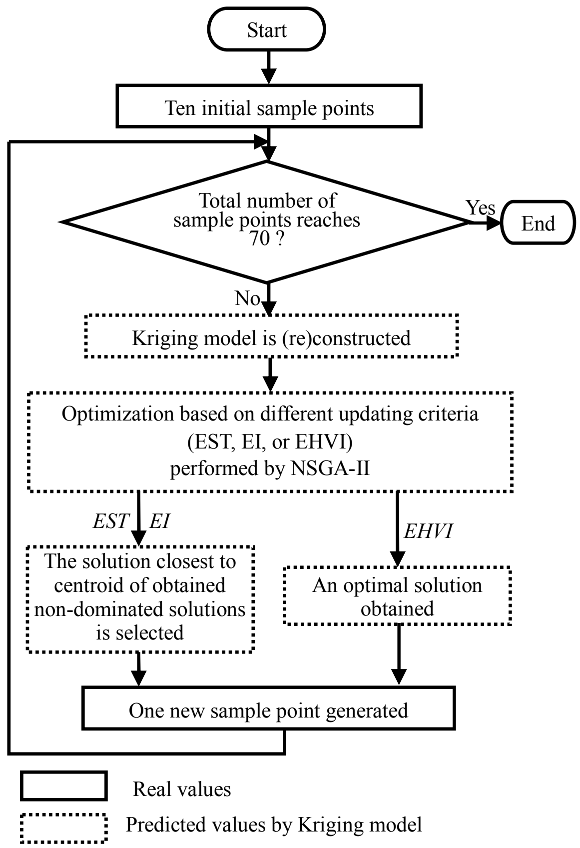


Fig. 2. Flowchart of many-objective optimization [14]

For each test problem, 20 runs are carried out with a different initial datasets of ten sample points. The initial sample points are generated using Latin hypercube sampling method [25], and distribute uniformly in the design space. The optimization of determining the locations of additional sample points are performed using NSGA-II [26]. The parameter values used in NSGA-II are given in Table 2.

Table 2. Parameter values used in NSGA-II

Parameters	NSGA-II
Population size	100
Number of generations	200
Simulated binary crossover probability	1
Polynomial mutation probability	$1/n$
η_c [27]	30
η_m [27]	20

In the optimization based on EHVI, a single solution with the maximum EHVI is obtained, and employed as the additional sample point. This indicates that the sample points are added one by one. On the other hand, a set of multiple non-dominated solutions is obtained in the optimization based on EI or EST. For a fair comparison, only the solution closest to the centroid of the obtained non-dominated solutions is chosen as the additional sample point. The termination condition of updating the Kriging model is that the total number of the initial sample points plus the additional sample points reaches 70.

C. Performance metric

The inverted generational distance (IGD) [28] is considered as the performance metric in the present numerical experiments. IGD can provide the combined information about the convergence and diversity of the obtained solutions. IGD is defined as follows:

$$IGD\left(T, A\right) = \frac{1}{|T|} \sum_{f \in T} \min_{f' \in D} \left| f - f' \right| \tag{11}$$

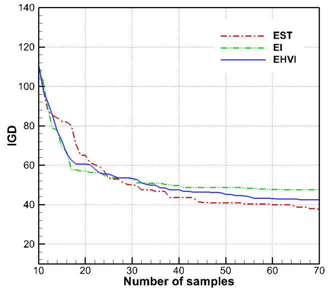
where T denotes the set of the points located on the true Pareto-optimal front. D is the non-dominated solution set obtained by the optimization. Smaller IGD indicates better performance in multi-objective optimization. The solutions in T distribute uniformly on the true Pareto-optimal front. The size of T varies with the number of objectives, and is shown in Table 3.

Table 3. The size of T with the number of objectives

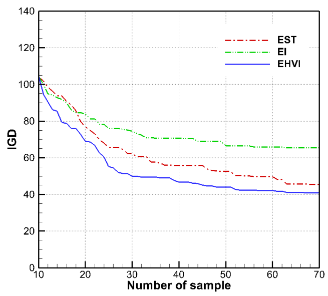
M	3	5	8	10	12
$\|T\|$	21^{M-1}	11^{M-1}	4^{M-1}	3^{M-1}	3^{M-1}
	441	14641	16384	19683	177147

D. Results and Discussion

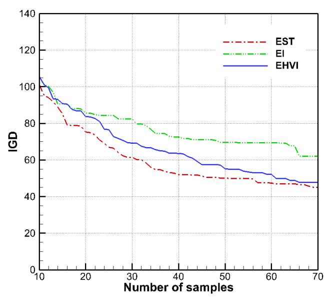
Figure 3 gives the IGD histories during the Kriging model update based on EHVI, EI and EST criteria in 3 to 12-objective DTLZ1 problems. The IGD values are averaged on 20 trials with different initial sample points. EST achieves faster reduction of IGD than EHVI and EI except for the case of $M = 5$, EHVI achieves the fastest IGD reduction. EST always obtains faster reduction of IGD than EI for any number of objectives. It indicates that the updating criterion EST without considering estimation errors works better than the updating criterion EI considering estimation error.



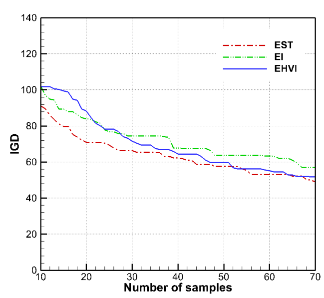
(a) M = 3



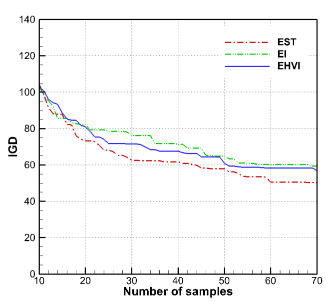
(b) M = 5



(c) M = 8

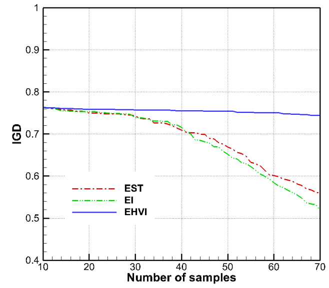


(d) M = 10

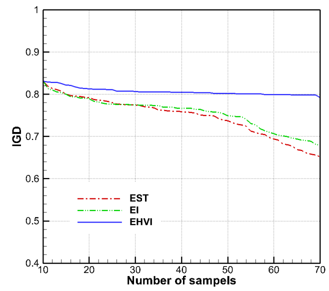


(e) M = 12

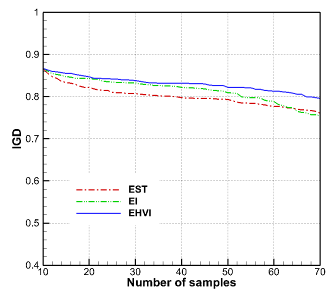
Fig. 3. IGD history during the Kriging model update in the DTLZ1



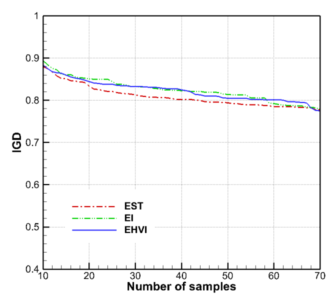
(a) $M = 3$



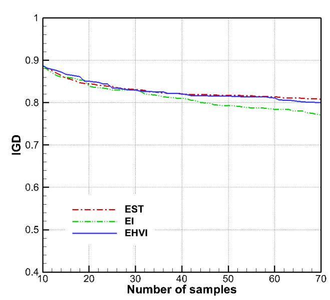
(b) $M = 5$



(c) $M = 8$

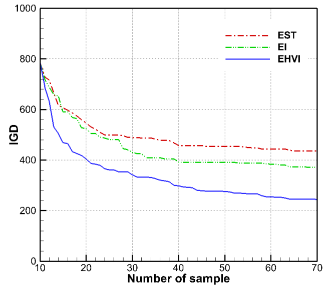


(d) $M = 10$

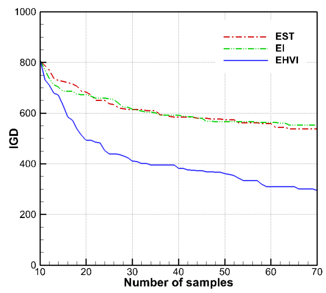


(e) $M = 12$

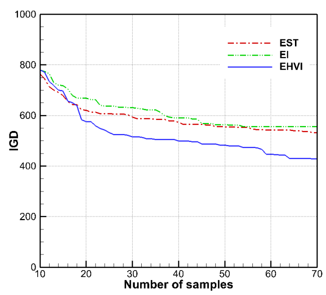
Fig. 4. IGD history during the Kriging model update in the DTLZ2



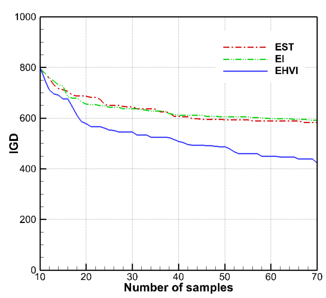
(a) $M = 3$



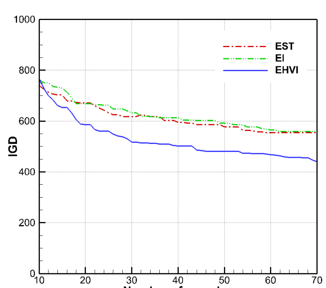
(b) $M = 5$



(c) $M = 8$

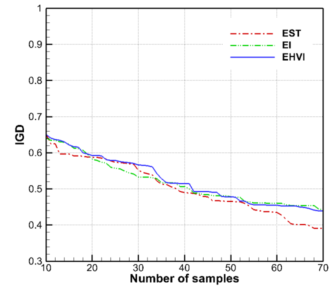


(d) $M = 10$

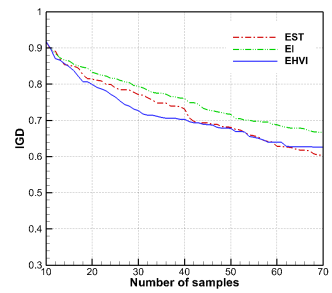


(e) $M = 12$

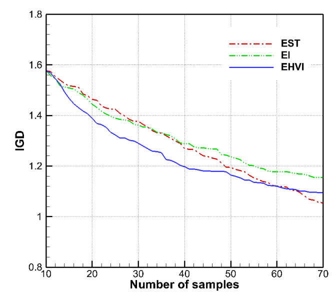
Fig. 5. IGD history during the Kriging model update in the DTLZ3



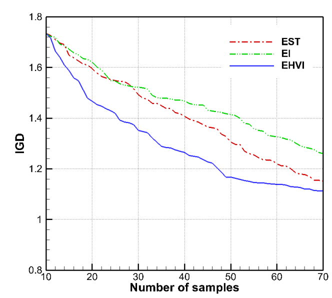
(a) $M = 3$



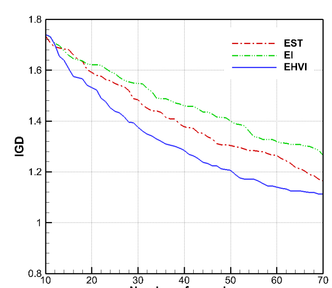
(b) $M = 5$



(c) $M = 8$



(d) $M = 10$



(e) $M = 12$

Fig. 6. IGD history during the Kriging model update in the DTLZ4

The IGD histories during the Kriging model update based on EHVI, EI and EST criteria in DTLZ2 are shown in Fig. 4. When $M = 3$ and 5, the optimization performance of using EHVI as the updating criterion is poor. This is because the updating strategy based on EHVI may result in additional sample points falling into a local optimal region. As M increases larger than 5, the optimization performance based on EHVI is gradually improved. It is noted that EI suddenly obtains much faster reduction of IGD than EHVI and EST when $M = 12$.

The IGD histories during the Kriging model update based on EHVI, EI and EST criteria in DTLZ3 are presented in Fig. 5. EHVI always achieves better convergence and diversity performances than EI and EST for any number of objectives. When $M = 3$, EI obtains faster IGD reduction than EST. When M is larger than 3, on the other hand, EST shows a little better optimization performance than EI.

Figure 6 shows the IGD histories during the Kriging model update based on EHVI, EI and EST criteria in 3 to 12-objective DTLZ4 problems. EST obtains faster IGD reduction than EI for any number of objectives. As M increases, convergence and diversity performances of using EHVI as updating criterion become better. In addition, when $M = 10$ and 12, EHVI outperforms EI and EST.

IV. CONCLUSION

Many-objective optimizations with the Kriging model based on EHVI, EI and EST updating criteria were conducted in 3 to 12-objective DTLZ1, DTLZ2, DTLZ3 and DTLZ4 test problems, and those performances are compared in this study. Thereinto, the fastest WFG exact algorithm and the approximate algorithm based on Monte Carlo sampling were adopted for hypervolume calculation.

In DTLZ1 problem, EHVI had faster IGD reduction than EI and EST when $M = 5$, but EHVI did not keep the advantage when M was larger than 5. On the other hand, EST always obtains faster reduction of IGD than EI for any number of objectives. In DTLZ2 and DTLZ4 problems, the advantage of EHVI was shown gradually as M increased, and EHVI obtained faster reduction of IGD than EST when M is larger than 8. In DTLZ3 problem, EHVI always obtained better convergence and diversity performances than EI and EST for any number of objectives. As a whole, EHVI is a highly competitive updating criterion of the Kriging model for many-objective optimization.

REFERENCES

[1] J. Bader and E. Zitzler, "Hype: An algorithm for fast hypervolume-based many-objective optimization," TIK-Report No.286, November 26, 2008.

[2] M. Garza-Fabre, G. T. Pulido, and C. A. C. Coello, "Ranking methods for many-objective optimization," in Proceedings of Mexican International Conference on Artificial Intelligence (MICAI 2009), Springer, pp. 633-645, 2009.

[3] S. F. Adra and P. J. Fleming, "Diversity management in evolutionary many-objective optimization," IEEE Transactions on Evolutionary Computation, vol. 15, no. 2, pp. 183-195, 2011.

[4] D. Hadka and P. Reed, "Borg: an auto-adaptive many-objective evolutionary computing framework," Evolutionary Computation, vol. 21, no. 2, pp. 231-259, 2013.

[5] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point based non-dominated sorting approach, part I: solving problems with box constraints," IEEE Transactions on Evolutionary Computation, accepted for publication in a future issue.

[6] Y. Lian and M. Liou, "Multiobjective optimization using coupled response surface model and evolutionary algorithm," AIAA Journal, vol. 43, no. 6, pp. 1316-1325, 2005.

[7] H. Fang, M. Rais-Rohani, and M. Horstemeyer, "Multiobjective crashworthiness optimization with radial basis functions," Proceedings of the Tenth AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, New York, AIAA-2004-4487, 2004.

[8] K. Shimoyama, S. Yoshimizu, S. Jeong, S. Obayashi, and Y. Yokono, "Multi-objective design optimization for a steam turbine stator blade using LES and GA," Journal of Computational Science and Technology, vol. 5, no. 3, pp. 134-147, 2011.

[9] M. Papadrakakis, N. Lagaros, and Y. Tsompanakis, "Optimization of large-scale 3D trusses using evolution strategies and neural networks," Int. J. Space Struct., vol. 14, no. 3, pp. 211-223, 1999.

[10] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box function," Journal of Global Optimization, vol. 13, pp. 455-492, 1998.

[11] S. Jeong, Y. Minemura, and S. Obayashi, "Optimization of combustion chamber for diesel engine using Kriging model," Journal of Fluid Science and Technology, vol. 1, no. 2, pp. 138-146, 2006.

[12] M. T. M. Emmerich, K. C. Giannakoglou, and B. Naujoks, "Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels," IEEE Transactions on Evolutionary Computation, vol. 10, no. 4, pp. 421-439, 2006.

[13] K. Shimoyama, K. Sato, S. Jeong, and S. Obayashi, "Updating Kriging surrogate models based on the hypervolume indicator in multi-objective optimization," Journal of Mechanical Design, Transactions of the ASME, vol. 135, no. 9, pp. 094503-1-7, 2013.

[14] K. Shimoyama, S. Jeong, and S. Obayashi, "Kriging-surrogate-based optimization considering expected hypervolume improvement in non-constrained many-objective test problems," in Proceeding of the 2013 IEEE Congress on Evolutionary Computation. Piscataway, NJ, USA: IEEE Press, pp. 658-665, 2013.

[15] N. A. C. Cressie, Statistics for spatial data, rev., Wiley, New York, 1993.

[16] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms - a comparative case study," PPSN 1998, LNCS, vol. 1498, pp. 292-301. Springer, Heidelberg, 1998.

[17] J. Wu and S. Azarm, "Metrics for quality assessment of a multiobjective design optimization solution set," Journal of Mechanical Design, vol. 123, pp. 18-25, 2001.

[18] M. Fleischer, "The measure of Pareto optima: applications to multiobjective metaheuristics," in Evolutionary Multi-objective Optimisation, ser. Lecture Notes on Computer Science, vol. 2632, Springer-Verlag, pp. 519-533, 2003.

[19] L. While, P. Hingston, L. Barone, and S. Huband, "A faster algorithm for calculating hypervolume," IEEE Transactions on Evolutionary Computation, vol. 10, no. 1, pp. 29-38, 2006.

[20] C. M. Fonseca, L. Paquete, and M. Lopez-Ibanez, "An improved dimension-sweep algorithm for the hypervolume indicator," in Congress on Evolutionary Computation, pp. 3973-3979, 2006.

[21] L. While, L. Bradstreet, and L. Barone, "A fast way of calculating exact hypervolumes," IEEE Transactions on Evolutionary Computation, vol. 16, no. 1, pp. 86-95, 2012.

[22] K. Bringmann and T. Friedrich, "Approximating the least hypervolume contributor: NP-hard in general, but fast in practice," in Evolutionary Multi-objective Optimisation, ser. Lecture Notes on Computer Science, vol. 5467, Springer-Verlag, pp. 6-20, 2009.

[23] H. Ishibuchi, N. Tsukamoto, Y. Sakane, and Y. Nojima, "Hypervolume approximation using achievement scalarizing functions for evolutionary many-objective optimization," in Congress on Evolutionary Computation, pp. 530-537, 2009.

[24] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multi-objective optimization," TIK-Technical Report No. 112, July 2001.

[25] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," Technometrics, vol. 21, no. 2, pp. 239 -245, May 1979.

[26] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, vol. 6, no. 2, pp. 182-197, 2002.

[27] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," Complex Systems, vol. 9, no. 2, pp. 115-148, 1995.

[28] D. A. Van Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithm research: a history and analysis," Air Force Institute of Technology, Dayton, OH, USA, TR-98-03, October 1998.