Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Constrained Optimization using the Lagrangian Method and the Improved Discrete Gradient Chaos Model

Takashi Okamoto
Graduate School of Engineering
Chiba University
Chiba, Japan
takashi@faculty.chiba-u.jp

Hironori Hirata
Graduate School of Engineering
Chiba University
Chiba, Japan
hiro@faculty.chiba-u.jp

***Abstract*—In this study, we propose a new chaotic global optimization method using the Lagrangian method to solve a nonlinear constrained optimization problem. Firstly, we explain the convergence behavior of the first order method regarding convexity of the Lagrangian with respect to decision variables in terms of linear stability theory. Further, we propose a new optimization method in which the convergence behavior of the first order method is improved by two techniques. One is the introduction of a coupling structure. The second is the introduction of objective function weighting. Then, we apply a multipoint type chaotic optimization method so that global search is implemented to find feasible global minima. We then confirm the effectiveness of the proposed method through applications to the coil spring design problem and benchmark problems used in the special session on constrained real parameter optimization in CEC2006.**

***Index Terms*—Constrained Optimization, Global Optimization, Lagrangian Method, Chaos, Gradient Dynamics, Coupled Dynamics**

I. Introduction

The development of the global optimization method, which is able to obtain global minima without being trapped at local minima, has been investigated extensively. In [1], we proposed a multipoint type chaotic optimization method, which is called the "Elite Coupling type Chaotic Optimization Method (EC-COM)", for unconstrained optimization problems. EC-COM is an unconstrained global optimization method in which multiple search points which implement global search driven by a chaotic gradient dynamic model are synchronized to their elite search points by using a coupling model. EC-COM successfully achieves diversification and intensification, which are reported to be important strategies for global optimization in the Meta-heuristics research field. Its superior global capability has been confirmed through applications to several unconstrained multi-peaked optimization problems with 100 variables and 1000 variables. However, actual optimization problems often have constraint conditions.

In this study, we propose a new chaotic global optimization method using the Lagrangian method to solve a nonlinear constrained optimization problem:

$$\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) \tag{1a}$$

$$\text{s.t. } \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0} \tag{1b}$$

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \tag{1c}$$

$$\text{where } x_i^l \leq x_i \leq x_i^u, \tag{1d}$$

where the vector of decision variables $\boldsymbol{x} = [x_1, \cdots, x_N]^T \in R^N$. The objective function $f(\boldsymbol{x})$, the inequality constraint vector $\boldsymbol{g}(\boldsymbol{x}) = [g_1(\boldsymbol{x}), \cdots, g_M(\boldsymbol{x})]^T$, and the equality constraint vector $\boldsymbol{h}(\boldsymbol{x}) = [h_1(\boldsymbol{x}), \cdots, h_Q(\boldsymbol{x})]^T$ are each twice continuously differentiable. The gradient of the objective function $\nabla f(\boldsymbol{x})$ is a column vector, and $n = 1, \cdots, N$, $m = 1, \cdots, M$, and $q = 1, \cdots, Q$ unless otherwise stated. Eq. (1d) is the bounded search space. The feasible region which satisfies Eq. (1b) and Eq. (1c) is located within the bounded search space, and optimization methods perform within the space. Generally, an equality constraint cannot be satisfied exactly using computational optimization. Consequently, in this study, a solution $\boldsymbol{x}$ is regarded as feasible if $g_m(\boldsymbol{x}) \leq 0$ and $|h_q(\boldsymbol{x})| - \epsilon \leq 0$. Specifically, ϵ is set to 0.0001.

The Lagrangian method is a standard constrained optimization method [2]. In this method, a Lagrangian, which consists of the objective function $f(\boldsymbol{x})$ and constrained conditions $\boldsymbol{g}(\boldsymbol{x})$ and $\boldsymbol{h}(\boldsymbol{x})$ in the original problem Eq. (1) is used as an augmented objective function to solve the original problem Eq. (1). It is known that an optimal solution of the original problem is a saddle point of the Lagrangian. However, existence of a saddle point is not necessarily guaranteed for all problems. It is known that a local saddle point exists if the second order sufficiency condition [2] is satisfied, in other words, convexity with respect to the decision variables $\boldsymbol{x}$ is assumed adjacent to an optimal solution [3], [4]. For a number of optimization methods that utilize gradient dynamics, such as the "first order method" [2], which is a primitive gradient based method used to explore a solution satisfying the first order necessary condition, and chaotic optimization methods that utilize the Lagrangian method [5]–[7], convexity of the Lagrangian with respect to decision variables is a necessary

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

assumption for their application or to guarantee convergence of their search trajectories. In addition, these methods do not have effective global search capability.

In this study, we explain the convergence behavior of the first order method regarding convexity of the Lagrangian with respect to decision variables in terms of linear stability theory. Furthermore, we propose a new optimization method in which the convergence behavior of the first order method is improved by two techniques. One is the introduction of a coupling structure. The other is the introduction of objective function weighting proposed in [8], [9]. Then, we apply EC-COM, which has been proposed in [1] so that global search is implemented to find feasible global minima. We then confirm the effectiveness of the proposed method through applications to the coil spring design problem [10] and the benchmark problems used in the special session on constrained real parameter optimization in CEC2006 [11], [12].

II. Lagrangian Method and Gradient Dynamics

A. Lagrangian Method

The Lagrangian method is known as a classical method for constrained optimization [2]. In the Lagrangian method, a Lagrangian for the original problem Eq. (1) is considered, and the original problem is solved using the Lagrangian as an augmented objective function for Eq. (1). In this study, we use an augmented Lagrangian [2], which is given by

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\phi}) =& f(\boldsymbol{x}) \\ &+ \frac{1}{2}\sum_{m=1}^{M}\left[(\max\{0,\lambda_m+g_m(\boldsymbol{x})\})^2-\lambda_m^2\right] \\ &+ \sum_{q=1}^{Q}\phi_q h_q(\boldsymbol{x}) + \frac{1}{2}\sum_{q=1}^{Q}h_q^2(\boldsymbol{x}), \end{aligned} \tag{2}$$

where $\boldsymbol{\lambda} \in R^M$ and $\boldsymbol{\phi} \in R^Q$ are Lagrangian multipliers. Let us write $X = [\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\phi}]$ for simpler notation. We choose the augmented Lagrangian because it provides better numerical stability [8].

It is known that the following "first order necessary condition" and "saddle point theorem" hold regarding local minima of the constrained optimization problem and Lagrangians [2].

Theorem 1 (First order necessary condition). *Let $\boldsymbol{x}^o$ be a local minimum of Eq.* (1)*. Then, there exists $\boldsymbol{\lambda}^o \in R^M$ and $\boldsymbol{\phi}^o \in R^Q$, and the following equations hold:*

$$\nabla_{\boldsymbol{x}} L(X^o) = \mathbf{0},\ \nabla_{\boldsymbol{\lambda}} L(X^o) = \mathbf{0},\ \nabla_{\boldsymbol{\phi}} L(X^o) = \mathbf{0}, \tag{3}$$

where $\nabla_{\boldsymbol{x}} L(X)$, $\nabla_{\boldsymbol{\lambda}} L(X)$, and $\nabla_{\boldsymbol{\phi}} L(X)$ are gradients of $L(X)$ w.r.t. $\boldsymbol{x}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\phi}$ respectively.

Theorem 2 (Saddle point theorem). *Let us define $\boldsymbol{x}^o, \boldsymbol{\lambda}^o, \boldsymbol{\phi}^o$ to be a saddle point of the Lagrangian of Eq.* (2)*, if the following inequalities hold:*

$$L(\boldsymbol{x}^o, \boldsymbol{\lambda}, \boldsymbol{\phi}) \le L(X^o) \le L(\boldsymbol{x}, \boldsymbol{\lambda}^o, \boldsymbol{\phi}^o). \tag{4}$$

Then, if there exists $\boldsymbol{\lambda}^o, \boldsymbol{\phi}^o$ such that X^o constitutes a saddle point of the Lagrangian $L(X)$, then $\boldsymbol{x}^o$ is a local minimum of the original problem given by Eq. (1).

B. First Order Method and Its Convergence Behavior

A saddle point that satisfies Eq. (4) is a solution of the following saddle point problem:

$$\min_{\boldsymbol{x}} L(X),\ \ \max_{\boldsymbol{\lambda},\boldsymbol{\phi}} L(X). \tag{5}$$

Therefore, one natural way to find saddle points is to descend in the $\boldsymbol{x}$ space and ascend in the $\boldsymbol{\lambda}$ space and the $\boldsymbol{\phi}$ space; that is, we can consider the following gradient dynamics:

$$\frac{d\boldsymbol{x}(t)}{dt} = -\nabla_{\boldsymbol{x}} L(X(t)) \tag{6a}$$

$$\frac{d\boldsymbol{\lambda}(t)}{dt} = \nabla_{\boldsymbol{\lambda}} L(X(t)),\ \ \frac{d\boldsymbol{\phi}(t)}{dt} = \nabla_{\boldsymbol{\phi}} L(X(t)). \tag{6b}$$

Then, let us consider the following discrete gradient dynamics, generated by using Euler's differentiation technique:

$$\boldsymbol{x}(k+1) = \boldsymbol{x}(k) - \Delta T \nabla_{\boldsymbol{x}} L(X(k)) \tag{7a}$$

$$\boldsymbol{\lambda}(k+1) = \boldsymbol{\lambda}(k) + \Delta T \nabla_{\boldsymbol{\lambda}} L(X(k)) \tag{7b}$$

$$\boldsymbol{\phi}(k+1) = \boldsymbol{\phi}(k) + \Delta T \nabla_{\boldsymbol{\phi}} L(X(k)), \tag{7c}$$

where $\Delta T > 0$ is the sampling parameter which appears in Euler's differentiation procedure. Obviously, an equilibrium point of the dynamics of Eq. (7) satisfies the first order necessary condition (Theorem 1). Therefore, the optimization method using the dynamics of Eq. (7) is called the "first order method" [2]. Its convergence behavior has been studied using a merit function under the assumption of positive definiteness of the Hessian of $L(X)$ w.r.t. $\boldsymbol{x}$ in [2], [7]. In this study, we explain its convergence behavior in terms of linear stability theory.

Let us consider the orbital stability of an equilibrium point X^* (which satisfies the first order necessary condition). Let the dynamics of Eq. (7) be expressed as

$$\begin{pmatrix} \boldsymbol{x}(k+1) \\ \boldsymbol{\lambda}(k+1) \\ \boldsymbol{\phi}(k+1) \end{pmatrix} = \boldsymbol{\gamma}(X(k)) = \begin{pmatrix} \boldsymbol{x}(k) - \nabla_{\boldsymbol{x}} L(X(k)) \\ \boldsymbol{\lambda}(k) + \nabla_{\boldsymbol{\lambda}} L(X(k)) \\ \boldsymbol{\phi}(k) + \nabla_{\boldsymbol{\phi}} L(X(k)) \end{pmatrix}. \tag{8}$$

Then, the Jacobian of $\boldsymbol{\gamma}(X)$ is given by

$$D\boldsymbol{\gamma}(X) = I - \Delta T A(X) \tag{9a}$$

$$\text{where } A(X) = \begin{pmatrix} \nabla^2_{\boldsymbol{xx}} L(X) & \nabla^2_{\boldsymbol{\lambda x}} L(X) & \nabla^2_{\boldsymbol{\phi x}} L(X) \\ -\nabla^2_{\boldsymbol{x\lambda}} L(X) & -\nabla^2_{\boldsymbol{\lambda\lambda}} L(X) & -\nabla^2_{\boldsymbol{\phi\lambda}} L(X) \\ -\nabla^2_{\boldsymbol{x\phi}} L(X) & -\nabla^2_{\boldsymbol{\lambda\phi}} L(X) & -\nabla^2_{\boldsymbol{\phi\phi}} L(X) \end{pmatrix}. \tag{9b}$$

$\nabla^2_{\boldsymbol{xx}} L(X), \nabla^2_{\boldsymbol{\lambda x}} L(X), \nabla^2_{\boldsymbol{\phi x}} L(X), \nabla^2_{\boldsymbol{x\lambda}} L(X), \cdots, \nabla^2_{\boldsymbol{\phi\phi}} L(X)$ are the Hessians of $L(X)$ w.r.t. $\boldsymbol{x}$ and $\boldsymbol{x}$, $\boldsymbol{\lambda}$ and $\boldsymbol{x}$, $\boldsymbol{\phi}$ and $\boldsymbol{x}$, $\boldsymbol{x}$ and $\boldsymbol{\phi}$, $\cdots$, and $\boldsymbol{\phi}$ and $\boldsymbol{\phi}$ respectively. For example, $\nabla^2_{\boldsymbol{\lambda x}} L(X)$ is given by

$$\nabla^2_{\boldsymbol{\lambda x}} L(X) = \begin{pmatrix} \dfrac{\partial^2 L(X)}{\partial \lambda_1 \partial x_1} & \cdots & \dfrac{\partial^2 L(X)}{\partial \lambda_M \partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 L(X)}{\partial \lambda_1 \partial x_N} & \cdots & \dfrac{\partial^2 L(X)}{\partial \lambda_M \partial x_N} \end{pmatrix}. \tag{10}$$

Let the eigenvalues of $D\boldsymbol{\gamma}(X^*)$ and $A(X^*)$ be $\mu_1(X^*), \cdots, \mu_{N+M+Q}(X^*)$ and $\mu_1^a(X^*), \cdots, \mu_{N+M+Q}^a(X^*)$ the corresponding absolute values in descending order, and $l = 1, \cdots, N+M+Q$ unless otherwise stated. According to linear stability theory, if $|\mu_l(X^*)| < 1$, then the trajectory of the gradient dynamics of Eq. (7) converges to the equilibrium point stably. It is required that $\mu_l^a(X^*) > 0$, that is, $A(X^*)$ is a positive definite matrix so that the trajectory of the gradient dynamics converges to the equilibrium point stably, because

$$\mu_l(X^*) = 1 - \Delta T \mu_l^a(X^*). \tag{11}$$

We can derive the following theorem regarding positive definiteness of $A(X^*)$.

Theorem 3. *If $\nabla^2_{\boldsymbol{xx}} L(X^*)$ is a positive definite matrix and $\nabla^2_{\boldsymbol{\lambda\lambda}} L(X^*)$ and $\nabla^2_{\boldsymbol{\phi\phi}} L(X^*)$ are semi-negative definite matrices, then $A(X^*)$ is a positive definite matrix.*

Proof: Let $\boldsymbol{a} \in R^N$, $\boldsymbol{b} \in R^M$, and $\boldsymbol{c} \in R^Q$. Suppose $\nabla^2_{\boldsymbol{xx}} L(X^*)$ is a positive definite matrix and $\nabla^2_{\boldsymbol{\lambda\lambda}} L(X^*)$ and that $\nabla^2_{\boldsymbol{\phi\phi}} L(X^*)$ are semi-negative definite matrices. Then

$$\begin{aligned}
&\left(\boldsymbol{a}^T \boldsymbol{b}^T \boldsymbol{c}^T\right) A(X^*) \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{b} \\ \boldsymbol{c} \end{pmatrix} \\
&= \boldsymbol{a}^T \nabla^2_{\boldsymbol{xx}} L(X^*)\boldsymbol{a} + \boldsymbol{a}^T \nabla^2_{\boldsymbol{\lambda x}} L(X^*)\boldsymbol{b} + \boldsymbol{a}^T \nabla^2_{\boldsymbol{\lambda x}} L(X^*)\boldsymbol{c} \\
&\quad - \boldsymbol{b}^T \nabla^2_{\boldsymbol{x\lambda}} L(X^*)\boldsymbol{a} - \cdots - \boldsymbol{c}^T \nabla^2_{\boldsymbol{\phi\phi}} L(X^*)\boldsymbol{c} \\
&= \boldsymbol{a}^T \nabla^2_{\boldsymbol{xx}} L(X^*)\boldsymbol{a} - \boldsymbol{b}^T \nabla^2_{\boldsymbol{\lambda\lambda}} L(X^*)\boldsymbol{b} - \boldsymbol{c}^T \nabla^2_{\boldsymbol{\phi\phi}} L(X^*)\boldsymbol{c} > 0
\end{aligned}$$

■

Furthermore, $\nabla^2_{\boldsymbol{\lambda\lambda}} L(X^*)$ is a semi-negative definite matrix, because

$$\begin{aligned}
&\text{if } \max\{0, \lambda_i + g_i(\boldsymbol{x})\} = 0 \text{ , then} \\
&\quad \frac{\partial^2 L(X)}{\partial \lambda_i^2} = -1, \ \frac{\partial^2 L(X)}{\partial \lambda_j \lambda_i} = 0, \\
&\text{if } \max\{0, \lambda_i + g_i(\boldsymbol{x})\} = \lambda_i + g_i(\boldsymbol{x}) \text{ , then} \\
&\quad \frac{\partial^2 L(X)}{\partial \lambda_i^2} = 0, \ \frac{\partial^2 L(X)}{\partial \lambda_j \lambda_i} = 0, \\
&\qquad\qquad\qquad (i = 1, \cdots, M, \ \ j = 1, \cdots, M).
\end{aligned}$$

In addition, $\nabla^2_{\boldsymbol{\phi\phi}} L(X^*)$ is a semi-negative definite matrix, because

$$\frac{\partial^2 L(X)}{\partial \phi_i^2} = \frac{\partial^2 L(X)}{\partial \phi_j \partial \phi_i} = 0 \qquad (i = 1, \cdots, Q, \ \ j = 1, \cdots, Q).$$

Therefore, we can easily obtain the following result.

Theorem 4. *If $\nabla^2_{\boldsymbol{xx}} L(X^*)$ is a positive definite matrix and ΔT is sufficiently small*[1]*, then the trajectory of the gradient dynamics of Eq.* (7) *converges to the equilibrium point X^*.*

Positive definiteness of $\nabla^2_{\boldsymbol{xx}} L(X^*)$ is guaranteed for general convex quadratic programming problems [2]. However, positive definiteness of $\nabla^2_{\boldsymbol{xx}} L(X^*)$ is not always guaranteed for general nonlinear optimization problems. In this case, it is expected that the trajectory will not converge to any solution. This is a shortcoming of the optimization method using gradient dynamics of Eq. (7).

[1] Specifically, ΔT is given as $\Delta T < \dfrac{2}{\lambda_1^a(X^*)}$, please see the article [1].

III. The Proposed Method

In this section, we propose a new constrained optimization method in which the convergence property of the gradient dynamics of Eq. (7) is improved by two techniques. One is the introduction of a coupling structure. The other is the introduction of an objective function weighting method. The improved method does not have global search capability. Therefore, we apply the multipoint type chaotic optimization method (EC-COM) that was proposed in [1] so that global search can be implemented to find feasible global minima.

A. Coupled Gradient Dynamics

Let us consider the following coupled gradient dynamics:

$$\begin{aligned}
\boldsymbol{x}(k+1) &= (1-c)\left\{\boldsymbol{x}(k) - \Delta T \nabla_{\boldsymbol{x}} L(X(k))\right\} \\
&\qquad + c\boldsymbol{x}^{tg}(k) \quad &(12a) \\
\boldsymbol{\lambda}(k+1) &= \boldsymbol{\lambda}(k) + \Delta T(1-c)\nabla_{\boldsymbol{\lambda}} L(X(k)) \quad &(12b) \\
\boldsymbol{\phi}(k+1) &= \boldsymbol{\phi}(k) + \Delta T(1-c)\nabla_{\boldsymbol{\phi}} L(X(k)), \quad &(12c)
\end{aligned}$$

where $c > 0$ is a coupling coefficient. The coupling structure of Eq. (12) is a stable coupling structure for a chaotic discrete time system [13]. In this coupling dynamics, a search point driven by the gradient dynamics of Eq. (7) is strongly advected to a coupling target $\boldsymbol{x}^{tg}(k)$.

We discuss the dynamical characteristics of the coupled gradient dynamics of Eq. (12). The coupled dynamics can be rewritten as

$$\begin{aligned}
\boldsymbol{x}(k+1) &= \boldsymbol{x}(k) - C\Delta T\Big\{\nabla_{\boldsymbol{x}} L(X(k)) \\
&\qquad - \frac{c}{C\Delta T}(\boldsymbol{x}^{tg}(k) - \boldsymbol{x}(k))\Big\} \quad &(13a) \\
\boldsymbol{\lambda}(k+1) &= \boldsymbol{\lambda}(k) + C\Delta T \nabla_{\boldsymbol{\lambda}} L(X(k)) \quad &(13b) \\
\boldsymbol{\phi}(k+1) &= \boldsymbol{\phi}(k) + C\Delta T \nabla_{\boldsymbol{\phi}} L(X(k)) \quad &(13c) \\
&\text{where } C = 1 - c. \quad &(13d)
\end{aligned}$$

Let the equilibrium point of the dynamics of Eq. (13) be X^*. Then, the following equations hold at the equilibrium point:

$$\begin{aligned}
&\begin{cases} \nabla_{\boldsymbol{x}} L(X^*) = 0, & (\boldsymbol{x}^* = \boldsymbol{x}^{tg}) \\ \nabla_{\boldsymbol{x}} L(X^*) - \dfrac{c}{C\Delta T}\left(\boldsymbol{x}^{tg} - \boldsymbol{x}^*\right) = 0 & (\boldsymbol{x}^* \neq \boldsymbol{x}^{tg}) \end{cases} \quad &(14a) \\
&\nabla_{\boldsymbol{\lambda}} L(X^*) = 0, \ \ \nabla_{\boldsymbol{\phi}} L(X^*) = 0. \quad &(14b)
\end{aligned}$$

Obviously, the first order necessary condition (Theorem 1) is satisfied in the former case in Eq. (14a). In the latter case, the first order necessary condition is not satisfied. However, the latter case occurs rarely, especially if $\dfrac{c}{C\Delta T}$ is variable w.r.t. k. Therefore, we suppose that the former case holds from now on. Let us consider the orbital stability of an equilibrium point of the coupled gradient dynamics Eq. (13). Let the right-hand side of the dynamics of Eq. (13) be written as $\boldsymbol{\gamma}_2(X)$ as per the discussion in section II-B. Then, the Jacobian of $\boldsymbol{\gamma}_2$ is given by

$$D\boldsymbol{\gamma}_2(X) = I - C\Delta T A_2(X) \tag{15a}$$

$$\text{where } A_2(\boldsymbol{x},\boldsymbol{\lambda}) = \tag{15b}$$

$$\begin{pmatrix} \nabla^2_{\boldsymbol{xx}}L(X) + \dfrac{c}{C\Delta T}I & \nabla^2_{\boldsymbol{\lambda x}}L(X) & \nabla^2_{\boldsymbol{\phi x}}L(X) \\ -\nabla^2_{\boldsymbol{x\lambda}}L(X) & -\nabla^2_{\boldsymbol{\lambda\lambda}}L(X) & -\nabla^2_{\boldsymbol{\phi\lambda}}L(X) \\ -\nabla^2_{\boldsymbol{x\phi}}L(X) & -\nabla^2_{\boldsymbol{\lambda\phi}}L(X) & -\nabla^2_{\boldsymbol{\phi\phi}}L(X) \end{pmatrix}. \tag{15c}$$

As described in section II-B, if $A_2(X^*)$ is a positive definite matrix, then a trajectory of the gradient dynamics of Eq. (13) converges to the equilibrium point stably. By Theorem 4, it is necessary that $\nabla^2_{\boldsymbol{xx}}L(X^*) + \frac{c}{C\Delta T}I$ be a positive definite matrix. Clearly, if $\frac{c}{C\Delta T}$ is set to an appropriately large value, then $\nabla^2_{\boldsymbol{xx}}L(X^*) + \frac{c}{C\Delta T}I$ becomes a positive definite matrix. Hence, we expect that the coupled gradient dynamics of Eq. (13) stably converges to its equilibrium point, given that the first order necessary condition (Theorem 1) is satisfied when $\frac{c}{C\Delta T}$ is made appropriately large.

B. Objective Function Weighting

Let us consider the following Lagrangian, in which the objective function is weighted:

$$\begin{aligned} L'(X;w) =& wf(\boldsymbol{x}) \\ &+ \frac{1}{2}\sum_{m=1}^{M}\left[(\max\{0,\lambda_m + g_m(\boldsymbol{x})\})^2 - \lambda_m^2\right] \\ &+ \sum_{q=1}^{Q}\phi_q h_q(\boldsymbol{x}) + \frac{1}{2}\sum_{q=1}^{Q}h_q^2(\boldsymbol{x}), \end{aligned} \tag{16}$$

where $w > 0$. The objective function weighting method was proposed in [8], [9] in order to improve the speed of convergence of the gradient dynamics. Let us consider the positive definiteness of $\nabla_{\boldsymbol{xx}}L'(X;w)$ in order to consider the convergence behavior of a gradient dynamics using the Lagrangian with objective function weighting Eq. (16). Let us define

$$\begin{aligned} \eta(X) = \frac{1}{2}\sum_{m=1}^{M}&\left[(\max\{0,\lambda_m + g_m(\boldsymbol{x})\})^2 - \lambda_m^2\right] \\ &+ \sum_{q=1}^{Q}\phi_q h_q(\boldsymbol{x}) + \frac{1}{2}\sum_{q=1}^{Q}h_q^2(\boldsymbol{x}) \end{aligned} \tag{17}$$

Then, $\nabla_{\boldsymbol{xx}}L'(X;w)$ is given by

$$\nabla^2_{\boldsymbol{xx}}L'(X;w) = w\nabla^2 f(\boldsymbol{x}) + \nabla^2_{\boldsymbol{xx}}\eta(X). \tag{18}$$

Let $\boldsymbol{y} \in R^N$, then

$$\boldsymbol{y}^T\nabla^2_{\boldsymbol{xx}}L'(X;w)\boldsymbol{y} = w\boldsymbol{y}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{y} + \boldsymbol{y}^T\nabla^2_{\boldsymbol{xx}}\eta(X)\boldsymbol{y}. \tag{19}$$

Hence, if $\nabla^2 f(\boldsymbol{x}^*)$ is not a positive definite matrix at an equilibrium point X^*, then setting w to a small value $(w < 1)$ improves the positive definiteness of $\nabla^2_{\boldsymbol{xx}}L'(X;w)$, and the convergence behavior of the gradient dynamics is improved. Meanwhile, if the search trajectory does not converge stably in spite of the positive definiteness of $\nabla^2 f(\boldsymbol{x}^*)$, then w should be set to a large value $(w > 1)$ in order to eliminate a negative effect arising from $\nabla^2_{\boldsymbol{xx}}\eta(X)$. When neither $\nabla^2 f(\boldsymbol{x})$ nor $\nabla^2_{\boldsymbol{xx}}\eta(X)$ are positive definite matrices, we cannot expect the objective function weighting method to be effective. However, we have not encountered such a case in our experience, except when the Hessian of the objective function cannot be defined.

C. Elite Coupling Type Chaotic Optimization Method

We apply the notion of EC-COM proposed in [1] to the proposed method so that global search can be implemented to find feasible global minima.

Let us consider the following multipoint type optimization model in which P search points are driven by the gradient dynamics of Eq. (7):

$$\boldsymbol{x}^p(k+1) = \boldsymbol{x}^p(k) - \Delta T\nabla_{\boldsymbol{x}}L(X^p(k)) \tag{20a}$$

$$\boldsymbol{\lambda}^p(k+1) = \boldsymbol{\lambda}^p(k) + \Delta T\nabla_{\boldsymbol{\lambda}}L(X^p(k)) \tag{20b}$$

$$\boldsymbol{\phi}^p(k+1) = \boldsymbol{\phi}^p(k) + \Delta T\nabla_{\boldsymbol{\phi}}L(X^p(k)) \tag{20c}$$

where p denotes the index of a search point, and $p = 1,\cdots,P$ unless otherwise stated. Search trajectories generated by Eq. (20) are confined to the following bounded search spaces which are set expediently[2]:

$$x_n^l \le x_n \le x_n^u,\ \ \lambda_m^l \le \lambda_m \le \lambda_m^u,\ \ \phi_q^l \le \phi_q \le \phi_q^u, \tag{21}$$

and then, we expect that search trajectories will be destabilized by setting the sampling parameter ΔT to a large value and chaotic search trajectories by which global search can be implemented will be generated as with the gradient dynamics for unconstrained optimization. We use toroidalization of search spaces [1] (see Eq. (23h)) in order to confine the search trajectories to the bounded search spaces. Next, we introduce a coupling structure to the multipoint type gradient dynamics of Eq. (20). As for coupled targets at each search point, we use the pbest and gbest defined by DOCHM [14]. In this method, the pbest $\boldsymbol{x}^p_{pb}(k)$ is defined by

$$\boldsymbol{x}^p_{pb}(k) = \begin{cases} \text{if } F^p(k) \neq \emptyset \text{ then} \\ \quad \operatorname*{argmin}_{\boldsymbol{x}^p(\kappa)}\{f(\boldsymbol{x}^p(\kappa)) \mid \boldsymbol{x}^p(\kappa) \in F^p(k)\} \\ \text{if } F^p(k) = \emptyset \text{ then} \\ \quad \operatorname*{argmin}_{\boldsymbol{x}^p(\kappa)}\{\Phi(\boldsymbol{x}^p(\kappa))\}, \end{cases} \tag{22}$$

where $\Phi(\boldsymbol{x}) = \sum_{m=1}^{M}(\max\{0,g_m(\boldsymbol{x})\})^2 + \sum_{q=1}^{Q}h_q^2(\boldsymbol{x})$, $F^p(k)$ consists of the set of feasible solutions obtained up to time k at the pth search point, and $\kappa = 1,\cdots,k$. The gbest $\boldsymbol{x}^{gb}(k)$ is the best of the pbest among all search points at time k.

The proposed model, which is a multipoint type chaotic optimization model in order to solve the constrained optimization

[2]Note that the bounded search space for $\boldsymbol{x}$ in Eq. (21) is equal to the bounded search space of Eq. (1d).

problem, is given by

$$\tilde{\boldsymbol{x}}^p(k+1) = C\left\{\boldsymbol{x}^p(k) - \Delta T(k)\nabla_{\boldsymbol{x}} L'(X^p(k);w)\right\} + c_1\boldsymbol{x}^p_{pb}(k) + c_2\boldsymbol{x}^{gb}(k) \quad (23a)$$

$$\tilde{\boldsymbol{\lambda}}^p(k+1) = \boldsymbol{\lambda}^p(k) + C\Delta T(k)\nabla_{\boldsymbol{\lambda}} L'(X^p(k);w) \quad (23b)$$

$$\tilde{\boldsymbol{\phi}}^p(k+1) = \boldsymbol{\phi}^p(k) + C\Delta T(k)\nabla_{\boldsymbol{\phi}} L'(X^p(k);w) \quad (23c)$$

$$x^p_n(k+1) = \tau(\tilde{x}^p_n(k+1), x^l_n, x^u_n) \quad (23d)$$

$$\lambda^p_m(k+1) = \tau(\tilde{\lambda}^p_m(k+1), \lambda^l_m, \lambda^u_m) \quad (23e)$$

$$\phi^p_q(k+1) = \tau(\tilde{\phi}^p_q(k+1), \phi^l_q, \phi^u_q) \quad (23f)$$

where

$$\Delta T(k) = \begin{cases} \Delta T_{\max}\cos^2\left(\frac{\pi k}{T}\right) & (k \le k_{\max} - \frac{T}{2}) \\ 0.5 \times \Delta T_{\max}\cos^2\left(\frac{\pi k}{T}\right) & (k > k_{\max} - \frac{T}{2}) \end{cases} \quad (23g)$$

$$\tau(\tilde{y}, a, b) = \begin{cases} \tilde{y} & (a < \tilde{y} < b) \\ (\tilde{y} - a)\mathrm{mod}(b-a) + a & (\tilde{y} \ge b) \\ (\tilde{y} - b)\mathrm{mod}(b-a) + b & (\tilde{y} \le a) \end{cases} \quad (23h)$$

$$C = 1 - c_1 - c_2. \quad (23i)$$

In Eq. (23g), the sampling parameter ΔT is given as a time varying coefficient using trigonometric functions. This aims at implementation of repetition of local and global searches.

IV. Numerical Simulations

In this section, we confirm the effectiveness of the proposed method through applications to various benchmark problems.

A. Benchmark Problems

We use the coil spring design problem [10] and benchmark problems used in the special session on constrained real parameter optimization in CEC2006 [11], [12]. In Table I and Table II, the former problem is labeled “coil” and the later problems are labeled “gxx” (xx is a problem number). We do not use g20 because it has no feasible solution.

B. Local Search

We implement local searches with Sequential Quadratic Programming (SQP) to find local minima exactly when $\Delta T(k) = 0$ (this holds every $T/2$ steps) and $k = k_{\max}$. Each local search starts from the gbest. The result of each local search is “not” fed back into the search points. In this study, we use the SQP function in Octave 3.0.3 [15] for the SQP implementation. For g10, line searches for $\boldsymbol{x}$ are not used.

C. Gradient Scaling

In some problems, since the range of search spaces or the sensitivity of the gradient varies considerably among different decision variable components, search trajectories of the proposed method keep diverging, even if ΔT approaches zero. Consequently, in these cases, we use a modified method in which the relevant component of the gradient is scaled. In this study, 0.000001 is multiplied by the relevant component of the gradient. This scaling procedure is applied to $\partial L'(X;w)/\partial x_3$ and $\partial L'(X;w)/\partial x_4$ in g05, $\partial L'(X;w)/\partial x_6$ in g17, $\partial L'(X;w)/\partial x_i (i = 2, \cdots, 7)$ in g21, $\partial L'(X;w)/\partial x_i (i = 2, \cdots, 22)$ in g22, and $\partial L'(X;w)/\partial x_9$ in g23.

D. Brake Function

In the proposed method, we assume feasible global minima (we hereinafter refer to feasible global minima as “optimal solutions”) are located within the bounded search space. However, optimal solutions are on or adjacent to the boundary in several problems. In these cases, we need to evaluate solutions that are on or adjacent to the boundary, therefore, we cannot obtain the optimal solution only using the toroidalization of search spaces. Consequently, in these cases, we use a modified method in which a brake function

$$b(x_n) = \frac{(x_n - x^l_n)(x^u_n - x_n)}{x^u_n - x^l_n} \quad (24)$$

is multiplied by $\nabla_{\boldsymbol{x}} L'(X^p(k);w)$ in Eq. (23a) in order to evaluate such solutions. The brake function is applied to coil, g01, g02, g04, g10, g14, g16, g19, g21, and g23.

E. Parameter Settings

We tune the following parameters for each problem:

- $\boldsymbol{x}^l, \boldsymbol{x}^u$: They are given preliminarily for each benchmark problem.
- $\boldsymbol{\lambda}^l, \boldsymbol{\lambda}^u, \boldsymbol{\phi}^l, \boldsymbol{\phi}^u$: $\boldsymbol{\lambda}^u$ and $\boldsymbol{\phi}^u$ are set to suitable positive values which are much larger than prospective Lagrangian multipliers of local minima. $\boldsymbol{\lambda}^l = -\boldsymbol{\lambda}^u$ and $\boldsymbol{\phi}^l = -\boldsymbol{\phi}^u$.
- $\Delta T_{\max}$: This is set to a large value with which global searches are implemented.
- w : They are selected experimentally from 0.01, 1.0, or 10.0. Generally, $w = 1.0$ is suitable.

The other parameters are fixed for all problems. $P = 20, k_{\max} = 5000, T = 1000$. Coupling coefficients are set to 0.01, if they are used.

F. Computational Cost

In order to compute the gradient, we use automatic differentiation (AD) with the ADOL-C tapeless forward mode [16], [17]. Automatic differentiation computes the target function and its gradient simultaneously. Therefore, computational cost is evaluated using the amount of AD calls. In SQP, line searches are implemented only using objective function calls. However, these function calls are counted as AD calls for simplicity.

G. Results

We apply three different versions of the proposed method in terms of the use of coupling and the use of objective function weighting (see the “Method” in the following explanation). The proposed method is performed 100 times, randomly resetting the initial points. The results are shown in Table I and Table II. Each column in the results denotes

- PB : the name of the problem.

TABLE I
RESULTS OF NUMERICAL SIMULATIONS (1)

PB	Method	SR (%)	FR (%)	Average	Best	AD Calls	Optimal Solution	Parameters
coil	Plain	**100**	**100**	0.0127216	**0.0126652**	5724	0.0126652	$\Delta T_{\max}$= 0.4, λ'' = **5**, w= 1.0
	GB/PB/W	**100**	**100**	**0.0127166**	**0.0126652**	5061		
	PB/W	**100**	**100**	0.0127189	**0.0126652**	4811		
g01	Plain	**100**	**100**	**-15.000**	**-15.000**	13934	-15.000	$\Delta T_{\max}$= 0.4, λ'' = **10**, w= 1.0
	GB/PB/W	**100**	**100**	**-15.000**	**-15.000**	25958		
	PB/W	**100**	**100**	**-15.000**	**-15.000**	15698		
g02	Plain	0	**100**	-0.46705	-0.73240	N/A	-0.80362	$\Delta T_{\max}$= 25.0, λ'' = **5**, w= 1.0
	GB/PB/W	**16**	**100**	**-0.77214**	**-0.80362**	64695		$\Delta T_{\max}$= 2.5, λ'' = **5**, w= **10.0**
	PB/W	1	**100**	-0.71032	-0.80354	13832		
g03	Plain	1	**100**	-0.15289	-1.00048	10055	-1.00050	$\Delta T_{\max}$= 1.0, ϕ'' = **10**, w= 1.0
	GB/PB/W	**100**	**100**	**-1.00044**	**-1.00050**	7737		$\Delta T_{\max}$= 1.0, ϕ'' = **10**, w= **0.01**
	PB/W	**100**	**100**	-1.00045	**-1.00050**	8807		
g04	Plain	**100**	**100**	**-30665.54**	**-30665.54**	10014	-30665.54	$\Delta T_{\max}$= 5.0, λ'' = **1000**, w= 1.0
	GB/PB/W	**100**	**100**	**-30665.54**	**-30665.54**	10014		
	PB/W	**100**	**100**	**-30665.54**	**-30665.54**	10014		
g05	Plain	**100**	**100**	**5126.4977**	5126.4973	69142	5126.49805*	$\Delta T_{\max}$= 0.35, λ'' = **10**, ϕ'' = **10**, w= 1.0
	GB/PB/W	3	**100**	5130.4740	**5126.4972**	86913		
	PB/W	67	**100**	5126.5449	5126.4973	72269		
g06	Plain	**100**	**100**	**-6961.8139**	**-6961.8139**	10031	-6961.8139	$\Delta T_{\max}$= 0.001, λ'' = **2000**, w= 1.0
	GB/PB/W	**100**	**100**	**-6961.8139**	**-6961.8139**	10422		
	PB/W	**100**	**100**	**-6961.8139**	**-6961.8139**	10027		
g07	Plain	**100**	**100**	**24.30621**	**24.30621**	10045	24.30621	$\Delta T_{\max}$= 0.005, λ'' = **5**, w= 1.0
	GB/PB/W	**100**	**100**	**24.30621**	**24.30621**	10044		
	PB/W	**100**	**100**	**24.30621**	**24.30621**	10044		
g08	Plain	**100**	**100**	-0.09576	**-0.09583**	9217	-0.09583	$\Delta T_{\max}$= 10.0, λ'' = **10**, w= 1.0
	GB/PB/W	**100**	**100**	**-0.09577**	**-0.09583**	9123		
	PB/W	**100**	**100**	**-0.09577**	**-0.09583**	9061		
g09	Plain	**100**	**100**	**680.63006**	**680.63006**	10071	680.63006	$\Delta T_{\max}$= 0.001, λ'' = **5**, w= 1.0
	GB/PB/W	**100**	**100**	**680.63006**	**680.63006**	10069		
	PB/W	**100**	**100**	**680.63006**	**680.63006**	10070		
g10	Plain	84	84	**7049.24802**	**7049.24802**	32045	7049.24802	$\Delta T_{\max}$= 0.01, λ'' = **1000**, w= 1.0
	GB/PB/W	**99**	**100**	7072.32215	**7049.24802**	18307		
	PB/W	94	98	7382.66361	**7049.24802**	21979		
g11	Plain	**100**	**100**	0.74995	**0.74990**	371	0.74990	$\Delta T_{\max}$= 0.5, ϕ'' = **10**, w= 1.0
	GB/PB/W	**100**	**100**	**0.74994**	**0.74990**	466		
	PB/W	**100**	**100**	0.74995	**0.74990**	377		

- Method
  "Plain" denotes the proposed method without any coupling structure and the objective function weighting $(c_1 = c_2 = 0.0)$. "GB/PB/W" is the proposed method using the coupling structure with the gbest and pbest and the objective function weighting $(c_1 = c_2 = 0.01)$. "PB/W" is the proposed method using the coupling structure with the pbest and the objective function weighting $(c_1 = 0.01, c_2 = 0.0)$.
- SR (Success Rate)
  Let $\boldsymbol{x}^o$ be the known optimal solution. A successful solution $\hat{\boldsymbol{x}}$ is a feasible solution satisfying $f(\hat{\boldsymbol{x}}) - f(\boldsymbol{x}^o) \leq 0.0001$. A trial is deemed successful if a successful solution is obtained in the trial. "SR" is rate of successful trials.
- FR (Feasible Rate) : rate of feasible trials. A trial is deemed to be feasible, if at least one feasible solution is found in the trial.
- Average
  Let $\boldsymbol{x}^*$ be the best solution among the feasible solutions obtained in each trial. "Average" is the average of $f(\boldsymbol{x}^*)$.
- Best: the best value of $f(\boldsymbol{x}^*)$ obtained as a result from all trials.
- AD Calls : average of AD calls needed in each trial for finding a successful solution.
- Optimal Solution : objective function value of the known optimal solution. In g05, g13, g14, g17, g21, and g23, described values are different from the values reported in [11]. The proposed method tends to obtain more exact solutions w.r.t. equality constraints. These are objective

TABLE II
RESULTS OF NUMERICAL SIMULATIONS (2)

PB	Method	SR (%)	FR (%)	Average	Best	AD Calls	Optimal Solution	Parameters
g12	Plain	**100**	**100**	-0.99993	**-1.00000**	979	-1.00000	$\Delta T_{\max}=20.0$, $\boldsymbol{\lambda''}=\mathbf{10}$, $w=1.0$
	GB/PB/W	**100**	**100**	**-0.99994**	-0.99999	1108		
	PB/W	**100**	**100**	**-0.99994**	-0.99999	845		
g13	Plain	**100**	**100**	**0.05418**	**0.05398**	60202	0.05410*	$\Delta T_{\max}=0.01$, $\boldsymbol{\phi''}=\mathbf{5}$, $w=1.0$
	GB/PB/W	11	49	0.35721	0.05419	82787		
	PB/W	**100**	**100**	0.05419	0.05400	63180		
g14	Plain	**100**	**100**	**-47.76109**	**-47.76109**	10235	-47.76019*	$\Delta T_{\max}=0.2$, $\boldsymbol{\phi''}=\mathbf{50}$, $w=1.0$
	GB/PB/W	**100**	**100**	-47.76108	**-47.76109**	10230		
	PB/W	**100**	**100**	**-47.76109**	**-47.76109**	10388		
g15	Plain	**100**	**100**	961.71507	961.71503	71854	961.71502	$\Delta T_{\max}=0.005$, $\boldsymbol{\phi''}=\mathbf{10}$, $w=1.0$
	GB/PB/W	98	99	961.71506	**961.71502**	79634		
	PB/W	**100**	**100**	**961.71505**	**961.71502**	74920		
g16	Plain	97	**100**	-1.90172	**-1.90516**	13490	-1.90516	$\Delta T_{\max}=1.0$, $\boldsymbol{\lambda''}=\mathbf{5}$, $w=1.0$
	GB/PB/W	**100**	**100**	**-1.90516**	**-1.90516**	12082		
	PB/W	**100**	**100**	**-1.90516**	**-1.90516**	13288		
g17	Plain	56	94	8878.79822	8853.53985	51814	8853.53981*	$\Delta T_{\max}=0.3$, $\boldsymbol{\phi''}=\mathbf{5}$, $w=1.0$
	GB/PB/W	86	**100**	8860.79702	8853.53979	28884		
	PB/W	**98**	**100**	**8854.28091**	**8853.53978**	34835		
g18	Plain	**100**	**100**	**-0.86603**	**-0.86603**	23638	-0.86603	$\Delta T_{\max}=0.5$, $\boldsymbol{\lambda''}=\mathbf{10}$, $w=1.0$
	GB/PB/W	**100**	**100**	**-0.86603**	**-0.86603**	21475		
	PB/W	**100**	**100**	**-0.86603**	**-0.86603**	20375		
g19	Plain	**100**	**100**	**32.65559**	**32.65559**	10051	32.65559	$\Delta T_{\max}=0.001$, $\boldsymbol{\lambda''}=\mathbf{5}$, $w=1.0$
	GB/PB/W	**100**	**100**	**32.65559**	**32.65559**	10048		
	PB/W	**100**	**100**	**32.65559**	**32.65559**	10052		
g21	Plain	75	98	193.82051	193.78336	36062	193.78692*	$\Delta T_{\max}=5.0$, $\boldsymbol{\lambda''}=\mathbf{100}$, $\boldsymbol{\phi''}=\mathbf{1000}$, $w=1.0$
	GB/PB/W	**91**	**100**	**193.78645**	**193.78215**	35456		
	PB/W	75	98	193.78693	193.78308	37531		
g22	Plain	**2**	46	**370.05309**	**236.37033**	100623	236.43098	$\Delta T_{\max}=0.003$, $\boldsymbol{\lambda''}=\mathbf{1000}$, $\boldsymbol{\phi''}=\mathbf{1000}$, $w=1.0$
	GB/PB/W	0	36	561.75119	237.23443	N/A		
	PB/W	0	46	1203.82862	236.60736	N/A		
g23	Plain	98	98	**-400.00015**	-400.00048	16834	-400.00000*	$\Delta T_{\max}=0.1$, $\boldsymbol{\lambda''}=\mathbf{10000}$, $\boldsymbol{\phi''}=\mathbf{10000}$, $w=1.0$
	GB/PB/W	**100**	**100**	-400.00013	**-400.00049**	15187		
	PB/W	98	98	-400.00013	**-400.00049**	14985		
g24	Plain	**100**	**100**	**-5.50801**	**-5.50801**	10011	-5.50801	$\Delta T_{\max}=0.02$, $\boldsymbol{\lambda''}=\mathbf{10}$, $w=1.0$
	GB/PB/W	**100**	**100**	**-5.50801**	**-5.50801**	10011		
	PB/W	**100**	**100**	**-5.50801**	**-5.50801**	10011		

function values of more exact solutions which are adjacent to solutions reported in [11].

- Parameter : parameter settings.

The bold font in the results denotes that its evaluation is the best among all methods.

As can be seen from Table I and Table II, optimal solutions are obtained frequently by using the proposed method except for g02 and g22. For g02 and g22, the proposed method obtains optimal solutions in 100 trials (SR $>$ 0). Furthermore, for g22, the proposed method finds a new solution superior to the optimal solution reported in [11]. The new solution is $\boldsymbol{x}^o =$ (236.3703263131, 135.432887751, 200.4284109782, 6462.5506931403, 3000008.50528427, 4000000.5141733, 32999990.9805424, 130.0000850528, 170.0000901946, 299.9999149472, 399.9999948583, 330.0000901946, 184.5937961784, 249.4656629923, 127.6585381975, 269.9999098054, 160, 5.2983169413, 5.1357979367, 5.5984216249, 5.4380789168, 5.0751738152).

As for the introduction of the coupling structure, its effectiveness is confirmed in solving g02, g10, g17, g21, and g23. However, it is not effective for g05. In this problem, the optimal solution is obtained without the introduction of coupling, and the introduction of the coupling structure is redundant. As for the introduction of objective function weighting, its effectiveness is confirmed for g02 and g03. In particular, its effectiveness is definitely apparent for g03.

V. CONCLUSION

In this study, we propose a new chaotic global optimization method using the Lagrangian method to solve a nonlinear constrained optimization problem. In the proposed method, the convergence behavior of the first order method is improved by the introduction of a coupling structure and introduction of the objective function weighting method. Then, we apply the multipoint type chaotic optimization method which has been proposed in [1] to the proposed coupled first order method so that global search can be implemented to find feasible global minima. As described in section IV, local searches are utilized in the proposed method. Therefore, the proposed method may be considered as one of the memetic algorithms that utilize gradient based local searches as reported in [18] or proposed in [19]. In these methods, local searches are organically incorporated, and global search procedures exploit results of local searches. Meanwhile, in the method proposed here, local search is used to obtain exact solutions, and they are independent from the main search procedure. Note that the proposed method performs to some degree without the local search procedure. Indeed, we can confirm that the proposed method obtains optimal solutions or quite good approximately optimal solutions without local search in most of the benchmark problems used in this study, such as coil, g01, g02, g03, g04 (in the case of $w = 0.01$), g05, g07, g08, g09, g11, g12, g13, g14, g15, g16, g18, and g24.

We confirm that the proposed method frequently obtains feasible global minima of the benchmark problems. The success rate of the proposed method for the entire set of problems is somewhat less than the winners of the special session on constrained real parameter optimization in CEC2006 [19], [20]. However, the proposed method is superior to the winners in that at least one optimal solution is obtained in all trials for all problems. In addition, the proposed method tends to obtain more exact solutions. We consider that the proposed method obtains a new solution for g22 due to this tendency. As for computational cost, simple comparisons are difficult since the proposed method mainly makes use of the gradient. The winners also use a gradient-based local search method or a gradient based operator. However, it is not clear how the computational cost of determining the gradient is evaluated as function calls in the winners. In the future, we will compare the proposed method with other methods with respect to computational cost. In addition, we will investigate applications of other strategies for the discrete gradient chaos model proposed in [21] to improve global search capability.

Acknowledgement

This work was supported by a Grant-in-Aid for Scientific Research (KAKENHI) for Young Scientists (B) (20700206) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. We appreciate the anonymous reviewers' very useful comments that have enhanced the quality of this paper.

References

[1] T. Okamoto and E. Aiyoshi: "Global optimization using a synchronization of multiple search points autonomously driven by a chaotic dynamic model", *J. Global Optimiz.*, Vol.41, No.2, pp.219-244 (2008)

[2] D. G. Luenberger: "Linear and nonlinear programming second edition", *Springer* (2003)

[3] X. L. Sun and D. Li, K. I. M. McKinnon: "On saddle points of augmented lagrangians for constrained nonconvex optimization", *SIAM J. Optim.*, Vol.15, No.4, pp.1128-1146 (2005)

[4] T. Li, Y. Wang, Z. Liang, and P. Pardalos: "Local saddle point and a class of convexification methods for nonconvex optimization problems", *J. Global Optimiz.*, Vol.38, No.3, pp.405-419 (2007)

[5] F. Tian and L. Wang: "Chaotic simulated annealing with augmented lagrange for solving combinatorial optimization problems", *Proc. of IECON 2000*, pp.2722-2725 (2000)

[6] H. Niitsuma, S. Ishii, and M. Ito: "Chaotic optimization method combined with coordinate transformation", *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, Vol.84, No.2, pp.12-20 (2001)

[7] X. Li and D. Chen: "Application of the transiently chaotic neural network to nonlinear constraint optimization problems", *Proc. of Intelligent Systems Design and Applications 2006*, pp.90-94 (2006)

[8] B. W. Wah, T. Wang, Y. Shang, and Z. Wu: "Improving the performance of weighted lagrange-multiplier methods for nonlinear constrained optimization", *Information Sciences*, Vol.124, No.1-4, pp.241-272 (2000)

[9] B. W. Wah, T. Wang: "Efficient and adaptive lagrange-multiplier methods for nonlinear continuous global optimization", *J. Global Optimiz.*, Vol.14, No.1, pp.1-25 (1999)

[10] J. S. Arora: "Introduction to optimum design", *McGraw-Hill* (1989)

[11] J. J. Liang, T. P. Runarsson, M. Clerc, P. N. Suganthan, C. A. Coello Coello, and K. Deb: "Problem definitions and evaluation criteria for the CEC 2006 special session on constrained real-parameter optimization", *Tech. Rep.*, Nanyang Technological University (2006)

[12] P. N. Suganthan: "P. N. suganthan's homepage", URL: http://www3.ntu.edu.sg/home/EPNSugan/, Accessed on 31 Mar. 2009.

[13] A. Pikovsky, M. Rosenblum, J. Kurths, and R. C. Hilborn: "Synchronization: A universal concept in nonlinear science", *Cambridge Univ. Pr.* (2002)

[14] H. Lu and W. Chen: "Self-adaptive velocity particle swarm optimization for solving constrained optimization problems", *J. Global Optimiz.*, Vol.41, No.3, pp.427-445 (2008-7)

[15] W. J. Eaton: "GNU octave", URL: http://www.octave.org/, Accessed on 31 Mar. 2009.

[16] A. Griewank, D. Juedes, and J. Utke: "Algorithm 755; ADOL-C: A package for the automatic differentiation of algorithms written in C/C++", *ACM Trans. on Mathematical Software*, Vol.22, No.2, pp.131-167 (1996)

[17] A. Walther and A. Griewank: "ADOL-C", URL: http://www.math.tu-dresden.de/~adol-c/, Accessed on 31 Mar. 2009.

[18] B. Li, Y. Ong, M. Le, and C. Goh: "Memetic gradient search", *Proc. of IEEE Congress on Eveolutionaly Computation 2008 (CEC2008)*, pp.2894-2901 (2008)

[19] J. J. Liang and P. N. Suganthan: "Dynamic multi-swarm particle swarm optimizer with a novel constraint-handling mechanism", *Proc. of IEEE Congress on Evolutionary Computation 2006 (CEC2006)*, pp.9-16 (2006)

[20] T. Takahama and S. Sakai: "Constrained optimization by the ϵ constrained differential evolution with gradient-based mutation and feasible elites", *Proc. of IEEE Congress on Evolutionary Computation 2006 (CEC2006)*, pp.1-8 (2006)

[21] T. Okamoto and E. Aiyoshi: "The improved draining method and its application to proper benchmark problems", *Proc. of SICE-ICCAS 2006*, pp.2190-2195 (2006)