Black-Box Optimization and Its Applications

Wataru Kumagai * and Keiichiro Yasuda **

Abstract: Black-Box Optimization (BBO) is a class of optimization problem using only design-variable values and objective-function values; i.e., assuming to an environment where solutions can be evaluated by numerical simulations or measurements. BBO is frequently used in various engineering fields and can use an advancement of numerical simulation and measurements technologies. It will play an important role in the optimization for large and complex real systems, such as "Systems of Systems (SOS)". From this background, this article reviews the present and future of BBO research and its application based on recent research trends. This article also provides a description of direct-search optimization algorithm to solve BBO especially metaheuristics and focuses on how to handle constraints in BBO.

Keywords: Black-box optimization, sequential model-based optimization, metaheuristics, invariance, covariance matrix adaptation evolution strategy, particle swarm optimization, and constrained optimization.

* Innovation Center, Marketing Headquarter, Yokogawa Electric Corporation, Ltd., 2-9-32 Nakacho, Musashino-shi, Tokyo, Japan

** Graduate School of Systems Design, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji-shi, Tokyo, Japan

*E-mail: Wataru.Kumagai@yokogawa.com

**E-mail: k-yasuda@tmu.ac.jp

Adapted from Keiichiro Yasuda and Wataru Kumagai.
"Black-Box Optimization and Its Applications (written in Japanese)."
Journal of The Society of Instrument and Control Engineers, 59-12, 914/917 (2020).
Partly translated by permission of The Society of Instrument and Control Engineers.

1 Introduction

Super Smart Society (Society 5.0)[1] and Bioeconomy[2] get attention recently as smarter or more advanced concept to realize pre-symptomatic state, healthy life expectancy, circular economy, and energy efficiency and conservation. To adapt to this trend, industry transformation is required, such as product design has more efficiency or manufacturing process is fundamental reformed by smart cells or genetics. Conversely, many systems become more large-scale and complex according to an appearance of *System of Systems* (SoS), which is a new kind of system consisting of connected various systems with operational autonomy and control autonomy (e.g., distribution system, medical system, power and energy system, or railway system).

From this background, the need for practical optimization is growing more urgent. Achieving this needs an integrated methodology by combination of not only on optimization theory but also surrounding technologies, such as simulation, artificial intelligence-based (AI) modeling, and computing technologies; e.g., general-purpose graphics processing units (GPGPU). This type of optimization is data-driven approach using on-line data obtained from simulator or sensor, called black-box optimization (BBO). BBO uses only design-variable value and its objective-function value because the objective function is black-box or expensive function with its unknown landscape. Mathematical programming, which is a classical optimization algorithm's class using analytical information about the objective function (e.g., gradient or Hessian matrix) and its properties (e.g., convexity or variable dependency), cannot be applied to BBO. Therefore, BBO technology with an ability to adapt to quickly change of the environment surrounding optimization is required. This article focuses some representative approaches of BBO especially metaheuristics[3]. This article reviews its overview, desirable properties, constraint-handling techniques, and its applications based on recent research trends.

Mathematical notations used in this article are as follows; $\mathbb{R}$ denotes the set of real numbers, $\mathbb{N}$ denotes the set of natural numbers, $\emptyset$ denotes the empty set, respectively. $[a, b], (a, b)$ denote the closed interval and the open interval between a and b, where $a, b \in \mathbb{R}$ $(a < b)$. The probability distributions $\mathcal{N}$,$\mathcal{U}(a, b)$ denote the multi-variable standard normal (or gaussian) distribution and the continuous uniform distribution within $[a, b]$.

2 Black-box optimization

This article focuses on an unconstrained optimization problem minimizing an objective function $f(\boldsymbol{x})$, but Section 4 does on a constrained optimization problem minimizing the objective function $f(\boldsymbol{x})$ such that constraints $g_\kappa(\boldsymbol{x}) \leq 0$ $(\kappa = 1,2, \dots, K)$[1]; where $\boldsymbol{x} \in \mathbb{R}^N$ denotes the design variable, $f: \mathbb{R}^N \to \mathbb{R}$ denotes the objective function, and $g_\kappa(\boldsymbol{x}): \mathbb{R}^N \to \mathbb{R}$ $(\kappa = 1,2, \dots, K)$ denotes the K constraint functions, respectively.

Direct-search or derivative-free method, and sequential model-based optimization (SMBO) are global optimization or BBO approach. They search the global optima efficiently in black-box or expensive function using only design-variable value and its objective-function value, not the gradient to reduce a risk of getting trapped at low-grade local optima. While the direct-search or derivative-free method search the optima in the objective function directly (e.g., metaheuristics or evolutionary algorithm), SMBO does it indirectly using a

[1] Constraint condition is classified as unequally and equally constraints, but this article only focuses on unequally constraints because equally constraints are often relaxed to unequally constraints by introducing small tolerance.

surrogate function sequentially, i.e., response surface methodology[4]. The procedure of SMBO is the following:

(i) Step 1: generating a surrogate function to approximate the objective function based on obtained data set of the design variables $\boldsymbol{x}$ and the objective-function values $f(\boldsymbol{x})$.
(ii) Step 2: obtaining the temporary optima in the surrogate function.
(iii) Step 3: sampling based on the obtained optima.

For example, Bayes optimization or Kriging method[5] is a well-known SMBO and iterates the following steps: generating a probabilistic surrogate function from sample data (e.g., Gaussian process regression); constructing an acquisition function corresponding to the surrogate model and searching its maxima (e.g., expected improvement); and sampling next candidate solutions around the maxima. SMBO is expected to advance by combination of AI-based modeling (e.g., the kernel method, ensemble learning, or deep learning); global optimization algorithm or the direct-search method[2]; and sampling techniques[6].

This article reviews the direct-search method as a core BBO technique, especially metaheuristics. Metaheuristics (or evolutionary algorithm) is a heuristic algorithm framework inspired by nature or physical phenomena; e.g., genetic algorithm and particle swarm optimization[3]. In metaheuristic algorithms, multiple search points (or population) search the global optima by updating or sampling iteratively in the solution space. $i = 1,2, \dots, m$ denotes the index of search points, $k \in \mathbb{N}$ denotes the iteration counter, $m \in \mathbb{N}$ denotes the number of search points, $\boldsymbol{x}^i(k) \in \mathbb{R}^N$ denotes the position of the i-th search point at k, respectively. They have the following features:

- **Approximation method**: which finds approximate solutions with highly optimality in a practical amount of time;
- **Stochastic method**: which uses pseudo-random numbers and can deal with uncertain values;
- **Multi-point type search method**: which has multiple search points interacting with each other.

Especially, they are expected as an effective approach to BBO from the viewpoint of the stochastic and multi-point type search method having advantages of reducing a risk of getting trapped at low-grade local optima in multimodal objective function, and benefiting from the evolution of parallel computing using GPGPU.

3 Robustness and adaptability for black-box optimization

This section provides (1) robustness and adaptability of metaheuristics as effective properties in BBO and (2) example algorithms with their properties.

3.1 Robustness and adaptability of metaheuristics

It is important or required for users to select an appliable and appropriate algorithm and set its tunable parameters so that they are adapted to the class and analytical properties of the optimization problem. These properties make some difficulties as the following:

[2] It needs global optimization for the surrogate function in SMBO to be complex or non-linear assuming that the objective function is non-convex or multimodal.

- **Separability (Decomposability) / Non-separability (Dependency)**: whether there is an existence of cross term between each coordinate of the search space in the objective function.
- **Well-Scaling (Well-Conditioning) / Ill-Scaling (Ill-Conditioning)**: whether different directions in the search space show a largely different sensitivity in their contribution to the objective function.
- **Uni-Modal (Convex) / Multi-Modal (likely Non-Convex)**: whether there are multiple local optimal solutions in the landscape of the objective function.
- **Origin-Independency / Origin-Dependency**: whether the search depends on translation of the origin in the solution space.

Conversely, BBO makes it impossible for users to know and use this information in advance. When applying metaheuristics to BBO, the search performance and parameter setting are affected by these difficulties. From this reason, users need to perform inefficient trial-and-error engineering. Therefore, it is desirable that metaheuristics for solving BBO have robustness and adaptability. In this context, the meaning of robustness and adaptability are the following inspired by modern control theory:

- **Robustness**: is the algorithm's properties which the search performance is maintained as the problem's properties changing even with its parameter being fixed during the search process.
- **Adaptability**: is the algorithm's ability of dynamically tuning its parameter so that it adapts to the problem's properties gradually during the search process.

Transformation invariance has been proposed as a property producing robustness. If we assume that a pseudo-random number sequence during the search process holds the same, transformation invariance is the quality of the search performance being reproducible with respect to transformation of the solution space or objective function (i.e., producing the same or similar results)[7]. Whether metaheuristic algorithm has transformation invariance depends on the case. When an algorithm lacks invariance to a certain transformation, it is shown that significant performance changes depending on this transformation. Transformation invariance is distinguished between two-types: invariance to transformations of the solution space and the objective function value f. Invariances giving an effectiveness of BBO are the following:

- **Similarity invariance:** which is invariant property to similarity transformation $T: \mathbb{R}^N \to \mathbb{R}^N$ of the search space; i.e., $T: \boldsymbol{x} \mapsto \alpha \boldsymbol{x}$, where $\alpha \in \mathbb{R}$.
- **Scale invariance:** which is invariant property to scaling $T: \mathbb{R}^N \to \mathbb{R}^N$ of the search space, including similarity transformation; i.e., $T: \boldsymbol{x} \mapsto \boldsymbol{D}\boldsymbol{x}$, where $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ denotes a diagonal matrix.
- **Rotational invariance:** which is invariant property to rotation $T: \mathbb{R}^N \to \mathbb{R}^N$ of the search space; i.e., $T: \boldsymbol{x} \mapsto \boldsymbol{D}\boldsymbol{x}$, where $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ denotes a rotation or orthonormal matrix.
- **Linear invariance:** which is invariant property to linear transformation $T: \mathbb{R}^N \to \mathbb{R}^N$ of the search space, including scaling and rotation; i.e., $T: \boldsymbol{x} \mapsto \boldsymbol{D}\boldsymbol{x}$, where $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ denotes a representation matrix.
- **Translation invariance:** which is invariant property to translation $T: \mathbb{R}^N \to \mathbb{R}^N$ of the search space; i.e., $T: \boldsymbol{x} \mapsto \boldsymbol{x} - \boldsymbol{t}$, where $\boldsymbol{t} \in \mathbb{R}^N$ denotes a translation vector.

- **Affine invariance:** which is invariant property to affine transformation $T: \mathbb{R}^N \to \mathbb{R}^N$ of the search space, including linear transformation and translation; i.e., $T: \boldsymbol{x} \mapsto \boldsymbol{D}\boldsymbol{x} - \boldsymbol{t}$, where $\boldsymbol{D} \in \mathbb{R}^{N \times N}, \boldsymbol{t} \in \mathbb{R}^N$ denote a linear-transformation matrix and a translation vector, respectively.
- **Monotonic invariance of f:** which is invariant property to monotonic or order preserving transformation $T: \mathbb{R} \to \mathbb{R}$ of the objective function value f; e.g., T satisfies $T(f_1) \leq T(f_2)$ whenever $f_1 \leq f_2$, where $f_1, f_2 \in \mathbb{R}$.

To deal with difficulties or necessary properties (e.g., ill-scalability, non-separability, and origin-dependency in the solution space), affine invariance (including scale invariance, rotational invariance, and translation invariance) is important because the search efficiency preserves regardless of them. Monotonic invariance of the objective function also is important because the search efficiency may preserve between non-convex (uni-modal) and non-continuous objective function, and convex continuous objective function. Invariance is extremely useful in evaluating the performance of heuristic methods because it can generalize to the complete class of problems induced by the invariance.

Adaptive parameter tuning has been proposed as a function producing adaptability. *Intensification* and *diversification* are well-known search strategies in metaheuristics and lead to outstanding performance[3]. It is important to set and tune parameters appropriately because search dynamics can change depending on parameter. Parameter tuning is distinguished between the following two-types inspired by system control theory[8]:

- **Feedforward tuning:** which gives a parameter-tuning schedule before searching and tunes parameter following this.
- **Feedback tuning:** which gives a parameter-tuning rule in response to the search state or dynamics during the search process.

The search dynamics can be evaluated based on the search history of the set of search points. Therefore, adaptive parameter-tuning rule based on the search history is expected to reduce the effort in parameter-setting and tuning and lead to excellent performance. Kanemasa and Aiyoshi (2012)[8] shows the above concept of parameter tuning and designs a new adaptive parameter rule providing excellent performance by genetic programming.

Additionally, parameter-tuning rules may have an influent on an algorithm's invariance. For this reason, invariance-based design guidelines for parameter-tuning rules are the following:

(i) Guideline A: when an algorithm has a certain invariance, we add tuning rules so as to preserve the invariance;
(ii) Guideline B: when an algorithm lacks a certain invariance, we add tuning rules so as to compensate for the invariance.

Accordingly, metaheuristics with transformation invariance and adaptive parameter-tuning rules are thought to result in good robustness and adaptability for BBO.

3.2 Example metaheuristics with robustness and adaptability

This subsection describes example metaheuristics following the above guidelines; the covariance matrix adaptation evolution strategy and the adaptive particle swarm optimization with rotational invariance.

3.2.1 Covariance matrix adaptation evolution strategy (CMA-ES)

Evolution strategy (ES) is a metaheuristic algorithm inspired by the evolution process in biology or nature. In ES, a single search point moves through random walk distributed according to the normal distribution. The update rules of ES are expressed by Equations 1 and 2.

$$\widehat{\boldsymbol{x}} = \boldsymbol{x}(k) + \sigma \boldsymbol{s} \tag{1}$$

$$\boldsymbol{x}(k+1) = \begin{cases} \widehat{\boldsymbol{x}};\ f(\widehat{\boldsymbol{x}}) \leq f\big(\boldsymbol{x}(k)\big) \\ \boldsymbol{x}(k); \text{otherwise} \end{cases} \tag{2}$$

Here, $\boldsymbol{s} \in \mathbb{R}^N$ denotes a random vector distributed according to the multi-variable standard normal distribution $\boldsymbol{\mathcal{N}}$, $\sigma > 0$ denotes the scale parameter. ES lacks similarity and scale invariance, so it is typically used with an adaptive parameter-tuning rule called *one-fifth success rule*, tuning the scale parameter σ according to the improvement frequency of the search point.

Conversely, Covariance Matrix Adaptation Evolution Strategy (CMA-ES) has been developed as a multi-point ES and various versions of CMA-ES have been proposed; e.g., $(\mu/\mu_w, \lambda)$-CMA-ES[9]. A search steps of a typical CMA-ES using weighted recombination, covariance matrix adaptation (CMA), and step-size adaptation (SSA) is the following:

(i) Step 1: sampling the next position of search points from the normal distribution according to Equation 3, which has a covariance matrix $\boldsymbol{C}(k) \in \mathbb{R}^{N \times N}$ and a scale parameter $\sigma(k) > 0$.

$$\boldsymbol{x}^i(k+1) = \boldsymbol{m} + \sigma(k)\boldsymbol{A}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{s}^i \tag{3}$$

(ii) Step 2: deriving a mean vector $\boldsymbol{m} \in \mathbb{R}^N$ using weighted recombination of the search points with superior objective-function value $f(\boldsymbol{x}^i(k+1))$.

(iii) Step 3: updating the parameters $\boldsymbol{C}(k)$ and $\sigma(k)$ according to the search state using CMA and SSA.

Here, $\boldsymbol{s}^i \in \mathbb{R}^N$ denotes a random vector distributed according to the multi-variable standard normal distribution $\boldsymbol{\mathcal{N}}$, $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_N) \in \mathbb{R}^{N \times N}$ denotes the basis transformation matrix, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) \in \mathbb{R}^{N \times N}$ denotes the eigen matrix, respectively. $\{(\boldsymbol{a}_1, \lambda_1), (\boldsymbol{a}_2, \lambda_2), \dots, (\boldsymbol{a}_N, \lambda_N)\}(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N)$ are pairs of the eigenvector and eigenvalue of the covariance matrix $\boldsymbol{C}(k)$.

CMA-ES has the following meaningful features for BBO:

- To have affine transformation invariance and monotonic invariance of the objective function f;
- To realize intensification and diversification-based search strategy by adaptive parameter-tuning rules so as to the search points move to a superior region obtained through the search process;
- To be provided recommended values for all parameters.

Parameter-tuning rules of ES following the design guideline B in Subsection 3.1 give an equivalent effect of adding affine invariance to ES; e.g., the one-fifth success rule, CMA, and SSA. It is expected that CMA-ES has high search performance according to the fact that an advanced version of it is highly ranked in BBO competitions called *Black-Box Optimization Benchmarking workshop* (BBOB)[10].

3.2.2 Adaptive particle swarm optimization with rotational invariance

Particle swarm optimization (PSO)[9] is a metaheuristic algorithm inspired by swarm intelligence in biology. The update rules of PSO are expressed by the following equations.

$$\boldsymbol{v}^i(k+1) = w\boldsymbol{v}^i(k) + c_1\boldsymbol{R}_1\left(\boldsymbol{p}^i(k) - \boldsymbol{x}^i(k)\right) + c_2\boldsymbol{R}_2\left(\boldsymbol{p}^g(k) - \boldsymbol{x}^i(k)\right) \tag{4}$$

$$\boldsymbol{x}^i(k+1) = \boldsymbol{x}^i(k) + \boldsymbol{v}^i(k+1) \tag{5}$$

$$\boldsymbol{p}^i(k) = \underset{\boldsymbol{x}^i(\kappa)}{\operatorname{argmin}}\{f(\boldsymbol{x}^i(\kappa))|\kappa = 1,2,\cdots,k\} \tag{6}$$

$$\boldsymbol{p}^g(k) = \underset{\boldsymbol{p}^i(k)\in\mathcal{P}(k)}{\operatorname{argmin}} f\left(\boldsymbol{p}^i(k)\right) \tag{7}$$

$$\mathcal{P}(k) = \{\boldsymbol{p}^i(k)|i = 1,2,\cdots,m\} \tag{8}$$

Here, $w, c_1, c_2 > 0$ denote PSO's parameters, $\boldsymbol{R}_\ell = \mathrm{diag}(R_{\ell 1}, R_{\ell 2}, \ldots, R_{\ell N}) \in \mathbb{R}^{N\times N}(\ell = 1,2)$ denotes a random number matrix, and $R_{\ell n}(n = 1,2,\ldots,N)$ denotes random number distributed according to the uniform distribution $\mathcal{U}(0,1)$, respectively. Especially, $\boldsymbol{p}^i(k) \in \mathbb{R}^N$ denotes a personal-best solution called *p-best*, $\boldsymbol{p}^g(k) \in \mathbb{R}^N$ denotes the group-best solution called *g-best*.

PSO has room for improvement in both of adaptability and robustness to various conditions. Yasuda et al. (2008)[11] have proposed the swarm activity P as a metrics of the search state in the terms of intensification and diversification, and derived accurately the stable and unstable regions in the PSO's parameter space using it. The swarm activity P is expressed by Equation 9.

$$P(k) = \frac{1}{m\sqrt{N}}\sum_{i=1}^{m}\left\|\boldsymbol{v}^i(k)\right\| \tag{9}$$

Moreover, the activity feedback PSO (AFPSO) is proposed as a new-type PSO with an adaptive parameter-tuning rule to control the swarm activity P during the search process[11]. The update rule of AFPSO consists of PSO and a tuning rule for the inertia parameter w. The tuning rules for $w(k)$ are expressed by Equations 10 and 11.

$$w(k+1) = w(k) + \mathrm{sgn}\big(P(k) - P_t(k)\big)\Delta w \tag{10}$$

$$P_t(k) = \frac{\varepsilon_{\mathrm{start}}}{\sqrt{N}}\left(\frac{\varepsilon_{\mathrm{end}}}{\varepsilon_{\mathrm{start}}}\right)^{\frac{k}{k_{\max}}}\|\boldsymbol{\gamma}_{\max} - \boldsymbol{\gamma}_{\min}\| \tag{11}$$

Here, $\boldsymbol{\gamma}_{\max}, \boldsymbol{\gamma}_{\min} \in \mathbb{R}^N$ denote the vectors determined from the initial search points, $k_{\max} \in \mathbb{N}$ denotes the iteration counter max, and $\mathrm{sgn}\colon \mathbb{R} \to \{-1,1\}$ denotes the sign function, respectively. After updating $w(k)$ with Equation 10, we revise it so that $w(k) \in [w_{\min}, w_{\max}]$. This tuning rule adjusts $w(k)$ so that the swarm activity $P(k)$ follows a target value $P_t(k)$. It is shown that AFPSO has better search performance than PSO with various parameter-tuning rules[11].

However, Kumagai and Yasuda (2017 and 2019)[12), 13)] have pointed out that AFPSO has scale invariance, translation invariance, and monotonic invariance of f, but lacks rotational invariance such as PSO with various parameter-turning rules through mathematical proofs and numerical simulations. On that point, the adaptive PSO with rotational invariance using correlativity (adaptive CRI-PSO) have proposed as a new-type PSO with both of rotational invariance and the adaptive parameter-tuning rule[13]. The update rules of the adaptive CRI-PSO are basically the same as that of AFPSO, except that Equation 4 is revised to Equation 12.

$$\boldsymbol{v}^i(k+1) = w(k)\boldsymbol{v}^i(k) + c_1\boldsymbol{A}\boldsymbol{R}_1\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{p}^i(k) - \boldsymbol{x}^i(k)\right) + c_2\boldsymbol{A}\boldsymbol{R}_2\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{p}^g(k) - \boldsymbol{x}^i(k)\right) \tag{12}$$

Here, $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_N) \in \mathbb{R}^{N\times N}$ denotes the basis transformation matrix, and $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_N\}$ are the eigenvectors of the covariance matrix based on the p-best distribution $\mathcal{P}(k)$. In the tuning rules for $w(k)$ of the adaptive CRI-PSO, parameters $\Delta w, \varepsilon_{\text{start}}, \varepsilon_{\text{end}}, w_{\min}, w_{\max} > 0$ are added[3].

The adaptive CRIPSO controls the swarm activity $P(k)$ while determining the perturbation direction from the p-best distribution $\mathcal{P}(k)$. In the above the parameter-tuning rule, controlling the search state according to the ideal ensures that the intensification and diversification-based search strategy is realized. This strategy is expected to give the adaptive CRI-PSO highly adaptation. Additionally, the adaptive CRIPSO consists of Equation 12 and the above parameter-tuning rule; and has similarity, rotational, and translation invariance in the solution space and monotonic invariance of f. But, because adding parameter-tuning rule preserves these invariances, this satisfies the design guideline A in Subsection 3.1. Thus, it is expected to have high robustness and adaptability. In fact, the search performance of the adaptive CRI-PSO is shown to be superior to that of PSO and AFPSO through numerical experiments[11].

CMA-ES and the adaptive CRI-PSO are common in the following points:

(i) They get transformation-invariance by using the covariance matrix based on the solution set obtained in the search process.
(ii) They get adaptability by parameter-tuning according to the intensification and diversification-based search strategy.

4 Constrained black-box optimization

The direct-search method including metaheuristics is considered to be unconstrained optimization problems, but we need to consider constraints in real-world use. This section describes classes of constraints in BBO and methods for handling them.

4.1 Classes of constraints in BBO

This section focuses on constrained optimization problem in BBO. A solution satisfying all constraints is called *feasible solution*, and a solution not satisfying a constraint is called *infeasible solution* or *constraint-violating solution*. The notation $\mathcal{F}$ denotes the set of feasible solutions called *feasible region*.

Constraints can be classified according to their analytic properties and formal properties. In terms of analytic properties, we can classify constraints as either linear constraints including non-negative constraints and upper/lower bound constraints, or non-linear constraints[4]. In terms of formal properties, we can classify constraints including simulation-based optimization or BBO based on the form. QRAK[14] is a taxonomy rule with four following perspectives:

---

[3] Kumagai and Yasuda (2019)[13] gives the following recommended values to for these parameters. $\boldsymbol{\gamma}_{\max}, \boldsymbol{\gamma}_{\min}$ are determined by the region for initial search points. $w_{\max}, w_{\min}$ should be to across the boundary line between stable and unstable regions in PSO's parameter space. It is pointed out that the boundary line is located inside the region $w \in [0.5, 1.0]$ while $c_1 = c_2 = 1.4955$ through parameter analysis.

[4] Besides linear constraints and non-linear constraints, there are discrete constraints and level-sets constraints determined by objective-function values.

- **Quantifiable (Q) / Nonquantifiable (N):** whether the degree of constraint violation can be quantified;
- **Relaxable (R) / Unrelaxable (U):** whether a constraint-violating solution can be evaluated;
- **A priori (A) / Simulation-based (S):** whether a simulation is needed to evaluate a solution;
- **Known (K) / Hidden (H):** whether the existence of constraints is known.

Constraints can be classified to 16 types according to the QRAK rule[5].

This includes situations such as when constraint-violating solutions cannot be evaluated due to reasons on simulator's side, or when the simulation does not have completed successfully. In practical use, many cases are QRAK or QRSK constraints. The reason for this is that a binary constraint meaning whether a solution is feasible or infeasible can be quantified as numerical constraint violation, or constraint-violating solutions can be evaluated in some reasonable way. It would be meaningful to know classes of constraints in BBO more broadly, and to consider handling methods matching them. Subsections 4.2 and 4.3 will explain methods for handling QRAK or QRSK constraints.

4.2 Classical method for constraint handling

Classical method for constraint handling is a type of problem-transformation method called the penalty-function method. If the class and properties of objective or constraint functions are clear, you can apply a method for constraint handling that matches the class of problem. For example, if the objective function and constraint functions are convex, a convex optimization method that uses analytic information about their functions (gradient or Hessian matrix) can solve this problem by constructing a new convex objective function introduced penalty function; and converting the problem to an unconstrained optimization, which called the augmented Lagrangian function. Equation 13 shows the augmented Lagrangian function $L\colon \mathbb{R}^N \to \mathbb{R}$ with a penalty function.

$$L(\boldsymbol{x}) = f(\boldsymbol{x}) + \lambda\phi(\boldsymbol{x}) \tag{13}$$

Here, $f\colon \mathbb{R}^N \to \mathbb{R}$ denotes the objective function, $\phi\colon \mathbb{R}^N \to \mathbb{R}$ denotes the penalty function, and $\lambda > 0$ denotes the penalty coefficient, respectively. Many definitions for penalty function ϕ are possible, but a typical definition is formulated as Equations 14 and 15.

$$\phi(\boldsymbol{x}) = \sum_{\kappa=1}^{K} \Omega_\kappa(\boldsymbol{x}) \tag{14}$$

$$\Omega_\kappa(\boldsymbol{x}) = \max\{g_\kappa(\boldsymbol{x}), 0\} \tag{15}$$

An amount of constraint violation $v(\boldsymbol{x})$, which amount of violating a constraint or all constraints, is basically defined by $v(\boldsymbol{x}) = \phi(\boldsymbol{x})$ and $v_\kappa(\boldsymbol{x}) = \Omega_\kappa(\boldsymbol{x})$.

4.3 Methods for constraint handling in BBO

4.3.1 Constraints handling techniques

[5] Actually, the QRAK rule classifies constraints into nine types because it is not possible to calculate or quantify numerical constraint-violation value if the existence of them is unknown.

Table 1 Summary of studies on constraint handling techniques

–	Penalty-based CHT	Separatist-based CHT	Multi-objective-based CHT
None	Death penalty[16]	–	–
Inexplicit	Static penalty[18], Dynamic penalty[19], Adaptive penalty[20]	Feasibility rule[16], Stochastic ranking[22], ε constraint method[23], GCR[25], HSR[26], MCR[27]	–
Explicit	ASCHEA[21]	MCODE[24], TNSDM[28]	Two-phase framework[30], IDEA[17], DeCODE[32], Adaptive Weighted MOEA/D[33]

The classical penalty-function method is not effective for constrained optimization problem including the black-box or non-convex objective function and constraint functions. Difficulties in this class of problem are the following:

(i) Because there are multiple local optimal solutions in the feasible region $\mathcal{F}$ being non-convex or disconnected, it needs a global search capability.
(ii) Typically, because the global optimal solution is located on the boundary of the feasible region $\mathcal{F}$, it needs an efficient search on this boundary.

The constraint handling techniques (CHTs) have been developed as methods expanding the applicable scope of metaheuristics to this class of problems and cope with the above issues[15, 16]. Metaheuristic algorithm searches the global optimal solution by selection superior solutions. The selection operation evaluates and compares solutions obtained through multi-point search according to a metrics called fitness function $S(\boldsymbol{x})$. While the selection for unconstrained optimization uses the fitness S defined by the objective-function value f, CHTs use the fitness S defined by the objective-function value f and amount of constraint violation v. Additionally, it has been pointed out that using constraint-violating solutions is important for improving search efficiency in constrained BBO[17]. According to the above discussion, CHTs can be largely classified depending on the following viewpoint:

(i) CHTs can be categorized as penalty-based, separatist-based, or multi-objective-based method depending on the definitions of fitness S.
(ii) CHTs can be categorized as "none", "inexplicit", or "explicit" class depending on degree of utilization of constraint-violating solutions.

Table 1 shows summary of studies on CHTs categorized based on these criteria. We will provide a description of each approach in next subsection and beyond.

4.3.2 Penalty-based CHT

The penalty-based CHT uses the augmented Lagrangian function L defined in Equation 13 as the fitness, but it differs from classical penalty functions in terms of the way it treats the penalty coefficient λ. The death penalty[16], static penalty[18], and dynamic penalty[19] have been proposed as the penalty-based CHT. The death penalty assigns an extremely bad value (or an infinitely large value) to the penalty coefficient λ for constraint-violating solutions, and searches only using feasible solutions. The static penalty assigns a constant to λ and the dynamic penalty increases λ according to a given schedule during the search process. However, the search

performance is deteriorated, and a feasible solution cannot be obtained without a properly set or tuning for the penalty coefficient λ. For example, the fitness function L highly depends on scale differences between constraint functions because of lacking monotonic invariance of the objective function and constraint functions if the amount of constraint violation v is defined in Equation 14.

On that point, the adaptive penalty[20] and ASCHEA[21] have been proposed as the penalty-based CHT for adaptively tuning the penalty coefficient during the search process. In the adaptive penalties, a fitness function S of a constraint-violating solution $\boldsymbol{x}$ in a solution set $\mathcal{X}$ is formulated as Equations 16, 17, and 18.

$$S(\boldsymbol{x}) = \hat{f}(\boldsymbol{x}) + \sum_{k=1}^{K} \lambda_k v_k(\boldsymbol{x}) \tag{16}$$

$$\hat{f}(\boldsymbol{x}) = \max\{f(\boldsymbol{x}), \bar{f}\} \tag{17}$$

$$\lambda_k = |\bar{f}| \frac{\overline{v_k}}{\sum_{k=1}^{K} \overline{v_k}^2} \tag{18}$$

Here, the notations $\bar{f}$, $\overline{v_k}$ denote the mean objective function value f and the mean amount of constraint violation v_k, respectively; and their mean values are averaged of f and v_k of solutions in the solution set $\mathcal{X}$. Thus, it is possible to mitigate the above deficiency by assigning different penalty coefficients for each constraint function and adaptively tuning in response to the objective-function value and each amount of constraint violation obtained in the search process. However, because many the penalty-based CHTs do not make use of the order relationship between the amount of constraint violation of solutions, it is thought that the usefulness of constraint-violating solutions be low.

4.3.3 Separatist-based CHT

The separatist-based CHT separates objective-function values f and amount of constraint violation v, and evaluates and compares solutions; they can be classified as either switching-based or ranking-based CHT. The switching-based CHT selects and uses either of the objective-function value or the amount of constraint violation as a fitness S. Feasibility rule[16], stochastic ranking[22], ε constraint method (εCM)[23], and MCODE[24] have proposed in this category. In εCM, when comparing two solutions $\boldsymbol{x}$ and $\boldsymbol{y}$, the order relation of fitness S is defined in Equation 19.

$$S(\boldsymbol{x}) \leq S(\boldsymbol{y}) \Leftrightarrow \begin{cases} f(\boldsymbol{x}) \leq f(\boldsymbol{y});\ v(\boldsymbol{x}), v(\boldsymbol{y}) < \varepsilon \\ v(\boldsymbol{x}) \leq v(\boldsymbol{y}); \text{otherwise} \end{cases} \tag{19}$$

Here, $\varepsilon > 0$ denotes a threshold parameter for the amount of constraint violation and decreases to almost 0 gradually during the search process, since the search points improve the objective-function value after they move to near the feasible region early on.

The ranking-based CHT uses a fitness S based on a total rank in a solution set. The total rank consists of rank based on objective-function values and rank based on amount of constraint violation. Global competitive ranking (GCR)[25], Ho-Shimizu ranking (HSR)[26], multiple constraint ranking (MCR)[27], and TNSDM[28] have proposed in this category. In MCR, the fitness function $S(\boldsymbol{x})$ of a solution $\boldsymbol{x}$ in a solution set $\mathcal{X}$ is formulated as Equation 20.

$$S(\boldsymbol{x}) = \begin{cases} R_n + \sum_{\kappa=1}^{K} R_{v_\kappa}\,;\ \mathcal{F} \cap \mathcal{X} = \emptyset \\ R_f + R_n + \sum_{\kappa=1}^{K} R_{v_\kappa}\,;\text{otherwise} \end{cases} \tag{20}$$

Here, R_f, R_{v_κ}, R_n denote a rank of objective-function value f, ranks for each amount of constraint violation v_κ, and a rank of constraint violation count, respectively; and their ranks are determined from an order relation in the solution set $\mathcal{X}$. Because the separatist-based CHTs make use of the order relation between the amount of constraint violation, it is thought that they make better use of constraint-violating solutions than penalty-based CHT.

4.3.4 Multi-objective-based CHT

The multi-objective-based CHT converts a single-objective constrained optimization problem into a two-objective (objective-function value f and amount of constraint violation v) unconstrained optimization problem; and uses evaluations and comparisons of the solution in multi-objective optimization. The terms of multi-objective optimization in the multi-objective-based CHT are explained below. Multi-objective optimization finds a Pareto optimum set in the objective function space considering trade-off relationship between each objective function while the multi-objective-based CHT finds a feasible global optima in the space (f, v). *Superiority relation* (*Pareto dominance*) gives superiority or inferiority relation between two solutions in the space (f, v). When two solutions $\boldsymbol{x}, \boldsymbol{y}$ satisfy Equation 21, an order relation between $\boldsymbol{x}$ and $\boldsymbol{y}$ is called "$\boldsymbol{x}$ dominates $\boldsymbol{y}$" or "$\boldsymbol{x}$ is superior to $\boldsymbol{y}$".

$$\big(f(\boldsymbol{x}) \le f(\boldsymbol{y}) \wedge v(\boldsymbol{x}) \le v(\boldsymbol{y})\big) \wedge \big(f(\boldsymbol{x}) < f(\boldsymbol{y}) \vee v(\boldsymbol{x}) < v(\boldsymbol{y})\big) \tag{21}$$

In a solution set $\mathcal{X}$, a solution $\boldsymbol{x} \in \mathcal{X}$ to be not dominated by all other solutions $\boldsymbol{y} \in \mathcal{X}$ is called *non-dominated solution* or *Pareto solution*. A set of non-dominated solutions in the solution space is called *Pareto optimum set* or *Pareto frontier*. A feasible solution is located on the f-axis in the space (f, v) and means *weakly Pareto solution*. Note that the multi-objective-based CHT aims to find the only feasible global optima located on the intersection of the Pareto frontier and the f-axis in the space (f, v), not the entire of this region.

The multi-objective-based CHT can be classified as either the Pareto ranking-based or decomposition-based CHT in the viewpoint of multi-objective optimization approaches. *Pareto ranking* gives candidate solutions a rank-based fitness constructed by superiority relation and crowded distance in the objective-function space, and is used in NSGA-II[29] for multi-objective optimization. The Pareto ranking-based CHT uses the fitness based on Pareto ranking in the space (f, v). However, even if feasible solutions are obtained in the search process, they are easily eliminated because they are weakly Pareto solutions. On that point, as the Pareto ranking-based CHT preserving some feasible solutions during the search process, the two-phase framework[30] and the infeasibility driven evolutionary algorithm (IDEA)[17] have been proposed. IDEA preserves both of feasible solutions and constraint-violating solutions in the search points, and assigns two-types fitness for each solution set. One fitness of feasible solutions is the objective-function value f, and the other fitness of constraint-violating solutions is a rank R_p based on Pareto ranking. In IDEA, when comparing

two feasible solutions or two constraint-violating solutions, the order relation of fitness S is defined in Equation 22.

$$S(\boldsymbol{x}) \leq S(\boldsymbol{y}) \Leftrightarrow \begin{cases} f(\boldsymbol{x}) \leq f(\boldsymbol{y});\ \boldsymbol{x}, \boldsymbol{y} \in \mathcal{F} \\ R_p(\boldsymbol{x}) \leq R_p(\boldsymbol{y}); \boldsymbol{x}, \boldsymbol{y} \notin \mathcal{F} \end{cases} \tag{22}$$

Decomposition decomposes a multi-objective optimization problem into a set of single-objective optimization subproblems and solves them in parallel by a scalar-based fitness function with a different weight parameter $w^i \in [0,1]$ assigned for each search point $\boldsymbol{x}^i$; and is used in MOEA/D[31] for multi-objective optimization. The decomposition-based CHT uses the above problem decomposition in the space (f, v). The scalarizing function S of the weighted sum is formulated as Equation 23.

$$S(\boldsymbol{x}; w) = wf(\boldsymbol{x}) + (1 - w)v(\boldsymbol{x}) \tag{23}$$

Note that MOEA/D for multi-objective optimization assigns the weights to constant value during the search process and distributes them uniformly in the objective function space to find the entire of the Pareto frontier, but this is not effective of constrained optimization to find the only feasible global optima finally. On that point, as the decomposition-based CHT with weight-tuning rule dynamically during the search process, DeCODE[32] and the adaptive weighted MOEA/D[33] have been proposed. DeCODE uses a weight-tuning rule dynamically so that the search region is limited from the violation region to the feasible region gradually. The adaptive weighted MOEA/D uses a weight-tuning rule adaptively so that the search region is balanced to cover the boundary of the feasible region $\mathcal{F}$, and the tuning rules for w^i are expressed by Equations 24 and 25.

$$w^i(k) = \alpha(k)\frac{i-1}{m-1} \tag{24}$$

$$\alpha(k+1) = \begin{cases} \gamma_{\mathrm{u}}\alpha(k);\ \boldsymbol{x}^t(k) \in \mathcal{F} \\ \gamma_{\mathrm{d}}\alpha(k); \text{ohterwise} \end{cases} \tag{25}$$

Here, $\alpha \in [0,1]$ denotes a variable parameter for controlling the distribution of weights w^i in the space (f, v), $\boldsymbol{x}^t(k)$ denotes the search point having t-th weight w^t. $\gamma_{\mathrm{u}} > 1, \gamma_{\mathrm{d}} \in (0,1)$ denote an increasing coefficient and a decreasing coefficient, respectively. After updating w^i with Equation 24, we revise it so that $w^i(k) \in [\delta, 1]$, and after updating α with Equation 25, we revise it so that $\alpha(k) \in [0,1]$[6].

The multi-objective-based CHT makes the most utilization of constraint-violating solutions than the penalty-based and the separatist-based CHTs because an effective search works on trade-off regions between objective-function value f and amount of constraint violation v by superiority relation. The Pareto ranking-based and the decomposition-based CHTs have the following advantages and disadvantages:

- **Pareto ranking-based CHT:**
  - Advantage: it is possible to search the trade-off region while not being affected by a difference between f and v scales; or a shape of the Pareto frontier because the fitness S is determined only from ranks in the space (f, v).

[6] Yasuda et al. (2022)[33] gives the following recommended values to for these parameters; $\alpha(1) = 1.0, t = \lfloor 0.8m \rfloor, \gamma_{\mathrm{u}} = 1.001, \gamma_{\mathrm{d}} = 0.999, \delta = 10^{-15}$.

- Disadvantage: it is difficult to gain feasible solutions if the Pareto frontier is large in the solution space, which the feasible region is narrower than the violation region, because it searches the entire of the region uniformly.

• **Decomposition-based CHT:**

- Advantage: it is possible to improve an ability to gain feasible solutions by weight tuning so that the search region is limited to a part of the Pareto frontier because solutions converges almost surely and the weights corresponds to the Pareto solutions basically.
- Disadvantage: it is difficult to maintain the limitation in the space (f, v) as ideal and the corresponding relation between the weights and the Pareto solutions if the difference between f and v scales is large[34]; or a scalarization function S does not match the shape of the Pareto frontier near the optimal solution[7].

Both approaches are complementary, but there is a difference in effectiveness on constrained optimization. The Pareto ranking-based CHT preserves some feasible solutions in the search process, but it is relatively disadvantaged in finding the only feasible global optima because the search region cannot be limited to the part of the Pareto frontier. In contrast, the decomposition-based CHT can search the limited region, and its disadvantage is considered less serious than the Pareto ranking-based CHT according to the following reasons:

(i) The difference between f and v scales can be suppressed by normalization or standardization[34].
(ii) Even if the uniformity of approximation to the Pareto frontier is reduced a little, there is some effectiveness.

In fact, Yasuda et al. (2022)[33] confirmed that the adaptive weighted MOEA/D is superior to the separatist-based CHT and the Pareto ranking-based CHT in terms of both of convergence on a feasible solution and global optimization performance thorough numerical simulation. From the above, the decomposition-based CHT is considered advantageous in constrained optimization in terms of the ability to gain feasible solutions.

4.3.5 General comment

According to the above review, because CHTs use the order relation in the solution set obtained in the search process, they have a high affinity for metaheuristics with multi-point and direct-search method. As showing Table1, CHTs can be largely classified based on the viewpoint of definition of fitness function and degree of utilization of constraint-violating solutions; and especially the multi-objective based CHT makes the most utilization of them. But the fitness function S in CHTs is deeply related to intensification and diversification-based search strategy, and transformation invariance; e.g., the monotonic invariance of objective function f and each constraint function g_κ. In the future research, it is desirable to develop and combine metaheuristic algorithms and CHTs for constrained BBO considering the above strategy or property totally.

[7] The scalarizing function of the weighted sum cannot approximate the Pareto frontier uniformly, whose shape is not convex (non-convex or disconnected)[31].

5 Application of BBO

This section reviews three important examples of BBO applications. First example is on-line operation optimization or configuration design for SoS. SoS consists of connected various systems with operational autonomy and control autonomy (e.g., distribution system, medical system, power and energy system, or railway system). This needs total and data-driven optimization using obtained on-line and uncertain data from sensor and state of each system in real time. Demand response maintains the supply-demand balance efficiently by providing people with data and financial incentives to indirectly change their behavior and demand, such as dynamic pricing in energy management systems[35]. Even if a state of SoS involves uncertainly such as potential to change people's behavior or demand, BBO can be applied to it. And there is an approach to find and design configuration system of SoS[36]. These problems are formulated as constrained BBO.

Second example is hyperparameter tuning or optimization for expensive machine learning models. Automated machine learning (AutoML) frameworks and deep learning (deep neural networks) have many tunable hyperparameters such as very complex parameter architecture[37]. On that point, parameter-tuning techniques has proposed for automating their engineering process. It is shown that these techniques are superior to manually tuning or engineering by human in the viewpoint of performance and time comsuming[5, 38]; especially neural architecture search is the most successful example[39]. Hyperparameter tuning problem is formulated as constrained BBO.

Third example is material discovery, design, and optimization for multi-range application. Design of new and highly-functional material including raw material or chemical compounds can change all manufacturing areas, such as healthcare, medical care, and energy. Industrial biotechnology can realize the production of renewable biological resources and the conversion of these resources and waste streams into value added products, which alternative energy sources, biopharmaceutical, and high-quality food, by smart genetics and cells. The bio-industry subcommittee of Ministry of Economy, Trade, and Industry (METI) in Japan shows the following examples as their use cases[40]:

(i) Regenerative medicine or gene therapy may cause radical treatments for diseases.
(ii) In manufacturing process, redesign of cell's function may improve efficiency of industrial production; e.g., transformation from glucose to raw material for high-quality plastics.

Material discovery problem for these processes is formulated as constrained BBO[41].

6 Conclusion

This article reviewed the main approaches to BBO, especially metaheuristics and constraint handling techniques, and provided some applications of BBO. As we shift toward Super Smart Society (Society 5.0) or Bioeconomy, real-world systems will grow in scale and complexity such as SOS. Conversely, practical system optimization will realize to combine surrounding technologies advancing rapidly. It is desirable to develop BBO with an ability to adapt to all of these changes. Given these trends, we believe that SMBO for constrained BBO will be increasingly important, and these technologies should be able to adapt to various applications such as SOS in the future.

References

1) Cabinet Office, Government of Japan (2021) Society 5.0. https://www8.cao.go.jp/cstp/english/society5_0/index.html. Accessed 15 May 2022
2) European Commission (2021) What is the Bioeconomy. https://ec.europa.eu/research/bioeconomy/policy/bioeconomy_en.htm. Accessed 15 May 2022
3) Yang XS (2010) Nature-Inspired Metaheuristic Algorithms. Second Edition. Luniver Press
4) Jones DR, Schonlau M, Welch WJ (1998) Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization 13:455-492
5) Snoek J, Larochelle H, Adams RP (2012) Practical Bayesian Optimization of Machine Learning Algorithms. Advances in Neural Information Processing Systems 25 2:2951-2959
6) Kawarabayashi M, Yasuda K (2008) Integrated Optimization Method Using Particle Swarm Optimization and Modeling (written in Japanese). Transactions of the Society of Instrument and Control Engineers 44(11):855-862
7) Hansen N, Ros R, Mauny N, Schoenauer M, Auger A (2011) Impacts of Invariance in Search: When CMA-ES and PSO Face Ill-Conditioned and Non-Separable Problems. Applied Soft Computing. 11(8):5755-5769
8) Kanemasa M, Aiyoshi E (2014) Algorithm Tuners for PSO Methods and Genetic Programming Techniques for Learning Tuning Rules. IEEJ Transactions on Electrical and Electronic Engineering 9(4):407-414
9) Hansen N, Arnold DV, Auger A (2015) Evolution Strategies. In Kacprzyk J, Pedrycz W (eds): Springer Handbook of Computational Intelligence. Springer, Berlin, Heidelberg, chapter 44, pp. 871-898
10) The Black-box Optimization Benchmarking (BBOB) Workshop. http://numbbo.github.io/workshops/index.html. Accessed 29 March 2022
11) Yasuda K, Iwasaki N, Ueno G, Aiyoshi E (2008) Particle Swarm Optimization: A Numerical Stability Analysis and Parameter Adjustment Based on Swarm Activity. IEEJ Transactions on Electrical and Electronic Engineering 3(6):642-659
12) Kumagai W, Yasuda K (2017) Making Rotational Invariance of Particle Swarm Optimization Based on Correlativity. IEEJ Transactions on Electrical and Electronic Engineering 12(S2):S131-132
13) Kumagai W, Yasuda K (2019) Adaptive Particle Swarm Optimization with Rotational Invariance (written in Japanese). Transactions of the Institute of Electrical Engineers of Japan Part C 139(10):1201-1214
14) Digabel SL, Wildy SM (2015) A Taxonomy of Constraints in Simulation-Based Optimization. Les cahiers du GERAD, Technical Report G-2015-57.https://doi.org/10.48550/arXiv.1505.07881
15) Coello CAC (2002) Theoretical and Numerical Constraint-Handling Techniques Used with Evolutionary Algorithms: A Survey of the State of the Art. Computer Methods in Applied Mechanics and Engineering 191(11-12):1245-1287
16) Mezura-Montes E, Coello CAC (2011) Constraint-Handling in Nature-Inspired Numerical Optimization: Past, Present and Future. Swarm and Evolutionary Computation 1(4):173-194

17) Ray T, Singh HK, Isaacs AA, Smith W (2009) Infeasibility Driven Evolutionary Algorithm for Constrained Optimization. Constraint-Handling in Evolutionary Optimization 145-165

18) Homaifar A, Qi CX, Lai SH (1994) Constrained Optimization via Genetic Algorithms. Simulation 62(4):242-253

19) Joines J, Houck C (1994) On the Use of Non-Stationary Penalty Functions to Solve Nonlinear Constrained Optimization Problems with GA's. In: Proceedings of the First IEEE Conference on Evolutionary Computation, pp. 579-584

20) Lemonge A, Barbosa H (2004) An Adaptive Penalty Scheme for Genetic Algorithms in Structural Optimization. International Journal for Numerical Methods in Engineering 59(5):703-736

21) Hamida SB, Schoenauer M (2000) An Adaptive Algorithm for Constrained Optimization Problems. In: Proceedings of the International Conference on Parallel Problem Solving from Nature, pp. 529-538

22) Runarsson TP, Yao X (2000) Stochastic Ranking for Constrained Evolutionary Optimization, IEEE Transactions on Evolutionary Computation 4(3):284-294

23) Takahama T, Sakai S (2005) Constrained Optimization by ε Constrained Particle Swarm Optimizer with ε-Level Control. In: Proceedings of the 4th IEEE International Workshop on Soft Computing as Transdisciplinary Science and Technology, pp. 1019-1029

24) Liu B, Ma H, Zhang X, Zhou Y (2007) A Memetic Coevolutionary Differential Evolution Algorithm for Constrained Optimization. In: Proceedings of the 2007 IEEE Congress on Evolutionary Computation, pp. 2996-3002

25) Runarsson TP, Yao X (2003) Constrained Evolutionary Optimization - The Penalty Function Approach. In: Sarker R, Mohammadian M. Yao X, Evolutionary Optimization. International Series in Operations Research & Management Science, vol 48. Springer, Boston, MA, pp 87-113

26) Ho PY, Shimizu K (2007) Evolutionary Constrained Optimization Using an Addition of Ranking Method and a Percentage-Based Tolerance Value Adjustment Scheme. Information Sciences 177(14):2985-3004

27) Garcia R, Lima B, Lemonge A, Jacob B (2017) A Rank-Based Constraint Handling Technique for Engineering Design Optimization Problems Solved by Genetic Algorithms. Computers and Structures 187:77-87

28) Miyakawa M, Takadama K, Sato H (2013) Two-Stage Nondominated Sorting and Directed Mating for Solving Problems with Multi-Objectives and Constraints. In: Proceedings of the 2013 Genetic and Evolutionary Computation Conference, pp. 647-654

29) Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2):182-197

30) Venkatraman S, Yen GG (2005) A Generic Framework for Constrained Optimization Using Genetic Algorithms. IEEE Transactions on Evolutionary Computation 9(4):424-435

31) Zhang Q, Li H (2007) MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. IEEE Transactions on Evolutionary Computation 11(6):712-731

32) Wang B, Li H, Zhang Q, Wang Y (2021) Decomposition-Based Multiobjective Optimization for Constrained Evolutionary Optimization. IEEE Transactions on Systems, Man, and Cybernetics: Systems 51(1):574-587

33) Yasuda Y, Kumagai W, Tamura K, Yasuda K (2022) An Extension of MOEA/D to Constrained Optimization and Adaptive Weight Adjustment (written in Japanese). Transactions of the Institute of Electrical Engineers of Japan Part C 142(1):108-109

34) Ishibuchi H, Doi K, Nojima Y (2017) On the Effect of Normalization in MOEA/D for Multi-Objective and Many-Objective Optimization. Complex & Intelligence Systems 3(4):279-294

35) Yasuda K, Tokoro K (2020) Smartification of Social Infrastructure for Efficient Power and Energy Use. In Kaihara T, Kita H, Takahashi S (eds): Innovative Systems Approach for Designing Smarter World. Springer, Singapore, pp. 133-145

36) Agarwal S, Pape LE, Dagli CH, Ergin NK, Enke D, Gosavi A, Qin R, Konur D, Wang R, Gottapu RD. (2015) Flexible and Intelligent Learning Architectures for SoS (FILA-SoS): Architectural Evolution in Systems-of-Systems. Procedia Computer Science. 44:76-85

37) Feurer M, Hutter F (2019) Hyperparameter optimization. In Hutter F, Kotthoff L, Vanschoren J (eds): Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham, pp. 3-33

38) Bergstra J, Bardenet R, Bengio Y, Ke'gl B (2011) Algorithms for Hyper-Parameter Optimization. Advances in Neural Information Processing Systems 24, 2546-2554.

39) Elsken T, Metzen JH, Hutter F (2019) Neural Architecture Search. In Hutter F, Kottohoff L, Vanschoren J (eds): Automated Machine Learning. The Springer Series on Challenges in Machine Learning. Springer, Cham, pp. 63-77

40) Ministry of Economy, Trade, and Industry (METI), Bio-Industry Subcommittee's Report (2021) Fifth Industrial Revolution. Cultivated with Biotechnology. https://www.meti.go.jp/english/press/2021/0202_001.html. Accessed 2 May 2022

41) Terayama K, Sumita M, Tamura R, Tsuda K. (2021) Black-Box Optimization for Automated Discovery. Accounts of Chemical Research 54(6):1334-1346. https://doi.org/10.1021/acs.accounts.0c00713

Wataru Kumagai

Born in 1991, Wataru Kumagai completed his doctoral program in Electrical and Electronic Engineering at the Graduate School of Science and Engineering, Tokyo Metropolitan University, Japan, in 2020. Since 2016, he has been working at Yokogawa Electric Corporation, Ltd., Japan. He is engaged in research on systems optimization (evolutionary computation), machine learning, and energy optimization. He holds a Doctor of Engineering and is a member of The Institute of Electrical Engineers of Japan.

Keiichiro Yasuda

Born in 1960, Keiichiro Yasuda completed his doctoral program in Electrical Engineering at the Graduate School of Engineering, Hokkaido University, in 1989. The same year, he became an Assistant Professor in the Faculty of Engineering at the Tokyo Metropolitan University. In 1991, he became an Associate Professor in the Faculty of Engineering at the Tokyo Metropolitan University and, since 2006, he has been a Professor in the Graduate School of Science and Engineering. He is engaged in research on systems optimization and power systems engineering. He holds a Doctor of Engineering and is a member of societies such as The Society of Instrument and Control Engineers, The Institute of Electrical Engineers of Japan, The Japanese Society for Evolutionary Computation, and the IEEE.