A Perturbation Based Chaotic System Exploiting the Quasi-Newton Method for Global Optimization

Keiji Tatsumi* and Tetsuzo Tanino†

*Graduate School of Engineering, Osaka University,
Yamada-oka 2-1, Suita, Osaka 565-0871, Japan*

*tatsumi@eei.eng.osaka-u.ac.jp

†tanino@eei.eng.osaka-u.ac.jp

Received September 7, 2016; Revised March 8, 2017

The chaotic system has been exploited in metaheuristic methods of solving continuous global optimization problems. Recently, the gradient method with perturbation (GP) was proposed, which was derived from the steepest descent method for the problem with additional perturbation terms, and it was reported that chaotic metaheuristics with the GP have good performances of solving some benchmark problems. Moreover, the sufficient condition of its parameter values was theoretically shown under which its updating system is chaotic. However, the sufficient condition of its chaoticity and the width of strange attractor around each local minimum, which are important properties for exploiting the chaotic system in optimization, deeply depend on the eigenvalues of the Hessian matrix of the objective function at the local minimum. Thus, if the eigenvalues of different local minima are widely different from each other, or if it is different in different problems, such properties can cause the difficulty of selecting appropriate parameter values for an effective search. Therefore, in this paper, we propose modified GPs based on the quasi-Newton method instead of the steepest descent method, where their chaoticities and the width of strange attractor do not depend on the eigenvalue of the Hessian matrix at any local minimum due to the scale invariant of the quasi-Newton method. In addition, we empirically demonstrate that the parameter selection of the proposed methods is easier than the original GP, especially with respect to the step-size, and the chaotic metaheuristics with the proposed methods can find better solutions for some multimodal functions.

Keywords: Metaheuristics; chaos dynamics; quasi-Newton method; global optimization; perturbations; snap-back repeller.

1. Introduction

Metaheuristic methods using chaotic systems, which is called the chaotic metaheuristics, have been investigated to solve continuous or discrete global optimization problems having a large number of local minima. Those methods make use of popular properties of the chaos, especially, the

unpredictable behavior caused by the sensitivity of initial conditions, in order to find a desirable solution within a practical time without being trapped at an undesirable local minimum. They were originally proposed as one of neural networks called the chaotic neural network [Chen & Aihara, 1985; Wang & Smith, 1998], which can be applied

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

to a certain kind of combinatorial optimization problem. Then, the method has been extended for the general continuous or discrete optimization. In this paper, we deal with the continuous global optimization problem, the objective function of which is multimodal and continuously differentiable.

The chaotic metaheuristics have two important elements: what kind of chaos is used and how to use the chaotic system. From the latter viewpoint, many kinds of methods have been investigated such as chaotic simulated annealing (CA) [Hasegawa *et al.*, 2002; Ji & Tang, 2004; Wang *et al.*, 2011], chaotic particle swarm optimization (chaotic PSO) [Alatas *et al.*, 2009; Gandomi *et al.*, 2013; Hamazia *et al.*, 2012; Tatsumi *et al.*, 2013, 2015; Yang *et al.*, 2012], chaotic ant swarm optimization [Li *et al.*, 2012], chaotic harmony search algorithm [Alatas, 2010] and chaotic metaheuristics based on a descent method [Okamoto & Aiyoshi, 2008; Okamoto & Hirata, 2013; Tatsumi *et al.*, 2009; Tokuda *et al.*, 1998].

On the other hand, from the former viewpoint, the desirable properties of chaotic systems for the continuous global optimization have not been sufficiently discussed, as long as we know. In this paper, we focus on such properties of the chaotic system and analyze theoretically and empirically. From the viewpoint of what kind of chaos is used, the chaotic metaheuristics can be mainly classified into two groups. The first group uses a well-known function such as the logistic function which can generate chaotic sequences. Most of them use the chaotic system instead of randomized numbers in evolutionary algorithms such as the chaotic PSO, as mentioned above. It was reported that this kinds of method can find better solutions, while it is difficult to keep a good performance for any problem because they use the same function for different problems. In contrast, a chaotic system is derived from a descent method for the problem in the second group, which makes use of the property that a descent system for a minimization problem with a sufficiently large step-size is chaotic [Hata, 1982]. The method can be expected to search for solutions in the descent direction of the objective function, while it can also avoid the stagnation of the search due to its chaotic behavior [Okamoto & Aiyoshi, 2008; Okamoto & Hirata, 2013; Tatsumi *et al.*, 2009; Tokuda *et al.*, 1998]. In addition, the chaotic PSOs proposed in [Tatsumi *et al.*, 2013, 2015] use chaotic systems

derived by applying the steepest descent method to a temporal objective function of which local minima are the best solutions obtained so far. Since this kind of method uses a chaotic system which can reflect the information about each optimization problem which is obtained during the search process, we believe that they have a potential to be more suitable for global optimization than the first group.

Hence, we focus on the second group, in particular, the gradient method with perturbations (GP) [Tatsumi *et al.*, 2009], which was derived from the steepest descent method for the objective function with perturbation terms. The sufficient conditions of the chaoticity of the derived system, which are given by the inequalities with respect to the parameter values, were theoretically shown in the sense of Li-Yorke [Tatsumi & Tanino, 2013]. The conditions can be used for selecting parameter values, and the theoretical results were verified by analyzing the bifurcation structure of the chaotic system. Some desirable properties for global optimization were reported such that the chaotic system has only vanishingly small windows, and the overconcentration and drastic crisis are not observed through numerical experiments. Moreover, it was demonstrated that the performances of several metaheuristics with the GP are superior to those with the existing chaotic metaheuristics based on the descent method in solving some benchmark problems [Tatsumi *et al.*, 2009].

However, the GP has some drawbacks. The shown theory indicates that the sufficient condition of its chaoticity depends on the maximal eigenvalue of the Hessian matrix of the objective function at each local minimum of the problem, and thus, appropriate parameter values should be selected according to the maximal eigenvalues for an effective search. In addition, if the maximal eigenvalues of different local minima of the same problem are widely different from each other, it is not easy to select suitable parameter values. Furthermore, since the size of the strange attractor of a local minimum, that denotes the area within which chaotic orbits move, can be controlled by tuning parameter values, it is ideal for the effective search if the parameter value can be selected such that the strange attractor of each local minimum has an intersection with at least another. However, it can be more difficult because such conditions also depend on the Hessian matrix at each local minimum.

Therefore, in this paper, we propose modified GPs which use the quasi-Newton method instead of the steepest descent method. Even though the updating system has perturbations, the proposed methods can generate approximation Hessian matrices or approximated inverse Hessian matrices which are symmetric and positive definite, and satisfy the modified secant condition similarly to the standard quasi-Newton method. Moreover, we approximately show that the chaoticity of the proposed systems do not depend on the maximal eigenvalue of Hessian matrix at any local minimum due to the scale invariant of the quasi-Newton method, where the next point is basically selected along the quasi-Newton direction with perturbations independent of the scale of Hessian matrix. In addition, we can expect that the size of the strange attractor around any local minimum is also independent of its Hessian matrix, and that the proposed updating systems generate a sequence which can quickly approach a local minimum on the basis of the super-linear convergence property of the quasi-Newton method. Through numerical experiments, we analyze properties of the modified chaotic system by depicting its bifurcation diagrams, and we compare the ease of selecting parameter values of the chaotic metaheuristic methods with the modified and existing chaotic systems, and evaluate the performance of the methods for global optimization by applying them to some benchmark problems.

This paper is organized as follows: in Sec. 2, we introduce the GP for chaotic metaheuristics and theoretical conditions of its chaoticity, and point out its drawbacks. In Sec. 3, we show examples of drawbacks by depicting bifurcation diagrams for the GP for some problems. Next, in Sec. 4, we propose modified GPs based on the quasi-Newton method, and show theoretical conditions of their chaoticities, and in Sec. 5, we verify the theoretical results through its bifurcation diagrams. In Sec. 6, we apply the chaotic metaheuristics with the original and modified GPs to some benchmark problems. Finally, we conclude this paper in Sec. 7.

2. Chaotic Systems Used in Metaheuristic Methods

In this section, we introduce the gradient method with perturbations (GP) and a chaotic metaheuristic method exploiting the GP.

2.1. Gradient method with perturbations

The gradient method with perturbations (GP) [Tatsumi *et al.*, 2009] was proposed for an unconstrained global optimization problem as follows:

$$(P) \quad \min f(x),$$

where the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is multimodal and continuously twice differentiable. In addition, the GP can be applied to many constraint optimization problems by transforming the problem into the unconstrained one with an appropriate penalty function. For example, if a feasible region is a rectangle such as $[-c_B, c_B]^n$, we can select an appropriate exterior penalty function:

$$p(x) = c_P \sum_{i=1}^n (\max\{0, x_i - c_B, -x_i - c_B\})^3, \quad (1)$$

where c_P is an appropriate positive constant. Thus, we focus on the unconstrained problem in this paper. The system used in the GP is derived from the steepest descent method with perturbations for (P) as follows:

$$(S1) \quad x(t+1) := x(t) - \alpha \nabla f(x(t)) - \beta \omega c_\omega(x(t)),$$

$$c_\omega(x) := \begin{pmatrix} \cos(\omega x_1) \\ \vdots \\ \cos(\omega x_n) \end{pmatrix},$$

where α denotes a step-size, and ω and $\beta\omega$ denote an angular frequency and an amplitude of the perturbations, respectively. In [Tatsumi & Tanino, 2013], the sufficient condition of the chaoticity of (S1) for parameter values (α, β, ω) were theoretically shown in the sense of Li-Yorke [Li & Chen, 2003; Marotto, 1978, 2005], which is summarized as follows. Hereinafter, let us focus on a local minimum $x^* \in \Omega$, where Ω denotes the set of all local minima of the problem (P).

Assumption 1. For a local minimum x^* of (P), there exists a positive constant r such that $\nabla^2 f(x)$ is positive semidefinite for all $x \in B_r(x^*) := \{x \mid \|x - x^*\|_\infty \leq r\}$.

This assumption guarantees that all eigenvalues of $\nabla^2 f(x)$ are non-negative and f is convex in a neighborhood of the local minimum. Many multimodal objective functions of global optimization problems can be expected to have this property.

Moreover, since $\nabla f(x)$ is continuously differentiable, we can easily show that $\nabla^2 f(x)$ are bounded for $x \in B_r(x^*)$. Then, we define $\lambda_\infty(x^*)$ for x^* as follows:

$$\lambda_\infty(x^*) = \max_{x \in B_r(x^*)} \|\nabla^2 f(x)\|_\infty, \quad (2)$$

where, in this paper, we use the matrix norm induced by the vector norm: for any matrix $A \in \mathbb{R}^{m \times n}$, it is defined by

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}, \quad p = 1, 2 \text{ and } \infty.$$

Next, let us consider the assumption that the cycle of the cosine function in (S1) is less than the size of neighborhood $B_r(x^*)$.

Assumption 2. Suppose that ω satisfies

$$r > \frac{2\pi}{\omega}, \quad (3)$$

for a positive constant r satisfying Assumption 1.

If ω is sufficiently large, the assumption is satisfied. Thus, Assumption 2 guarantees that points and sets in $B_{2\pi/\omega}(x^*)$ are also included in $B_r(x^*)$ where $\nabla^2 f(x)$ is positive semidefinite and $\|\nabla^2 f(x)\|_\infty \leq \lambda_\infty(x^*)$. Finally, we introduce the following relations between $\lambda_\infty(x^*)$ and parameters in (S1), α , β and ω :

Assumption 3. Suppose that positive parameters α , ω and β satisfy the following inequalities:

$$\alpha \lambda_\infty(x^*) \leq \frac{1}{3\pi^2} \beta \omega^2, \quad (4)$$

$$2\pi \leq \beta \omega^2. \quad (5)$$

Under these assumptions, (S1) is chaotic in the sense of Li–Yorke [Li & Chen, 2003; Marotto, 1978, 2005] as follows:

Theorem 1. *If Assumptions 1–3 are satisfied for a local minimum x^* of (P), there exists a snap-back repeller x^{sb} of (S1) such that $\|x^{\text{sb}} - x^*\|_\infty < (\pi + 0.16)/2\omega$.*

Theorem 1 means that there exists an unstable fixed point x^{sb} close to the local minimum x^* , and that there exist many homoclinic orbits, which start from a point within a neighborhood of x^{sb} and go far away from x^{sb} , but come back to x^{sb} . Such an orbit is attracted to x^{sb} , but it is not trapped because of its chaoticity. By using the property of the orbits, we can search for solutions around x^{sb} ,

that is, around the local minima x^* . Hence, at the same time, we can expect that the search is extensive. In addition, if Assumptions 1–3 are satisfied for all $x^* \in \Omega$, then (S1) is chaotic around all local minima as follows:

Corollary 2.1. *If Assumptions 1–3 are satisfied for any local minimum $x^* \in \Omega$ of (P), then for any x^* there exists a snap-back repeller x^{sb} of (S1) such that $\|x^{\text{sb}} - x^*\|_\infty < (\pi + 0.16)/2\omega$.*

The corollary indicates that the sufficient condition depends on $\lambda_\infty(x^*)$, and from Lemma 1 in [Tatsumi & Tanino, 2013], we have

$$\frac{1}{\sqrt{n}} \|\nabla^2 f(x^*)\|_\infty \leq \sigma_{\max}(\nabla^2 f(x^*)) \leq \|\nabla^2 f(x^*)\|_\infty,$$

where $\sigma_{\max}(A)$ denotes the maximal eigenvalue of the matrix A . Hence, the sufficient condition also depends on the maximal eigenvalue of the Hessian matrix of the objective function at each local minimum. These facts imply that appropriate parameter values should be selected according to the maximal eigenvalues for an effective search. If the maximal eigenvalues of different local minima of the same problem are widely different from each other, it is not easy to select suitable parameter values.

Furthermore, the corollary means that there exists a strange attractor including each local minimum, where the strange attractor denotes the area within which homoclinic orbits move. The size of each strange attractor is also important to solve the global optimization by using (S1), as discussed at the next subsection. As the preparation of the discussion, we derive a rough estimation of the size of each strange attractor. The region of points generated by (S1) around x^* can be estimated as follows:

Theorem 2. *Suppose that $\nabla f(x^*)$ is nonsingular, and $f(x)$ is approximated by a quadratic function around a local minimum x^* as follows:*

$$\begin{aligned} f(x) &\simeq f(x^*) + \nabla f(x^*)^T (x - x^*) \\ &\quad + \frac{1}{2} (x - x^*)^T \nabla^2 f(x^*) (x - x^*) \\ &= f(x^*) + \frac{1}{2} (x - x^*)^T \nabla^2 f(x^*) (x - x^*). \end{aligned} \quad (6)$$

If a current point $x(t)$ generated by (S1) satisfies the inequality

$$\|x(t) - x^*\|_2 > \frac{\beta\omega\sqrt{n}}{\alpha\sigma_{\min}(\nabla^2 f(x^*))}, \quad (7)$$

where $\sigma_{\min}(A)$ denotes the minimal eigenvalue of A , then for the next point $x(t+1)$,

$$-(x^* - x(t))^T(x(t+1) - x(t)) < 0$$

holds.

Proof. From (6) we have that $\nabla f(x) = \nabla^2 f(x^*) \times (x - x^*)$, which, together with (S1), yields that

$$\begin{aligned} x(t+1) - x(t) &= -\alpha \nabla f(x(t)) - \beta\omega c_\omega(x(t)) \\ &= -\alpha \nabla^2 f(x^*)(x(t) - x^*) - \beta\omega c_\omega(x(t)). \end{aligned}$$

In addition, by using the property of the cosine function, we obtain that for any $v, u \in \mathbb{R}^n$,

$$(u)^T c_\omega(v) \leq \|u\|_2 \sqrt{n}. \quad (8)$$

Now, these results and (7) imply that

$$\begin{aligned} (x(t) - x^*)^T(x(t+1) - x(t)) &= -\alpha(x(t) - x^*)^T \nabla^2 f(x^*)(x(t) - x^*) \\ &\quad - \beta\omega(x(t) - x^*)^T c_\omega(x(t)) \\ &\leq -\alpha\sigma_{\min}(\nabla^2 f(x^*))\|x(t) - x^*\|_2^2 \\ &\quad + \beta\omega\|x(t) - x^*\|_2 \sqrt{n} \\ &= \|x(t) - x^*\|_2(-\alpha\sigma_{\min}(\nabla^2 f(x^*))\|x(t) - x^*\|_2 \\ &\quad - \beta\omega\sqrt{n}) < 0. \end{aligned}$$

Theorem 2 shows that the distance from the next point of $x(t)$ satisfying (7) to x^* is less than the distance from $x(t)$ to x^* , which implies that any point generated by (S1) around x^* after a sufficient number of iterations is included in

$$\|x - x^*\|_2 \leq \frac{\beta\omega\sqrt{n}}{\alpha\sigma_{\min}(\nabla^2 f(x^*))}. \quad (9)$$

This inequality indicates that the size of a strange attractor depends on the eigenvalue of $\nabla^2 f(x^*)$ for the corresponding local minima.

2.2. Chaotic metaheuristics

In this subsection, we discuss the desirable properties of the chaotic system for solving global optimization. Firstly, we introduce the basic concept of the chaotic metaheuristics, in which the balance between diversification and intensification of the search is often controlled. For diversification of the search, the extensive search is executed by making use of a chaotic system, and for intensification of the search, some kind of local search is used such as the steepest descent method with a sufficiently small step-size. In this paper, we focus on the chaotic simulated annealing method (CA) which is one of the most popular methods as mentioned in Sec. 1. In particular, we discuss the CA using (S1). In this method, the balance between diversification and intensification of the search is controlled by tuning the parameter β . At the early stages, the method searches for solutions by using (S2) with a sufficiently large β as an extensive search by the chaotic system, in which (α, β, ω) satisfies the conditions (4) and (5). Then, β is gradually decreased as the search progresses, and it finally becomes sufficiently small. Thus, we can expect that a sequence generated by (S1) converges to a local minimum.

Next, let us consider the desirable properties of the chaotic system for solving the problem. At the final stages of the search, intensification of the search is important to find a local minima. Therefore, the step-size α also should be appropriately selected for each problem. In addition, at the early stages, it is ideal that there exist many orbits which pass close to most of the local minima. Thus, β should be selected so that any strange attractor has at least one intersection with another strange attractor. If β is extraordinarily large, we can always expect the existence of such kind of strange attractors. However, too large β can cause a uniformly randomized search for the whole feasible region, which means a coarse and useless search. Hence, β should be appropriately large such that the sizes of strange attractors are suitable for each problem. On the other hand, as shown in the previous subsection, Theorem 2 and (9) indicate that the size of strange attractor of each local minimum can also be expected to depend on the eigenvalues of the corresponding Hessian matrix, and in general, the eigenvalue for different local minima can be different from each other, and furthermore, they can be diversely different for different problems. Therefore, from the above discussion, it is not easy to

select appropriate parameter values (α, β, ω) , and the preparatory experiment is required. Here, note that the desirable properties are required for other chaotic metaheuristics. These drawbacks of the GP will be verified through numerical experiments in the next section.

3. Bifurcation Diagram of GP

In this section, we apply the GP to two one-dimensional functions and show the bifurcation diagrams with different α . Under some different α , as β is increased gradually, we update points by (S1) from randomized initial points and plot them after a certain number of iterations, where ω was fixed at 100 for all experiments in this section.

3.1. Example 1

First, we applied the GP to one-dimensional $2n$ -minima function, which is shown in Table 1 in Sec. 6. Its function value at the two local minima are different from each other, while the corresponding second order differential values are relatively similar

at the local minima. We used three variations of the $2n$ -minima function, which are given by $f_{2n}(x)$, $f_{2n}^L(x) = 20f_{2n}(x)$ and $f_{2n}^S(x) = f_{2n}(x)/20$. Namely, the function values of f_{2n}^L and f_{2n}^S are 20 and 0.05 times than that of f_{2n} , which means that $(f_{2n}^L)'(x) = 20(f_{2n})'(x)$ and $(f_{2n}^S)'(x) = (f_{2n})'(x)/20$.

The results are shown in Figs. 1(a)–1(f). Figure 1(b) is the enlarged figure of a part of Fig. 1(a). Figures 1(a), 1(d) and 1(f) indicate that if an appropriate α is selected for each problem, namely $\alpha = 0.01$ for f_{2n} , $\alpha = 0.2$ for f_{2n}^L and $\alpha = 0.0005$ for f_{2n}^S , respectively, there exist two fixed points which correspond to two local minima at small β , respectively, and as β increases, the system becomes chaotic. Furthermore, with the increase of β , the size of each strange attractor also increases, and the two attractors merge at a sufficiently large β . Thus, under the appropriate selection of α , we can execute the extensive search by using (S1) with a sufficiently large β , and the local search by using (S1) with $\beta = 0$.

On the other hand, when $\alpha = 0.01$ for f_{2n}^L or f_{2n}^S , the bifurcation diagram is widely different from

Fig. 1. Bifurcation diagrams of (S1) for one-dimensional $2n$ -minima functions. (a) f_{2n} with $\alpha = 0.01$, (b) f_{2n} with $\alpha = 0.01$, $x \in (2.6, 2.9)$ and $\beta \in (0, 0.0001)$, (c) f_{2n}^L with $\alpha = 0.01$, (d) f_{2n}^L with $\alpha = 0.0005$, (e) f_{2n}^S with $\alpha = 0.01$ and (f) f_{2n}^S with $\alpha = 0.2$.

Fig. 1. (Continued)

the case that $\alpha = 0.01$ for f_{2n} . In particular, when $\alpha = 0.01$ for f_{2n}^L , there is no fixed point, and the distribution area of generated points is considerably wide, which cannot be used for global optimization. In the case of f_{2n}^S , the size of the strange attractor becomes wider at each β than that for f_{2n} . These results mean that preparatory experiments are required to select an appropriate α even if the second order differential values, that is, the eigenvalues of the Hessian matrices, are relatively similar at the local minima.

3.2. Example 2

Next, let us consider the following one-dimensional function:

$$f_{\text{ex2}}(x) = 0.1x^2 - \frac{200}{\exp(0.05(x+4)^2)} - \frac{100}{\exp((x-4)^2)} - \frac{500}{\exp((x-8)^2)}, \quad (10)$$

which is shown in Fig. 2. This function has three local minima, $x^1 \simeq -3.960389$, $x^2 \simeq 3.962545$ and $x^3 \simeq 7.998221$. The second order differential values are $f''_{\text{ex2}}(x^1) \simeq 20.2$, $f''_{\text{ex2}}(x^2) \simeq 194.9$ and $f''_{\text{ex2}}(x^3) \simeq 1000.0$, which are widely different from each other. Thus, it is difficult for all local minima to select appropriate parameter values. Similarly to Example 1, we compared bifurcation diagrams of (S1) for f_{ex2} , which are shown in Fig. 3. Although it is theoretically guaranteed that there exist three fixed points and strange attractors of (S1) close to three local minima, x^1 , x^2 and x^3 , respectively, one or two of these fixed points cannot be observed in Figs. 3(a)–3(c), in which α is set to be 1.0, 0.1 or 0.01, respectively. Only if $\alpha = 0.0001$, as shown in Fig. 3(d), there exist three fixed points close to the corresponding local minima, respectively. Thus, if we apply the CA with (S1) under the selection that $\alpha = 1.0, 0.1$ or 0.01 to this function, then both local and extensive searches cannot be executed around x^3 .

Fig. 2. Function value of f_{ex2} .

Fig. 3. Bifurcation diagrams of (S1) for f_{ex2} . (a) f_{ex2} with $\alpha = 1.0$, (b) f_{ex2} with $\alpha = 0.1$, (c) f_{ex2} with $\alpha = 0.01$ and (d) f_{ex2} with $\alpha = 0.001$.

These results are mainly due to the fact that the size of the strange attractor of each local minimum can be expected to depend on the eigenvalues of the corresponding Hessian matrix, as mentioned in the previous section: since $f''_{\text{ex2}}(x^1)$ and $f''_{\text{ex2}}(x^3)$ are considerably small and large, respectively, the sizes of the strange attractors around x^1 and x^3 are extremely small and large, respectively, if $\alpha = 1.0$, 0.1 or 0.01 , which can be estimated by (9). Therefore, we can conclude that an appropriate selection of α is essential for the effective search, if the second order differential values, that is, the eigenvalues of the Hessian matrices, are considerably different at local minima.

4. GP Based on the Quasi-Newton Method

In this subsection, we propose two kinds of modified systems to overcome drawbacks of the GP as mentioned in the previous sections. The proposed method uses the updating system based on the quasi-Newton method, which updates basically a current point along the quasi-Newton direction

by using an approximated Hessian matrix or its inverse matrix of the objective function, while the system is chaotic because of its perturbation terms. It can be expected that the chaoticity of the modified systems or the size of the strange attractor are independent of the Hessian matrix of the objective function. We firstly introduce the quasi-Newton method, and then, propose the modified systems and discuss their desirable properties.

4.1. Quasi-Newton method

For the unconstrained nonlinear optimization problem, the quasi-Newton method is well-known, and has been used widely because of its good performance [Fletcher, 2000]. The quasi-Newton method is based on the Newton method to find the stationary point of the objective function, and can overcome drawbacks of the Newton method. Although the Newton method can guarantee the fast convergence to a local minimum called the quadratic convergence, it requires the Hessian matrix of the objective function, and it can be applied to only the problem of which the Hessian matrix at any

solution is always positive definite. On the other hand, the quasi-Newton method uses an approximated Hessian matrix which is updated by using the gradient vector at each iteration instead of the Hessian matrix. This method can be applied to any unconstrained optimization problem of which objective function is continuously differential. The approximated Hessian matrix $B(t+1)$ obtained at iteration t satisfies the following properties:

- (1) Secant condition: at each iteration t , we have

$$y(t) = B(t+1)s(t), \quad (11)$$

where $y(t)$ and $s(t)$ are defined by

$$\begin{aligned} y(t) &:= \nabla f(x(t)) - \nabla f(x(t-1)), \\ s(t) &:= x(t) - x(t-1). \end{aligned} \quad (12)$$

The condition (11) is corresponding to the property of the Hessian matrix $\nabla^2 f$ such that

$$\begin{aligned} \nabla f(x(t)) - \nabla f(x(t-1)) \\ \simeq \nabla^2 f(x(t))(x(t) - x(t-1)), \end{aligned}$$

which is derived from the first order Taylor expansion of $\nabla f(x)$ about $x(t)$ and substitution $x = x(t-1)$.

- (2) $B(t)$ is symmetric and positive definite.

In addition, if an initial point is sufficiently close to a local minimum, the fast convergence to the local minimum is guaranteed, which is called the super-linear convergence. Although some variations of approximating the Hessian matrix have been proposed, we deal with the most popular one, the BFGS method. It can be summarized as follows: At the first iteration $t = 1$, we have $B(1) = I$, where I denotes an identity matrix, and set $s(1)$ and $y(1)$ by an appropriate initialization process. At each iteration t , $B(t)$ is updated by

$$B(t+1) := B(t) + \frac{y(t)y(t)^T}{y(t)^T s(t)} - \frac{B(t)s(t)s(t)^T B(t)}{s(t)^T B(t)s(t)} \quad (13)$$

and the next point is updated by

$$x(t+1) := x(t) - \alpha(t)(B(t+1))^{-1} \nabla f(x(t)), \quad (14)$$

where the step-size $\alpha(t)$ is determined by an exact linear search in the direction of $(B(t+1))^{-1} \times$

$\nabla f(x(t))$ such as

$$\alpha(t) := \underset{\alpha}{\operatorname{argmin}} f(x(t) - \alpha(B(t+1))^{-1} \nabla f(x(t))), \quad (15)$$

at each iteration t . This search ensures the inequality

$$s(t+1)^T y(t+1) > 0, \quad (16)$$

which is an important condition because it guarantees the above two properties: the secant condition and the positive definiteness of $B(t+1)$. In addition, it is well known that if the objective function can be approximated by a quadratic function around a local minimum x^* , the current point $x(t)$ is sufficiently close to x^* , and its Hessian matrix is properly approximated by B , then x^* is found by an iteration with $\alpha = 1$. Thus, under similar conditions, a fast convergence can be expected.

On the other hand, if $B(t)$ is updated by (13), a system of linear equations is required to be solved to obtain $x(t)$ from (15), which needs $O(n^3)$ operations. Thus, the reduced method was proposed, in which an approximated inverse Hessian matrix $B^{-1}(t+1)$ is directly updated by using $B^{-1}(t)$, $y(t)$ and $s(t)$. Now, we define $H(t) := B^{-1}(t)$, then $H(t)$ is updated by

$$\begin{aligned} H(t+1) &:= \left(I - \frac{s(t)y(t)^T}{y(t)^T s(t)} \right) H(t) \left(I - \frac{y(t)s(t)^T}{y(t)^T s(t)} \right) \\ &\quad + \frac{s(t)s(t)^T}{y(t)^T s(t)}. \end{aligned} \quad (17)$$

Since (17) requires only $O(n^2)$ operations, this method has been used [Fletcher, 2000]. Then, update system (14) as

$$x(t+1) := x(t) - \alpha(t)H(t+1)\nabla f(x(t)). \quad (18)$$

For this method, the secant condition and the positive definiteness of $H(t+1)$ always hold if the step-size $\alpha(t)$ is determined by

$$\alpha(t) := \underset{\alpha}{\operatorname{argmin}} f(x(t) - \alpha H(t+1) \nabla f(x(t))), \quad (19)$$

at each iteration t , similarly to the method described above.

Furthermore, a modified approximation technique of the Hessian matrix is used for some algorithms such as the successive quadratic programming method where the relation (16) cannot be necessarily guaranteed because an exact line search (15) is difficult to execute. Those methods

use the updating system called the Powell's modification [Fletcher, 2000; Powell, 1977]. If

$$s(t)^T y(t) < \gamma s(t)^T B(t) s(t) \quad (20)$$

holds, then we find $\hat{y}(t)$ by

$$\hat{y}(t) := \eta_t y(t) + (1 - \eta_t) B(t) s(t), \quad (21)$$

$$\eta_t := \frac{(1 - \gamma) s(t)^T B(t) s(t)}{s(t)^T B(t) s(t) - s(t)^T y(t)}, \quad (22)$$

where γ is constant such that $0 < \gamma < 1$, and $B(t)$ is updated by (13) with $\hat{y}(t)$. Otherwise, $B(t)$ is updated in the ordinary way. Then, the modified secant condition

$$B(t+1)s(t) = \hat{y}(t) \quad (23)$$

and $s(t)^T \hat{y}(t) > 0$ hold, which guarantees the positive definiteness of $B(t)$. In the next subsection, we propose chaotic systems based on the quasi-Newton method with Powell's modification.

4.2. Proposed system based on the quasi-Newton method

Now, we propose the two kinds of chaotic systems based on (14) and (18), respectively.

The first system is given by

$$(S2B) \quad x(t+1) = x(t) - \alpha(B(t+1))^{-1} \nabla f(x(t)) - \beta \omega c_\omega(x(t)).$$

The matrix $B(t)$ is updated by the following algorithm:

Perturbation-based chaotic system using the quasi-Newton method 1

Step 0. Initialization:

The initial point $x(0)$ is set by randomized numbers, and update $x(0)$:

$$x(1) := x(0) - \alpha u^b(\nabla f(x(0))), \quad (24)$$

$$s(1) := x(1) - x(0), \quad (25)$$

$$y(1) := u^b(\nabla f(x(1))) - u^b(\nabla f(x(0))). \quad (26)$$

Set $t := 1$ and $B(1) = I$.

Step 1. Update $B(t)$:

If $\|s(t)\|_2^2 < \varepsilon n$ or $s(t)^T y(t) < \varepsilon n$, then $B(t+1) := B(t)$. Otherwise, if $s(t)^T y(t) < \gamma s(t)^T B(t) s(t)$, then

calculate $\hat{y}_t$ by using (21) and (22), and $y(t) = \hat{y}_t$. Obtain $B(t+1)$ by using (13).

Step 2. Update $x(t)$ by using $B(t+1)$ and add perturbations:

$$\bar{x}(t+1) := x(t) - \alpha(B(t+1))^{-1} u^b(\nabla f(x(t))), \quad (27)$$

$$s(t+1) := \bar{x}(t+1) - x(t), \quad (28)$$

$$y(t+1) := u^b(\nabla f(\bar{x}(t+1))) - u^b(\nabla f(x(t))), \quad (29)$$

$$x(t+1) := \bar{x}(t+1) - \beta \omega c_\omega(\omega x(t)). \quad (30)$$

Step 3. $t := t+1$ and go to Step 1.

Here, at Steps 0 and 2, we set the upper and lower limits on each component of ∇f by using a sufficiently large positive constant c_u :

$$u_i^b(v) := \begin{cases} c_u & \text{if } v_i \geq c_u, \\ -c_u & \text{if } v_i \leq -c_u, \\ v_i & \text{otherwise,} \end{cases} \quad i = 1, \dots, n. \quad (31)$$

due to the numerical stability. Notice the two exception handlings for $\|s(t)\|_2^2 < \varepsilon$ or $s(t)^T y^t < \varepsilon$ at Step 1, where ε denotes a sufficiently small positive constant, and the calculation of $\bar{x}$ at Step 2. Such handlings are not required for the standard quasi-Newton method. The two exception handlings at Step 1 can be explained as follows: The standard optimization algorithm such as the quasi-Newton method usually generates a sequence $\{x(t)\}$ which converges to a stationary point. Thus, the algorithm is terminated when $\|s^t\|_2 = \|\bar{x}(t+1) - x(t)\|_2$ is sufficiently small. However, the proposed system generates an orbit which goes near many local minima, where $\|s^t\|_2 \simeq 0$ can hold if $\|x^* - x(t)\|_2 \simeq 0$, while the orbit does not converge to any point. Therefore, the first exception handling is important to keep the numerical stability. In the second exception handling, we deal with the case that $s(t)^T y(t) < n\varepsilon$. Even if $s(t)^T y(t) < n\varepsilon$ holds, the equality $s(t)^T y(t) > 0$ is kept at each iteration due to Powell's modification. However, if $s(t)^T y(t)$ is sufficiently small, the amount of change of $B(t)$ by (13), (21) and (22) is also sufficiently small. Thus, such a wasteful computation is reduced by the second exception handling. On the one hand, at Step 2, the gradient ∇f is calculated at $\bar{x}(t+1)$,

which is obtained by updating $x(t)$ along the quasi-Newton direction (27) without perturbations, in order to approximate precisely the Hessian matrix around a local minimum. By the discussion as mentioned in the previous subsection, it can be easily shown that $B(t)$ generated by the proposed method is symmetric, positive definite, and satisfies the modified secant condition in the similar way as in [Powell, 1977].

Secondly, we propose the second system based on $H(t)$, which requires less computational complexity than (S2B):

$$(S2H) \quad x(t+1) = x(t) - \alpha H(t+1) \nabla f(x(t)) - \beta \omega c_\omega(x(t))$$

and the $H(t)$ is updated as follows:

Perturbation-based chaotic system using the quasi-Newton method 2

Step 0. Initialization:

The initial point $x(0)$ is set by randomized numbers, and update $x(0)$ by (24)–(26). Set $t := 1$ and $H(1) = I$.

Step 1. Update $H(t)$:

If $\|s(t)\|_2^2 < \varepsilon n$ or $s(t)^T y(t) < \varepsilon n$, then $H(t+1) := H(t)$. Otherwise, if the following inequality

$$s(t)^T y(t) < \gamma y(t)^T H(t) y(t), \quad (32)$$

holds, then calculate $\hat{s}_t$ by using

$$\hat{s}(t) := \zeta_t s(t) + (1 - \zeta_t) H(t) y(t), \quad (33)$$

$$\zeta_t := \frac{(1 - \gamma) y(t)^T H(t) y(t)}{y(t)^T H(t) y(t) - s(t)^T y(t)}, \quad (34)$$

where γ is constant such that $0 < \gamma < 1$ similarly to (21) and (22), and $s(t) = \hat{s}_t$. Obtain $H(t+1)$ by using (17).

Step 2. Update $x(t)$ by using $H(t+1)$ and add perturbations by

$$\bar{x}(t+1) := x(t) - \alpha H(t+1) u^b(\nabla f(x(t))) \quad (35)$$

and (28)–(31).

Step 3. $t := t + 1$ and go to Step 1.

The modification technique of $s(t)$ used in Step 1 for (S2H) is derived from the similar concept to Powell's modification, as mentioned in the previous section. We can easily show that $H(t+1)$, $y(t)$

and $\hat{s}(t)$ obtained by this method satisfy another modified secant condition:

$$\hat{s}(t) = H(t+1) y(t), \quad (36)$$

and $\hat{s}(t)^T y(t) > 0$, which guarantees the positive definiteness of $H(t)$. On the other hand, even under the same situation, $B^{-1}(t)$ generated by (S2B) is not necessarily equal to $H(t)$ generated by (S2H), which is caused by the difference between secant conditions (23) and (36), and the difference between conditions (20) and (32).

4.3. Chaoticity of proposed system

In this subsection, we approximately show the chaoticities of (S2B) and (S2H). Firstly, we focus on a local minimum $x^* \in \Omega$ similarly to Sec. 3, and suppose that Assumptions 1 and 2 are satisfied. Moreover, we introduce a small strong assumption under which we can easily derive conditions of the chaoticities of (S2B) and (S2H).

Assumption 4. If a subsequence of $\{x(t)\}$ generated by (S2B) or (S2H) is included in $B_r(x^*)$, then for the corresponding t , the equality $B(t) = \nabla^2 f(x^*)$ or $H(t) = (\nabla^2 f(x^*))^{-1}$ holds.

This assumption means that if $\{x(t)\}$ approaches sufficiently close to x^* , then $B(t)$ also approaches $\nabla^2 f(x^*)$ or $H(t)$ approaches $(\nabla^2 f(x^*))^{-1}$. If $\beta\omega = 0$ and α is determined by an exact line search (15) or (19), Assumption 4 is asymptotically satisfied. Otherwise, in general, it is not necessarily hold. However, we suppose the assumption because it is useful to derive approximately a sufficient condition of the chaoticity of (S2B) or (S2H). Here, note that Assumption 4 guarantees the positive definiteness of $\nabla^2 f(x^*)$. Then, we can prove the following theorem:

Theorem 3. If Assumptions 1, 2 and 4 are satisfied for any local minimum $x^* \in \Omega$ of (P) , ω is sufficiently large, and α , β and ω satisfy the following inequalities:

$$\alpha < \frac{1}{3\pi^2} \beta \omega^2, \quad (37)$$

$$2\pi \leq \beta \omega^2, \quad (38)$$

then there exists a snap-back repeller x^{sb} of (S2B) or (S2H) for any x^* such that

$$\|x^{sb} - x^*\|_\infty \leq \frac{(\pi + 0.16)}{2\omega}. \quad (39)$$

Proof. We prove this theorem by applying Theorem 1, which shows the existence of a snap-back repeller close to the local minimum x^* by proving the existence of a homoclinic orbit included in $B_r(x^*)$. Thus, we deal with points $\{x(t)\}$ which are included in $B_r(x^*)$ in this proof. Moreover, since ω is sufficiently large, we have that for any $x(t) \in B_r(x^*)$, $-c_u \leq \nabla f_i(x(t)) \leq c_u$, $i = 1, \dots, n$.

Then, under Assumption 4, (S2B) and (S2H) can be regarded as equivalent to the following system:

$$(S3) \quad x(t+1) = x(t) - \alpha(\nabla^2 f(x^*))^{-1} \nabla f(x(t)) - \beta\omega c_\omega(x(t)),$$

for any $x(t) \in B_r(x^*)$. Moreover, (S3) can be considered to be a system which is derived from the steepest descent method with perturbations for the following problem:

$$(P2) \quad \min g(x) = (\nabla^2 f(x^*))^{-1} f(x).$$

Namely, the relation between (S3) and (P2) is the same as that between (S1) and (P). At the same time, due to the positive definiteness of $\nabla^2 f(x^*)$, we can easily show that any local minimum x^* of (P) is also a local minimum of (P2), and vice versa.

In addition, since $\nabla^2 g(x) = (\nabla^2 f(x^*))^{-1} \times \nabla^2 f(x)$ is bounded in $x \in B_r(x^*)$, we define $\lambda_r^Q(x^*)$ as follows:

$$\begin{aligned} \lambda_r^Q(x^*) &= \max_{x \in B_r(x^*)} \|\nabla^2 g(x)\|_\infty \\ &= \max_{x \in B_r(x^*)} \|(\nabla^2 f(x^*))^{-1} \nabla^2 f(x)\|_\infty \end{aligned} \quad (40)$$

and we have

$$\|\nabla^2 g(x^*)\|_\infty = \|(\nabla^2 f(x^*))^{-1} \nabla^2 f(x^*)\|_\infty = 1. \quad (41)$$

Now, if ω is sufficiently large, then we can select a sufficiently small r such that Assumption 2 holds. Therefore, because of the continuousness of $\nabla^2 f(x)$ and (37), we have that

$$\alpha \leq \alpha \lambda_r^Q(x^*) < \frac{1}{3\pi^2} \beta\omega^2. \quad (42)$$

Finally, we apply Theorem 1 to system (S3) for problem (P2). Then, from (38), (40) and (42) we obtain that there exists a snap-back repeller x^{sb} of (S3) for the local minimum x^* of (P2) such as (39). Then, x^{sb} is also a snap-back repeller of (S2B) or (S2H) for the local minimum x^* of (P) such as (39). ■

This condition implies that the chaoticity of (S2B) or (S2H) around x^* is independent to the Hessian matrix $\nabla^2 f$. Thus, it is easier to select parameter values of (S2B) or (S2H) than (S1). On the other hand, the systems have the same desirable properties as (S1) with respect to the homoclinic orbits. Namely, there exist many kinds of orbits attracted to a neighborhood of the snap-back repeller close to each local minimum without being trapped at the snap-back repeller or the local minimum, which is useful for global optimization. Next, we show a rough estimation of the width of a strange attractor of (S2B) or (S2H) in the similar way to Theorem 2.

Theorem 4. Suppose that Assumptions 1 and 2 are satisfied, and Assumption 4 is satisfied for $B_{\bar{r}}(x^*)$ with a sufficiently large $\bar{r} > r$. If $f(x)$ is approximated by a quadratic function (6) around a local minimum x^* , and a current point $x(t)$ generated by (S2B) or (S2H) satisfies the inequality

$$\|x(t) - x^*\|_2 > \frac{\beta\omega\sqrt{n}}{\alpha}, \quad (43)$$

then for the next point $x(t+1)$,

$$-(x^* - x(t))^T(x(t+1) - x(t)) < 0 \quad (44)$$

holds.

Proof. From Assumption 4, (6) and (S2B) or (S2H), we have

$$\begin{aligned} x(t+1) - x(t) &= -\alpha(\nabla^2 f(x^*))^{-1} \nabla f(x(t)) - \beta\omega c_\omega(x(t)) \\ &= -\alpha(x(t) - x^*) - \beta\omega c_\omega(x(t)). \end{aligned}$$

Then, from the property of sine function (8) and (43), we can easily derive the inequality (44). ■

Theorem 4 means that any point generated by (S2B) or (S2H) around x^* after a sufficient number of iterations is included in

$$\|x - x^*\|_2 \leq \frac{\beta\omega\sqrt{n}}{\alpha}. \quad (45)$$

Thus, we can expect that the size of any strange attractor depends on only α , β and ω , and not on the Hessian matrix for each local minima. Therefore, this property is suitable for the search, and it overcomes the drawbacks of (S1) which are pointed out in the previous section. These results represent that the proposed systems (S2B) and (S2H) are useful for global optimization.

5. Bifurcation Diagram of the Proposed System

In this section, we verify the theoretical results shown in the previous section through numerical experiments, where we evaluate bifurcation diagrams of the proposed system (S2B) for the two functions, Examples 1 and 2, used in Sec. 3.

First, we applied the proposed system (S2B) to three variations of one-dimensional $2n$ -minima function f_{2n} , f_{2n}^L and f_{2n}^S , which are shown in Figs. 4(a)–4(d). When $\alpha = 1$, a similar bifurcation diagram is obtained for three different variations of the function, and it is also similar to Figs. 1(a), 1(d) or 1(f). Namely, there exist two fixed points close to two local minima at a small β , respectively, and the system becomes chaotic as β increases. Finally, the two attractors merge at a sufficiently large β . Hence, even if the scale of the objective function is varied, (S2B) with $\alpha = 1$ can be used for both local and extensive searches.

Next, let us consider the one-dimensional function f_{ex2} which is given by (10) in Sec. 3.2, which

has three local minima, and the second order differential values are widely different from each other at the minima. The diagrams of (S2B) with different $\alpha \in \{1.0, 0.1, 0.01, 0.001\}$ for f_{ex2} are shown in Figs. 5(a)–5(d). In all diagrams, three fixed points are observed, which differ significantly from those of (S1). Hence, even if the second order differential values, namely, the eigenvalues of the Hessian matrices, are considerably different at local minima, (S2B) with $\alpha = 1$ can be used for both local and extensive searches. In addition, in almost all diagrams, we can see the same linear increase of the sizes of three strange attractors regardless of the second order differential value, which implies that the estimation (45) is valid.

These results coincide with the properties of (S2B) shown in the previous section, which means that the chaoticity of (S2B) and the width of the strange attractor are independent of the eigenvalues of the Hessian matrix at any local minimum. In addition, the results also demonstrate that selecting α is considerably easy.

Fig. 4. Bifurcation diagrams of (S2B) for the one-dimensional $2n$ -minima function (Example 1). (a) f_{2n} with $\alpha = 1.0$, (b) f_{2n} with $\alpha = 1.0$, $x \in (2.6, 2.9)$ and $\beta \in (0, 0.001)$, (c) f_{2n}^L with $\alpha = 1.0$ and (d) f_{2n}^S with $\alpha = 1.0$.

Fig. 5. Bifurcation diagrams of (S2B) for $f_{\text{ex}2}$. (a) $f_{\text{ex}2}$ with $\alpha = 1.0$, (b) $f_{\text{ex}2}$ with $\alpha = 0.1$, (c) $f_{\text{ex}2}$ with $\alpha = 0.01$ and (d) $f_{\text{ex}2}$ with $\alpha = 0.001$.

6. Application of the Proposed Method

In this section, we applied the metaheuristic method exploiting (S1), and the proposed systems (S2B) and (S2H) to some benchmark problems. As a metaheuristic method, we used a hybrid of chaotic simulated annealing (CA) [Wang & Smith, 1998] and particle swarm optimization (PSO) method [Kennedy & Eberhart, 1995].

6.1. Used chaotic metaheuristics

The CA and its variations have been used for global optimization, as mentioned in Secs. 1 and 2. However, it is well known that the simple CA is sometimes ineffective for some kind of problems because its search efficiency deeply depends on the bifurcation structure of a used chaotic system, and the search does not exploit the best solution obtained until the current iteration [Tatsumi *et al.*, 2009; Tokuda *et al.*, 1998]. Therefore, recently, the modified methods have been proposed. One of them is a hybrid of the CA and PSO, which makes use of the best solution [Okamoto & Hirata, 2013]. Although

this kind of method usually uses multiple searching points for the diversification, in this paper, we used only one chaotic orbit for the search because our aim is to evaluate the difference between the three chaotic systems, (S1), (S2B) and (S2H). Here, note that the desirable properties of a chaotic system for the CA, which are discussed in Sec. 2, are also required for the hybrid method. It is summarized as follows:

Chaotic metaheuristic method with proposed system (S2B): CM-S2B

Step 0. Set parameters $\beta := \beta_{\text{st}}$ and $\Delta\beta := \beta_{\text{st}}/t_{\text{max}}$. The initial point $x(0)$ is set by randomized numbers, and update $x(0)$ by using (24). Calculate $s(1)$ and $y(1)$ by using (25) and (26). Set $t := 1$, $B(1) := I$ and $x_b(1) := \text{argmin}\{f(x) | x \in \{x(0), x(1)\}\}$.

Step 1. If $\|s(t)\|_2^2 < \varepsilon n$ or $s(t)^T y(t) < \varepsilon n$, then $B(t+1) := B(t)$. Otherwise, if $s(t)^T y(t) < \gamma s(t)^T B(t)s(t)$, then calculate $\hat{y}(t)$ by using (21) and $y(t) = \hat{y}(t)$. Obtain $B(t+1)$ by using (13).

Step 2. Obtain $x(t+1)$ by using $B(t+1)$ and add perturbations by (27)–(31).

Step 3. Update $x(t+1)$ by using a PSO-like term:

$$x(t+1) := x(t+1) + c_{\text{ps}}(x_b(t) - x(t+1)). \quad (46)$$

If $f(x(t+1)) < f(x_b(t))$, then $x_b(t+1) := x(t+1)$.

Step 4. If $t < t_{\text{max}}$ then $t := t+1$, $\beta := \beta - \Delta\beta$ and go to Step 1. Otherwise, terminate algorithm.

Chaotic metaheuristic method with proposed system (S2H): CM-S2H

Step 0. Set parameters $\beta := \beta_{\text{st}}$ and $\Delta\beta := \beta_{\text{st}}/t_{\text{max}}$. The initial point $x(0)$ is set by randomized numbers, and update $x(0)$ by using (24). Calculate $s(1)$ and $y(1)$ by using (25) and (26). Set $t := 1$, $H(1) := I$ and $x_b(1) := \text{argmin}\{f(x) | x \in \{x(0), x(1)\}\}$.

Step 1. If $\|s(t)\|_2^2 < \varepsilon n$ or $s(t)^T y(t) < \varepsilon n$, then $B(t+1) := B(t)$. Otherwise, if $s(t)^T y(t) < \gamma y(t)^T H(t) y(t)$, then calculate $\hat{s}(t)$ by using (33) and $s(t) = \hat{s}(t)$. Obtain $H(t+1)$ by using (17).

Step 2. Obtain $x(t+1)$ by using $H(t+1)$ and add perturbations by (35), (28)–(31).

Step 3. Update $x(t+1)$ by using (46). If $f(x(t+1)) < f(x_b(t))$, then $x_b(t+1) := x(t+1)$.

Step 4. If $t < t_{\text{max}}$ then $t := t+1$, $\beta := \beta - \Delta\beta$ and go to Step 1. Otherwise, terminate algorithm.

Here, t_{max} is the maximal number of iterations, and $x_b(t)$ denotes the best solution found until

iteration t , and $0 < c_{\text{ps}} < 1$ is a positive constant. At Step 3, the current point $x(t)$ is updated along the direction to $x_b(t) - x(t+1)$ in a similar way to the PSO [Kennedy & Eberhart, 1995], which can intensify the search around the best solution. The parameter β_{st} denotes an initial value of β , and β is decreased by $\Delta\beta$ in this method.

In addition, the above chaotic metaheuristic method in which $x(t)$ is updated by (S1) at Steps 1 and 2 is called CM-S1.

6.2. Experimental conditions

We applied CM-S1, CM-S2B and CM-S2H to the 2-, 10-, 50- and 100-dimensional benchmark problems, Rastrigin, Rosenbrock, Griewank and $2n$ -minima functions, as shown in Table 1, and their two kinds of variations whose function values are 20 and 0.05 times each original function value, respectively, such that $f^{(L)} := 20f(x)$ and $f^{(S)} := f(x)/20$. The Rastrigin, Griewank and $2n$ -minima functions have a large number of local minima, while Rosenbrock function has only a few local minima and the basin of a local minimum is considerably large and curved, which greatly differs from other functions. Although our main aim is solving a global optimization problem which has many local minima, we applied the CM-S1, CM-S2B and CM-S2H to Rosenbrock function because the results demonstrate the difference between performances of the three methods.

Table 1. Four benchmark problems.

Function Name	Objective Function and Constraints	Optimal Solution and Optimal Value
Rastrigin	$\min f_{\text{RA}}(x) = 10n + \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i))$	$x^* = (0, \dots, 0)^T$
	s.t. $-5.12 \leq x_i \leq 5.12, i = 1, \dots, n.$	$f_{\text{RA}}(x^*) = 0$
Rosenbrock	$\min f_{\text{RO}}(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2$	$x^* = (1, \dots, 1)^T$
	s.t. $-2.048 \leq x_i \leq 2.048, i = 1, \dots, n.$	$f_{\text{RO}}(x^*) = 0$
Griewank	$\min f_{\text{GR}}(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	$x^* = (0, \dots, 0)^T$
	s.t. $-512 \leq x_i \leq 512, i = 1, \dots, n.$	$f_{\text{GR}}(x^*) = 0$
$2n$ -minima	$\min f_{2n}(x) = \sum_{i=1}^n (x_i^4 - 16x_i^2 + 5x_i) + 78.319n$	$x^* = (-2.9, \dots, -2.9)^T$
	s.t. $-5 \leq x_i \leq 5, i = 1, \dots, n.$	$f_{2n}(x^*) = 0$

In all methods, the maximum numbers of iterations $t_{\max}$ were set to be 1000, 3000, 15 000 and 30 000 for the 2-, 10-, 50- and 100-dimensional problems, respectively. For the Rastrigin and Griewank functions, we used a penalty function (1) with $c_P = 1$. Parameters γ in (22), c_u in (31), ε in the exception handling, and c_{pso} in (46) were set to be 0.8, $1.0\text{E}+8$, $1.0\text{E}-24$ and 0.5, respectively. All numerical experiments were executed on a PC (CPU: Core (TM) i7-2600 3.40 GHz, Memory: 16 GB) and coded by optimization tools in MathWorks Matlab 2010a.

6.3. Numerical experiments

First of all, we executed preparatory experiments in which we compared the means of objective function values obtained by three methods, CM-S1, CM-S2B and CM-S2H, on 50 trials for 2-, 10-, 50- and 100-dimensional four functions, as α and β_{st} were varied in $\{0.0001, 0.001, 0.01, 0.05, 0.1, 0.5, 1.0, 1.5, 2.0\}$ and $\{0.001, 0.01, 0.1, 1.0, 2.0\}$, respectively. Table 2 shows a part of the results, which are the mean function values obtained by CM-S1 and CM-S2B with each pair of $(\alpha, \beta_{\text{st}})$ for three 50-dimensional

Table 2. Means of objective function values obtained by CM-S1 and CM-S2B with varied α and β_{st} for three 50-dimensional functions.

			β_{st}				
α			2	1	0.1	0.01	0.001
CM-S1	f_{RA}	2.0	8.705E+02	8.654E+02	8.617E+02	8.964E+02	8.490E+02
		1.5	8.707E+02	8.692E+02	8.664E+02	8.630E+02	8.743E+02
		1.0	8.766E+02	8.563E+02	8.667E+02	8.603E+02	8.465E+02
		0.5	8.533E+02	8.631E+02	8.676E+02	8.733E+02	8.443E+02
		0.1	8.613E+02	8.529E+02	8.576E+02	8.534E+02	8.777E+02
		0.05	8.597E+02	8.571E+02	8.564E+02	7.233E+02	7.119E+02
		0.01	8.409E+02	8.279E+02	5.063E+02	4.314E+02	3.651E+02
		0.001	3.798E+02	3.637E+02	2.794E+02	2.171E+02	3.445E+02
		0.0001	7.061E+02	6.266E+02	3.571E+02	2.842E+02	3.464E+02
CM-S1	f_{RO}	2.0	1.781E+04	1.718E+04	1.664E+04	1.646E+04	1.599E+04
		1.5	1.695E+04	1.805E+04	1.666E+04	1.712E+04	1.742E+04
		1.0	1.609E+04	1.767E+04	1.649E+04	1.715E+04	1.667E+04
		0.5	1.688E+04	1.679E+04	1.638E+04	1.680E+04	1.792E+04
		0.1	1.581E+04	1.608E+04	1.726E+04	1.732E+04	1.610E+04
		0.05	1.714E+04	1.753E+04	1.754E+04	1.664E+04	1.714E+04
		0.01	1.691E+04	1.648E+04	1.762E+04	1.698E+04	1.586E+04
		0.001	4.241E+03	4.174E+03	4.083E+03	4.466E+03	2.605E+03
		0.0001	9.844E+03	1.057E+04	1.051E+04	5.530E+01	4.874E+01
CM-S1	f_{GR}	2.0	1.219E+03	1.172E+03	4.321E−02	2.837E−02	4.975E−02
		1.5	1.238E+03	1.211E+03	7.132E−03	6.998E−03	3.649E−03
		1.0	1.213E+03	1.167E+03	6.732E−03	6.785E−03	3.764E−02
		0.5	1.172E+03	1.171E+03	1.007E−02	7.126E−03	9.693E−01
		0.1	1.191E+03	1.088E+03	2.610E−01	1.226E+01	2.073E+02
		0.05	1.173E+03	2.360E+02	4.099E−01	6.112E+01	4.572E+02
		0.01	4.561E+01	1.375E+00	5.017E−01	2.093E+02	7.659E+02
		0.001	2.469E+00	1.352E+00	5.316E−01	2.577E+02	1.071E+03
		0.0001	2.528E+00	1.353E+00	5.354E−01	2.599E+02	1.049E+03
CM-S2	f_{RA}	2.0	8.684E+02	8.534E+02	8.695E+02	8.538E+02	8.542E+02
		1.5	6.226E+02	5.107E+02	3.284E+02	1.537E+02	3.750E+02
		1.0	6.914E+02	2.948E+02	1.632E+02	9.515E+01	3.750E+02
		0.5	3.920E+02	3.184E+02	2.182E+02	1.398E+02	3.806E+02
		0.1	5.938E+02	5.138E+02	3.018E+02	2.263E+02	3.862E+02
		0.05	6.573E+02	5.983E+02	3.402E+02	2.682E+02	3.774E+02
		0.01	7.464E+02	7.010E+02	4.651E+02	2.876E+02	3.108E+02
		0.001	7.700E+02	7.426E+02	5.363E+02	3.606E+02	3.666E+02
		0.0001	7.659E+02	7.418E+02	5.609E+02	3.193E+02	3.571E+02

Table 2. (Continued)

			β_{st}				
α			2	1	0.1	0.01	0.001
CM-S2	f_{RO}	2.0	1.714E+04	1.707E+04	1.662E+04	1.645E+04	1.751E+04
		1.5	1.712E+04	1.799E+04	1.708E+04	1.774E+04	1.636E+04
		1.0	1.663E+04	1.723E+04	1.704E+04	1.634E+04	1.634E+04
		0.5	1.645E+04	1.699E+04	1.656E+04	1.717E+04	1.738E+04
		0.1	1.795E+04	1.704E+04	1.716E+04	1.747E+04	1.674E+04
		0.05	1.724E+04	1.845E+04	1.667E+04	1.673E+04	1.681E+04
		0.01	1.750E+04	1.717E+04	1.220E+04	9.366E+03	9.377E+03
		0.001	1.668E+04	1.557E+04	2.625E+03	1.848E+03	9.288E+01
		0.0001	1.506E+04	1.365E+04	2.203E+03	1.260E+02	6.051E+01
CM-S2	f_{GR}	2.0	1.192E+03	1.168E+03	5.184E+00	1.053E+00	5.252E−01
		1.5	2.309E+00	1.320E+00	5.665E−01	3.645E−03	1.534E−04
		1.0	2.226E+00	1.313E+00	4.141E−03	3.352E−03	3.006E−04
		0.5	2.302E+00	1.318E+00	7.445E−03	4.582E−03	4.930E−04
		0.1	2.298E+00	1.322E+00	4.389E−03	5.244E−03	2.468E−04
		0.05	2.336E+00	1.298E+00	6.741E−03	7.074E−03	9.907E−04
		0.01	2.404E+00	1.357E+00	2.775E−02	9.210E−03	5.540E−03
		0.001	2.440E+00	1.366E+00	2.466E−01	1.795E−02	9.145E−03
		0.0001	2.587E+00	1.362E+00	4.419E−01	1.189E+00	2.953E+01

functions, Rastrigin, Rosenbrock and Griewank functions. The boldface in the table denotes the least mean function value obtained by each method for each function. Table 3 shows the best parameter set (α, β_{st}) in which the least mean of objective function value was obtained by CM-S1, CM-S2B or CM-S2H for each function. Table 4 shows the mean of the objective function values at the best parameter set for each function and their standard deviations, respectively, where the boldface and italics denote the first and second least mean function values obtained by three methods for each function, respectively. Each standard deviation is given in parenthesis. Moreover, we counted the number of (α, β_{st}) which is the best parameter set in each method for all problems, as shown in Fig. 6. In addition, we compared the search processes of three methods for four 50-dimensional functions, where the mean of the least function value $f(x_b)$ on 50 trials at each iteration is plotted, as shown in Fig. 7. Furthermore, we estimated the mean of the maximal eigenvalues of $B^{-1}(t)$ or $H(t)$ obtained in a trial of CM-S2B and CM-S2H, which is shown by Table 5. We compared the mean computational time of CM-S1, CM-S2B and CM-S2H with the best parameter set on five trials for each problem, shown in Table 6.

Now, we compare the properties of CM-S2B and CM-S2H. Table 5 shows the mean eigenvalues

of approximated inverse Hessian matrices generated in a trial of CM-S2B or CM-S2H, in which the mean eigenvalues of $H(t)$ in CM-S2H are greater than those of $B^{-1}(t)$ in CM-S2B for many problems, especially, for Rosenbrock function or low-dimensional functions. The result shows the difference between two modification techniques of keeping the positive definiteness of $B(t)$ and $H(t)$ used in CM-S2B and CM-S2H, respectively, which use the different modified secant conditions (23) and (36), and the different conditions (20) and (32), respectively.

Next, we focus the parameter values (α, β_{st}) in CM-S1, CM-S2B and CM-S2H. Table 2 indicates that the obtained mean function values are widely varied for each (α, β_{st}) , which displays that an appropriate parameter selection is important for the effective search. Table 3 shows the best parameter set (α, β_{st}) of CM-S1, CM-S2B or CM-S2H for each function, in which we can observe that the best parameter sets of CM-S2B or CM-S2H are relatively the same for different scale variations of same benchmark function except Rosenbrock function, while those of CM-S1 are varied for them. In addition, Table 3 and Fig. 6 show that the best parameter value α of the CM-S2B and CM-S2H is confined within a narrow range, while the range of α in CM-S1 is relatively wide: The best parameter set of (α, β_{st}) in the CM-S2B is mainly given by

Table 3. Parameter values (α, β_{st}) of CM-S1, CM-S2B and CM-S2H under which the least mean function value is obtained.

Func.	Method	Dim.			
		2	10	50	100
f_{RA}	CM-S1	0.001, 0.01	0.001, 0.01	0.001, 0.01	0.001, 0.01
	CM-S2B	1.0, 0.001	1.0, 0.01	1.0, 0.01	1.0, 0.01
	CM-S2H	0.1, 0.01	0.0001, 0.01	0.1, 0.001	0.05, 0.001
f_{RA}^L	CM-S1	0.0001, 0.01	0.0001, 0.01	0.0001, 0.01	0.0001, 0.01
	CM-S2B	1.0, 0.01	1.0, 0.01	1.0, 0.01	1.0, 0.01
	CM-S2H	0.05, 0.01	0.1, 0.01	0.5, 0.01	0.001, 0.001
f_{RA}^S	CM-S1	0.05, 0.01	0.1, 0.01	0.1, 0.01	0.1, 0.01
	CM-S2B	1.0, 0.01	1.0, 0.01	1.0, 0.01	1.0, 0.01
	CM-S2H	0.1, 0.01	0.5, 0.01	1.0, 0.001	1.0, 0.001
f_{RO}	CM-S1	0.001, 0.001	0.0001, 0.001	0.0001, 0.001	0.0001, 0.001
	CM-S2B	0.001, 0.01	0.001, 0.001	0.0001, 0.001	0.0001, 0.001
	CM-S2H	0.001, 0.001	0.001, 0.001	0.05, 0.001	0.05, 0.001
f_{RO}^L	CM-S1	0.0001, 2.0	1.5, 2.0	0.05, 0.001	1.5, 0.01
	CM-S2B	0.0001, 0.01	0.0001, 0.1	0.0001, 0.1	0.0001, 0.001
	CM-S2H	0.0001, 0.01	0.0001, 0.001	0.05, 0.001	0.05, 0.001
f_{RO}^S	CM-S1	0.01, 0.001	0.01, 0.001	0.01, 0.001	0.01, 0.001
	CM-S2B	0.05, 0.001	0.01, 0.001	0.0001, 0.001	0.0001, 0.001
	CM-S2H	0.01, 0.001	0.01, 0.001	0.05, 0.001	0.1, 0.001
f_{GR}	CM-S1	2.0, 0.1	0.5, 0.1	1.5, 0.001	0.5, 0.001
	CM-S2B	1.5, 0.1	1.5, 0.01	1.5, 0.001	1.0, 0.001
	CM-S2H	0.05, 1.0	0.05, 0.1	0.05, 0.001	0.1, 0.01
f_{GR}^L	CM-S1	0.1, 0.1	0.01, 0.1	0.05, 0.01	0.05, 0.01
	CM-S2B	1.0, 0.1	1.0, 0.001	0.1, 0.001	0.1, 0.001
	CM-S2H	0.1, 1.0	0.05, 0.1	0.05, 0.001	0.1, 0.001
f_{GR}^S	CM-S1	0.05, 1.0	2.0, 0.1	2.0, 0.1	2.0, 0.01
	CM-S2B	1.0, 1.0	1.5, 0.01	1.5, 0.001	1.0, 0.001
	CM-S2H	0.1, 1.0	0.05, 0.1	0.05, 0.01	0.05, 0.01
f_{2n}	CM-S1	0.001, 0.1	0.001, 0.1	0.001, 0.1	0.001, 0.01
	CM-S2B	1.0, 0.1	0.5, 0.1	0.5, 0.1	0.5, 0.1
	CM-S2H	0.01, 0.1	0.05, 0.1	0.5, 0.1	0.5, 0.1
f_{2n}^L	CM-S1	0.001, 0.001	0.001, 0.01	0.0001, 0.01	0.0001, 0.01
	CM-S2B	1.0, 0.1	0.5, 0.1	0.5, 0.1	0.5, 0.1
	CM-S2H	0.001, 0.1	0.05, 0.1	0.5, 0.1	0.5, 0.1
f_{2n}^S	CM-S1	0.01, 0.1	0.01, 0.1	0.05, 0.01	0.05, 0.01
	CM-S2B	1.0, 0.1	0.5, 0.1	0.5, 0.1	0.5, 0.1
	CM-S2H	0.1, 0.1	0.05, 0.1	0.5, 0.1	0.5, 0.1

$0.1 \leq \alpha \leq 1.5$ and $0.0001 \leq \beta_{st} \leq 1$, and that in the CM-S2H is mainly given by $0.05 \leq \alpha \leq 0.5$ and $0.0001 \leq \beta_{st} \leq 1$, respectively, for three functions except Rosenbrock function. In contrast, the set of CM-S1 is given by $0.0001 \leq \alpha \leq 2.0$ and $0.0001 \leq \beta_{st} \leq 2.0$, respectively. Here, the exceptional results for Rosenbrock function can be explained as follows: Since its largest basin is too curved, the appropriate

selection of a fixed step-size is difficult in three methods, which can account for the dispersion of best parameter sets. The difference of the range of the set between CM-S1 and CM-S2B or CM-S2H can be considered to be caused by the independence of its chaoticity and the width of the strange attractor of the proposed systems (S2B) and (S2H) from the eigenvalues of the Hessian matrix at any local

Table 4. Means of objective function values and their standard deviations obtained by CM-S1, CM-S2B and CM-S2H.

Func.	CM	Dim.			
		2	10	50	100
f_{RA}	S1	$6.14E-02$ (2.39E-01)	$2.54E+01$ (1.19E+01)	$2.17E+02$ (5.45E+01)	$4.98E+02$ (9.05E+01)
	S2B	3.50E-13 (3.17E-13)	1.45E+01 (6.93E+00)	9.51E+01 (2.05E+01)	1.60E+02 (2.25E+01)
	S2H	$1.24E-01$ (3.26E-01)	$2.75E+01$ (1.27E+01)	$1.35E+02$ (2.81E+01)	$2.27E+02$ (3.02E+01)
f_{RA}^L	S1	$1.06E-03$ (5.52E-04)	$3.60E+02$ (1.74E+02)	$3.36E+03$ (9.58E+02)	$6.28E+03$ (1.06E+03)
	S2B	6.25E-12 (6.51E-12)	2.90E+02 (1.58E+04)	1.78E+03 (3.38E+02)	3.20E+03 (4.66E+02)
	S2H	$4.89E+00$ (9.51E+00)	$5.91E+02$ (3.18E+02)	$3.20E+03$ (7.43E+02)	$4.79E+03$ (7.67E+02)
f_{RA}^S	S1	$1.57E-09$ (1.22E-09)	4.63E-01 (1.48E-01)	4.84E+00 (1.11E+00)	$8.97E+00$ (9.87E-01)
	S2B	1.63E-14 (1.46E-14)	$6.68E-01$ (3.37E-01)	$4.98E+00$ (1.10E+00)	8.31E+00 (1.48E+00)
	S2H	$3.21E-03$ (1.20E-02)	$1.40E+00$ (6.29E-01)	$7.53E+00$ (1.84E+00)	$1.26E+01$ (1.82E+00)
f_{RO}	S1	3.22E-07 (4.13E-07)	7.38E+00 (1.98E+00)	$4.87E+01$ (8.73E+00)	$9.94E+01$ (1.15E+01)
	S2B	$1.79E-03$ (3.49E-03)	$7.59E+00$ (2.04E+00)	$6.05E+01$ (2.46E+01)	$1.58E+02$ (4.53E+01)
	S2H	$1.93E-03$ (3.30E-03)	$7.76E+00$ (1.36E+00)	4.20E+01 (1.22E+01)	7.91E+01 (1.93E+01)
f_{RO}^L	S1	$1.59E+03$ (3.73E+03)	$5.43E+04$ (2.89E+04)	$3.19E+05$ (6.76E+04)	$6.60E+05$ (9.47E+04)
	S2B	2.80E-01 (1.70E+00)	$7.24E+03$ (1.58E+04)	$5.51E+03$ (2.95E+04)	$2.69E+03$ (7.79E+02)
	S2H	$6.63E+02$ (4.62E+03)	2.56E+03 (1.14E+04)	8.31E+02 (1.77E+02)	1.70E+03 (3.59E+02)
f_{RO}^S	S1	1.81E-05 (8.47E-05)	2.68E-01 (1.39E-01)	$2.35E+00$ (3.71E-01)	$4.83E+00$ (3.74E-01)
	S2B	$6.62E-05$ (2.88E-04)	$3.58E-01$ (1.06E-01)	$3.52E+00$ (1.37E+00)	$7.67E+00$ (3.12E+00)
	S2H	$4.06E-05$ (1.26E-04)	$3.32E-01$ (9.71E-02)	1.99E+00 (4.45E-01)	3.67E+00 (8.77E-01)
f_{GR}	S1	4.94E-03 (2.34E-02)	$2.34E-01$ (1.26E-01)	$3.65E-03$ (5.20E-03)	$3.29E-03$ (4.43E-03)
	S2B	$5.19E-03$ (1.04E-02)	7.87E-02 (1.15E-01)	1.53E-04 (1.05E-03)	0.00E+00 (0.00E+00)
	S2H	$3.04E-02$ (2.99E-02)	$1.96E-01$ (9.19E-02)	$2.41E-03$ (1.08E-02)	$2.03E-03$ (4.27E-03)
f_{GR}^L	S1	$1.03E-01$ (3.10E-01)	$4.17E+00$ (2.54E+00)	$1.28E-01$ (1.53E-01)	$9.25E-02$ (1.26E-01)
	S2B	6.71E-02 (1.44E-01)	1.26E+00 (9.21E-01)	1.68E-02 (5.14E-02)	$5.92E-03$ (2.93E-02)
	S2H	$6.22E-01$ (5.14E-01)	$3.99E+00$ (2.89E+00)	$2.18E-02$ (6.76E-02)	2.41E-06 (1.91E-07)
f_{GR}^S	S1	$1.99E-03$ (1.57E-03)	$1.09E-02$ (5.99E-03)	$1.15E-02$ (4.12E-03)	$5.66E-04$ (2.95E-04)
	S2B	2.38E-04 (2.72E-04)	5.37E-03 (6.96E-03)	2.28E-05 (1.16E-04)	0.00E+00 (0.00E+00)
	S2H	$1.28E-03$ (1.19E-03)	$1.00E-02$ (3.92E-03)	$3.21E-04$ (3.78E-04)	$1.65E-04$ (2.33E-04)
f_{2n}	S1	$4.42E-02$ (6.03E-02)	$8.59E+01$ (3.72E+01)	$5.16E+02$ (1.02E+02)	$1.10E+03$ (1.40E+02)
	S2B	5.98E-05 (3.96E-04)	7.07E+01 (3.44E+01)	$2.68E+02$ (8.18E+01)	$4.49E+02$ (1.39E+02)
	S2H	$6.64E-02$ (1.06E-01)	$9.32E+01$ (4.50E+01)	2.21E+02 (6.59E+01)	3.32E+02 (9.82E+01)
f_{2n}^L	S1	$4.75E+02$ (4.18E+02)	$2.50E+03$ (1.84E+03)	$1.15E+04$ (3.23E+03)	$2.03E+04$ (2.99E+03)
	S2B	5.55E-03 (3.75E-02)	1.27E+03 (8.36E+02)	$5.64E+03$ (1.79E+03)	$9.87E+03$ (2.42E+03)
	S2H	$1.43E+00$ (2.06E+00)	$1.79E+03$ (8.50E+02)	4.27E+03 (1.77E+03)	6.85E+03 (1.84E+03)
f_{2n}^S	S1	$2.22E-03$ (2.28E-03)	$4.73E+00$ (2.02E+00)	$2.85E+01$ (8.99E+00)	$5.21E+01$ (7.65E+00)
	S2B	1.22E-07 (1.27E-07)	2.72E+00 (1.89E+00)	$1.33E+01$ (4.71E+00)	$2.17E+01$ (5.82E+00)
	S2H	$3.93E-03$ (9.88E-03)	$4.76E+00$ (2.03E+00)	1.27E+01 (4.08E+00)	1.60E+01 (4.72E+00)

minimum. The results represent that the parameter selection of CM-S2B or CM-S2H is easier than that of CM-S1. On the other hand, we can see that the best parameter values α of CM-S2H are relatively smaller than that of CM-S2B, which is caused by different modification techniques, as mentioned above.

Thirdly, let us discuss the function values obtained by three methods which are shown in

Table 4 and Fig. 7. The mean function values obtained by the CM-S2B or CM-S2H are less than those by the CM-S1 for all functions except six low-dimensional functions. The CM-S2H is inferior to the other two methods in finding high quality solutions for two- and ten-dimensional problems, while both of CM-S2B or CM-S2H are superior to the CM-S1 for 50- and 100-dimensional functions. These results show that for large-scale problems,

Fig. 6. The number of pairs of (α, β_{st}) with the best parameter pair in each method for 48 problems.

Fig. 7. Mean of the least function value at each iteration on 50 trials for CM-S1, CM-S2B and CM-S2H for 50-dimensional functions. (a) Rastrigin function, (b) Rosenbrock function, (c) Griewank function and (d) $2n$ -minima function.

Table 5. Mean of the maximal eigenvalues of $B^{-1}(t)$ or $H(t)$ obtained in CM-S2B and CM-S2H on a trial.

Dim.	CM-S2B				CM-S2H			
	f_{RA}	f_{RO}	f_{GR}	f_{2n}	f_{RA}	f_{RO}	f_{GR}	f_{2n}
100	0.0530	0.7834	88.5855	0.0307	0.0266	0.0715	3082.8313	0.0754
50	0.0507	0.0492	51.2607	0.0283	0.0263	0.0710	788.9136	0.0472
10	0.0250	0.0726	78.0665	0.0275	0.0429	0.0770	3616.8734	0.0741
2	0.0047	0.0064	12.4108	0.0121	0.1730	1.2474	12.6041	0.0938

the inverse Hessian matrix can be regarded as properly approximated around each local minimum in CM-S2B or CM-S2H so that the systems (S2B) and (S2H) generate a sequence which approaches the local minimum.

In comparing CM-S2B and CM-S2H further, CM-S2H obtained the least mean function values by three methods, for high-dimensional Rosenbrock and $2n$ -minima functions, while CM-S2B did so for Griewank and Rastrigin functions, which shows that the CM-S2B and CM-S2H have high search capabilities. The different performances of CM-S2B and CM-S2H are caused by different modification techniques used in them. The results imply that the modification technique is very important not only in order to keep the positive definite of approximated matrices but also for an effective search. Moreover, Fig. 7 shows that CM-S2B or CM-S2H obtains less function values than CM-S1. These results coincide with the above results. In the figures, we can see that CM-S2B and CM-S2H can effectively improve the function value at the early and final stages of the search, which represents that both the extensive search based on the proposed chaotic systems and the local search based on the quasi-Newton method can contribute to improving solutions.

Finally, we compare the computational time of the three methods. Table 6 indicates that the computational time of CM-S2B is at most 3.8 times greater than that of CM-S1, and that of CM-S2B is at most 2.5 times greater than it, which can be considered not to reach the worst case complexity, $O(n^3)$ and $O(n^2)$. In addition, compared with CM-S2B, CM-S2H can reduce the computational time for 50- and 100-dimensional functions because it directly updates an approximated inverse Hessian matrix. Although CM-S2B and CM-S2H require more computational time than CM-S1, it is an acceptable increase. Because the best parameter set of the CM-S2B or CM-S2H is confined within a considerably narrow range than that in the CM-S1, as mentioned above, it means that the two methods require a smaller number of preparatory experiments than CM-S1 to select appropriate parameter values for each problem.

From these results, we can comprehensively conclude that the proposed systems (S2B) and (S2H) are useful in both the parameter selection and the search capability for solving global optimization problems, and in particular, (S2H) is more effective for large-scale problems.

Table 6. Mean computational time (sec) on five trials of CM-S1, CM-S2B and CM-S2H.

Dim. Func.	100				50			
	f_{RA}	f_{RO}	f_{GR}	f_{2n}	f_{RA}	f_{RO}	f_{GR}	f_{2n}
CM-S1	6.4480	5.6841	8.5159	6.2776	3.1137	2.7707	4.1033	2.8893
CM-S2B	24.4275	19.3408	25.8484	23.6784	9.7016	8.2512	10.604	9.0878
CM-S2H	15.2046	14.2564	18.3344	15.3399	7.4933	6.8235	9.6498	7.3944

Dim. Func.	10				2			
	f_{RA}	f_{RO}	f_{GR}	f_{2n}	f_{RA}	f_{RO}	f_{GR}	f_{2n}
CM-S1	0.5875	0.5339	0.7186	0.5314	0.5879	0.4685	0.6794	0.4835
CM-S2B	1.4719	1.3122	1.6756	1.2979	1.3547	1.2153	1.5814	1.2310
CM-S2H	1.2879	1.1757	1.5635	1.1976	1.3695	1.0579	1.6113	1.2366

7. Conclusion

In this paper, we have focused on chaotic systems used in metaheuristic methods for the continuous global optimization, especially, the gradient method with perturbations (GP). Although the condition of its chaoticity was theoretically shown and some numerical experiments demonstrated a good performance of the GP in solving problems, the GP has some drawbacks: its chaoticity and the size of the strange attractor of a local minimum depend on the Hessian matrix of the objective function at the local minimum, which causes the difficulty of selecting appropriate parameter values for each problem. Therefore, in this paper, we have proposed two kinds of modified GPs based on the quasi-Newton method: One method updates an approximated Hessian matrix, and the other updates its inverse matrix. We have approximately shown that the chaoticity of the proposed systems do not depend on the maximal eigenvalue of the Hessian matrix at any local minimum due to the scale invariant of the quasi-Newton method. In addition, we can expect that the size of the strange attractor around any local minimum is also independent of its Hessian matrix, and that the proposed updating systems generate a sequence which can quickly approach a local minimum on the basis of the super-linear convergence of the quasi-Newton method. Through numerical experiments, we have verified the desirable properties of the proposed chaotic systems by depicting their bifurcation diagrams, and applying metaheuristic methods with the proposed chaotic system to some benchmark problems. We have observed that the parameter selection of the proposed systems is easier than the original GP, and that the chaotic metaheuristic methods with the proposed system, often find better solutions than the method with the original GP for some global optimization problems. Therefore, we can conclude that the proposed chaotic system is more useful for solving global optimization problems.

Through the numerical experiments, we have observed that the modification technique of keeping the definite positiveness of approximated matrices plays an important role for effective search. Therefore, as further research, we would like to analyze modified secant conditions or approximated Hessian matrices used in the technique in order to enhance the efficient metaheuristics for global optimization.

References

- Alatas, B., Akin, E. & Ozer, A. B. [2009] "Chaos embedded particle swarm optimization algorithms," *Chaos Solit. Fract.* **40**, 1715–1734.
- Alatas, B. [2010] "Chaotic harmony search algorithms," *Appl. Math. Comput.* **216**, 2687–2699.
- Chen, L. & Aihara, K. [1985] "Chaotic simulated annealing by a neural network model with transient chaos," *Neur. Netw.* **8**, 915–930.
- Fletcher, R. [2000] *Practical Methods of Optimization*, 2nd edition (John Wiley & Sons, NY).
- Gandomi, A. H., Yun, G. J., Yang, X.-S. & Talatahari, S. [2013] "Chaos-enhanced accelerated particle swarm optimization," *Commun. Nonlin. Sci. Numer. Simul.* **18**, 327–340.
- Hamaizia, T., Lozi, R. & Hamri, N. [2012] "Fast chaotic optimization algorithm based on locally averaged strategy and multifold chaotic attractor," *Appl. Math. Comput.* **219**, 188–196.
- Hasegawa, M., Ikeguchi, T. & Aihara, K. [2002] "Solving large scale traveling salesman problems by chaotic neurodynamics," *Neur. Netw.* **15**, 271–283.
- Hata, M. [1982] "Euler's finite difference scheme and chaos in $\mathbb{R}^n$," *Proc. Japan Acad. Ser. A* **58**, 178–181.
- Ji, M. & Tang, H. [2004] "Application of chaos in simulated annealing," *Chaos Solit. Fract.* **21**, 933–941.
- Kennedy, J. & Eberhart, R. C. [1995] "Particle swarm optimization," *Proc. IEEE Int. Joint Conf. Neural Networks*, pp. 1942–1948.
- Li, C. & Chen, G. [2003] "An improved version of the Marotto theorem," *Chaos Solit. Fract.* **18**, 69–77.
- Li, Y., Wen, Q. & Zhang, B. [2012] "Chaotic and swarm optimization with passive congregation," *Nonlin. Dyn.* **68**, 129–136.
- Marotto, F. R. [1978] "Snap-back repellers imply chaos in $\mathbb{R}^n$," *J. Math. Anal. Appl.* **63**, 199–223.
- Marotto, F. R. [2005] "On redefining a snap-back repeller," *Chaos Solit. Fract.* **25**, 25–28.
- Okamoto, T. & Aiyoshi, E. [2008] "Global optimization using a synchronization of multiple search points autonomously driven by a chaotic dynamic model," *J. Global Optimiz.* **41-2**, 219–244.
- Okamoto, T. & Hirata, H. [2013] "Global optimization using a multipoint type quasi-chaotic optimization method," *Appl. Soft Comput.* **13**, 1247–1264.
- Powell, M. J. D. [1977] "The convergence of variable metric methods for nonlinearly constrained optimization calculations," *Nonlinear Programming*, Vol. 3, eds. Mangasarian, O. L., Meyer, R. R. & Robinson, S. M. (Elsevier), pp. 27–63.
- Tatsumi, K., Obita, Y. & Tanino, T. [2009] "Chaos generator exploiting a gradient model with sinusoidal perturbations for global optimization," *Chaos Solit. Fract.* **42**, 1705–1723.

- Tatsumi, K. & Tanino, T. [2013] "A sufficient condition for chaos in the gradient model with perturbation method for global optimization," *Int. J. Bifurcation and Chaos* **23**, 1350102-1–21.
- Tatsumi, K., Ibuki, T. & Tanino, T. [2013] "A chaotic particle swarm optimization exploiting a virtual quartic objective function based on the personal and global best solutions," *Appl. Math. Comput.* **219**, 8991–9011.
- Tatsumi, K., Ibuki, T. & Tanino, T. [2015] "Particle swarm optimization with stochastic selection of perturbation-based chaotic updating system," *Appl. Math. Comput.* **269**, 904–929.
- Tokuda, I., Onodera, K., Tokunaga, R., Aihara, K. & Nagashima, T. [1998] "Global bifurcation scenario for chaotic dynamical systems that solve optimization problems and analysis of their optimization capability," *Electron. Commun. Japan* **81**, 1–12.
- Wang, L. & Smith, K. [1998] "On chaotic simulated annealing," *IEEE Trans. Neur. Netw.* **9**, 716–718.
- Wang, L., Tian, F., Soong, B. H. & Wan, C. [2011] "Solving combinatorial optimization problems using augmented Lagrange chaotic simulated annealing," *Diff. Eqs. Dyn. Syst.* **19**, 171–179.
- Yang, C.-H., Tsai, S.-W., Chuang, L.-Y. & Yang, C.-H. [2012] "An improved particle swarm optimization with double-bottom chaotic maps for numerical optimization," *Appl. Math. Comput.* **219**, 260–279.