



Design by adaptive infill sampling with multi-objective optimization for exploitation and exploration

Jae-Young Choi ^{a,1}, Jangho Park ^{a,1}, Seulgi Yi ^{b,2}, Yeongmin Jo ^{b,2}, Seongim Choi ^{c,*,3}

^a Virginia Polytechnic Institute and State University, Blacksburg, VA, 24060, United States of America

^b Korea Advanced Institute of Science and Technology, Daejeon, 34141, South Korea

^c Gwangju Institute of Science and Technology, Gwangju, 61005, South Korea

ARTICLE INFO

MSC:
00-01
99-00

Keywords:

Multi-objective optimization
Multivariate Gaussian process
Infill sampling criteria

ABSTRACT

A surrogate-based design optimization with an adaptive sampling technique based on the innovative infill sampling criteria (ISC) is developed, where key issues are mathematically formulated on where in the design space and how many sample points are infilled to guarantee desirable accuracy at minimal computational cost. The ISC developed in the study involve multi-objective optimization (MOO) to determine infilling sample points separately for exploitation and exploration of the design space, which represent the global and the local accuracy of the surrogate model, respectively. The infill samples are found by the MOO on the Pareto front in terms of variance of estimation uncertainty and a predicted function value. To dynamically control the location and the count of the infilling points per iteration for the sample infilling, two criteria of the balancing and the dynamic switching approach are developed.

The balancing approach selects infill sample points equally from the two far ends of the Pareto front as well as on the middle of it. The dynamic switching approach uses cut-off variance of uncertainty estimation to dynamically switch the ISC exclusively from the exploration to the exploitation, or vice versa adaptively to the accuracy of the model. Solution optimality and computation efficiency of the present method for the EGO, are compared for two analytic functions with those of the EGO with a conventional, multi-point Expected Improvement (q-EI) ISC and a Latin Hypercube Sampling (LHS) method. The gradient-based optimization without using the surrogate model was also carried out independently for the comparison purpose on the solution accuracy and efficiency. The proposed method shows the greatest efficiency, requiring the smallest number sample points in the training set and becomes even compatible with the gradient-based optimization method. For the practical design problem, high-life multi-element airfoil is chosen to maximize a lift coefficient with non-increasing drag constraints. The proposed method showed about 18% increase of lift force.

1. Introduction

Design optimization problems in aerospace engineering are becoming more complicated as they demand higher precision modeling and simulation (M&S) to solve underlying physics which involves a large number of design parameters and coupled with other physics. High-fidelity analysis of computational fluid dynamics (CFD), for example, takes several hours to solve RANS (Reynolds-Averaged Navier–Stokes) equations even with the massively parallel computations and efficient numerical algorithms. The direct use of the high-fidelity analyses alone in a gradient-free design optimization framework is practically impossible due to high cost in computation time and memory. A gradient-based optimization (GBO) algorithm can be effective for large-scale design

problems using the state-of-the-art adjoint sensitivity analysis. However, the solution optimality is highly sensitive to initial guesses and the adjoint solver is function-specific, i.e., a new development is needed to solve different physics. The GBO is also easily trapped in the local optima.

An efficient global optimization (EGO) method for design is advantageous for the following reasons: a gradient-free search algorithm is better in finding global optima, a surrogate model replaces highly expensive physics simulation, and an effective adaptive sampling strategy further improves the efficiency by reducing computational cost. The adaptive sampling strategy becomes a key to the successful EGO as it directly determines the accuracy of the surrogate model and the efficiency of a design process, and it decides the locations and the

* Corresponding author.

E-mail addresses: holv@vt.edu (J.-Y. Choi), astrox@vt.edu (J. Park), ysg0221@kaist.ac.kr (S. Yi), yminjo@kaist.ac.kr (Y. Jo), schoi1@vt.edu (S. Choi).

¹ Ph.D. Candidate, Aerospace and Ocean Engineering Department.

² Ph.D. Candidate, Aerospace Engineering Department.

³ Assistant Professor, School of mechanical engineering, AIAA Member.

counts of sample points of the training set. Since Jones et al. suggested infill sampling criteria (ISC) to add more sample points in search of the global optimum efficiently and accurately, several ISC methods including expected improvement (EI) [1], Kushner's predicted improvement (PI) [2] and lower confidence bounding (LCB) [3] have been developed. Those conventional ISC methods add only a single sample point per design iteration to update the surrogate model, which continues until the design termination criteria are satisfied. In an effort to reduce the number of iterations for the adaptive sampling, the concept of selecting multiple sample points has been proposed parallel computation of those sample points. Haftka et al. [4] suggested the selection of multiple sample points where high estimation uncertainty is predicted by the Kriging surrogate model. Ginsbourger et al. [5] derived the analytic formulation of the multi-point EI as an extension of a single EI concept to a higher dimension. They pointed out that the highly expensive computational method such as Monte-Carlo simulation is needed to calculate the analytic form of the multi-point EI. However, due to high computational cost associated with the MCS runs, as an alternative to the analytic form of the multi-point EI, they proposed a concept of approximated multi-point EI (q-EI) with various heuristics to reduce the computation burden.

In the present study, a new method of the multi-point infill sampling criteria is developed for the EGO method using the Kriging surrogate model [6]. The Kriging model approximates relatively well noisy and multi-modal responses, and provides probabilistic estimation uncertainty associated with a predicted function value. During the design process, the additional sample points are infilled into the training set for two major purposes: (1) exploration of design space to improve the global accuracy of the model and (2) exploitation to improve the local accuracy of the model around the point of interest. The main idea is the effective utilization of the trade-off between the exploration and the exploitation using the multi-point ISC. The variance value of estimation uncertainty identifies regions less explored in the design space, while the prediction of output responses can locate the potential function minimum. The multi-objective optimization (MOO) using a genetic algorithm (GA) [7,8] is ideal to obtain infill sample points contributing to the two objectives in the separate sample sets. Two separate approaches to choose points on the Pareto-front are developed: a balancing and a dynamic switching approach. The balancing approach chooses two points on the far ends of the Pareto-front to balance two competing objectives and one additional point in between the end points according to the ratio of two coefficients of variable (COVs) [9] based on the distribution of estimation uncertainty and predicted function value. On the other hand, the dynamic switching approach weighs one objective more than the other by bookkeeping the global accuracy of the model over the iterations, and switch or alternate the infill criteria between the exploration and the exploitation, and selects multiple points from only at one end of Pareto front based on a pre-specified cut-off value. The covariance matrix used in the Kriging model becomes ill-conditioning if point clustering in the design space results in highly uneven distribution of the training sample points. A series of numerical experiments on the parameters in the adaptive sampling process using the MPMOISC are carried out to investigate the effects on the numerical stability.

For the validation, two-dimensional analytic functions of six-hump camel back and Matlab peak functions are used to demonstrate the accuracy and efficiency in finding the global optimum of the EGO framework using the MPMO ISC. The total number of sample points are compared with those from the three approximated q-EI ISC methods: Kriging believer (KB), minimum constant liar (CLmin), and maximum constant liar (CLmax). EGO with the Latin hypercube sampling (LHS) method [10] is also carried out for a comparison purpose. It was shown that the MPMO ISC method with the balancing and dynamic switching approaches required the least number of sample points to achieve solution accuracy, and the efficiency was as good as that of the gradient-based optimization of the SLSQP algorithm [11] which does not use the surrogate model for function evaluation.

For a practical design application, the shape of high-lift multi-element NLR 7301 airfoil [12] is optimized for maximal lift with respect to a flap angle, flap gap, overlap of the flap and angle of attack. The high-fidelity RANS solutions are obtained by the CFD method for aerodynamic analysis. The MPMO ISC method with the balancing approach shows the best efficiency at given solution accuracy requiring half the sample points needed for the approximated q-EI methods. These results are compatible with those of the gradient-based optimization in efficiency and solution optimality.

The organization of the paper is as follows. In Section 2, the EGO framework with various ISC methods are explained. A mathematical formulation of the ordinary Kriging model is briefly summarized. Existing ISC methods are described including the multi-point EI method as well as a conventional single point EI method. The MPMO ISC with the two sampling approaches are also explained. Validations using the analytic function are shown in Section 3, and the solution optimality and efficiency of the EGO framework are compared those of the multi-point EI are used. The practical design results of high-lift, multi-element NLR 7301 airfoil are discussed in Section 4, followed by the conclusions and future work in Section 5.

2. Efficient global optimization with infill sampling criteria

The computational efficiency of the global optimization (EGO) method [13,14] is attributed to the following factors: 1) expensive function evaluation such as CFD analysis is replaced by a surrogate model, and 2) the surrogate model is updated adaptively with the additional sample points such that the size of the training sample set changes dynamically from the initial set. In this section, a brief overview of the Kriging is described, followed by the proposed ISC method in detail on the mathematical formulation and physical meaning. Conventional ISC methods are also introduced for a comparison purpose: the single point EI and various multi-point EI methods [5].

2.1. A Kriging surrogate model for the EGO

Unlike surrogate models based on the least square polynomial regression (PR) [14] and support vector regression (SVR) [14,15], the Kriging model [6] provides a probabilistic measure of estimation uncertainty as well as the prediction of the function value. It is a stochastic model using the weighted sum of correlations of all sample points with their weight coefficients determined by the maximum likelihood estimation (MLE) [16]. The Kriging model is known to be accurate for responses with nonlinearity, multiple modes, discontinuities and various noise sources [17,18]. It allows various types of correlation functions. As a black-box function analysis tool in the design framework, its characteristics of interpolation better suits the Kriging model rather than polynomial regression types which are more appropriate for the experimental data containing random errors [18]. Modified Kriging models with variable-fidelity function analysis can further enhance the accuracy and efficiency of the EGO [19,20] by including more high-accuracy data. Gradient values can be included in the training set as additional information on the slope as well as the ion. [20–22]. In the recent work of Jo et al. [23], the regression process is integrated into the gradient-enhanced Kriging model with variable-fidelity function analysis and mitigates noise influences from function values and gradient values.

As the mathematical formulation of the Kriging model is well described in the literature [6,16–23], only a brief summary is shown. A low-order, polynomial-based trend term $\bar{f}(\mathbf{x})$ in Eq. (1) is used in the Kriging model, where $f_j(\mathbf{x})$ is a basis function and the likelihood estimation coefficient of the trend term β is defined as $\hat{\beta}$ by the generalized least square (GLS) method [17].

$$\hat{y}(\mathbf{x}) = \bar{f}(\mathbf{x}) + Z(\mathbf{x}) = \sum_{j=1}^m \beta_j f_j(\mathbf{x}) + Z(\mathbf{x}) \quad (1)$$

To improve the numerical stability of the method, a distribution parameter, θ_j , which directly controls the covariance of any given two sample points in a correlation matrix, \mathbf{R} , is determined by penalized maximum likelihood estimation (PMLE), first suggested by Li et al. [24]. The PMLE utilized in the current study was recently proposed by Kwon et al. [25] and uses a data-driven and cross-validation method for the parameter estimation. The final Kriging prediction value and the mean squared error (MSE) representative of prediction uncertainty at an arbitrary location, \mathbf{x} , are defined in Eq. (2) and (3), respectively.

$$\hat{y}(\mathbf{x}) = \bar{f}(\mathbf{x}) + Z(\mathbf{x}) = \sum_{j=1}^m \beta_j f_j(\mathbf{x}) + Z(\mathbf{x}) = \mathbf{f}_x^T \hat{\boldsymbol{\beta}} + \mathbf{r}_x^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{F} \hat{\boldsymbol{\beta}}) \quad (2)$$

$$\text{where } \hat{\boldsymbol{\beta}} = (\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^T \mathbf{R}^{-1} \mathbf{y}$$

$$MSE(\mathbf{x}) = \hat{s}^2(\mathbf{x}) = \sigma_z^2 \left[1 - \begin{bmatrix} \mathbf{f}_x^T & \mathbf{r}_x \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{F}^T \\ \mathbf{F} & \mathbf{R} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f}_x \\ \mathbf{r}_x \end{bmatrix} \right], \quad (3)$$

Relatively short distance between sample locations makes the correlation matrix, \mathbf{R} , ill-conditioned and causes numerical instability and inaccurate function prediction. To improve the numerical stability, a small non-zero constant term, which is called a nugget, is added along the diagonal elements of correlation matrix. This reduces the condition number of the matrix \mathbf{R} and the numerical stability is enhanced and improves the accuracy of the Kriging prediction. From the parametric study, the nugget value is set as 1.0E-11.

2.2. Existing methods of infill sampling criteria

As the accuracy and computational efficiency of the EGO method greatly depends on the performance of the infill sampling criteria (ISC), various types of ISC have been proposed and a couple of them are introduced in this section: conventional, single point ISC with the exact expected improvement (EI) value, multiple point ISC using the exact EI, and multiple point ISC using approximated EI (q-EI). For the consistency, the optimization problem is set as the minimization problem hereinafter.

2.2.1. Single point ISC

One of the most used ISC is the EI [1] which literally represents an expected value of improvement, where the improvement is defined as difference between a predicted minimum and the true minimum found so far, and is considered as a random variable following a normal distribution of $N(\hat{y}, \hat{s}^2)$. The mathematical formulation of the EI is represented in closed form in Eq. (4). The first term of the RHS of Eq. (4) denoted as A represents the exploitation with the probability to find the smaller function value than the current minimum \mathbb{Y}_{min} , and the second term of B indicates exploration with the probability to find the high estimation uncertainty [26,27]. To find a point with the maximum EI is a single objective optimization problem where two objectives of exploitation and exploration are summed with weights represented by cumulative density function and probability density functions, respectively. The mathematical form is shown in Eq. (4). Throughout the adaptive sampling process of the EGO method, the ratio of the terms of A and B varies. The term B is larger than A at the beginning of the design iterations and the term A grows larger towards the later design stage when the ISC focus on searching the minimum.

$$EI(\mathbf{x}) = \begin{cases} \overbrace{\left(\mathbb{Y}_{min} - \hat{y}(\mathbf{x}) \right) \Phi \left(\frac{\mathbb{Y}_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})} \right)}^A + \overbrace{\hat{s}(\mathbf{x}) \phi \left(\frac{\mathbb{Y}_{min} - \hat{y}(\mathbf{x})}{\hat{s}(\mathbf{x})} \right)}^B, & \hat{s} > 0 \\ 0, & \hat{s} = 0 \end{cases} \quad (4)$$

A potential problem of the EI-based ISC method is that the additional point may be trapped in around the local minimum incumbent

at each iteration. If the sample points are clustered with proximity to one other, the method of regression and re-interpolation of the Kriging model can be used to avoid ill-conditioning of the correlation matrix [23], as described at the end of previous Section 2.1.

2.2.2. Multiple point ISC

The method of multiple point-based ISC for the EGO framework is suggested by Haftka et al. [4] and Ginsbourger et al. [5]. It adds multiple infill points at each design iteration and becomes greatly efficient with parallel computation which became nearly standard for high-fidelity physics calculation. Details of the computation of the multivariate EI values are referred to [5], and the assumption of the multivariate Gaussian process for the multiple functions and the resultant mathematical formulations of the multivariate EI value are shown in Eq. (5) and (6). The definitions of conditional covariance of S_q is also made in the reference papers and omitted here.

$$\begin{aligned} EI \left[(\mathbf{x}^{n+1}, \dots, \mathbf{x}^{n+q}) \right] &= E \left[\max \{ I(\mathbf{x}^{n+1}), \dots, I(\mathbf{x}^{n+q}) \} \mid \mathbb{Y} \right] \\ &= E \left[\max \{ (\mathbb{Y}_{min} - Y(\mathbf{x}^{n+1})), \dots, (\mathbb{Y}_{min} - Y(\mathbf{x}^{n+q})) \} \mid \mathbb{Y} \right] \\ &= E \left[\mathbb{Y}_{min} - \min \{ Y(\mathbf{x}^{n+1}), \dots, Y(\mathbf{x}^{n+q}) \} \mid \mathbb{Y} \right] \end{aligned} \quad (5)$$

where $\mathbb{Y} = \{ Y(\mathbf{x}^1), \dots, Y(\mathbf{x}^n) \}$

$$\{ \{ Y(\mathbf{x}^{n+1}), \dots, Y(\mathbf{x}^{n+q}) \} \} \sim N \left((\hat{y}(\mathbf{x}^{n+1}), \dots, \hat{y}(\mathbf{x}^{n+q})), S_q \right) \quad (6)$$

where an input \mathbf{x} is located in the design space with a dimension of $d \in \mathbb{R}^d$ and the exact function value at \mathbf{x} is $Y(\mathbf{x})$, then a set of known exact function values for n number of sample points are defined as $\mathbb{Y} = \{ Y(\mathbf{x}^1), \dots, Y(\mathbf{x}^n) \}$.

The algorithms of the approximated q-EI

q-EI algorithm with a Kriging Believer	q-EI algorithm with a Constant Liar
Do $i = 1, \dots, q$	Do $i = 1, \dots, q$
$\mathbf{x}^{n+i} = \operatorname{argmax}_{\mathbf{x} \in D} EI(\mathbf{x})$	$\mathbf{x}^{n+i} = \operatorname{argmax}_{\mathbf{x} \in D} EI(\mathbf{x})$
$\mathbb{X} = \mathbb{X} \cup \{ \mathbf{x}^{n+i} \}$	$\mathbb{X} = \mathbb{X} \cup \{ \mathbf{x}^{n+i} \}$
$\mathbb{Y} = \mathbb{Y} \cup \{ \hat{y}(\mathbf{x}^{n+i}) \}$	$\mathbb{Y} = \mathbb{Y} \cup \{ L \}$
	where, $L \in \{ \mathbb{Y}_{min}, \mathbb{Y}_{mean}, \mathbb{Y}_{max} \}$
END	END

However, due to prohibitive computational cost related to the Monte Carlo simulation (MCS) [28] to calculate the exact value of multivariate EI, Ginsbourger et al. suggested an alternative ISC method based on an approximate EI value. It finds the q number of sample points in a sequential manner using the same procedure of the single point EI, and a function response of a new point is assumed as an arbitrary value rather than the value from computation or simulation. Until a total of q points are selected, the distribution parameters, θ_j , of the Kriging surrogate model are fixed at the values determined by existing n sample points. Authors suggested two approaches depending on the arbitrary value: a Kriging Believer (KB) and a Constant Liar (CL). The KB method uses a Kriging response value temporarily for the exact function value of the new point. The CL method uses an arbitrary constant of L chosen from the existing function responses, which can be a minimum, an average, or a maximum value of the observed responses and denoted as \mathbb{Y}_{min} , \mathbb{Y}_{mean} , and \mathbb{Y}_{max} , respectively. The schematics of algorithms of q-EI using the KB and CL are shown in Table 1. Although the approximate q-EI method is more efficient than the exact q-EI method, depending on the choice of a constant value and a resultant set of different infill sample points, the efficiency of the design process can be enhanced or deteriorated.

Table 1

Pseudo-algorithm for dynamic switching approach.

Dynamic switching approach	
If $\mu_k(\hat{s}^2) > T$	Assign a rank on the individual in population in descending order of $(\hat{s}^2)_i$ ($i = 1, \dots, n_{pop}$)
	Select q points around the $P_{\hat{s}^2_{max}}$
else	Assign a rank on the individual in population in descending order of $(\hat{y})_i$ ($i = 1, \dots, n_{pop}$)
	Select q points around the $P_{\hat{y}_{min}}$
End if	

2.3. Multi-point and multi-objective infill sampling criteria (MPMO ISC)

The method of the MPMO ISC is developed as an alternative method to the multi-point EI explained in the previous section. The basic idea is to model the trade-off between two competing merits of the exploration and the exploitation in a form of the multi objective optimization problem. In this multi-objective approach, two metrics representing the exploration and the exploitation quantitatively are the (maximum) variance value of estimation uncertainty \hat{s}^2 and the (minimum) predicted response value \hat{y} from the Kriging surrogate model, respectively. The Pareto front from the multi-objective optimization provides a set of optimal candidate points with quantitative measure of each metric. As the design iteration proceeds, the goal of ISC generally changes such that the exploration is required during initial design iterations to increase the global accuracy of the model, and once a certain level of global accuracy is achieved, the exploitation is necessary to better approximate the local area around the optimum. One important feature is that various combinations are possible for selecting multiple points and criteria for such combination can change dynamically depending on the global versus local characteristics of the surrogate model at the current iteration.

Advantages of the method are several. First, as pointed out in the work of Haftka et al. [4] and Ginsbourger et al. [5], the parallel computation can be effective for simultaneous function evaluations of the multiple infill sample points. As the MPMO ISC selects q additional sample points per one execution of the multi-objective optimization, which increase design efficiency by a factor of $q \times n$, where q is the total number of additional sample points at each iteration and n is the number of total design iterations. In addition, unlike the approximated q-EI ISC method which selects the q sample points sequentially through a series of the incomplete Kriging models using one arbitrary constant or one of the Kriging prediction values, the MPMO ISC approach is solely relying on the Kriging model updated by sample points evaluated by true function analysis. Although the approximated q-EI method does not re-compute a distribution parameter θ_j during the infill sampling procedure, the re-construction and inversion of the Kriging correlation matrix is required during the q number of optimizations to find the maximum EI point.

However, the main issues of the MPMO ISC on the adaptive sampling strategy need to be addressed: (1) which should be more focused out of the exploration and the exploitation, and (2) how to determine the weights of the two metrics. The following subsections describe two different approaches to directly address the issues: a balancing approach and a dynamic switching approach. With three additional infill sample points per design iteration, the balancing approach adds two infill sample points that correspond to those on the opposite ends of the Pareto front, one for the exploration and the other for the exploitation, respectively, and one additional point in between the two end points. The dynamic switching approach selects all three points near the one far end out of two. For both approaches, local point clustering is checked by the metric of a clustering radius to prevent the repetitive selection of the nearby sample points of similar input parameters. This is because the clustered local points deteriorates the condition number of the Kriging covariance matrix and greatly lowers the accuracy of the Kriging predictions.

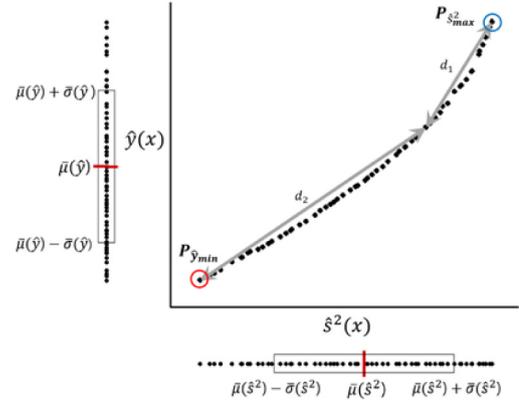


Fig. 1. Pareto-front for ISC.

2.3.1. A balancing approach

At a given design iteration and prior to updating the Kriging surrogate model, a sub-procedure of the multi-objective optimization is carried out and produces the Pareto front of the optimum candidates. Fig. 1 shows an example of the multi-objective Pareto front where two axes represent two objectives of finding the large estimation uncertainty (\hat{s}^2) to improve the global accuracy (exploration) and minimizing the predicted value of the response (\hat{y}) for the local accuracy (exploitation), respectively. The point $P_{\hat{s}^2_{max}}$ corresponds to the maximum estimation uncertainty on the Pareto-front and the point $P_{\hat{y}_{min}}$ to the minimum predicted responses, respectively. Therefore, points near the $P_{\hat{s}^2_{max}}$ tend to have large \hat{s}^2 values, while those around the $P_{\hat{y}_{min}}$ tend to have a lower function response than the current minimum. The balancing approach chooses the points of $P_{\hat{s}^2_{max}}$ and $P_{\hat{y}_{min}}$ at each design iteration regardless of the particular distribution of points on the Pareto front and tries to balance the objectives of the exploration and the exploitation. The balancing approach also selects one additional point somewhere on the Pareto-front between $P_{\hat{s}^2_{max}}$ and $P_{\hat{y}_{min}}$, so that the point improves the accuracy of the current Kriging surrogate model, whether local or global.

However, the exact quantification of the global and local accuracies requires the cross-validation with a large number of sample points, which is not practical and may become computationally prohibitive. An alternative way to determine which objective needs to be focused on is based on the statistics of optimal points ranked in a single objective only in terms of mean and variance. Low variance with a low mean value of an objective indicates the convergence with desired accuracy in finding minimum objective, and therefore variance is used to diagnose whether the current Kriging model requires exploration or exploitation. For example, large variance in the MSE values of the estimation uncertainty indicates poor convergence of the global accuracy and more sample points are needed for the exploration. Likewise, the small variance value in the minimum \hat{y} values implies that additional samples point need to be infilled for further refinement near the minimum. A direct comparison of the magnitude of the variance values is made by their ratio. For a fair comparison between two variance values irrespective of the units, the coefficient of variation (COV) [9], $[\hat{c}_v]_i$, is introduced, and those for the MSE value \hat{s}^2 and predicted response value \hat{y} are calculated using Eq. (7).

$$[\hat{c}_v]_{\hat{s}^2} = \frac{\bar{\sigma}(\hat{s}^2)}{\bar{\mu}(\hat{s}^2)}, \quad [\hat{c}_v]_{\hat{y}} = \frac{\bar{\sigma}(\hat{y})}{\bar{\mu}(\hat{y})} \quad (7)$$

Values with an overbar symbol in Eq. (8), represent that these are estimated from the existing training set:

$$\begin{aligned} \bar{\mu}(\hat{s}^2) &= \frac{1}{n_{pop}} \sum_{i=1}^{n_{pop}} (\hat{s}^2)_i, \quad \bar{\sigma}(\hat{s}^2) = \frac{1}{n_{pop}-1} \sum_{i=1}^{n_{pop}} [(\hat{s}^2)_i - \bar{\mu}(\hat{s}^2)]^2 \\ \bar{\mu}(\hat{y}) &= \frac{1}{n_{pop}} \sum_{i=1}^{n_{pop}} (\hat{y})_i, \quad \bar{\sigma}(\hat{y}) = \frac{1}{n_{pop}-1} \sum_{i=1}^{n_{pop}} [(\hat{y})_i - \bar{\mu}(\hat{y})]^2 \end{aligned} \quad (8)$$

where n_{pop} is the total number of optimal candidates on the Pareto-front.

The third point P_3 is selected based on the ratio of the two COVs of $[\hat{c}_v]_{\hat{s}^2}$ and $[\hat{c}_v]_{\hat{y}}$. To relate the COV ratio to a sample point on the Pareto-front, the Euclidean distances to a point P from the two extreme points of $P_{\hat{s}^2_{max}}$ and $P_{\hat{y}_{min}}$ are first computed as d_1 and d_2 as in Eq. (9):

$$d_1(P) = \sqrt{\left[\hat{s}^2(P) - \hat{s}^2(P_{\hat{s}^2_{max}})\right]^2 + \left[\hat{y}(P) - \hat{y}(P_{\hat{s}^2_{max}})\right]^2}$$

$$d_2(P) = \sqrt{\left[\hat{s}^2(P) - \hat{s}^2(P_{\hat{y}_{min}})\right]^2 + \left[\hat{y}(P) - \hat{y}(P_{\hat{y}_{min}})\right]^2}$$
(9)

The third point P_3 is chosen such that the ratio of d_1/d_2 is equivalent, or closest to that of the two COVs, and this relation is shown in Eq. (10). Note that the indices for two ratios are reversed.

$$d_1 : d_2 = [\hat{c}_v]_{\hat{y}} : [\hat{c}_v]_{\hat{s}^2}$$
(10)

In this way, a point close to $P_{\hat{s}^2_{max}}$ is selected to improve the global accuracy when the variance value of MSE $\bar{\sigma}^2(\hat{s}^2)$ is relatively large ($d_1 > d_2$), or the local accuracy vice versa. The three additional sample points, including two extreme points on the Pareto front and the one in between, are selected at each design iteration to improve the accuracy of the Kriging model. This procedure is iterated until the termination criteria are met.

2.3.2. A dynamic switching approach

The dynamic switching approach is developed to decide dynamically on which criterion should be more focused at each design iteration out of the two criteria of the exploration or the exploitation. From the Pareto front shown in Fig. 1, sample points near the one end of the front, rather from both ends, can be selected. The dynamic switching approach keeps track of the global accuracy of the Kriging model throughout the entire design iterations by the average value of MSE of candidate points on the Pareto front, $\mu_k(\hat{s}^2)$, where k represents a current iteration step. The ISC switch the infilling criteria between the exploitation and the exploration depending on the average value of MSE distribution of all candidate points, $\mu_k(\hat{s}^2)$, on the Pareto-front at the k th design iteration with the threshold value, T . The threshold value is calculated by multiplying the switching parameter, α , and the maximum value of $\mu_k(\hat{s}^2)$ found so far, and shown as

$$T = \alpha \times \max[\mu_k(\hat{s}^2)]$$
(11)

If $\mu_k(\hat{s}^2)$ is larger than T , then points around the $P_{\hat{s}^2_{max}}$ are selected for the exploration, otherwise; points around the $P_{\hat{y}_{min}}$ are sampled for the exploitation. Additional samples are infilled in the training set to avoid ill-conditioning of the covariance matrix and to guarantee the global accuracy.

Clustering parameters, θ_G and θ_L , are defined to avoid additional samples being clustered near the existing sample points. The geometric size of the training set, D_k , is defined by the L_2 norm of the two most distant points in the set of all sample points. If $\mu_k(\hat{s}^2)$ is larger than T , the distance between any pair in the neighboring points of the Pareto front is calculated, and the points whose distance is larger than $\theta_G D_k$ are selected around the $P_{\hat{s}^2_{max}}$. If $\mu_k(\hat{s}^2)$ is smaller than T , those points whose distance is larger than $\theta_L D_k$ are selected around $P_{\hat{y}_{min}}$. Three parameters of α , θ_G , and θ_L varies from 0 to 1.

2.4. The EGO framework with the MPMO ISC

The EGO design framework includes two optimization procedures: the primary optimization to find the minimum of the objective of the design problem and the secondary optimization to select multiple infill sample points to update the Kriging model as summarized in Table 2. The primary optimization problem is formulated as a single-objective problem with multiple constraints. The constraints are handled using a sub-problem approach suggested by Sasena et al. [26] that uses Kriging

Table 2

The multi-layered EGO design framework.

Primary optimization problem		Secondary optimization problem for MPMO ISC	
Single objective with multiple constraints		Multi objectives with multiple constraints	
Minimize	$\hat{y}(\mathbf{x})$	Minimize	$\hat{y}(\mathbf{x})$
subject to	$\hat{c}_j(\mathbf{x}) \leq 0$	Maximize	$\hat{s}^2(\mathbf{x})$
		subject to	$\hat{c}_j(\mathbf{x}) \leq 0$

models of constraint functions and the feasibility of the optimum candidates are checked by the model. Thus, the constrained optimization problem does not need to be re-formulated as the unconstrained one like in the penalty method or the probability method. Sasena et al. pointed out that those two methods may distort the constrained boundary of design space, leading to infeasible optimum results. The secondary optimization, which is formulated as the multi-objective problem for infill sampling also carried out with constraints. That is, the infill sampling process is carried out only in feasible regions, and the global accuracy of the surrogate model can be affected by the size of feasible region. An approach such as the trust region method [29] can be applied to consider both global accuracy in the infeasible regions and the solution feasibility, although the approach is not implemented in the present study.

The MPMO ISC method is directly integrated into the EGO framework to update the Kriging surrogate model at each design iteration, and the values of the objective and constraint functions are estimated through separate surrogate models. For the optimization process, the non-dominated sorting genetic algorithm (NSGA-II) [7] is utilized in the current study as it is known to be robust.

The overall design procedure of the EGO framework with the MPMO ISC can be seen in Fig. 2. A total of $(d + 1)(d + 2)/2$ initial sample points [18] are randomly selected using the LHS method [10], where d represents the number of design variables (or dimension of design space). The initial Kriging surrogate model is created correspondingly, and the primary optimization is initiated to search the optimum point. If the accuracy of the Kriging model and the optimality of the current design solution do not satisfy termination criteria, then the adaptive sampling is carried out using the MPMO ISC methods as shown in Fig. 2(b). The Pareto front from the secondary optimization is used to determine the additional sample points. Then, the high-fidelity simulation such as CFD analysis is carried out in parallel to calculate output/objective and constraint responses for those additional sample points. The secondary design optimization is carried out at each primary design iteration. This sequential, primary design continues until the termination criteria are satisfied. Parallel computation is used to maximize computational efficiency for function and constraints evaluations of the multiple infill sample points as well as of the initial training sample points.

3. Validation of the EGO with the MPMO ISC

This section shows the validation results of the proposed EGO framework which uses the MPMO ISC method of both the balancing and the dynamic switching approach. An optimization problem with simple constraints of bounds on design variables is set for two 2D analytic functions: a six-hump camel function and a Matlab peak function. For more complicated constraints, additional surrogate models should be constructed in explicit form for each constraint, and the current GA algorithm handles them effectively without any changes in the infill sample strategy.

For comparison purposes, the approximated q-EI method [5] and the LHS sampling method are also implemented for the EGO framework, respectively. For all methods, the additional sample points are selected to update the Kriging surrogate model sequentially and adaptively, and the accuracy of the model and the optimality of the solutions are compared one another. The gradient-based optimization method is also carried out to compare the efficiency and optimality of the proposed EGO results.

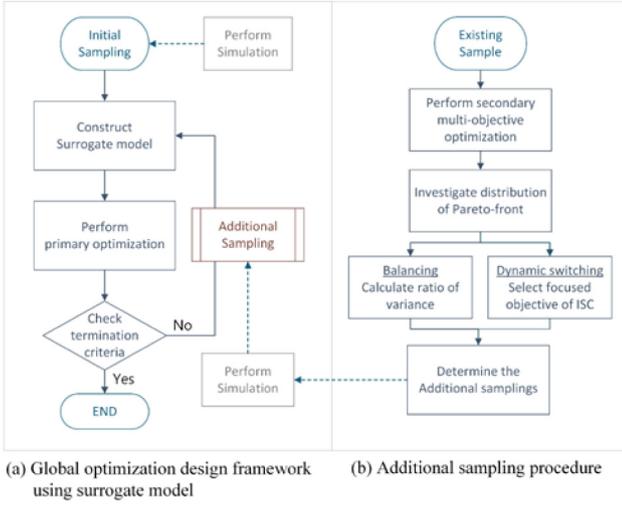


Fig. 2. The EGO design framework with MPMO ISC approach.

3.1. Problem description

3.1.1. Optimization problems

The test problems of two-dimensional six-hump camel back and Matlab peak functions have different characteristics in terms of local and global minima and the degree of non-linearity. The mathematical formulations and the exact minimum values are summarized in Table 3. The six-hump camel back function has two global minima while the Matlab peak function has one global and one local minima. Fig. 3 shows the functions and the 2D iso-contours, with the locations of minima by the symbols of ‘x’ in blue.

3.1.2. Performance metrics

The best performance of the EGO framework is defined as the ability to find accurate optimum/optima at computation cost as low as possible. As the overall quality of the surrogate model should be addressed by the local and global accuracies, two metrics are defined. First, to check the solution optimality, or the local accuracy, relative errors are computed: difference in x location between the optimizer and the true minimum, ϵ_x , and the difference in function value of y , ϵ_y . Those are formulated in Eq. (12), where the value of $\hat{y}^* = f_{krig}(\mathbf{x}^*)$ represents the predicted function minimum, and y_{min} is the true optimum at \mathbf{x}_{min} . Second, a total of 2500 validation points are pre-selected and evenly distributed in two-dimensional design space. The predicted function values are cross-validated at these points with the exact function values by the RMSE value as shown in Eq. (13).

$$\epsilon_x = \frac{\|\mathbf{x}^* - \mathbf{x}_{min}\|}{\|\mathbf{x}_{min}\|}, \quad \epsilon_y = \frac{|\hat{y}^* - y_{min}|}{|y_{min}|} \quad (12)$$

$$RMSE = \frac{1}{N_{test}} \sqrt{\sum_{i=1}^{N_{test}} [\hat{y}(x_i) - y(x)]^2} \quad (13)$$

3.2. A parameter study for the MPMO-ISC with a dynamic switching approach

The method of MPMO-ISC with the dynamic switching approach involves three parameters: a switching parameter, α , a clustering parameter for the exploration, θ_G , and a clustering parameter for the exploitation, θ_L . A parameter study to find optimal values of parameters leading to the least number of total samples is carried out with analytic functions used in Section 3. Due to the inherent randomness of initial training set for the sampling process and of the GA algorithm in populating individuals, fifty sets of six initial samples are randomly selected using the LHS method for the Monte Carlo Simulation (MCS)

runs and each set is used for separate optimization. The total number of samples needed to find the function minimum is averaged over the fifty MCS runs to compare computational cost. For both analytic functions, the termination criterion is set for the relative error of optimum function value between current and previous design steps to be less than $1.0E-5$. The average number of samples for the convergence is compared with that by the balancing approach.

Figs. 4 and 5 show the results of the parameter study for six hump camel back function and Matlab peak function, respectively, in terms of the average number of total samples for the convergence. In the parametric study, the switching parameter, α , and the clustering parameter for the exploration, θ_L , are varied between zero and one, while the clustering parameter for the exploitation, θ_G , is varied from 0.001 to 0.1, with two other parameters of θ_L and α are fixed. It is shown that an optimal range of parameters for the minimal number of required sample points exists for all parameters, and corresponds to $\alpha = 0.3$, $\theta_G = 0.3 \sim 0.4$, and $\theta_L = 0.05 \sim 0.005$. These values are used in the practical design application shown in Section 4. However, it should be noted that finding a constant value that is optimal for all kinds of design problems is difficult, and a certain amount of inefficiency is inevitable if non-optimal parameters are used for other design problems. However, the values ranging from 0.2 to 0.5 for α and θ_G , and from 0.002 to 0.1 for θ_L appear to be reasonable without much sensitivity, based on the numerical test with other analytical functions.

3.3. Cost-efficiency

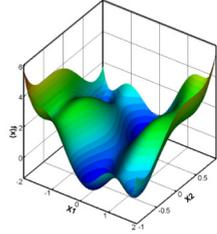
Three types of the ISC methods are compared: 1) the MPMO ISC method with the balancing approach, 2) the MPMO ISC method with the dynamic-switching approach, and 3) the approximated q-EI method. For the approximated q-EI method, three different approximations to the maximum EI value are used: Kriging Believer (KB), Constant Liars (CL) with a Kriging predicted minimum (CLmin), and CL with a Kriging predicted maximum (CLmax). For the MPMO ISC method with the dynamic switching approach, the optimal values of parameters which are found in the previous section is used. Starting from a total of six sample points, three additional sample points, or less if clustering is detected, are appended to the existing training sample point set at each design iteration and the Kriging surrogate model is updated. The performance metrics are computed correspondingly. The termination criteria are the values of ϵ_x and ϵ_y , to be less than 0.01%. To mitigate the randomness of the initial sample point set for the surrogate model construction and in the population of the genetic algorithm for the optimization, a total of 50 sets of different initial points are randomly chosen using the Latin Hypercube Sampling (LHS) method. For each set, the MPMO-ISC strategy is applied and the average number of samples for the design convergence is calculated. The values of average, standard deviation, and the minimum number of total required sample points for convergence, are summarized in Table 4 with respect to 5 an individual ISC method.

For both six-hump camel back function and Matlab peak function, the results of the MPMO ISC with the balancing approach and the dynamic switching approach demonstrate the best performance at the lowest computation cost as shown in Tables 4 and 5. For the Matlab peak function, the dynamic switching approach requires only one more design iteration than the balancing approach, which indicates no difference in the computational efficiency. Since the dynamic switching approach needs three parameters which can be varied case by case, the standard deviation of the total number of samples was larger than that of the balancing approach. However, once the parameter is set correctly, the total number of samples is dramatically reduced as shown in the row of minimum number of samples; the minimum number of samples for the balancing approach is 27, but the one for the dynamic switching approach is 12.

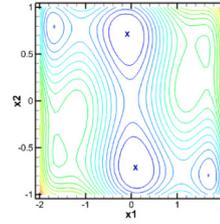
The approximated q-EI method with the Kriging Behavior, the minimum Constant Liar and the maximum Constant Liar show about 30%

Table 3
The analytic functions.

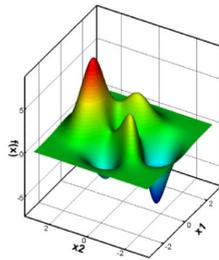
	Analytic Function	Bounds of the design variable
Six-hump camel back [30]	$f = (4 - 2.1x_1^2 + \frac{1}{3}x_1^4)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$	$x_1 \in [-2, 2]$ $x_2 \in [-1, 1]$
	$f_{min} = f(-0.0898, 0.7126) = f(0.0898, -0.7126) = -1.0316$	
Matlab Peak [27]	$f = 3(1 - x_1)^2 \exp[-x_1^2 - (x_2 + 1)^2]$ $-10(\frac{1}{5}x_1 - x_1^3 - x_2^5) \exp[-x_1^2 - x_2^2] - \frac{1}{3} \exp[-(x_1 + 1)^2 - x_2^2]$	$x_1 \in [-3, 3]$ $x_2 \in [-3, 3]$
	$f_{min} = f(0.2282, -1.6255) = -6.551133$	



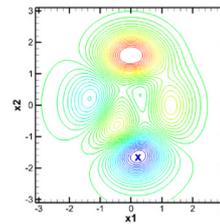
(a) Configuration of six-hump camel back function



(b) Iso-contour of six-hump camel back function



(c) Configuration of matlab peak function



(d) Iso-contour of matlab peak function

Fig. 3. Configurations and iso-contours of the analytic functions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

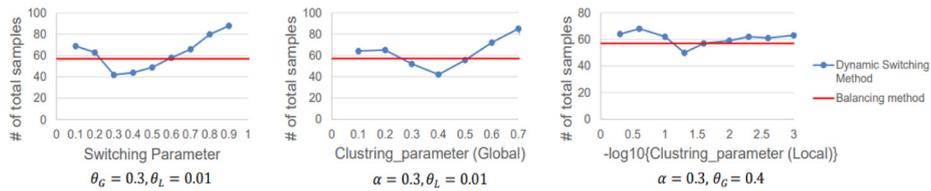


Fig. 4. Parameter study result for six hump camel back function.

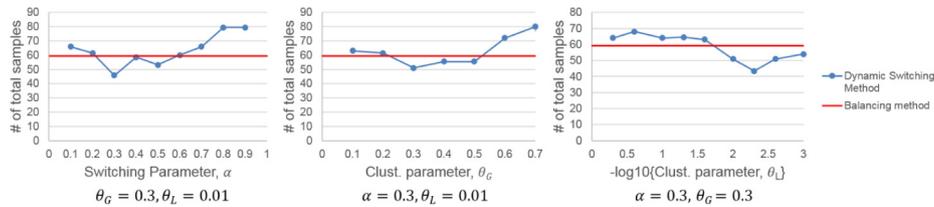


Fig. 5. Parameter study result for Matlab peak function.

Table 4
A total required number of sample points (Six hump camel back function) for the convergence.

	MPMO-ISC (Balancing)	MPMO-ISC (Dynamic switching)	Q-EI (KB)	Q-EI (Max. CL)	Q-EI (Min. CL)
Average number of total samples	54	48	69	74	79
Standard deviation of total samples	12	15	16	13	15
Minimum number of total samples	27	15	42	45	52

Table 5

A total required number of sample points (Matlab Peak Function) for the convergence.

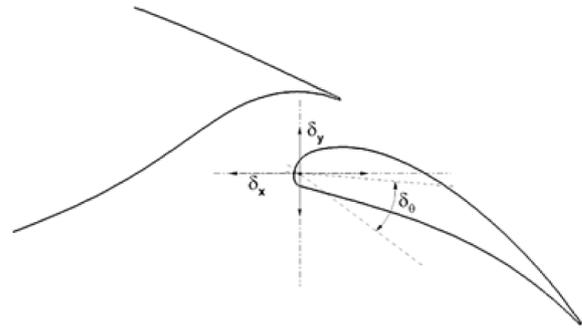
	MPMO-ISC (Balancing)	MPMO-ISC (Dynamic switching)	Q-EI (KB)	Q-EI (Max. CL)	Q-EI (Min. CL)
Average number of total samples	51	54	66	78	79
Standard deviation of total samples	12	19	16	13	15
Minimum number of total samples	27	12	36	45	48

to 55% higher computational cost for the design convergence. Based on the results from the two analytic functions, it is shown that the MPMO ISC method with both balancing and dynamic switching approaches shows relatively robust and accurate results. On the other hand, the accuracy of the approximated q-EI method is sensitive to the specific approximation value of the EI. It is not straightforward to decide a priori which approximation value should be used and predict the corresponding accuracy and efficiency. However, the MPMO ISC method also involves some degree of arbitrariness in determining the parameter values: the number of infill sample points at each iteration (three or less in the current case) and the parameter values in the dynamic switching approach.

In summary, for real engineering design applications where the characteristics of the design space are not known in advance, the choice of the MPMO ISC method becomes reasonable and robust especially when the computing resources are limited, and guarantees both the local and global accuracies.

For more fair comparisons, the efficiency of the MPMO ISC method is also compared with that of the optimization method which does not use the surrogate model. Both gradient-based and -free optimizations are conducted with direct function evaluation with no surrogate modeling. A sequential least squares programming (SLSQP) [11] is an algorithm for the gradient-based optimization where sensitivity information is calculated analytically. The NSGA-II algorithm [7] is employed for the gradient-free optimization with direct function evaluations. The results from the best performing approximated q-EI method and the MPMO ISC with the balancing approach are directly compared in Tables 6 and 7. The metric for the efficiency comparison is computational cost for the function evaluations and Kriging predictions to carry out the ISC method, and is summarized in Tables 6 and 7.

If the computation time for the function and its partial derivative (N_{df/dx_1} or N_{df/dx_2}) calculation is assumed to be similar, then the gradient-free optimization without a surrogate model requires as many as 3,200 function evaluations for the six-hump camel-back function, whereas the gradient-based optimization needs about 120 computations for function and derivative calculations depending on the location of the starting point. On the other hand, the current MPMO ISC method takes 54 function evaluations, although it needs a considerable number of the Kriging prediction for the primary optimization as well as the sub-optimization to choose the ISC (whether the EI or the MPMO ISC). The Kriging prediction of the function output response typically takes milliseconds to carry out, and if we consider the entire computation time to find a function minimum, then the current MPMO ISC method can be compatible, for simple analytic functions, to the gradient-based optimization in efficiency. However, the ratio of the computational cost between function (and derivative) computation to the Kriging prediction becomes huge if the design application uses high-fidelity CFD analysis for function evaluation, which may take up to several hours even with parallel computing resources. Computation of the function and/or constraint derivatives with respect to a large number of design variables is also challenging, although the state-of-the-art adjoint-based sensitivity analysis can be alternatively used [31] with its biggest advantage being almost no dependence on the number of design variables. In those design applications, the efficiency of the MPMO ISC method is significant. Also, the limitation of the local search by the gradient-based optimization does not exist in the EGO with the MPMO ISC method as it can explore larger design space. Although the associated computational cost saving from the parallel computation in the MPMO ISC method to carry out function evaluations of multiple infill sample points is not directly formulated in Table 6, it can further improve the computational efficiency.

**Fig. 6.** Configuration of Design variables.

4. Practical design application

As a practical design application of the EGO framework with the MPMO ISC approach, high-lift multi-element airfoil is optimized in shape to maximize lift at a take-off condition. The EGO design framework with the MPMO ISC method and the approximated q-EI method, as well as the gradient-based design method with direct high-fidelity CFD evaluations are used and their results are compared. The derivative values are computed by the finite-difference and the adjoint solution method.

4.1. Design problem description

A baseline airfoil is chosen as the NLR 7301 multi-element airfoil [12]. The objective of aerodynamic design is to maximize lift at drag force maintained as the baseline value at the take-off condition. The design variables are angle of attack, flap deflection angle, gap and overlap of flap. The design flow condition is at Mach number of 0.185, Reynolds number of 2.51 million and angle of attack of 13.1°. The baseline NLR 7301 multi-element airfoil has a flap of the 32% of the chord length, 2.6% of the gap and 5.3% of the overlap relative to the main airfoil. Baseline configuration of main airfoil and flap is shown in Fig. 6 and 7.

The mathematical formulation of the design problems is shown in Eq. (14) as a single objective optimization problem. The constraint of non-increasing C_d is included in the objective function, and consequently the design problem reduces to the unconstrained optimization and requires only a single surrogate model. The optimization process is iterated until the relative error in the objective function value between previous and the current design steps to be less than 0.01%.

$$\underset{\alpha, \delta_\theta, \delta_x, \delta_y}{\text{minimize}} \quad \left(1.0 - \frac{C_l^*}{C_l}\right)^2 + \left(1.0 - \frac{C_d^*}{C_d}\right)^2 \quad (14)$$

where, $C_l^* = 1.5C_{l, \text{baseline}}$, $C_d^* = C_{d, \text{baseline}}$

Instead of using conventional definitions of flap gap and overlap, the location of flap is defined by horizontal and vertical displacements in a cartesian coordinate system to conveniently handle geometry variation and CFD mesh generation. Mesh deformation for the CFD analysis is carried out automatically through the software journaling technique [32]. The flap geometry changes from the baseline configuration depicted in Fig. 6 corresponding to the variation of the design variables with their upper and lower bounds listed in Table 8.

Table 6
Performance comparison of Six-hump camel back function.

	Average number of actual function N_f and derivative N_{df/dx_i} evaluations	Number of function prediction
Gradient-based optimizer (SLSQP)	120	N/A
Gradient-free optimizer with direct function evaluations	3210	N/A
Gradient-free optimizer with the surrogate model using the approximated q-EI method with Kriging Behavior	69 – initial samples : 6 – infill samples : 63	10.75 million – Optimization : 256 pop. × 500 gen. × 21 iter. – Maximize EI : 256 pop. × 500 gen. × 21 iter. × 3 add. samples
Gradient-free optimizer with the surrogate model using the MPMO ISC with the balancing approach	54 – initial samples : 6 – infill samples : 48	4.10 million – Optimization : 256 pop. × 500 gen. × 16 iter. – MPMO ISC : 256 pop. × 500 gen. × 16 iter.

Table 7
Performance comparison of Matlab peak function.

	Average number of actual function N_f and derivative N_{df/dx_i} evaluations	Number of function prediction
Gradient-based optimizer (SLSQP)	221	N/A
Gradient-free optimizer with direct function evaluations	1920	N/A
Gradient-free optimizer with the surrogate model using the approximated q-EI method with Kriging Behavior	66 – initial samples : 6 – infill samples : 60	10.24 million – Optimization : 256 pop. × 500 gen. × 20 iter. – Maximize EI : 256 pop. × 500 gen. × 20 iter. × 3 add. samples
Gradient-free optimizer with the surrogate model using the MPMO ISC with the balancing approach	51 – initial samples : 6 – infill samples : 45	3.84 million – Optimization : 256 pop. × 500 gen. × 15 iter. – MPMO ISC : 256 pop. × 500 gen. × 15 iter.

Table 8
Design variable bound.

Design variable	Symbol	Baseline	Lower bound	Upper bound
Angle of attack	α	13.1°	7.0°	13.0°
Flap deflection angle	δ_θ	-14.5°	-30.0°	-12.0°
Horizontal displacement	δ_x	0.0%c	-5.0%c	10.0%c
Vertical displacement	δ_y	0.0%c	-2.0%c	2.0%c

4.2. High-fidelity aerodynamic analysis using CFD

The function evaluations of a set of initial sample points and adaptive infill sample points are carried out through high-fidelity CFD analysis using parallel CPUs. The Reynolds-averaged Navier–Stokes (RANS) equations are solved by a SU2 flow solver [33]. The flow solver uses an unstructured grid topology and includes a computational design suite with a sensitivity analysis module which is based on continuous adjoint formulation. Automatic mesh deformation using a torsional spring analogy or a free-form deformation (FFD) method for three-dimensional shape variation are available in the design framework along with the choice of various gradient-based optimization algorithms.

To validate the flow solver, flows around the baseline airfoil (NRL 7301 with the flap) are solved at the take-off condition and the results are compared with the experimental data. The O-type unstructured grid system with 64,201 elements and 32,531 nodes is generated using the T-REX model of Pointwise software [32] and shown in Fig. 7(a). For the spatial discretization, the 2nd order centered differencing scheme with

JST artificial dissipation [34] is used, and time is integrated using the second-order of Euler-implicit method [35]. Viscous flux is discretized using the centered differencing method with a Spalart–Allmaras turbulence model [36]. The two-level multi-grid technique [37] is used to accelerate the convergence. Pressure coefficients on the airfoil surface and near-field pressure coefficients are plotted in Figs. 7(b) and 7(b), respectively, showing good agreement with experimental data.

4.3. Design results

Initially, a set of 15 initial sample points are generated by the LHS method for the four dimensional design space. This number is notably small considering a general sampling rule based on $(d+1)(d+2)/2$ [18], but was sufficient for the design due to the efficiency of the MPMO ISC. Initial sample points are evaluated by the CFD analysis, and the initial Kriging model is constructed correspondingly. For infill sampling method, the MPMO ISC with the balancing and the dynamic switching approach as well as the approximated q-EI method with KB, CLmax, and CLmin are used. To save computational time considerably, an

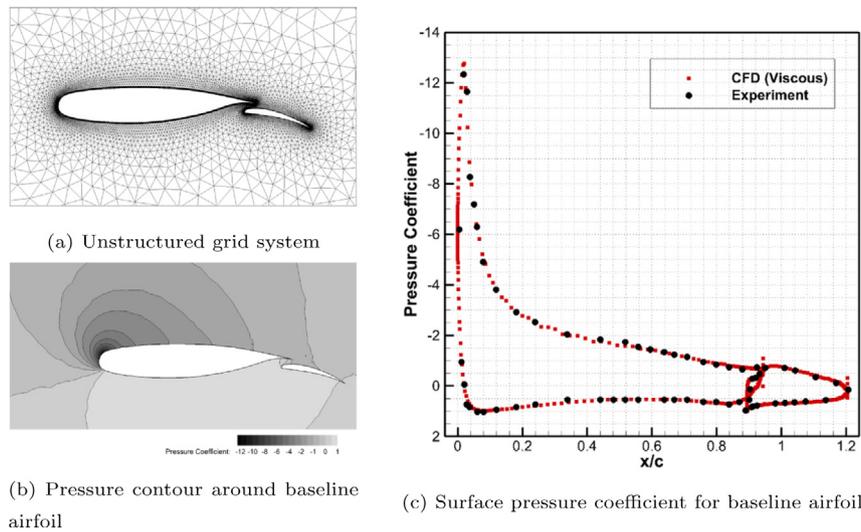


Fig. 7. Validation of CFD analysis for baseline airfoil.

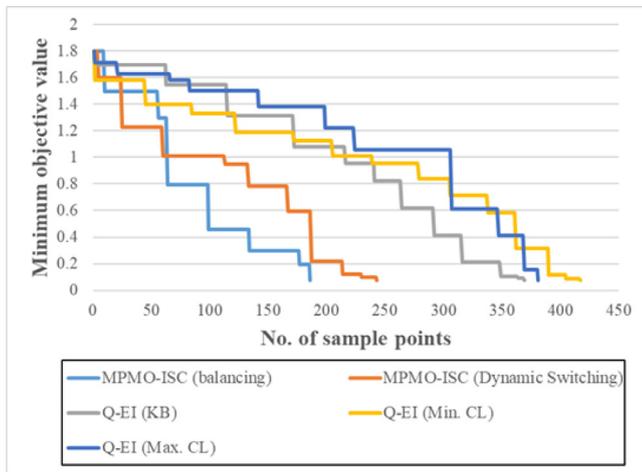


Fig. 8. Comparison of objective value variations w.r.t the number of sample points.

alternative to finding the optimal value by the GA with the Kriging model, the minimum objective value of existing sample points at each design iteration is regarded as the minimum point. This approach does not harm the original purpose of the EGO framework, because the optimal point searched by the optimization procedure and the minimum function value of sample points will converge to the same value, if the exploitation performance of the ISC method is good and

refines design space near the minimum. The number of additional sample points is set three.

Starting from the 15 initial sample points, more sample points are added at each design iteration by the MPMO ISC method with the balancing approach, the MPMO ISC method with the dynamic switching approach, and the q-EI method with CLmin. An optimum function value at each design iteration is plotted in Fig. 8 with respect to the number of sample points. The MPMO ISC method with the balancing approach is the most efficient, finding the minimum with 186 sample points after 57 design iterations. The second most efficient ISC is the MPMO ISC method with the dynamic switching approach and finds the minimum point with 243 sample points after 76 iterations. The optimum found by the q-EI method with the CLmax has a larger objective value than that by the MPMO ISC method, even using more than the twice the number of sample points.

The gradient-based optimization using the SLSQP algorithm [11] is conducted independently for the multi-element airfoil design. The gradient value can be calculated by the finite-difference method or the adjoint solution method. The adjoint-based gradient computation requires one CFD analysis and one adjoint solution which costs nearly equivalent computational time to that of the CFD analysis even with a large number of design variables. On the other hand, the finite-difference method, if the forward differencing is used, for example, calculates gradient values by five CFD computations with four design variables. As can be seen in the last column of Table 9, the number of CFD computation is 421 with 33 runs to compute each of $\partial C_l / \partial x_j$ and $\partial C_d / \partial x_j$ ($j = 1, \dots, 4$) as well as 157 runs for the objective function evaluation. Even if the adjoint method is assumed, the equivalent

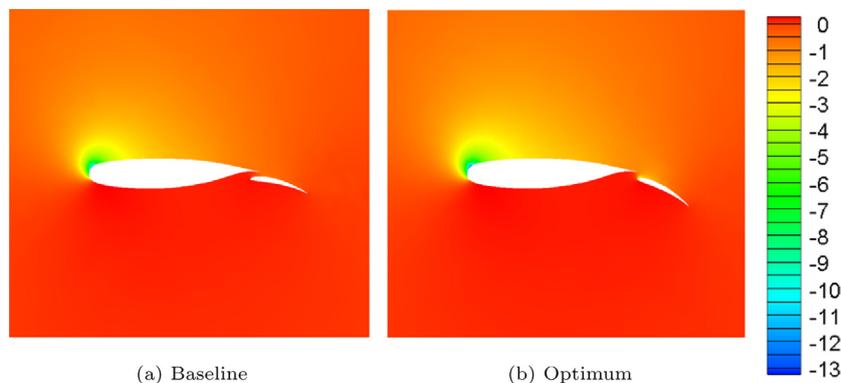


Fig. 9. Comparison of baseline and design results in pressure contour.

Table 9
Aerodynamic performance of the minimum point.

Design method	Obj. value	Lift coefficient (Increment, %)	Drag coefficient (Increment, % : counts)	Number of design iter.	Number of func. eval. (CFD)
Baseline	0.25	3.2089	0.06079		
Gradient-free optimization using MPMO ISC with balancing approach	0.07332	3.80822 (18.68%)	0.064703 (6.438% : 3.914 counts)	57	186
Gradient-free optimization using MPMO ISC with dynamic switching	0.07601	3.79125 (18.15%)	0.064512 (6.124% : 3.723 counts)	76	243
Gradient-free optimization using approx. q-EI KB	0.07333	3.81164 (18.78%)	0.0650356 (6.985% : 4.246 counts)	118	369
Gradient-free optimization using approx. q-EI CLmax	0.07480	3.80177 (18.48%)	0.064892 (6.749% : 4.103 counts)	134	417
Gradient-free optimization using approx. q-EI CLmin	0.07371	3.80618 (18.61%)	0.06472 (6.466% : 3.931 counts)	122	381
Gradient-based optimization with F.D. sensitivity	0.08213	3.7450 (16.71%)	0.0625 (2.814% : 1.711 counts)	n/a	421 = 157 N_{obj} + 4 × 33 $N_{dC_l/dx}$ + 4 × 33 $N_{dC_d/dx}$

Table 10
Design variable of the minimum point.

Design method	α	δ_θ	δ_x	δ_y
Baseline	10.26°	-29.66°	6.44%c	1.06%c
Gradient-free optimization using MPMO ISC with balancing approach	10.26°	-29.66°	6.44%c	1.06%c
Gradient-free optimization using MPMO ISC with dynamic balancing	10.24°	-29.84°	6.42%c	1.05%c
Gradient-free optimization using approx. q-EI KB	10.25°	-30.00°	6.47%c	1.08%c
Gradient-free optimization using approx. q-EI CLmax	10.24°	-29.92°	6.45%c	1.05%c
Gradient-free optimization using approx. q-EI CLmin	10.26°	-29.99°	6.41%c	1.06%c
Gradient-based optimization with F.D. sensitivity	8.9962°	-29.33°	4.272%c	0.645%c



Fig. 10. Comparison of baseline and design results in geometry — Black : Baseline, Red : Optimum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

number of CFD computation would be 190 with 33 runs for adjoint analysis to compute df/dx and 157 runs of CFD analysis, which involves more CFD computations than the MPMO ISC with balancing approach. Computational cost for the single CFD computation takes about 15 min using 60 CPUs of Authentic AMD@2300 MHz.

Improvements in the lift force are summarized in Table 9 by the EGO frameworks with different ISC methods as well as the gradient-based optimization. The lift coefficient is increased by about 18.5% by the EGO with the MPMO ISC and q-EI methods, while the drag coefficient is slightly increased by about 6.5%. Meanwhile, the lift coefficient is increased by 16.71% with slight increase in the drag coefficient by 2.814% as a result of the gradient-based optimization.

Optimal values of the design variables are listed in Table 10. The angle of attack is reduced to about 10.3 degree from 13.1 degree of the baseline airfoil. The flap angle is reduced so that the total chord length is increased and flows over the main airfoil move smoothly over the flap, which can be seen in Figs. 9 and 10. The increase of the lift force is also observed in the plot of surface pressure coefficient in Fig. 9.

5. Conclusions and future work

In this study, the multi-point and multi-objective infill sampling criteria (MPMO ISC) is developed and implemented in the EGO design framework to maximize the efficiency. The MPMO ISC is proposed in the consideration of where additional sampling should be placed, either

on the less explored area — exploration, or near the optimum point — exploitation. The trade-off between exploration and exploitation is resolved with the multi-objective optimization where two objectives are to find a point of the large prediction uncertainty, $\hat{s}^2(\mathbf{x})$ and to search a point of the minimum response, $\hat{y}(\mathbf{x})$. Two approaches of the balancing and dynamic switching are developed to choose infill sample points with flexibility and adaptivity along the Pareto-front set of the MPMO ISC method. They are based on the assumption that the reduction of uncertainty in design space is indicated by decreased variance of prediction uncertainty during the design iteration. Both approaches are observed to be more efficient than the approximated q-EI methods from the results of validation of two analytic functions — six-hump camel back function and Matlab peak function. The practical aerodynamic optimization is conducted for the NRL 7301 multi-element airfoil, where the design objective is to increase the lift force of airfoil, while maintaining the drag force of the baseline. The flap deflection angle, and the horizontal and vertical displacements of flap location are design variables as well as the angle of attack. Lift increase by about 18% is resulted by the proposed design method at the computational cost significantly lower than the existing approximated q-EI method and the gradient-based design optimization method. In the future work, current design framework with MPMO ISC will be extended to multidisciplinary design optimization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning. We acknowledge

the financial support from the Ministry of Science, ICT & Future Planning, South Korea, subjected to the project EDISON (Education-research Integration through Simulation On the Net, Grant No.: NRF-2014-039318). This research was also supported by National Science Foundation, United States of America under Grant No. 1849300 and the Climate Change Research Hub of KAIST, South Korea (Grant No. N01150026). One of the author, S. Choi, also acknowledges partial support provided by the National Research Foundation of Korea (NRF 2017R1A5A1015311).

References

- [1] M. Schonlau, *Computer Experiments and Global Optimization*, University of Waterloo, 1997.
- [2] H.J. Kushner, A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise, *J. Basic Eng.* 86 (1) (1964) 97–106.
- [3] D.D. Cox, S. John, SDO: A statistical method for global optimization, in: *Multidisciplinary Design Optimization: State-of-the-Art*, 1997, pp. 315–329.
- [4] F. Viana, R. Haftka, L. Watson, Why not run the efficient global optimization algorithm with multiple surrogates? in: *51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2010, pp. 3090.
- [5] D.R.R.L. Ginsbourger, L. Carraro (Eds.), *Computational intelligence in expensive optimization problems*, vol. 2, Springer Science & Business Media, New York, 2010, pp. 131–162.
- [6] D.G. Krige, A Statistical Approach to Some Mine Valuation and Allied Problems On the Witwatersrand: By DG Krige, (Ph.D. thesis), University of the Witwatersrand, 1951.
- [7] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2) (2002) 182–197.
- [8] D. Pham, G. Jin, A hybrid genetic algorithm, in: *Critical Technology: Proceedings of the Third World Congress on Expert Systems*, vol. 1, 1996, pp. 748–757.
- [9] A.J. Hayter, *Probability and Statistics for Engineers and Scientists*, Nelson Education, 2012.
- [10] J. Sacks, S.B. Schiller, W.J. Welch, Designs for computer experiments, *Technometrics* 31 (1) (1989) 41–47.
- [11] D. Kraft, A Software Package for Sequential Quadratic Programming, *Forschungsbericht, Deutsche Forschungs Und Versuchsanstalt Fur Luft Und Raumfahrt*, 1988.
- [12] B. Vandenberg, B. Oskam, Boundary layer measurements on a two-dimensional wing with flap and a comparison with calculations, in: *In AGARD Turbulent Boundary Layers vol. 14*, 1980, pp. (SEE N80-27647 18-34).
- [13] D.R. Jones, M. Schonlau, W.J. Welch, Efficient global optimization of expensive black-box functions, *J. Global Optim.* 13 (4) (1998) 455–492.
- [14] A. Forrester, A. Sobester, A. Keane, *Engineering Design Via Surrogate Modelling: A Practical Guide*, John Wiley & Sons, 2008.
- [15] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, *Stat. Comput.* 14 (3) (2004) 199–222.
- [16] M. Abt, W.J. Welch, Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes, *Canad. J. Statist.* 26 (1) (1998) 127–137.
- [17] J.D. Martin, T.W. Simpson, A study on the use of kriging models to approximate deterministic computer models, in: *ASME 2003 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, 2003, pp. 567–576.
- [18] T.W. Simpson, J. Poplinski, P.N. Koch, J.K. Allen, Metamodels for computer-based engineering design: survey and recommendations, *Eng. Comput.* 17 (2) (2001) 129–150.
- [19] Z. Han, R. Zimmerman, S. Görtz, Alternative cokriging method for variable-fidelity surrogate modeling, *AIAA J.* 50 (5) (2012) 1205–1210.
- [20] W. Yamazaki, D.J. Mavriplis, Derivative-enhanced variable fidelity surrogate modeling for aerodynamic functions, *AIAA J.* 51 (1) (2012) 126–137.
- [21] H.-S. Chung, J. Alonso, Using gradients to construct cokriging approximation models for high-dimensional design optimization problems, in: *40th AIAA Aerospace Sciences Meeting & Exhibit*, 2002, p. 317.
- [22] Z.-H. Han, S. Görtz, R. Zimmermann, Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function, *Aerosp. Sci. Technol.* 25 (1) (2013) 177–189.
- [23] Y. Jo, S. Choi, D. Lee, Variable-Fidelity Design Method Using Gradient-Enhanced Kriging Surrogate Model with Regression, in: *14th AIAA Aviation Technology, Integration, and Operations Conference*, 2014, pp. 2867.
- [24] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.* 96 (456) (2001) 1348–1360.
- [25] H. Kwon, S. Yi, S. Choi, Numerical investigation for erratic behavior of kriging surrogate model, *J. Mech. Sci. Technol.* 28 (9) (2014) 3697–3707.
- [26] M.J. Sasena, Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations, (Ph.D. thesis), University of Michigan Ann Arbor, 2002.
- [27] H.-G. Bae, A study on EGO Method for Aerodynamic Optimal Design, Korea Advanced Institute of Science and Technology, Mechanical Aerospace and Systems Engineering, (Ph.D. Dissertation), 2012.
- [28] K. Binder, D. Heermann, L. Roelofs, A.J. Mallinckrodt, S. McKay, Monte Carlo simulation in statistical physics, *Comput. Phys.* 7 (2) (1993) 156–157.
- [29] E. Omojokun, Trust region algorithms for optimization with nonlinear equality and inequality constraints, 1990.
- [30] A. Törn, A. Žilinskas, *Global Optimization*, vol. 350, Springer, 1989.
- [31] A. Bueno-Orovio, C. Castro, F. Palacios, E. Zuazua, Continuous adjoint approach for the Spalart-Allmaras model in aerodynamic optimization, *AIAA J.* 50 (3) (2012) 631–646.
- [32] I. Pointwise, Pointwise, URL www.pointwise.com.
- [33] F. Palacios, T.D. Economou, A. Aranake, S.R. Copeland, A.K. Lonkar, T.W. Lukaczyk, D.E. Manosalvas, K.R. Naik, S. Padron, B. Tracey, et al., Stanford university unstructured (SU2): Analysis and design technology for turbulent flows, in: *52nd Aerospace Sciences Meeting*, 2014, p. 0243.
- [34] A. Jameson, W. Schmidt, E. Turkel, Numerical solution of the Euler equations by finite volume methods using Runge Kutta time stepping schemes, in: *14th Fluid and Plasma Dynamics Conference*, 1981, pp. 1259.
- [35] Y. Saad, M.H. Schultz, Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Comput.* 7 (3) (1986) 856–869.
- [36] P. Spalart, S. Allmaras, A one-equation turbulence model for aerodynamic flows, in: *30th Aerospace Sciences Meeting and Exhibit*, 1992, p. 439.
- [37] D.J. Mavriplis, *Multigrid Techniques for Unstructured Meshes*, Tech. Rep., Institute for Computer Applications in Science and Engineering Hampton VA, 1995.