



Efficient global optimization method via clustering/classification methods and exploration strategy

Naohiko Ban¹ · Wataru Yamazaki¹

Received: 22 October 2019 / Revised: 22 June 2020 / Accepted: 23 June 2020 / Published online: 16 July 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

The objective of this research is to efficiently solve complicated high dimensional optimization problems by using machine learning technologies. Recently, major optimization targets have been changed to more complicated ones such as discontinuous and high dimensional optimization problems. It is necessary to solve the high-dimensional optimization problems to obtain an innovate design from topology design optimizations that have enormous numbers of design variables in order to express various topologies/shapes. In this research, therefore, an efficient global optimization method via clustering/classification methods and exploration strategy (EGOCCS) is developed to efficiently solve the high dimensional optimization problems without using probabilistic values as standard deviation, that are generally given/utilized in Gaussian process, and to reduce the construction cost of response surface models. Two optimization problems are solved to verify the usefulness of the developed method of EGOCCS. First optimization is executed to demonstrate the validity of the EGOCCS in 2, 10, 40, 80 and 160-dimensional analytic function problems that are also solved by the Bayesian optimization for comparison purposes. It is confirmed that the EGOCCS with radial basis function interpolation approach can obtain the best solutions in many analytic function problems with larger numbers of design variables. Second optimization is executed to examine the effect of the EGOCCS in high dimensional aerodynamic shape optimization problems for a two-dimensional biconvex airfoil that are also solved by a genetic algorithm for comparison purposes. It is confirmed that the EGOCCS can be efficiently used in the high dimensional aerodynamic shape optimization problems.

Keywords Efficient global optimization · Machine learning · High-dimensional optimization problem · Clustering · Classification · Regression

✉ Naohiko Ban
naohiko.b.78@gmail.com

¹ Department of Energy and Environment Science, Nagaoka University of Technology, Niigata 940-2188, Japan

1 Introduction

Design optimization techniques have already been used in actual designs such as a winglet shape design of Mitsubishi Regional Jet (Takenaka et al. 2008) and a nose shape design of the Bullet Train 700 series (Igarashi 2008). Recently, major optimization targets have been changed to more complicated ones such as discontinuous optimization problem, robust optimization problem, high dimensional optimization problem and so on. Such complicated optimization problems often appear in topology design optimizations due to the complexity of its topology/shape expressions. The topology optimization is a design approach to obtain an optimal layout including the changes of size, shape and topology. In the aircraft design, for example, the topology optimization can determine not only the optimal size and shape of aircraft, but also the number of blades, fuselages, engines and so on. By applying the topology optimization instead of the conventional shape optimization, it is possible to explore innovative optimal design configurations that exceed the designer's idea/experience.

The optimization methods in numerical simulation fields can be roughly divided into two groups, invasive and noninvasive types. The noninvasive optimization methods include evolutionary strategies such as genetic algorithms (GA) and differential evolution. These methods have the feature to explore global optimal solutions even with complicated objective functions. As one of famous noninvasive approaches for high dimensional functions, there is an intense stochastic search method which is based on evaluation and optimization of a hypercube and is called the hypercube optimization (HO) algorithm (Rahib and Mustafa 2015). The hypercube is used to describe searching areas, and the algorithm can explore a global optimal solution by mimicking the behavior of a dove discovering new areas for food in natural life. The HO algorithm was executed in optimization problems with 1000, 5000 and 10,000 dimensions, and then it was demonstrated that the algorithm was a potential candidate for optimization of both low and high dimensional functions. However, required numbers of functional evaluations were from 500,000 to 50 million. The invasive optimization method often includes adjoint-based approach. In the adjoint approach, the gradient of the objective function with respect to all design variables can be obtained at once by solving the adjoint equation, and optimal solutions are explored by using gradient based optimization methods. It has the feature to be able to explore a local optimal solution even in high dimensional optimization problems. Therefore, the topology design optimization is generally executed with the adjoint approach (Fujii et al. 2000; Lars et al. 2002) since the topology design optimization requires an enormous number of design variables in order to express various topologies/shapes. In Fujii et al. (2000), an effective gradient-based method for the topology optimization of 3D structures has been presented. A flat cantilever beam problem of 32,000 dimensions, MBB beam problem of 60,000 dimensions, and 3D beam design problem of 144,000 dimensions were solved by the adjoint-based optimization approach. However, the adjoint approach requires to develop the program code to solve the adjoint equation and is not possible to search global optimal

solutions. In addition, the adjoint approach cannot set integer/binary type constraints such as the acceptable number of objects. To tackle these issues, there is an example of topology design optimization using evolutionary algorithms (Payot et al. 2017). Payot et al. applied an R-snake method to the topology design optimization as a topology expression method, and then the topology design optimization was conducted in supersonic flow conditions using a differential evolutionary optimization method. As a result, the topology optimization revealed optimum profiles that have both features of the Busemann biplane and conventional single body optimal airfoils. However, there is a problem in terms of optimization costs since the differential evolutionary optimization approach was directly used in this research. In the topology design optimization in the field of aerodynamics, flow simulations are generally used to evaluate the objective function whose computational cost is not trivial. Therefore, it is necessary to reduce the number of evaluations of aerodynamic objective functions.

To realize efficient global topology design optimization at supersonic flow conditions, it is necessary to efficiently solve three challenging optimization problems, i.e. discontinuous optimization problem, optimization problem with infeasible regions and high dimensional optimization problem. With respect to the discontinuous optimization problem, the objective functions have discontinuities due to shock waves that have to be treated in optimization problems. With respect to the optimization problem with infeasible regions, the generation of large shock waves yields large infeasible regions of design space. In addition, the negative thickness of airfoil occurs depending on the definition of the ranges of design variables, and then the negative thickness of airfoil leads to an infinite loop in the optimization process by iteratively selecting a same additional sample point in the infeasible regions. In our previous research, an efficient global optimization method for discontinuous optimization problems with infeasible regions using classification method (EGODISC) (Ban and Yamazaki 2019) has been developed to efficiently solve discontinuous optimization problems as well as optimization problems with large infeasible regions by using machine learning technologies. By developing the EGODISC algorithm, an efficient global optimization could be executed without the infinite loop in the optimization process. However, it is still necessary to develop an efficient optimization approach for high-dimensional optimization problems since the topology design optimization generally has an enormous number of design variables in order to express various topologies/shapes.

There are several efficient optimization methods to solve the complicated optimization problems by using noninvasive approaches and machine learning techniques. For example, there is an optimization method using the Bayesian approach such as Gaussian process (Kriging) (Jones et al. 1998). The Bayesian optimization method has already been applied to a design optimization problem of supersonic transport, and it could obtain design knowledge of a twin-body/biplane-wing configuration (Ban et al. 2016, 2018). Although the Bayesian optimization isn't a state-of-the-art algorithm, it is still often utilized in many research fields. In addition, several advanced approaches using multi fidelity information have also been proposed (Han and Görtz 2012; Yamazaki and Mavriplis 2013). In the Bayesian optimization approaches, however, the Gaussian process is used to construct a response surface

model, and it takes large calculation cost to solve inverse matrices whose calculation cost is proportional to the cube of the number of sample points. In addition, it is necessary to optimize hyperparameters in the Bayesian optimization, and a gradient-based optimization method or GA is generally used for the optimization of the hyperparameters. Therefore, although the Bayesian optimization method enables to reduce the number of evaluations of objective functions (number of sample points), it requires huge construction costs of the response surface model especially with larger number of sample points. Since the number of sample points tend to become larger in higher dimensional optimization problems to explore higher dimensional design space, it is difficult to utilize the Bayesian method in high dimensional optimization problems. In addition, the prediction accuracy of the Gaussian process regression decreases at the high dimensional optimization problems and it was indicated that the upper limit of the dimensionality of the design variables space to efficiently obtain global optimal solutions was about 10 (Ziyu et al. 2013). There are several other methods to efficiently solve the high-dimensional optimization problems, such as GPEME and KPLS. The Gaussian Process surrogate model assisted Evolutionary algorithm for Medium-scale computationally Expensive optimization problems (GPEME) uses dimension reduction machine learning techniques for tackling the “curse of dimensionality”. In GPEME, the Sammon mapping is introduced to transform the design variables space to take advantage of the Gaussian process surrogate modeling in a low-dimensional space (Liu et al. 2014). However, since the dimension reduction method may lose some neighborhood information of the training data points in the original space, it is difficult to obtain comparable results with the direct Bayesian optimization. Another optimization method for the high-dimensional optimization problems has been proposed as the Kriging model with the Partial Least Squares technique to obtain a fast predictor (KPLS). The combination of Kriging and Partial Least Square (PLS) is abbreviated KPLS and allows us to build a fast Kriging model because it requires fewer hyper-parameters in its covariance function. The KPLS method is used for many academic and industrial verifications, and promising results have been obtained for problems with up to 100 dimensions (Bouhleb et al. 2016). However, it was also shown in (Bouhleb et al. 2016) that the KPLS has lower prediction accuracy than the Gaussian process (Kriging) model with a lot of sample points, which is not adequate to use in high dimensional optimization problems. As a state-of-the-art algorithm in the optimization, the reinforcement learning has been proposed (Volodymyr et al. 2016) to achieve the balance of “exploration at unknown region” and “utilization of prediction” that are similar principles to the Bayesian optimization. The reinforcement learning is a method to select actions by maximizing the assessment from environments such as recognitions, judgements and actions, and then it is used for the controls of complex systems such as automatic driving, game AI and robot. On the other hand, since the reinforcement learning needs an enormous number of iterations (Volodymyr et al. 2016), it is considered not to be appropriate for the design optimization problems discussed in this research.

Furthermore, there are several efficient optimization methods by combining the above-mentioned approaches such as gradient-enhanced Kriging and DynDGA. The gradient-enhanced Kriging method (Mohamed and Joaquim 2019; Rumpfkeil et al.

2011; Han et al. 2017) utilizes the gradient of the objective function with respect to all design variables obtained by solving the adjoint equation and more accurate surrogate models can be constructed. This approach can reduce the number of optimization iterations thanks to the accurate surrogate model. However, this approach still has an issue of the invasive optimization approach. A distributed genetic algorithm that divides and merges individuals dynamically (DynDGA) (Sugimoto et al. 2019) utilizes both of simulation and Kriging estimation to reduce the number of simulations for GA. In this approach, the number of simulation calls reduces to approximately 63% while maintaining the performance. However, this approach still has an issue of the Kriging method for the high dimensions.

In summary, the conventional approaches to solve the high dimensional optimization problems can be classified into five types even though the state-of-the-art algorithms. The first is to use the HO algorithm or evolutionary algorithms as GA by evaluating an enormous number of objective functions. The second is the gradient-based optimization using the adjoint sensitivity analysis whose solution is considered to be a local optimum. The third is to use the Bayesian optimization despite its difficulties. The fourth is to introduce some dimension reduction methods despite loss of neighborhood information of the training data points. The fifth is to combine the above-mentioned approaches. Our proposed approach is completely different from the five conventional approaches. A novel optimization method is developed using machine learning technologies to efficiently solve the high dimensional optimization problems without Gaussian process. The present paper is organized as follows. The optimization methods and performance evaluation methods used in this research are concisely described in Sect. 2. Then, several surrogate models used in this research are concisely introduced and construction/utilization costs of them are compared in Sect. 3. And then, optimization results obtained by the conventional Bayesian/GA approaches and the developed approach are compared/discussed in Sect. 4. Finally, concluding remarks are provided in Sect. 5.

2 Optimization and evaluation methods

2.1 Bayesian optimization

2.1.1 Algorithm of Bayesian optimization

Several surrogate model-based optimization methods have been proposed to efficiently explore global optimal solutions on blackbox functions such as the Bayesian approaches (Jones et al. 1998; Yamazaki and Mavriplis 2013; Forrester et al. 2008; Williams and Carl 1996) (Kriging response surface model-based approaches). The Gaussian process regression is used to construct a prediction model for mean and standard deviation values at unknown locations in design variables space, and then GA is used to explore promising locations in the design variables space in the general Bayesian optimization approach. The expected improvement (EI) is calculated from the mean and standard deviation values to achieve the balance of “exploration at unknown region (with larger standard deviation values)” and “utilization of

prediction by response surface model (with lower mean values)”, and an additional point is selected at the maximum location of the EI value.

2.1.2 Gaussian process regression

The Gaussian process regression was introduced in Williams and Carl (1996) and Rasmussen and Williams (2006). The regression problem is considered for an unknown blackbox function giving the input/output relationship when the observation values $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ and the input values $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ are given. Mean $\mu_{y^*|\mathbf{y}}$ and variance $\Sigma_{y^*|\mathbf{y}}$ of the posterior distribution of Gaussian distribution are expressed by the following equations.

$$\mu_{y^*|\mathbf{y}} = \mathbf{k}^T \mathbf{C}^{-1} \mathbf{y} = \mu, \quad \Sigma_{y^*|\mathbf{y}} = c - \mathbf{k}^T \mathbf{C}^{-1} \mathbf{k} = \sigma^2 \quad (2.1.2.1)$$

$$c = k(x^*, x^*), \quad \mathbf{k} = [k(x_1, x^*)k(x_2, x^*) \dots k(x_n, x^*)]^T \quad (2.1.2.2)$$

where \mathbf{C} is a gram matrix having $C_{i,j} = k(x_i, x_j)$ as an element, and $k(x_i, x_j)$ is referred to as a kernel function, and then the following ARD Matern 5/2 kernel is used in this research.

$$k(x_i, x_j) = \theta_0 \left(1 + \sqrt{5r^2(x_i, x_j)} + \frac{5}{3} r^2(x_i, x_j) \right) e^{-\sqrt{5r^2(x_i, x_j)}} \quad (2.1.2.3)$$

$$r^2(x_i, x_j) = \sum_{d=1}^D (x_{i,d} - x_{j,d})^2 / \theta_d^2$$

where D represents the number of design variables.

2.1.3 Genetic algorithm

GA is one of the metaheuristic methods inspired by the process of natural selection such as selection, crossover and mutation. GA simulates the selection of the individuals by considering the relationship between the applicability to the environment and the probability to survive (Selection). Then, it simulates the natural evolution of the individuals by generating child individuals from the selected parental individuals (Crossover). In addition, it maintains the diversity of the population by altering a part of the genes of the child individuals (Mutation). In this research, a tournament selection, the blend cross over (BLX- α) (Eshleman and Schaffer 1993), and a polynomial mutation (Deb and Deb 2014) are used as the selection, crossover, and mutation methods.

2.2 Efficient global optimization method via clustering/classification methods and exploration strategy, EGOCCS

In the conventional Bayesian optimization, the Gaussian process regression is used to construct a response surface model, and it takes large calculation cost to solve

inverse matrices whose size is $n \times n$ (n : number of sample points) since the calculation cost is proportional to the cube of the number of sample points. In addition, it is necessary to optimize hyperparameters in the Bayesian optimization [θ_i in Eq. (2.1.2.3)], and a gradient-based optimization method or GA is generally used for the optimization of the hyperparameters. Therefore, since the Bayesian optimization requires huge construction cost for the response surface model and the prediction accuracy of the Gaussian process regression decreases at high dimensional optimization problems, it is difficult to utilize for high dimensional optimization problems. To tackle these problems, a novel efficient global optimization method without using probabilistic values as the standard deviation given in the Gaussian process is proposed, and then its construction cost is also reduced to efficiently obtain global optimal solutions in the high dimensional design variables space. In the proposed method, optimization problems are solved only with predicted functional values without the uncertain metrics of the surrogate model (standard deviation) which is used for calculating EI value in the conventional Bayesian optimization approach. Therefore, the proposed approach will bring higher compatibility with arbitrary response surface models such as radial basis function interpolation, inverse distance weighting, neural network regression, support vector regression, random forest regression and so on. For example, the inverse distance weighting can be used when the optimization cost is necessary to be reduced, and the neural network can be used when big data (a very large number of design variables) has to be taken into consideration in the optimization.

2.2.1 Algorithm of EGOCCS

The flowchart of efficient global optimization method via clustering/classification methods and exploration strategy (EGOCCS) is shown in Fig. 1, and an outline diagram of EGOCCS is shown in Fig. 2. In this algorithm, firstly, initial sample

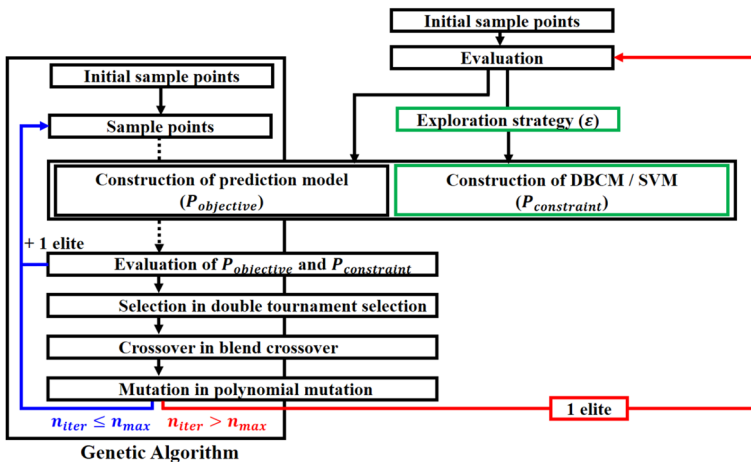


Fig. 1 Flowchart of EGOCCS

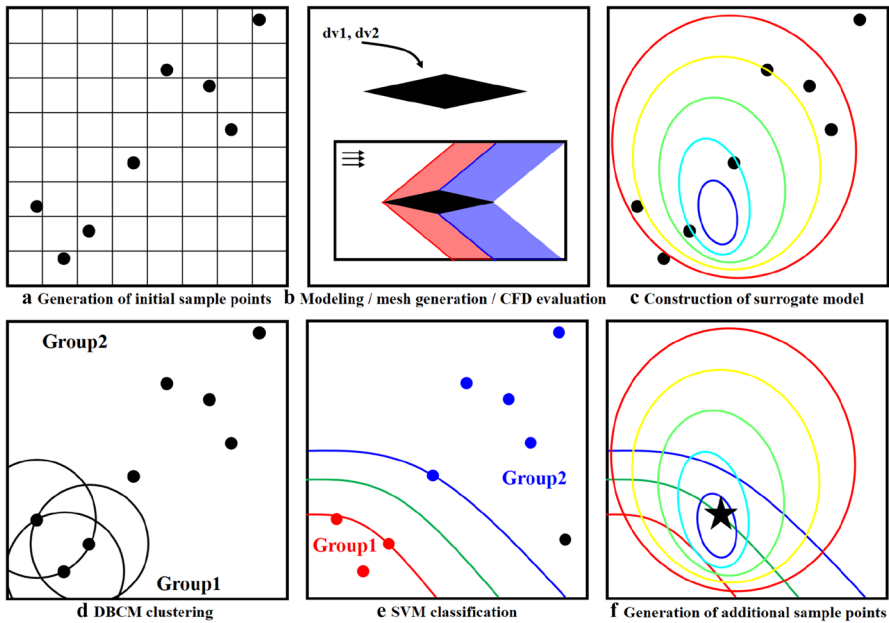


Fig. 2 Outline of EGOCCS

points are generated in the design variables space by the Latin hypercube sampling method (Fig. 2a). Then, generation of computation grids and evaluation of objective function are carried out for the models corresponding to the initial sample points (Fig. 2b). And then, the response surface model is constructed based on the objective function values (Fig. 2c). Sparse regions of sample points are considered as the candidates to generate an additional sample point and are defined by using the machine learning methods of the density-based clustering method (DBCM) (Fig. 2d) and support vector machine classification method (SVM) (Fig. 2e). By generating additional samples in the sparse regions, bias/concentration of the distributions of sample points can be reduced. The DBCM is a method to group sample points using a distance parameter ϵ , and it is used to make a group of sample points in which the current optimal solution belongs in. The SVM is a method to classify the design variables space, and it is used to classify the design variables space into two groups, one of which is the group obtained by the DBCM. The SVM has a feature to create a separating hyperplane by maximizing a margin between two groups. It means that the separating hyperplane represents the sparsest region of sample points between two groups. In other words, it is possible to efficiently generate additional sample points considering the predicted value of the objective function as well as the separating hyperplane. In this research, GA is used to search an additional sample point where the predicted objective function value is the lowest on the separating hyperplane (sparse region of sample points) (Fig. 2f).

2.2.2 Density-based clustering method (DBCM)

Density-based clustering method (DBCM) is used to classify input datasets into two groups via unsupervised learning. This method is similar to the density-based spatial clustering of applications with noise (DBSCAN) (Ester et al. 1996), and it is possible to classify sample points into two groups based on the distance between points. The outline of this algorithm is summarized in Fig. 3. In this algorithm, firstly, a starting point P_s and a reference distance ε are defined as a criterion of the clustering (Fig. 3a, b). The starting point corresponds to the current optimal sample point in this research. Then, sample points within the reference distance ε are classified into the group same with the starting point (Fig. 3c), and then other sample points are checked again whether they are within the reference distance ε from the sample points of the group (Fig. 3d). By iteratively performing this process until there are no more points to add (Fig. 3e–g), finally, the sample points can be divided into two groups whether the starting point belongs in or not (Fig. 3h). The reference distance ε can be changed at each iteration as an exploration strategy explained in Sect. 2.2.4.

2.2.3 Support vector machine classification method (SVM)

The support vector machine (SVM) is one of the two-classes classification methods proposed by Corinna and Vapnik (Corinna and Vladimir 1995) and one of the popular machine learning algorithms (Alex and Bernhard 2004). It is possible to obtain the optimal separating hyperplane even in the nonlinear data by mapping the data to a high dimensional space by using a kernel method. In addition, SVM separates the two-classes data with the idea of the maximum-margin hyperplane, and it is known to have high accuracy to predict for unknown data. The characteristic of the maximum-margin is important for EGOCCS algorithm.

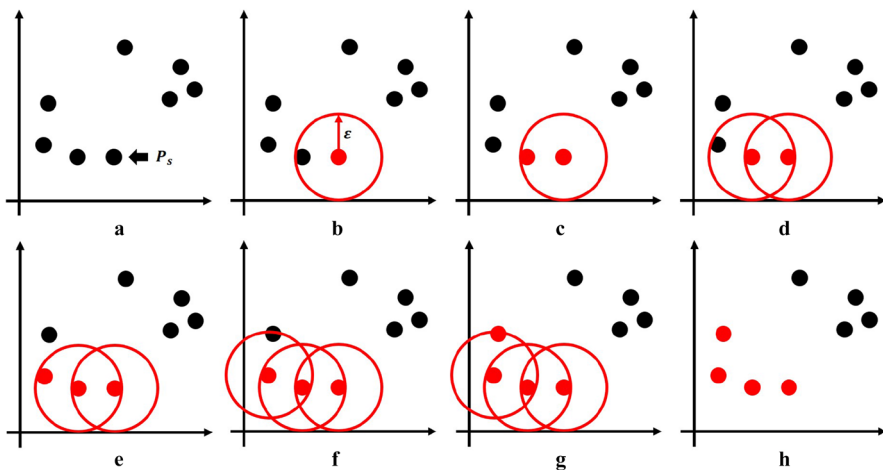


Fig. 3 Outline of density-based clustering method

2.2.4 Exploration strategy in EGOCCS

Consideration of “exploration at unknown regions” and “utilization of predictions by response surface models” with a proper balance is required in efficient global optimizations. With respect to the “exploration at unknown regions”, the idea of Bayes is to generate an additional sample point at the maximum standard deviation location on the probabilistic prediction model in the design variables space. The idea of EGOCCS is to generate an additional sample point at the sparsest region in the design variables space. Of course, it is difficult to obtain exact global optimal solutions when only the “exploration at unknown regions” is considered to obtain additional sample points, so that the “utilization of predictions by response surface models” should be considered to explore better solutions. In the minimization problem, the idea of Bayes is to generate an additional sample point at the minimum mean location on the probabilistic prediction model in the design variables space. The idea of EGOCCS is similar with that, to generate an additional sample point at the minimum location on the prediction model in the design variables space. However, it is difficult to obtain global optimal solutions by falling into local optimal solutions when only the “utilization of predictions by response surface models” is considered to obtain additional sample points. Therefore, the balance between them is important for the efficient global optimization. This can be achieved by using EI value in the Bayesian optimization, while this can be considered by using following strategy in EGOCCS.

The exploration strategy of GA is modeled into the EGOCCS algorithm as “exploration at unknown region” and “utilization of predictions by response surface models”. In the optimization process of GA, individuals are widely distributed in the design variables space at the initial stage of the optimization, and then the distribution becomes narrower with the advancement of generations and converges to global optimal solutions. In other words, the proper balance between “exploration at unknown regions” and “utilization of predictions by response surface models” can be expressed by modeling the exploring strategy of GA. It is modeled by using the parameter of ϵ used in DBCM which determines the details of exploration. ϵ should be set to a large value at the initial stage of the optimization for the “exploration at unknown regions”, and then it decreases with the advancement of optimizations for the exploration around the current optimal solution, in other words for the “utilization of predictions by response surface models”.

In this research, the DBCM is used to classify input datasets into two groups, and then the SVM is used to obtain a separating hyperplane by maximizing a margin between the two groups. It is possible to efficiently determine the location of an additional sample point by minimizing the predicted objective function value on the separating hyperplane. In the EGOCCS, it is necessary to set an appropriate value of ϵ to carry out an efficient iterative process of optimization. The balance of “exploration at unknown regions” and “utilization of predictions by response surface models”, in other words the balance between global/local search is modeled as follows by changing the parameter ϵ . Examples of the exploration strategy in this research are summarized in Fig. 4, and typical parameters

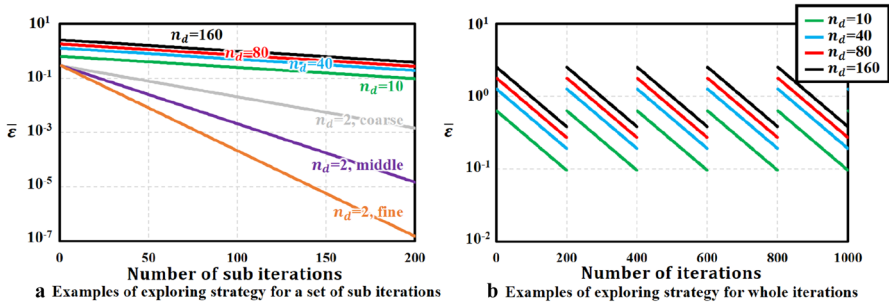


Fig. 4 Variable definition of reference distance ϵ

settings of this method are summarized in Table 1. Figure 4a is the general settings of ϵ among the total number of sub iterations (t_{max}) whose value is set to 200. The value of ϵ is represented as follows:

$$\epsilon = F(\text{mut}(r, f(t_i))) \tag{2.2.4.1}$$

where t_i and r are respectively the integer number of sub iterations between 1 to t_{max} and random real number between 0 to 1. The values of t_i and r change in each sub iteration. In Fig. 4a, $\bar{\epsilon}$ indicates the value of ϵ with $r = 0.5$. $f(t_i)$ represents a monotonically decreasing function with increasing of the sub iteration as following equation.

$$f(t_i) = 1 - \frac{t_i}{t_{max}} \tag{2.2.4.2}$$

$\text{mut}(r, p)$ is defined as the same formulation with the polynomial mutation of the GA (Deb and Deb 2014), and it models the mutation of the GA as following equation.

Table 1 Parameters settings for exploration strategy of EGOCCS

| # of dim (n_d) | Pattern | | |
|--------------------|---------|--|--|
| 2 | Coarse | $1.0 \times 10^{-3} \times \sqrt{n_d}$ | $2.0 \times 10^{-1} \times \sqrt{n_d}$ |
| 2 | Middle | $1.0 \times 10^{-5} \times \sqrt{n_d}$ | |
| 2 | Fine | $1.0 \times 10^{-7} \times \sqrt{n_d}$ | |
| 10 | Coarse | $1.0 \times 10^{-3} \times \sqrt{n_d}$ | |
| 40 | Coarse | $1.0 \times 10^{-3} \times \sqrt{n_d}$ | |
| 80 | Coarse | $1.0 \times 10^{-3} \times \sqrt{n_d}$ | |
| 160 | Coarse | $1.0 \times 10^{-3} \times \sqrt{n_d}$ | |

$$mut(r, p) = p + \begin{cases} (2r + (1 - 2r)(1 - p)^{\eta+1})^{\frac{1}{\eta+1}} - 1 & \text{if } r \leq 0.5 \\ 1 - (2(1 - r) + (2r - 1)p^{\eta+1})^{\frac{1}{\eta+1}} & \text{if } r > 0.5 \end{cases} \quad (2.2.4.3)$$

The value of $mut(r, p)$ is given between 0 to 1. The probabilistic distribution of the polynomial mutation can be adjusted by the additional parameter of η in which smaller η leads to lower probability of p , while larger η leads to higher probability of p . In other words, smaller η should be adopted when global search is necessary to focus as well as larger η should be adopted when local search is necessary to focus. In this research, η is set to 100 to aim to obtain exact optimal solutions. The example of the polynomial mutation is summarized in Fig. 5, in which the histogram is calculated from 1000 samples generated at even intervals between 0 and 1. $F(x)$ is defined as follows which determines the reference distance ϵ .

$$F(x) = 10^{(\log_{10} \epsilon_{min} + x(\log_{10} \epsilon_{max} - \log_{10} \epsilon_{min}))} \quad (2.2.4.4)$$

where ϵ_{min} and ϵ_{max} are respectively the user-specified minimum and maximum values of ϵ . In this equation, since the value of x is mainly decided by the iteration number t_i as seen in Eq. (2.2.4.2) and the value is fluctuated by the effect of r and η as shown in Fig. 6 and Eq. (2.2.4.3), $F(x)$ decreases exponentially with decreasing of x as well as increasing of the iteration number t_i . In the actual optimization process on the EGOCCS algorithm, since it is difficult to set an appropriate number of the optimization iterations beforehand, the total number of iterations is divided into

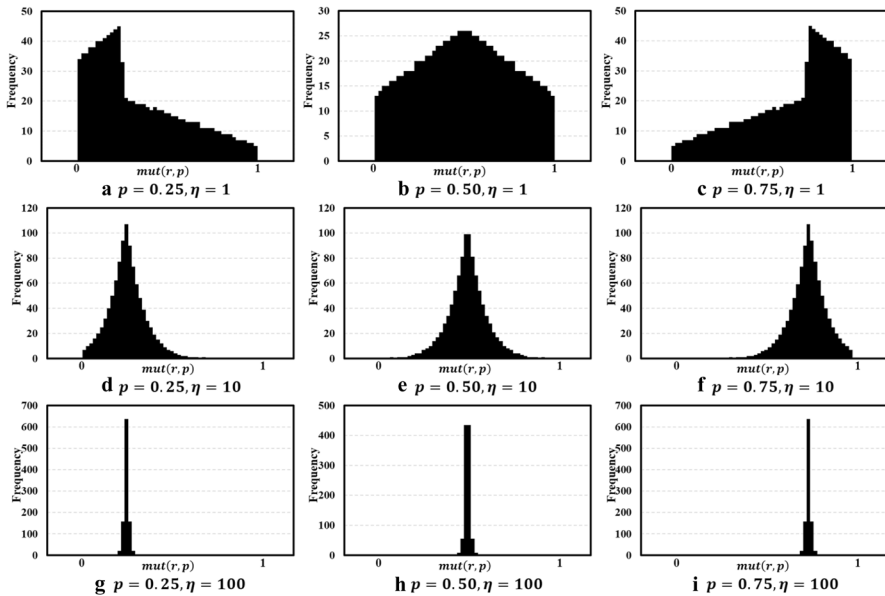


Fig. 5 Histogram of polynomial mutation

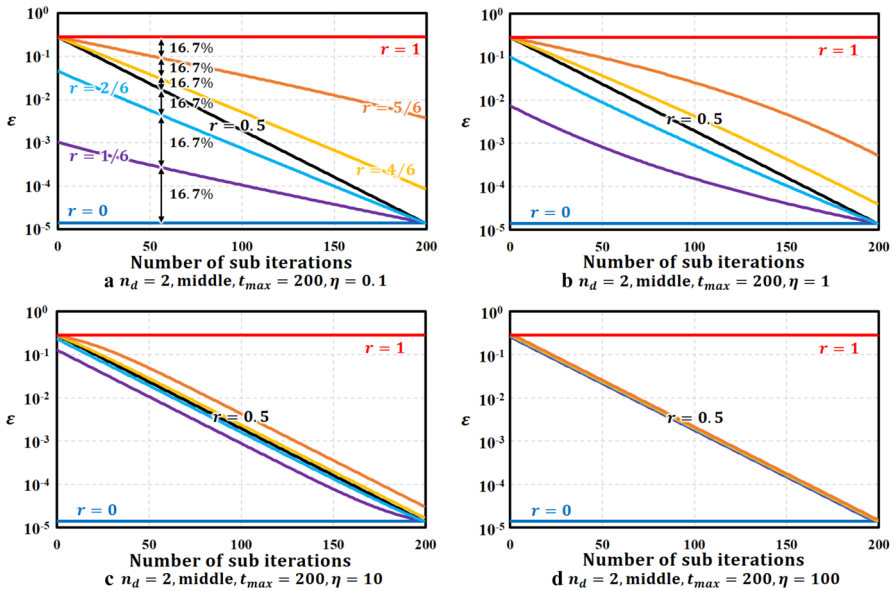


Fig. 6 Variation tendency of reference distance ϵ

several sets of the sub iterations as shown in Fig. 4b in which the total number of iterations is set to 1000 while the total number of sub iterations is set to 200.

The example of the global optimization by the EGOCCS is summarized in Figs. 7 and 8. Two dimensional Rosenbrock function is used as an analytic function to be minimized. When the relationship between the design variables (x and y coordinates) and output (color) is given at several locations as the white filled points in Figs. 7a-1 and 8a-1, functional values at arbitrary locations on the design variables space can be predicted by any interpolation (response surface) methods as shown in Figs. 7a-1 and 8a-1. In this example, the Gaussian process regression is used to construct the prediction model. Figures 7a-2 and 8a-2 indicate the distance to the separating hyperplane which is created by maximizing a margin between two groups of black filled points and the other (white filled) points in Figs. 7a-2 and 8a-2. An additional sample point is generated at the position where the predicted objective function value is the lowest on the separating hyperplane (red filled point). In the example of Fig. 7, the reference distance ϵ is fixed to 0.2 and the maximum number of iterations t_{max} is set to 10. In the example of Fig. 8, the reference distance ϵ is decreased monotonically from 0.2 to 0.1 and the maximum number of (sub) iterations t_{max} is set to 10. It is confirmed that the additional sample points in Fig. 7j-1 are widely distributed on lower objective function regions while the additional sample points in Fig. 8j-1 are widely/effectively distributed on lower objective function regions. In other words, the decreasing of the reference distance ϵ in the optimization process leads to smaller numbers of sample points in the group including the current optimal solution

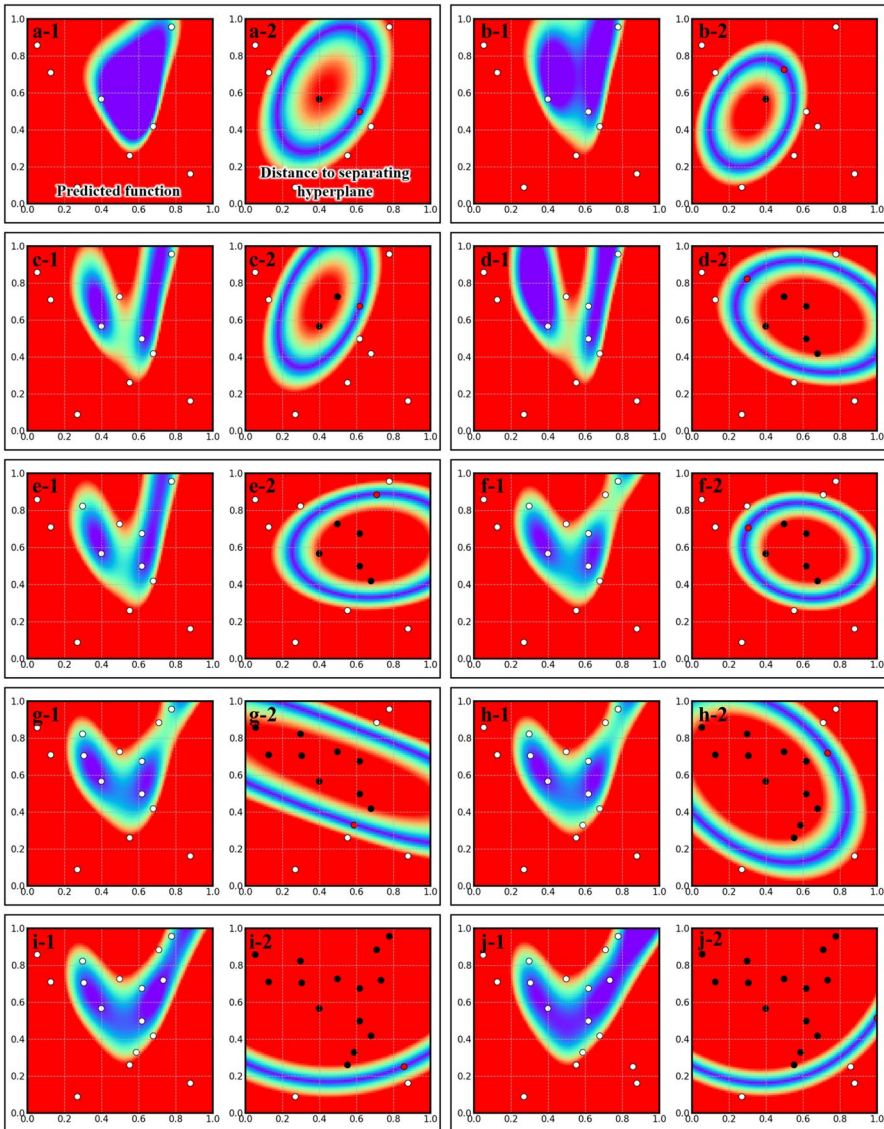


Fig. 7 Optimization process of EGOCCS-GPs approach with fixed ϵ

(black points), which results in that additional sample points are efficiently generated around the current optimal solution as a function of local search.

2.3 Computational fluid dynamics analysis method

An inhouse Computational Fluid Dynamics (CFD) code using a gridless method is used to evaluate aerodynamic objective functions in present optimization problems.

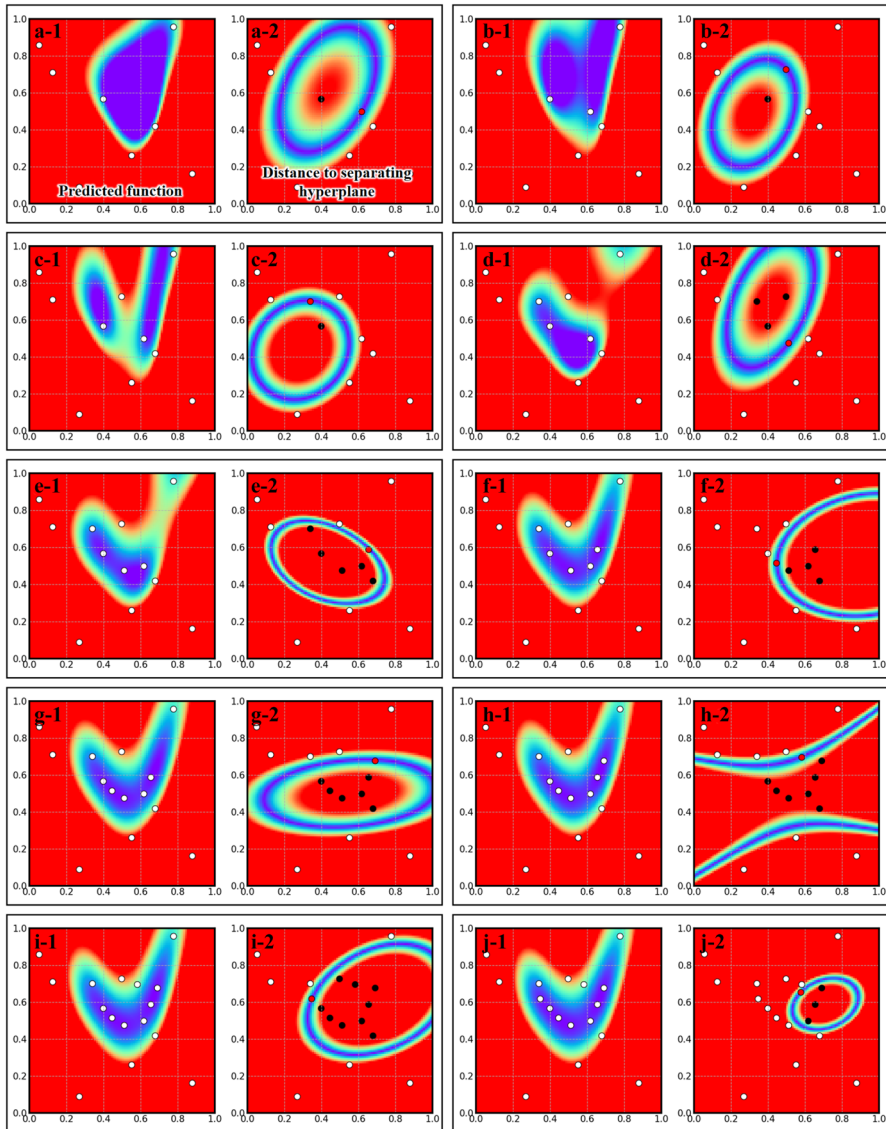


Fig. 8 Optimization process of EGOCCS-GPs approach with decreasing ϵ

In the gridless method (Ma et al. 2008; Suga and Yamazaki 2015), computational points are distributed in the calculation domain and a spatial differential term of a physical quantity around the computational point is approximated by a least squares method. Therefore, this method is unnecessary to construct the grid connectivity information and can analyze flow fields around complicated shapes. Two-dimensional inviscid compressible Euler equations are solved by the gridless method. The Euler equations express the conservation law of mass, momentum and

energy in inviscid and compressible fluids. The two-dimensional Euler equations are expressed as follows.

$$\frac{\partial \mathbf{W}}{\partial t} + \frac{\partial \mathbf{E}}{\partial x} + \frac{\partial \mathbf{F}}{\partial y} = 0$$

$$\mathbf{W} = \begin{bmatrix} \rho \\ \rho u \\ \rho v \\ e \end{bmatrix}, \mathbf{E} = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ u(e + p) \end{bmatrix}, \mathbf{F} = \begin{bmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ v(e + p) \end{bmatrix} \tag{2.3.1}$$

$$e = \frac{p}{\gamma - 1} + \frac{1}{2} \rho (u^2 + v^2)$$

where \mathbf{W} represents the vector of conservative variables, and \mathbf{E} and \mathbf{F} represent flux vectors. ρ, u, v, p, e respectively represent fluid density, velocity components in X/Y directions, pressure and total energy per unit volume. The specific heat ratio γ of the ideal gas is set to 1.4.

3 Comparison of regression methods

The EGOCCS algorithm has higher compatibility with any kind of surrogate models such as radial basis function interpolation, inverse distance weighting, neural network regression, support vector regression, random forest regression and so on. In this research, Gaussian process regression, radial basis function, and inverse distance weighting are considered. In this section, these methods are concisely introduced and then construction and utilization costs of them are compared.

3.1 Inverse distance weighting (IDW)

Inverse distance weighting (IDW) is a method to predict functional values at unknown points with a weighted average of the information of known sample points (Shepard 1968). The interpolated function is expressed as following equations.

$$f(\mathbf{x}) = \begin{cases} \frac{\sum_{i=1}^n w_i(\mathbf{x})y_i}{\sum_{i=1}^n w_i(\mathbf{x})}, & \text{if } d(\mathbf{x}, \mathbf{x}_i) \neq 0 \text{ for } \forall_i \\ y_i, & \text{if } d(\mathbf{x}, \mathbf{x}_i) = 0 \end{cases} \tag{3.1.1}$$

$$w_i(\mathbf{x}) = \frac{1}{d(\mathbf{x}, \mathbf{x}_i)^P} \tag{3.1.2}$$

where d is the distance from the known sample point \mathbf{x}_i to the unknown point \mathbf{x} and the power parameter P is set to 2 in this research. n and y_i are the total number of known sample points used in the interpolation and the output functional values of the known sample points \mathbf{x}_i . Since IDW has a feature that there is almost no construction cost of the surrogate model, it is expected that the optimization cost will be dramatically reduced in high dimensional problems by using the IDW as the prediction model of EGOCCS.

3.2 Radial basis function interpolation (RBF)

Radial basis function interpolation is one of the methods to predict functional values at unknown points (Press et al. 2007). The interpolated function can be expressed as following equation by using known output data y_i and input data $\mathbf{x}_i \{i = 1, 2, \dots, n\}$.

$$f(\mathbf{x}) = \sum_{i=1}^n \omega_i \phi(\|\mathbf{x} - \mathbf{x}_i\|) \tag{3.2.1}$$

where the approximating function $f(\mathbf{x})$ is represented as the sum of n radial basis functions, each associated with the center point \mathbf{x}_i , and weighted by an appropriate coefficient ω_i . ϕ is a kernel function, and there are various kernel functions as follows.

$$r = \|\mathbf{x} - \mathbf{x}_i\| \tag{3.2.2}$$

$$\phi_{\text{multiquadric}} = \sqrt{\left(\frac{r}{\beta}\right)^2 + 1}, \quad \phi_{\text{inverse multiquadric}} = \frac{1.0}{\sqrt{\left(\frac{r}{\beta}\right)^2 + 1}} \tag{3.2.3}$$

$$\phi_{\text{gaussian}} = e^{-\left(\frac{r}{\beta}\right)^2 + 1}, \quad \phi_{\text{linear}} = r$$

where β is a hyperparameter and it is generally determined by cross-validation approaches. In this research, since the optimization cost reduction is one of the aims of EGOCCS, the linear kernel ϕ_{linear} is used as the kernel function. There is no hyperparameter in the linear kernel. The weights ω_i can be estimated as follows.

$$\mathbf{y} = \Phi \boldsymbol{\omega}$$

$$\mathbf{y} = [y_1 \dots y_n]^T, \quad \Phi = \begin{bmatrix} \phi_{1,1} & \dots & \phi_{1,n} \\ \vdots & \ddots & \vdots \\ \phi_{n,1} & \dots & \phi_{n,n} \end{bmatrix} \tag{3.2.4}$$

$$\boldsymbol{\omega} = \Phi^{-1} \mathbf{y} \tag{3.2.5}$$

3.3 Calculation times of various regression methods

In the RBF, the calculation cost for the inverse matrix for $n \times n$ is also expensive which is proportional to the cube of the number of sample points. However, since the hyperparameters are not included in the linear kernel ϕ_{linear} contrary to the Gaussian process regression and the calculation of the inverse matrix is only once, it can dramatically reduce the construction cost of the surrogate model. For example, comparing with the Gaussian process regression in which the hyperparameters are optimized by GA with 100 population and 100 generation, the construction cost of the RBF becomes 1/10000.

The construction cost and utilization cost of the Gaussian process regression, RBF, and IDW are compared using the Rosenbrock analytic function as shown in Fig. 7. The Rosenbrock analytic function is expressed as follows.

$$f(\mathbf{x}) = \sum_{i=1}^{n_d-1} \left[100(x'_{i+1} - x'_i)^2 + (1 - x'_i)^2 \right], \quad x'_i = -2.048 + 4.096x_i, \quad 0 \leq x_i \leq 1 \tag{3.3.1}$$

In Fig. 9, the horizontal axis indicates the number of sample points to construct a surrogate model, and the vertical axis indicates the construction time for the surrogate model and utilization time for 40,000 predictions which corresponds to the execution of GA with 200 populations and 200 generations. All simulations were processed in a Processor Intel(R) Xeon(R) CPU E7-8867 v3 @ 2.50 GHz. The hyperparameters of the Gaussian process regression are optimized by a gradient-based optimization method in which optimization is executed with the scaled conjugate gradients (SCG) method of 500 iterations. The calculation times are investigated at various dimensions (n_d) of 2, 4, 8, 16, 32, 64, 128, and the color lines are averaged values of 10 times trials. It is confirmed that the construction time is increased with larger numbers of sample points and with larger numbers of dimensions in the Gaussian process regression and RBF, and it is also confirmed that the construction cost of RBF is much smaller than the Gaussian process regression whose ratio is about 1/1000. Therefore, the adoption of the RBF interpolation can reduce the construction cost. It is also confirmed that IDW has the lowest construction/prediction costs among the three surrogate model approaches.

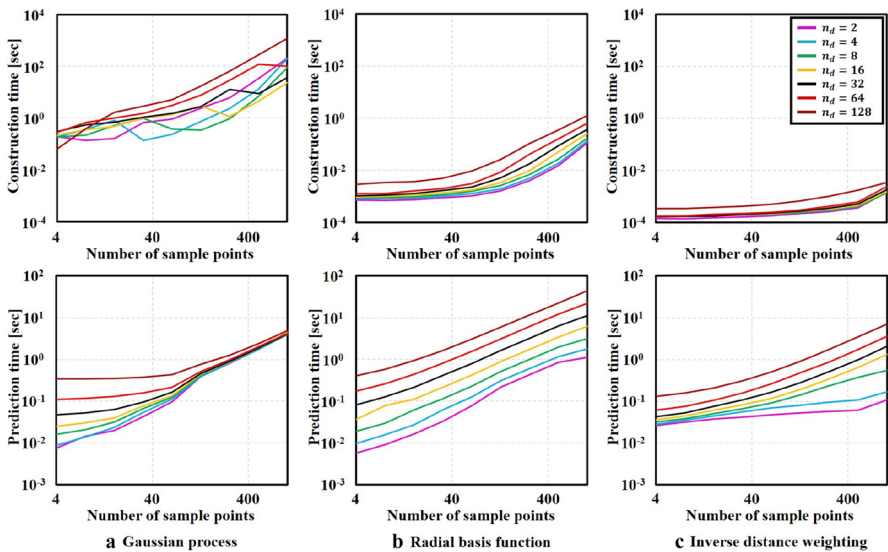


Fig. 9 Construction and prediction costs of representative surrogate models

4 Results and discussion

Two optimization problems are solved to verify the usefulness of the developed method of EGOCCS. First optimization is executed to demonstrate the validity of the EGOCCS in 2, 10, 40, 80, 160-dimensional analytic function problems within a limited number of functional evaluations. Second optimization is executed to examine the effect of the EGOCCS in 2, 10, 40, 80 dimensional aerodynamic shape optimization problems of a biconvex airfoil, and to demonstrate the applicability of the EGOCCS to more realistic optimization problems than Sect. 4.1.

4.1 Optimization of various analytic functions

In order to verify the usefulness of the EGOCCS, optimization problems of various analytic functions are solved by both of the conventional Bayesian optimization and the developed algorithm of EGOCCS. The analytic functions are Sumsquares function (Eq. 4.1.1), Quadric function (Eq. 4.1.2), Schwefel function (Eq. 4.1.3), Ackley function (Eq. 4.1.4), Rosenbrock function (Eq. 4.1.5), and Levy function (Eq. 4.1.6), and the two-dimensional analytical functions are visualized in Fig. 10. The Sumsquares/Quadric functions are concave analytic functions, the Schwefel function is a discontinuous analytic function, the Ackley function is a multi-modal function, and the Rosenbrock/Levy functions are complex functions (Crina and Ajith 2009). The geometric mean of 10 optimization histories are summarized in Figs. 11, 12, 13, 14 and 15. The optimization

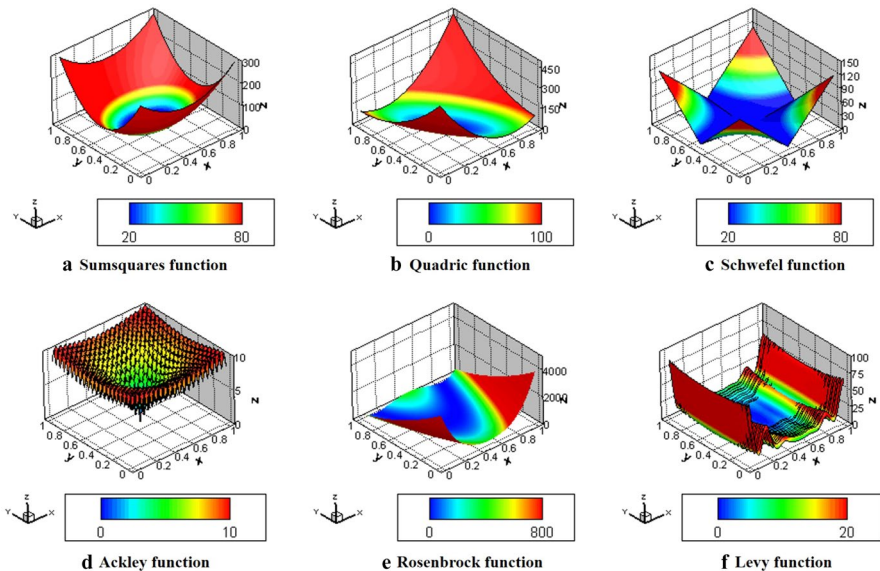


Fig. 10 Analytic functions in two-dimensional design variables space

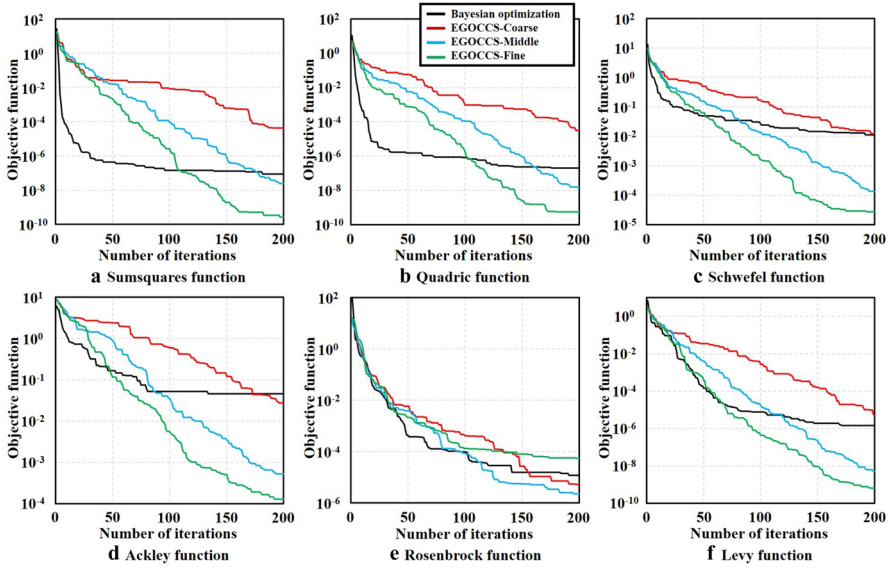


Fig. 11 Optimization histories of various function problems in 2-dimensional variable space

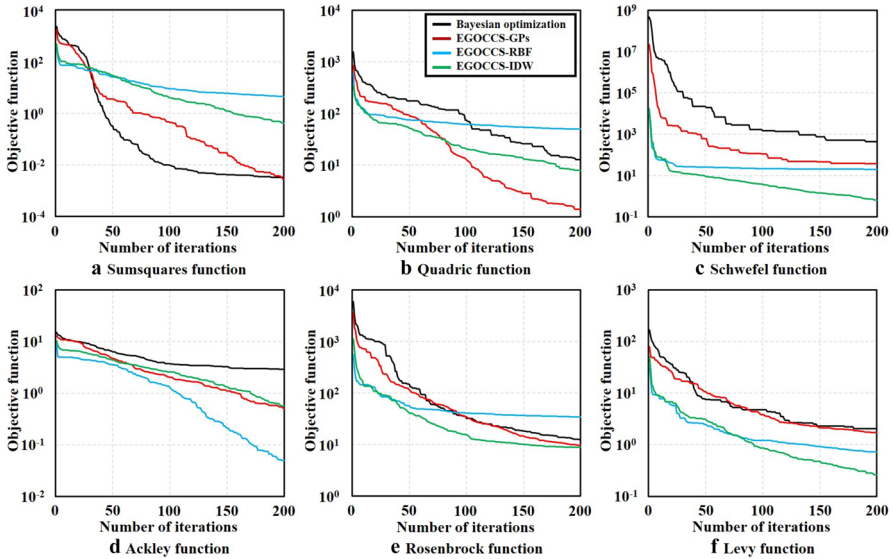


Fig. 12 Optimization histories of various function problems in 10-dimensional variable space

conditions of the 40, 80 160-dimensional variables space are almost same with that of the 10-dimensional variables space excepting the number of initial sample points (20 → 80, 160, 320).

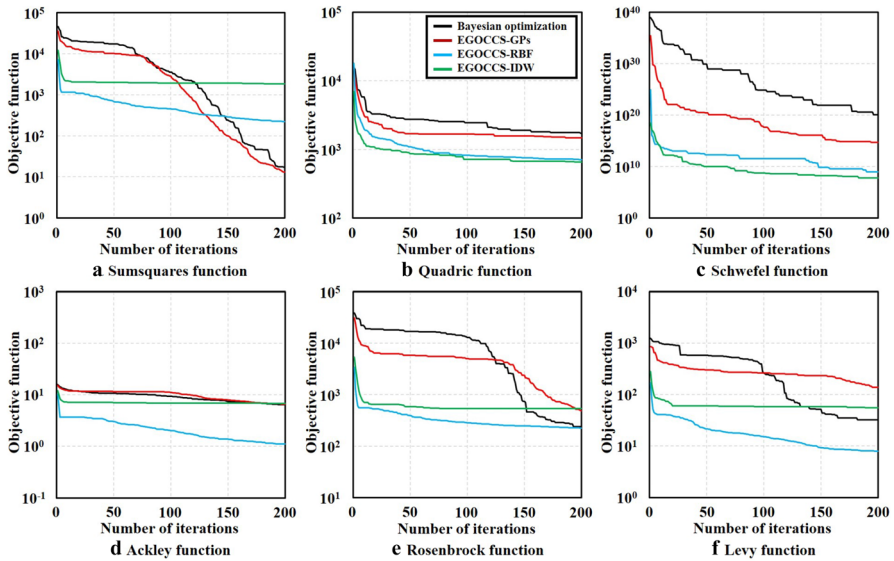


Fig. 13 Optimization histories of various function problems in 40-dimensional variable space

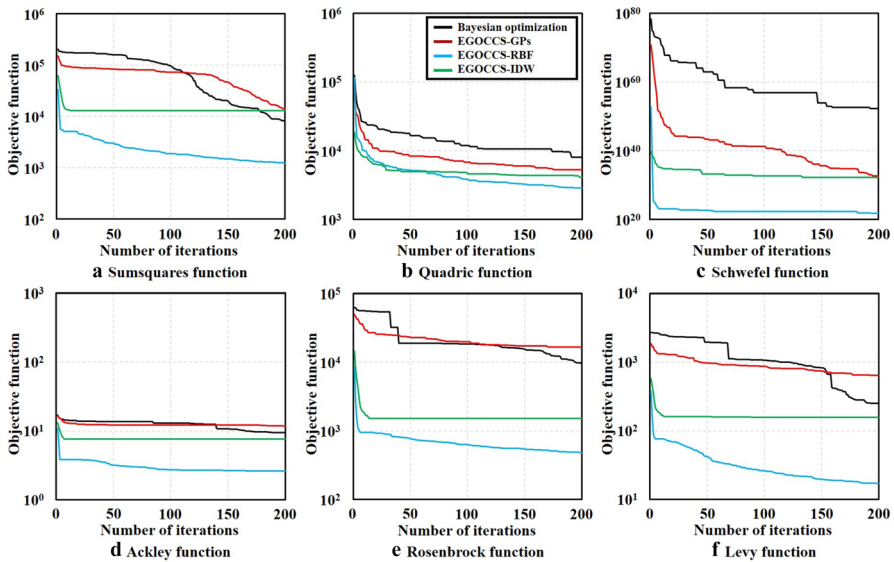


Fig. 14 Optimization histories of various function problems in 80-dimensional variable space

Firstly, the optimizations of the analytic functions are solved in the 2-dimensional variables space. Each optimization problem was solved 10 times by the Bayesian approach and EGOCCS with three different strategies of Coarse, Middle, Fine as shown in Table 1. The type of the surrogate model used in EGOCCS is the Gaussian

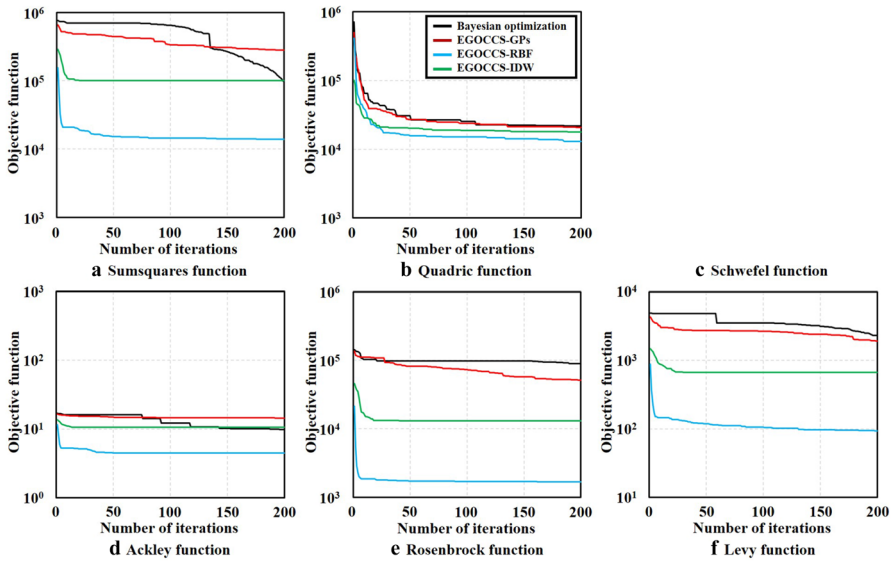


Fig. 15 Optimization histories of various function problems in 160-dimensional variable space

process regression to examine the pure effect of the difference between the algorithms. The number of initial sample points is 8, and then 200 additional sample points are evaluated. In higher dimensional optimization problems, initial surrogate models are necessary to be accurate for efficient optimizations and the number of initial sample points becomes larger. Then, the number of additional sample points is desired to be smaller to reduce the construction cost of surrogate models in the optimization. In many global optimization problems motivated by engineering applications, the number of function evaluations is severely limited by time or cost (Julien et al. 2009). In this research, the number of additional functional evaluations is limited to 200. The setting of major parameters of GA is the mutation rate of 0.1, population size of 100 and number of generations of 50. In the optimization history of Fig. 11, the horizontal axis represents the number of optimization iterations, and the vertical axis represents the objective function to be minimized. The black line is the optimization history of the Bayesian approach, and the red, blue and green lines are respectively the optimization histories of the EGOCCS approach with the strategies of Coarse, Middle and Fine. It is confirmed that EGOCCS-Fine can explore better optimal solutions in the Sumsquares, Quadric, Schwefel, Ackley and Levy functions compared with the EGOCCS-Middle and EGOCCS-Coarse, while the EGOCCS-Fine cannot reach exact optimal solutions in the Rosenbrock function. This result indicates the tradeoff relationship between local/global searches, and the Fine-strategy can reach exact optimal solutions while that is sometimes difficult to reach the vicinity of the global optimal solution. On the other hand, the Coarse-strategy can reach the vicinity of the global optimal solution while that is difficult to reach the exact optimal solutions. It is also confirmed that EGOCCS-Middle can obtain better solutions in all analytic functions compared with the conventional

Bayesian optimization, which indicates that the EGOCCS has the ability to explore better solutions in various optimization problems.

Secondly, in the 10-dimensional variables space. According to another report (Ziyu et al. 2013), the 10-dimensional variables space is almost the upper limit of the Bayesian approach to explore exact global optimal solutions. Each optimization problem was solved 10 times by the Bayesian optimization and EGOCCS with three different surrogate models of Gaussian process regression (GPs), radial basis function interpolation (RBF), and inverse distance weighting (IDW). The exploring strategy is set to $\epsilon_{min} = 1.0 \times 10^{-3} \times \sqrt{n_d}$ and $\epsilon_{max} = 2.0 \times 10^{-1} \times \sqrt{n_d}$ as seen in Fig. 4 and Table 1. The number of initial sample points is 20, and then 200 additional sample points are evaluated. The setting of major parameters of GA is the mutation rate of 0.1, population size of 200 and number of generations of 200. In the optimization history of Fig. 12, the black line is the optimization history of the Bayesian approach, and the red, blue and green lines are respectively optimization histories of the EGOCCS with GPs, RBF and IDW. It is confirmed that EGOCCS-GPs can obtain better solutions in all cases compared with the conventional Bayesian optimization, which indicates that the EGOCCS algorithm has ability to explore better solutions in various optimization problems in this strategy and also in 10-dimensional variables space. However, since the EGOCCS-GPs uses the Gaussian process regression for the surrogate model, the optimization cost is not reduced compared with the conventional Bayesian optimization. It is also confirmed that the EGOCCS-RBF can obtain better solutions in the Schwefel, Ackley and Levy functions compared with the Bayesian optimization. The EGOCCS-IDW can obtain the best solution in the Schwefel, Rosenbrock and Levy functions. These results indicate that the EGOCCS algorithm has the ability to obtain better solutions while reducing the construction cost for the surrogate model in several function cases. In the 10-dimensional variables space, it is considered that the EGOCCS-GPs is the most appropriate to obtain global optimal solutions since optimization problems in the 10-dimensional variables space don't require much cost for the construction and the EGOCCS-GPs has the ability to obtain better solutions in various optimization problems compared with the conventional Bayesian optimization.

Thirdly, in the 40-dimensional variables space. It is considered that the optimization in the 40-dimensional variables space is difficult to obtain the global optimal solutions with the Bayesian approach. It is confirmed from Fig. 13 that the EGOCCS-RBF can obtain better solutions in the Quadric, Schwefel, Ackley, Rosenbrock and Levy functions compared with the Bayesian optimization while it cannot obtain better solution in the Sumsquares function. Since the Sumsquares function is the simplest function and the Gaussian process regression can easily fit the exact functional shape, the Bayesian optimization can explore better solutions compared with the EGOCCS. It is also confirmed that the EGOCCS-IDW can obtain the best solutions in the Quadric and Schwefel, while it falls into local optimal solutions in the other functions. Since IDW is the simplest prediction method only using the neighboring sample points information, the prediction accuracy should be lower than the other response surface models so that the optimization falls into local optimal solutions in the analytic functions. In the 40-dimensional variables space, it is considered that the EGOCCS-RBF is the most appropriate to obtain the global optimal

solution robustly since the EGOCCS-RBF can obtain better solutions in the various analytic functions and can dramatically reduce the optimization cost.

Fourthly, in the 80-dimensional variables space. It is considered that to obtain global optimal solutions in the 80-dimensional variables space with the Bayesian optimization is impossible. It is confirmed from Fig. 14 that the EGOCCS-RBF can obtain the best solutions in all analytic functions. Therefore, it is considered that the EGOCCS-RBF is the most appropriate to obtain better solutions in the 80-dimensional variables space.

Fifthly, in the 160-dimensional variables space. It is considered that to obtain global optimal solutions in the 160-dimensional variables space with the Bayesian optimization is impossible. The optimization in the Schwefel function cannot be executed since the function value exceeds the maximum number of considerable digits. It is confirmed from Fig. 15 that the EGOCCS-RBF can obtain the best solutions in all analytic functions. However, it is also confirmed that the update of optimal solutions seldom occurs in the optimization process and then the obtained solutions by the EGOCCS-RBF are not at the vicinities of exact global optimal solutions. Therefore, it is considered that the EGOCCS-RBF should be used in the optimization problems whose number of design variables is less than 160, in which the update of optimal solutions continuously occurs in the optimization process. It can be expected that the EGOCCS-RBF can finally obtain exact optimal solutions by iteratively performing the sub iterations as Fig. 4b.

For more quantitative comparison, gradient-based optimizations were executed from the optimal solutions obtained by the Bayesian/EGOCCS in the cases of the Ackley and Levy functions that are multimodal functions. The conjugate gradient algorithm was used as the gradient-based optimization algorithm (Hestenes and Stiefel 1952). When the final objective function value was less than 10^{-5} , it was considered as reaching the global optimal solution. In Table 2, the numbers of cases reaching the global optimal solution (among 10 trials) were summarized. It was confirmed that most optimal solutions of the EGOCCS-RBF reached the global optimal solutions in 10, 40 and 80 design variables in the Ackley function while that of the Bayesian optimization falls into local optimal solutions. This means that the EGOCCS could reach the vicinities of the global optimal solutions within the limited number of functional evaluations. In other words, the EGOCCS would be utilized as the preprocess for the local search using the gradient-based optimization. It was confirmed that most optimal solutions of the Levy function

Table 2 Number of cases reaching the global optimal solutions

| | Ackley | | | | Levy | | | |
|-----------------------|--------|--------|--------|---------|--------|--------|--------|---------|
| | 10 dim | 40 dim | 80 dim | 160 dim | 10 dim | 40 dim | 80 dim | 160 dim |
| Bayesian optimization | 5 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| EGOCCS-GPs | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EGOCCS-RBF | 10 | 10 | 9 | 1 | 0 | 0 | 0 | 0 |
| EGOCCS-IDW | 7 | 0 | 0 | 0 | 4 | 0 | 0 | 0 |

couldn't reach the global optimal solutions. It is considered that the Levy function is too complex to reach the vicinities of the global optimal solutions within the limited number of functional evaluations. The optimization costs with various number of design variables are summarized in Fig. 16. The horizontal axis represents the number of dimensions, and the vertical axis represents the computational time required in each optimization algorithm. The computational time is calculated by averaging 60 times optimizations (6 analytic function \times 10 times optimizations). It is confirmed that the EGOCCS-GPs has higher computational cost compared with the Bayesian optimization due to additional computational cost for SVM classifier. It is also confirmed that the EGOCCS-RBF and EGOCCS-IDW have lower costs than the Bayesian optimization due to the cost reduction in the construction of surrogate models. It can be confirmed that the developed EGOCCS algorithm can obtain better optimal solutions inexpensively than the conventional Bayesian optimization in the high dimensional design variables space.

Sumsquares function

$$f(\mathbf{x}) = \sum_{i=1}^{n_d} [ix'_i]^2 \tag{4.1.1}$$

$$x'_i = -10 + 20x_i, \quad 0 \leq x_i \leq 1, \quad f_{min} = f(\mathbf{x}) = f(0.5, 0.5, \dots, 0.5) = 0$$

Quadric function

$$f(\mathbf{x}) = \sum_{i=1}^{n_d} \left(\sum_{j=1}^i x'_j \right)^2 \tag{4.1.2}$$

$$x'_i = -10 + 20x_i, \quad 0 \leq x_i \leq 1, \quad f_{min} = f(\mathbf{x}) = f(0.5, 0.5, \dots, 0.5) = 0$$

Schwefel function

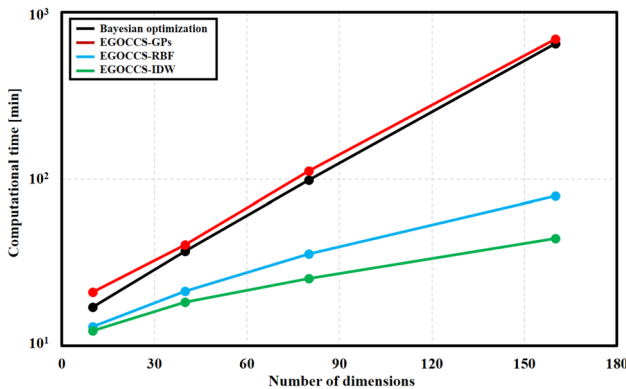


Fig. 16 Optimization costs with various numbers of dimensions

$$f(\mathbf{x}) = \sum_{i=1}^{n_d} |x'_i| + \prod_{i=1}^{n_d} |x'_i| \quad (4.1.3)$$

$$x'_i = -10 + 20x_i, \quad 0 \leq x_i \leq 1, \quad f_{\min} = f(\mathbf{x}) = f(0.5, 0.5, \dots, 0.5) = 0$$

Ackley function

$$f(\mathbf{x}) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{n_d} \sum_{i=1}^{n_d} x_i'^2}} - e^{\frac{1}{n_d} \sum_{i=1}^{n_d} \cos(2\pi x'_i)} \quad (4.1.4)$$

$$x'_i = -10 + 20x_i, \quad 0 \leq x_i \leq 1, \quad f_{\min} = f(\mathbf{x}) = f(0.5, 0.5, \dots, 0.5) = 0$$

Rosenbrock function

$$f(\mathbf{x}) = \sum_{i=1}^{n_d-1} \left[100(x'_{i+1} - x_i'^2)^2 + (1 - x'_i)^2 \right] \quad (4.1.5)$$

$$x'_i = -2.048 + 4.096x_i, \quad 0 \leq x_i \leq 1$$

$$f_{\min} = f(\mathbf{x}) = f(3.048/4.096, 3.048/4.096, \dots, 3.048/4.096) = 0$$

Levy function

$$f(\mathbf{x}) = \sin^2(\pi y_1) + \sum_{i=1}^{n_d-1} (y_i - 1)^2 (1 + 10 \sin^2(\pi y_i + 1)) + (y_{n_d} - 1)^2 (1 + \sin^2(2\pi x'_{n_d}))$$

$$x'_i = -10 + 20x_i, \quad y_i = 1 + \frac{x'_{n_d} - 1}{4}, \quad 0 \leq x_i \leq 1, \quad f_{\min} = f(\mathbf{x}) = f\left(\frac{11}{20}, \frac{11}{20}, \dots, \frac{11}{20}\right) = 0 \quad (4.1.6)$$

4.2 High dimensional aerodynamic shape optimization problem of biconvex airfoil

Since the performance of the EGOCCS has been validated in 4.1 with the various analytic function problems by comparing with the conventional Bayesian approach, high dimensional aerodynamic shape optimization problems in two-dimensional supersonic flow field are solved in this section to demonstrate the applicability of the EGOCCS to more realistic optimization problems. The baseline airfoil is set to a biconvex airfoil (Fig. 17a, c, black line). New airfoil shapes are given as symmetrical airfoils, and the shape of the one side surface is modified by 2, 10, 40, 80 Hicks-Henne bump functions arranged in even intervals (Fig. 17b). A Hicks-Henne bump function is expressed as following equation.

$$h_i(x) = h_{\max}(x_{\text{bump},i}) \times \sin^3 \left(\pi (1-x)^{\frac{\log(x_{\text{bump},i})}{\log(0.5)}} \right) \quad (4.2.1)$$

$$h_{\max}(x_{\text{bump},i}) = 0.1 \times tc(x_{\text{bump},i}) \times dv_i$$

where x_{bump} and tc are the bump apex position and the thickness of the biconvex airfoil at x_{bump} . h_{\max} indicates the maximum height of the bump, and it can be

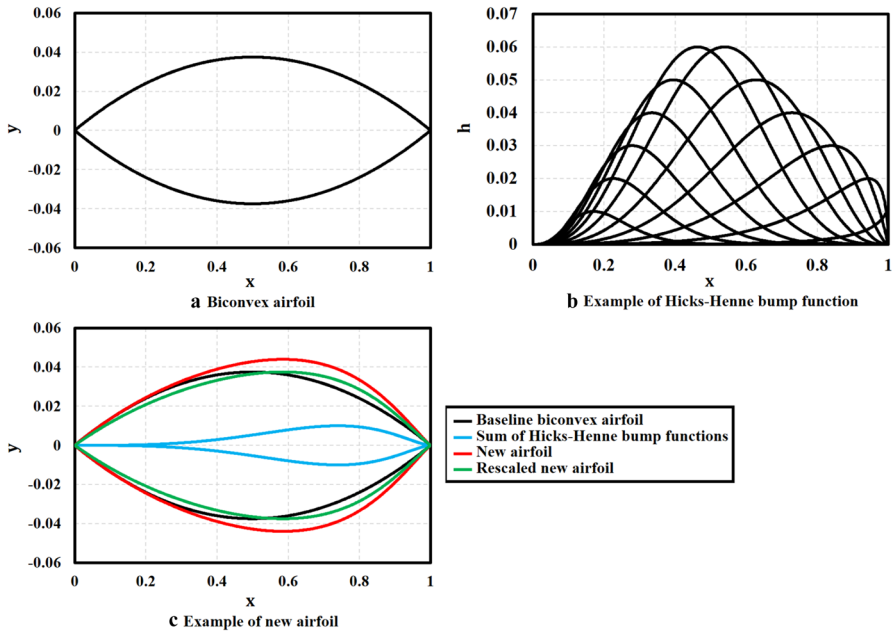


Fig. 17 Example of shape modification

adjusted by the design variable dv_i . The displacement of a new airfoil at a location x is expressed by a summation of 2, 10, 40, 80 bump functions as following equation.

$$\Delta y(x) = \sum_{i=1}^{n_d} h_i(x), \quad 0 < x < 1, \quad 0 < dv_i < 1 \tag{4.2.2}$$

A new airfoil is created by adding the displacement (Fig. 17c, blue line) on the baseline biconvex airfoil (Fig. 17c, red line), and then the y-coordinate of the new airfoil is rescaled to keep a constant airfoil area (Fig. 17c, green line). The negative thickness of airfoil never occurs with this definition of design variables. In this problem, the drag coefficient of airfoil is minimized. The aerodynamic performance is evaluated at the freestream Mach number of 1.7 and angle of attack of 0 degree by using inviscid CFD computations. This optimization problem is solved three times with different sets of initial sample points at all conditions of the number of design variables. The numbers of initial sample points are set to 8 at $n_d = 2$ and 200 at the other cases. New sample points are iteratively added 200 times within a set of sub iterations. Table 3 shows the objective function values of the obtained optimal solutions, and Fig. 18 shows the average of the three times optimization histories. The dotted lines indicate the results using the GA directly to the 10, 40 and 80 design variables problems, that were also executed for comparison purposes. The setting of major parameters of GA is the mutation rate of 0.1, population size of 200 and 3, 5 and 7 generations that correspond to the same numbers of functional evaluations with the EGOCCS cases. It is confirmed that the EGOCCS algorithm can obtain lower drag

Table 3 Objective function values of optimal solutions at every 200 iteration

| | init. | 200 | 400 | 600 | 800 | 1000 | 1200 |
|---------------------|---------|---------|---------|---------|---------|---------|---------|
| EGOCCS ($n_d=2$) | | | | | | | |
| Trial 1 | 0.02137 | 0.02133 | | | | | |
| Trial 2 | 0.02138 | 0.02133 | | | | | |
| Trial 3 | 0.02139 | 0.02133 | | | | | |
| Average | 0.02138 | 0.02133 | | | | | |
| EGOCCS ($n_d=10$) | | | | | | | |
| Trial 1 | 0.02076 | 0.02054 | 0.02052 | | | | |
| Trial 2 | 0.02071 | 0.02055 | 0.02053 | | | | |
| Trial 3 | 0.02069 | 0.02056 | 0.02053 | | | | |
| Average | 0.02072 | 0.02055 | 0.02053 | | | | |
| EGOCCS ($n_d=40$) | | | | | | | |
| Trial 1 | 0.02073 | 0.02018 | 0.02007 | 0.02002 | 0.01995 | | |
| Trial 2 | 0.02054 | 0.02023 | 0.02013 | 0.02006 | 0.02000 | | |
| Trial 3 | 0.02060 | 0.02021 | 0.02012 | 0.02003 | 0.01998 | | |
| Average | 0.02062 | 0.02021 | 0.02011 | 0.02003 | 0.01997 | | |
| EGOCCS ($n_d=80$) | | | | | | | |
| Trial 1 | 0.02064 | 0.02039 | 0.01981 | 0.01970 | 0.01957 | 0.01946 | 0.01939 |
| Trial 2 | 0.02062 | 0.01969 | 0.01957 | 0.01956 | 0.01952 | 0.01943 | 0.01940 |
| Trial 3 | 0.02076 | 0.01987 | 0.01957 | 0.01944 | 0.01936 | 0.01926 | 0.01917 |
| Average | 0.02067 | 0.01998 | 0.01965 | 0.01956 | 0.01948 | 0.01938 | 0.01932 |
| GA ($n_d=10$) | | | | | | | |
| Trial 1 | 0.02070 | 0.02060 | 0.02054 | | | | |
| Trial 2 | 0.02075 | 0.02066 | 0.02056 | | | | |
| Trial 3 | 0.02071 | 0.02066 | 0.02054 | | | | |
| Average | 0.02072 | 0.02064 | 0.02055 | | | | |
| GA ($n_d=40$) | | | | | | | |
| Trial 1 | 0.02053 | 0.02034 | 0.02018 | 0.02014 | 0.02011 | | |
| Trial 2 | 0.02067 | 0.02037 | 0.02037 | 0.02017 | 0.02011 | | |
| Trial 3 | 0.02056 | 0.02047 | 0.02035 | 0.02034 | 0.02023 | | |
| Average | 0.02059 | 0.02039 | 0.02030 | 0.02021 | 0.02015 | | |
| GA ($n_d=80$) | | | | | | | |
| Trial 1 | 0.02052 | 0.02049 | 0.02009 | 0.02009 | 0.02003 | 0.01998 | 0.01982 |
| Trial 2 | 0.02065 | 0.02032 | 0.02024 | 0.02003 | 0.01985 | 0.01985 | 0.01983 |
| Trial 3 | 0.02047 | 0.02031 | 0.02006 | 0.02001 | 0.01996 | 0.01971 | 0.01963 |
| Average | 0.02055 | 0.02037 | 0.02013 | 0.02004 | 0.01995 | 0.01985 | 0.01976 |

coefficient values than the GA in all cases. This result can emphasize the superiority of the developed EGOCCS algorithm over the conventional approach. Since the three optimizations by EGOCCS with different sets of initial sample points give almost same objective function values in all numbers of design variables, it could be confirmed that the developed EGOCCS algorithm worked robustly even in 80

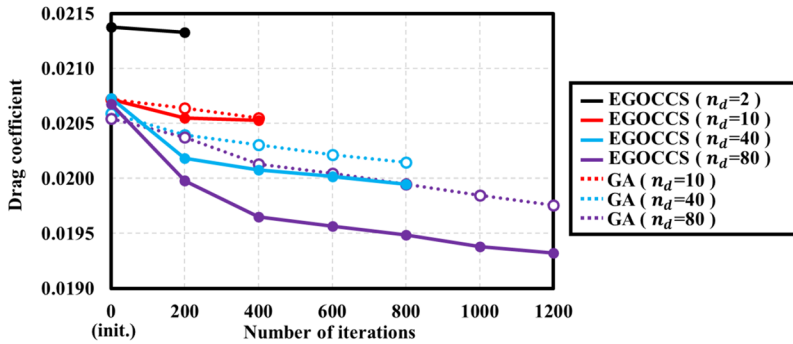


Fig. 18 Optimization histories in various numbers of design variables

dimensional design variables space. Figures 19 and 20 show the shapes and pressure visualizations around optimal designs obtained by the EGOCCS algorithm. With respect to the optimal designs, it was demonstrated that the theoretical optimal shape with minimum pressure drag has a blunt trailing edge (Chapman 1951; Klunker and Harder 1952). In addition, an optimal profile is shown in Payot et al. (2017) at the freestream Mach number of 2.0 (1.7 in this research) and angle of attack of 0 degree, and the optimal profile also has the blunt trailing edge. Therefore, it is considered that the obtained optimal designs have thicker trailing edge shapes to express the blunt trailing edge within the given degree of freedom. Larger numbers of design

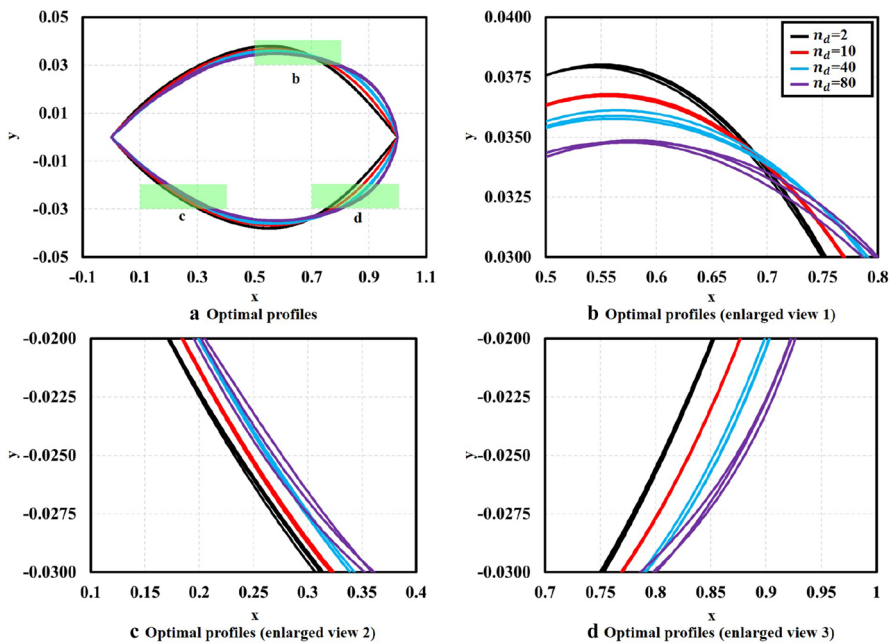


Fig. 19 Shapes of optimal solutions obtained by EGOCCS

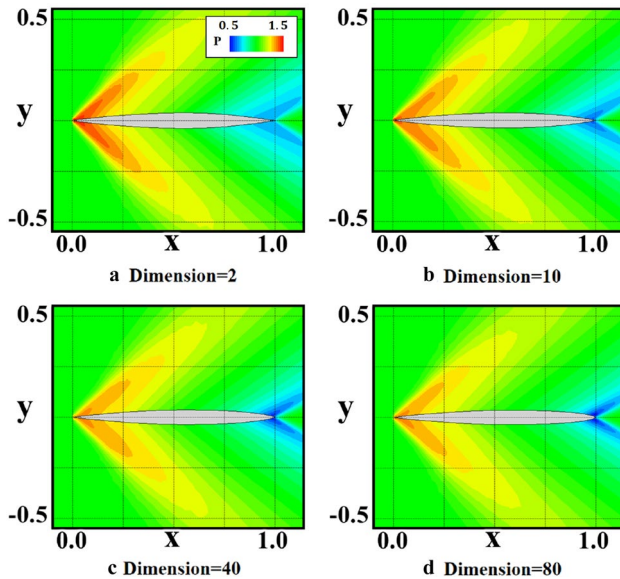


Fig. 20 Pressure visualizations around optimal solutions obtained by EGOCCS

variables enable to express the airfoil shape in more detail, and then the optimal airfoil should have better performance than that obtained with smaller numbers of design variables. It is confirmed from Figs. 18, 19 and 20 that the optimal designs with larger numbers of design variables have thinner leading edge and thicker trailing edge, and the static pressure around the leading edge becomes lower, and then the objective function value becomes lower. Since better solutions were obtained with larger numbers of design variables (with more degree of freedom) monotonically, it was confirmed that the EGOCCS algorithm could be appropriately used in the present high-dimensional aerodynamic shape optimization problems without to be caught in local optimal solutions.

5 Concluding remarks

In this research, an efficient global optimization method via clustering/classification methods and exploration strategy (EGOCCS) has been developed to efficiently solve high-dimensional optimization problems. The conventional Bayesian optimization method takes a huge calculation cost to construct a response surface model, and the accuracy of that decreases at high-dimensional cases. To tackle these issues, a novel efficient global optimization method without using probabilistic values as standard deviation obtained in the Gaussian process was proposed and aimed to reduce the construction cost of the response surface model and to efficiently obtain global optimal solutions in high-dimensional optimization problems. The developed algorithm has achieved better performance than

conventional approaches by using the machine learning methods of a density-based clustering method and a support vector machine classification method.

Two optimization problems were solved to verify the usefulness of the developed method of EGOCCS. First optimization was executed to demonstrate the validity of the EGOCCS in 2, 10, 40, 80, 160-dimensional analytic function problems that were solved by both of the Bayesian optimization and the EGOCCS. In the 2-dimensional analytic functions, the EGOCCS could obtain better solutions in all analytic functions compared with the conventional Bayesian approach by adopting an appropriate exploration strategy. In the 10-dimensional analytic functions, the EGOCCS on the Gaussian process regression had the ability to obtain better solutions in many problems compared with the conventional Bayesian optimization. In the 40, 80 and 160-dimensional analytic functions, the EGOCCS with the radial basis function interpolation was the most appropriate to obtain better optimal solutions. In addition, it was also confirmed that the EGOCCS with the radial basis function interpolation and inverse distance weighting could execute with much lower computational costs than the Bayesian optimization. Second optimization was executed to examine the effect of the EGOCCS in 2, 10, 40, 80-dimensional aerodynamic shape optimization problems of a biconvex airfoil. The biconvex airfoil was deformed by 2, 10, 40, 80 Hicks-Henne bump functions, to minimize its drag coefficient at the freestream Mach number of 1.7 and angle of attack of 0 degree. Larger numbers of design variables could express optimal airfoil shapes in more detail, and then the optimal airfoils had better objective function values compared with that with smaller numbers of design variables. Since better solutions were obtained with larger numbers of design variables (with more degree of freedom) monotonically, it was confirmed that the EGOCCS algorithm could be appropriately used in the present high-dimensional aerodynamic shape optimization problems without to be caught in local optimal solutions.

The proposed EGOCCS algorithm brings higher compatibility with other response surface models such as neural network regression, support vector regression, random forest regression and so on. For example, the neural network can be used when a very large number of design variables has to be taken into consideration. The compatibility with the other response surface models will be examined as future work. In addition, since the EGOCCS has been developed as noninvasive approaches, it can be applied to optimization problems of other disciplines. It is necessary to efficiently solve three challenging optimization problems that are discontinuous optimization problem, optimization problem with infeasible regions and high dimensional optimization problem, to realize efficient global topology optimizations in supersonic flow conditions. By using the developed EGODISC (Ban and Yamazaki 2019) and EGOCCS algorithms, an efficient global optimization method for the three challenging optimization problems can be constructed. Therefore, the topology optimization in supersonic flow conditions will be carried out as our future works.

References

- Alex JS, Bernhard S (2004) A tutorial on support vector regression. *Stat Comput Arch* 14(3):199–222
- Ban N, Yamazaki W (2019) Development of efficient global optimization method for discontinuous optimization problems with infeasible region via classification method. *J Adv Mech Des Syst Manuf* 13(1):p.JAMDSM0017
- Ban N, Yamazaki W, Kusunose K (2016) A fundamental design optimization of innovative supersonic transport configuration and its design knowledge extraction. *Trans Jpn Soc Aeronaut Space Sci* 59(6):340–348
- Ban N, Yamazaki W, Kusunose K (2018) Low-boom/low-drag design optimization of innovative supersonic transport configuration. *J Aircr* 55(3):1071–1081
- Bouhlef MA, Bartoli N, Otsmane A, Morlier J (2016) Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. *Struct Multidiscip Optim* 53(5):935–952
- Chapman DR (1951) Airfoil profiles for minimum pressure drag at supersonic velocities—general analysis with application to linearized supersonic flow. *NACA TN* 2264
- Corinna C, Vladimir V (1995) Support-vector networks. *Mach Learn* 20:273–297
- Crina G, Ajith A (2009) A novel global optimization technique for high dimensional functions. *Int J Intell Syst* 24:421–440
- Deb K, Deb D (2014) Analysing mutation schemes for real-parameter genetic algorithms. *Int J Artif Intell Soft Comput* 4(1):1–28
- Eshleman L, Schaffer JD (1993) Real-coded genetic algorithms and interval-schemata. *Found Genet Algorithms* 2:187–202
- Ester M, Kriegl HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the 2nd international conference on knowledge discovery and data mining*, Portland, OR, AAAI Press, pp 226–231
- Forrester AI, Sobester A, Keane AJ (2008) *Engineering design via surrogate modelling: a practical guide*. Wiley, Hoboken
- Fujii D, Suzuki K, Ohtsubo H (2000) Topology optimization of structures using the voxel finite element method. *Trans Jpn Soc Comput Eng Sci* 2:87–94
- Han ZH, Görtz S (2012) Hierarchical kriging model for variable-fidelity surrogate modeling. *AIAA J* 50(9):1885–1896
- Han ZH, Zhang Y, Song CX, Zhang KS (2017) Weighted gradient-enhanced kriging for high-dimensional surrogate modeling and design optimization. *AIAA J* 55(12):4330–4346
- Hestenes MR, Stiefel E (1952) Methods of conjugate gradients for solving linear systems. *J Res Natl Bureau Stand* 49(6):409–436
- Igarashi K (2008) Technological overview of the next generation Shinkansen high-speed train Series N700. *Jpn Soc Mech Eng, No*, pp 07–66
- Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Glob Optim* 13:455–492
- Julien V, Emmanuel V, Eric W (2009) An informational approach to the global optimization of expensive-to-evaluate functions. *J Glob Optim* 44(4):509–534
- Klunker EB, Harder K (1952) Comparison of supersonic minimum-drag airfoils determined by linear and nonlinear theory. *National Advisory Committee for Aeronautics Technical Note (NACA TN) 2623*, Washington
- Lars K, Alastair T, Gerrit R (2002) Application of topology, sizing and shape optimization methods to optimal design of aircraft components. Airbus UK Ltd, Altair Engineering Ltd
- Liu B, Zhang Q, Gielen GGE (2014) A Gaussian process surrogate model assisted evolutionary algorithm for medium scale expensive optimization problems. *IEEE Trans Evol Comput* 18(2):180–192
- Ma Z, Chen H, Zhou C (2008) A study of point moving adaptivity in gridless method. *Comput Methods Appl Mech Eng* 197(21–24):1926–1937
- Mohamed AB, Joaquim RRAM (2019) Gradient-enhanced kriging for high-dimensional problems. *Eng Comput* 35:157–173
- Payot AD, Rendall T, Allen CB (2017) Mixing and refinement of design variables for geometry and topology optimization in aerodynamics. In: *35th AIAA Applied Aerodynamics Conference, AIAA 2017-3577*, American Institute of Aeronautics and Astronautics, Reston, Virginia
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) *Radial basis function interpolation*. In: *Numerical recipes: the art of scientific computing*, 3 edn

- Rahib HA, Mustafa T (2015) Optimization of high-dimensional functions through hypercube evaluation. *Comput Intell Neurosci* 967320:2015
- Rasmussen CE, Williams CKI (2006) *Gaussian processes for machine learning*. MIT Press, Cambridge
- Rumpfkeil MP, Yamazaki W, Mavriplis DJ (2011) Countering the curse of dimensionality using higher-order derivatives. In: *SIAM conference on computational science and engineering*, Reno, Nevada
- Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of the 1968 ACM National Conference*, pp 517–524
- Suga Y, Yamazaki W (2015) Aerodynamic uncertainty quantification of supersonic biplane airfoil via polynomial chaos approach. *AIAA Paper* 2015-1815
- Sugimoto J, Yonemoto K, Fujikawa T (2019) A surrogate-assisted dynamically distributed genetic algorithm for multi-objective design optimization. *APISAT* 2019:2056–2065
- Takenaka K, Hatanaka K, Yamazaki W, Nakahashi K (2008) Multidisciplinary design exploration for a Winglet. *J Aircr* 45(5):1601–1611
- Volodymyr M, Adria PB, Mehdi M, Alex G, Timothy PL, Tim H, David S, Koray K (2016) Asynchronous methods for deep reinforcement learning. In: *Proceedings of the 33rd international conference on machine learning*, pp 1928–1937
- Williams CKI, Carl ER (1996) *Gaussian Processes for Regression*. *Adv Neural Inf Process Syst* 8:514–520
- Yamazaki W, Mavriplis DJ (2013) Derivative-enhanced variable fidelity surrogate modeling for aerodynamic functions. *AIAA J* 51(1):126–137
- Ziyu W, Masrour Z, Frank H, David M, Nando F (2013) Bayesian Optimization in high dimensions via random embeddings. In: *Proceedings of IJCAI'13*

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