Global Optimization Method Using Chaos of Discrete Inertial Gradient Dynamics

K. Masuda[1], E. Aiyoshi[1]

1 Faculty of Science and Technology, Keio University,
3-14-1 Hiyoshi, Kouhoku-ku, Hiyoshi, Yokohama, 223-8522, Japan
{kazuaki,aiyoshi}@sys.appi.keio.ac.jp

Abstract: In this paper, the optimization method based on "the chaotic annealing" applied to the trajectories generated in the discrete model of the inertial gradient dynamics is proposed. The new inertial gradient model is superior in (i) the easiness in the initial state configuration, and (ii) the assurance of the convergence to local optima. The efficiency is demonstrated through numerical simulations.

Keywords: Optimization, Inertial gradient model, Chaotic annealing

1. Introduction

In global optimization, there is a method using the high searching ability of the chaotic trajectories generated by the discrete gradient model[1, 2, 3, 4]. The global optimum is calculated with the aid of "the chaotic annealing" algorithm. It converges the trajectories to the global optimum by gradually stabilizing the dynamics from chaotic phase to descent phase.

On the other hand, the inertial gradient model is also used in global optimization because of the ability to overcome the unwilling local optima during the search[5].

In this paper, we consider to avail the both advantages in chaos and the inertial gradient model. We propose a new discrete inertial gradient models by simply modifying the existing non-inertial ones. At the same time, we also propose a little modefication of the chaotic annealing in order to avail it with the inertial models. They can solve two problems known in using the inertial gradient model at the same time, (i) the difficulty in the initial state configuration, and (ii) the assurance of the convergence to local optima.

2. Inertial Gradient Models for Constrained Optimization Problems

2.1 Model for Optimization Problem with Bounds Pair

We consider the constrained minimization problem with bounds pair

$$\min_{\boldsymbol{x}} \quad E(\boldsymbol{x}), \tag{1a}$$

$$\text{subj.to.} \quad x_i \in (p_i, q_i), \quad i = 1, \cdots, n, \tag{1b}$$

where $\boldsymbol{x} = [x_1, \cdots, x_n]$ is the decision variable, E is the continuous differenciable minimization function, and p_i, q_i are given real values correspondingly. This class of the optimization problems appear in various fields of engineering. In order to solve eq.(1), the steepest descent model with brake[1, 3)

$$\frac{dx_i(t)}{dt} = -(x_i(t) - p_i)(q_i - x_i(t))\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}. \tag{2}$$

Eq.(2) is equivalently expressed by introducing the inner state variable $\boldsymbol{u} = [u_1, \cdots, u_n]$,

$$x_i(t) = f_i(u_i(t)) \equiv \frac{q_i + p_i \exp(-u_i(t))}{1 + \exp(-u_i(t))}, \tag{3a}$$

$$\frac{du_i(t)}{dt} = -\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}. \tag{3b}$$

At the initial state $t = 0$, if $\boldsymbol{x}(0)$ is given, $\boldsymbol{u}(0)$ is calculated by the inverse transforms of eq.(3a). $\boldsymbol{x}$ stops at the equilibrium points in the constraints eq.(1b) where corresponding $\boldsymbol{u}$ satisfies $d\boldsymbol{u}/dt = \boldsymbol{0}$.

In this paper, we replace the gradient $\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}$ with the convoluted gradient $\int_0^t e^{-a(t-\tau)} \frac{\partial E(\boldsymbol{x}(\tau))}{\partial x_i} d\tau$ in (2)

$$\frac{dx_i(t)}{dt} = -(x_i(t) - p_i)(q_i - x_i(t)) \int_0^t e^{-a(t-\tau)} \frac{\partial E(\boldsymbol{x}(\tau))}{\partial x_i} d\tau, \tag{4}$$

where a is a real number. Because of the storage of past information, the convoluted gradient doesn't always get zero at the points which correspond to the local minima of E, where the simple gradient is zero. Eq.(4) also searches minima of eq.(1a), but can escape the unwilling minima thanks to the characteristic of the convoluted gradient. The inner state expression of eq.(4) is given in the way as

$$x_i(t) = f_i(u_i(t)) \equiv \frac{q_i + p_i \exp(-u_i(t))}{1 + \exp(-u_i(t))}, \tag{5a}$$

$$\frac{du_i(t)}{dt} = -\int_0^t e^{-a(t-\tau)} \frac{\partial E(\boldsymbol{x}(\tau))}{\partial x_i} d\tau. \tag{5b}$$

Then, the differencial equation (5) of the first order is transformed into the one of the second order

$$\frac{d^2 u_i(t)}{dt^2} + a\frac{du_i(t)}{dt} = -\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}, \tag{6}$$

which is named the inertial gradient model. Now, we introduce another variable $\boldsymbol{v} = [v_1, \cdots, v_n]$ so as to partition eq.(6) to two differencial equations of the first order

$$\frac{du_i(t)}{dt} + au_i(t) = v_i(t), \tag{7a}$$

$$\frac{dv_i(t)}{dt} = -\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}. \tag{7b}$$

The initial state of eq.(7) can be set in the following way. $\boldsymbol{u}(0)$ is the inverse of eq.(5a) if $\boldsymbol{x}(0)$ is given. Let the change $d\boldsymbol{u}(0)/dt$ be $\mathbf{0}$. Accordingly, due to eq.(7a), $\boldsymbol{v}(0) = a\boldsymbol{u}(0)$ is required.

The discrete model of eq.(5) and eq.(7) with the sampling time ΔT are derived by

$$x_i(t) = f_i(u_i(t)) \equiv \frac{q_i + p_i \exp(-u_i(t))}{1 + \exp(-u_i(t))}, \tag{8a}$$

$$u_i(t+1) = u_i(t) + \Delta T\{v_i(t) - au_i(t)\}, \tag{8b}$$

$$v_i(t+1) = v_i(t) - \Delta T \frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}. \tag{8c}$$

The dynamics eq.(8) gets chaotic as ΔT gets large due to the unstabilization of its equilibrium points[6].

2.2 Model for Optimization Problem Constrained on a Simplex

Then, we consider the constrained minimization problem constrained on the surface of a simplex

$$\min_{\boldsymbol{x}} \quad E(\boldsymbol{x}), \tag{9a}$$

$$\text{subj.to.} \quad \sum_{i=1}^{n} x_i = 1 \tag{9b}$$

$$x_i \geq 0, \quad i = 1, \cdots, n. \tag{9c}$$

Eq.(9b) and eq.(9c) define the inside of the polygon with n apexes $(1, 0, \cdots, 0)$, $\cdots$, $(0, \cdots, 0, 1)$, which is the surface of n-dimensional simplex. The problem eq.(9) is more difficult to solve than eq.(9) due to the existance of the relationship between each component of decision variable x_i. This class of the optimization problem may be applied for the matters of scheduling or resource allocation. If x_i is regarded as the possibility where state i appears, the possibility estimation problems may also be written in this form.

In order to solve eq.(9), the exclusive model for this class of the optimization problem, called "the variable metric gradient projection model", is known as below

$$\frac{dx_i(t)}{dt} = -x_i(t)\left\{\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i} - \sum_{j=1}^{n} x_j(t)\frac{\partial E(\boldsymbol{x}(t))}{\partial x_j}\right\}. \tag{10}$$

Its inner state expression, like eq.(3), is also given by[4]

$$x_i(t) = f_i(\boldsymbol{u}(t)) = \frac{\exp u_i(t)}{\sum_{j=1}^{n} \exp u_j(t)} \tag{11a}$$

$$\frac{du_i(t)}{dt} = -\frac{\partial E(\boldsymbol{x}(\tau))}{\partial x_i}. \tag{11b}$$

Here, in the same way as the previous section, we replace the gradient $\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}$ with the convoluted gradient $\int_0^t e^{-a(t-\tau)} \frac{\partial E(\boldsymbol{x}(\tau))}{\partial x_i} dt$ in (10)

$$\frac{dx_i(t)}{dt} = -x_i(t)\left\{\int_0^t e^{-a(t-\tau)} \frac{E(\boldsymbol{x}(\tau))}{\partial x_i} d\tau - \sum_{j=1}^{n} x_j(t) \int_0^t e^{-a(t-\tau)} \frac{\partial E(\boldsymbol{x}(\tau))}{\partial x_i} d\tau\right\}. \tag{12}$$

The inner state expression for eq.(12), the counterpart of eq.(4), is derived as below,

$$x_i(t) = f_i(\boldsymbol{u}(t)) = \frac{\exp u_i(t)}{\sum_{j=1}^{n} \exp u_j(t)} \tag{13a}$$

$$\frac{du_i(t)}{dt} = -\int_0^t e^{-a(t-\tau)} \frac{\partial E(\boldsymbol{x}(\tau))}{\partial x_i} d\tau. \tag{13b}$$

In comparison to eq.(5), we can note that eq.(13b) has all same form as eq.(5b), and that only difference lies in the form of f_i between eq.(13a) and eq.(5a). The f_i of eq.(13a) assures the constraints eq.(9b) and (9c) whatever value $\boldsymbol{u}$ takes.

Therefore, the transformations of equations originated from eq.(13) are done in the same way as those originated from eq.(5). In the process, the continuous model

$$x_i(t) = f_i(\boldsymbol{u}(t)) = \frac{\exp u_i(t)}{\sum_{j=1}^{n} \exp u_j(t)} \tag{14a}$$

$$\frac{du_i(t)}{dt} + au_i(t) = v_i(t), \tag{14b}$$

$$\frac{dv_i(t)}{dt} = -\frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}, \tag{14c}$$

is derived. In the long run, the discrete model for eq.(14) is described as below.

$$x_i(t) = f_i(\boldsymbol{u}(t)) = \frac{\exp u_i(t)}{\sum_{j=1}^{n} \exp u_j(t)}, \tag{15a}$$

$$u_i(t+1) = u_i(t) + \Delta T\{v_i(t) - au_i(t)\}, \tag{15b}$$

$$v_i(t+1) = v_i(t) - \Delta T \frac{\partial E(\boldsymbol{x}(t))}{\partial x_i}. \tag{15c}$$

The dynamics eq.(15) also gets chaotic as the enlargement of ΔT.

3. The Chaotic Annealing Method

The chaotic annealing is proposed to converge the chaotic trajectories into the global optimum. Regarding the sampling time in the discretized map of the gradient model as the temperature, whose concept is the same as that of the simulated annealing, a high temperature enough for the dynamical system to have chaotic characteristic is given during the earlier searching stage in order to generate various states as candidates for optima. Though, during the following searching stage, the temperature is gradually decreased

so as to stablize it and converge trajectories to the global optimum.

However, we must note that general chaotic annealing algorithms, which simply decrease the temperature according to certain cooling scheludes, don't always give the global optimum[2]. It can be because trajectories may be attracted more easily to certain local optima than the global optimum due to the inheritance of the chaotic characteristic of the dynamical system which generates states.

To clear the matter, we apply the modefied chaotic annealing method proposed by us which uses the threshold acception method together, called the hybrid type of the chaotic annealing method[3]. The newly generated state is accepted as a candidate for the optimum if the the increase of the value of the objective function at the new state from that of the currently held candidate is smaller than a given threshold value. Otherwize, the state is rejected for a candidate and another one is generated. In addition, we use the adaptive cooling schedule which decreases the temperature if such rejections continue to happen more than a given number of times C_1, or the generation of states at a temperature continues to happen more than a given number of times C_2. They are expected to effective for the improvement of the convergence rate and speed to the global optimum.

We show the detail algorithm as follows, noting that we describe the generated states by the discrete gradient models (eq.(8) or eq.(15)) as $\tilde{\boldsymbol{x}}$ and candidates for the optimum as $\boldsymbol{x}$.

[The algorithm of the hybrid type of the chaotic annealing]

Step 1 Set $t := 0$. Give the initial value of candidates $\boldsymbol{x}(0)$ and its corresponding inner states $\boldsymbol{u}(0)$, $\boldsymbol{v}(0) = a\boldsymbol{u}(0)$. Set the initial temperature $\Delta T := \Delta T_0$. The inital state is given by $\tilde{\boldsymbol{x}}(0) := \boldsymbol{x}(0)$, $\tilde{\boldsymbol{u}}(0) := \boldsymbol{u}(0)$, $\tilde{\boldsymbol{v}}(0) := \boldsymbol{v}(0)$. Set the decrease of the temperature in one cooling d, the threshold value for the difference of the objective function T, the maximum number of the continuous rejections C_1, and the maximum number of the state generations per one temperature C_2.

Step 2 Calculate a new state $\tilde{\boldsymbol{x}}(t+1)$ by substituting $\tilde{\boldsymbol{u}}(t)$, $\tilde{\boldsymbol{v}}(t)$, $\tilde{\boldsymbol{x}}(t)$ to the discrete gradient models and the value of the objective function $E(\tilde{\boldsymbol{x}}'(t+1))$.

Step 3 Define the difference between the values of the objective function as

$$\Delta E(t) = E(\tilde{\boldsymbol{x}}(t+1)) - E(\boldsymbol{x}(t)) \quad (16)$$

and accept the revision of candidate $\boldsymbol{x}(t+1) := \tilde{\boldsymbol{x}}(t+1)$ if $\Delta E(t) < T$. Otherwize, set $\boldsymbol{x}(t+1) := \boldsymbol{x}(t)$. The inner states $\boldsymbol{u}$ and $\boldsymbol{v}$ must also be renewed according to the change of $\boldsymbol{x}$.

Step 4 Cool the temperature $\Delta T := \Delta T - d$ if the number of the continuous rejections or the state generations per one temperature exceeds the bound.

Step 5 Finish searching when $\Delta T \leq 0$. Otherwize, set $t := t+1$, $\boldsymbol{v}(t) = a\boldsymbol{u}(t)$ and return to Step 2.

Note that we specially add the operation of revising $\boldsymbol{v}(t)$ in **Step 5**, even though $\boldsymbol{v}$ is once renewed in **Step 3**, just for the use of discrete inertial gradient dynamics. Through our investigation, applying the chaotic annealing to the inertial gradient dynamics, in the same way for non-inertial gradient dynamics, lacks the converge effect of $\boldsymbol{x}$ to local optima[7], so thet this operation is quite essential in order to converge $\boldsymbol{x}$ to optima.

4. Numerical Simulation

4.1 Results for Problems with Bounds Pair

In this section, we show the results of numerical simulations for some examples in order to evaluate the availability of newly proposing method. as an example, consider minimization function

$$E(x) = \frac{1}{50}(x-4.5)(x-3.8)(x-3)(x-1)(x+2) \cdot (x+3)(x+4)(x+5), \quad (17)$$

with bounds pair $x \in (-5.7, 5.7)$. The structure of eq.(17) is shown in Figure 1, and local optima are $x = -4.667$, -2.466, 1.925, 4.243. The global optimum is $x = 1.925$.

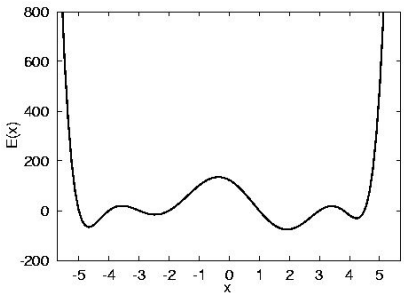

Figure 1: Structure of the minimization function eq.(17).

The global bifurcation diagram of for.(8) where eq.(17) is substituted as E is shown in Figure 2. It is drawn by plotting the passing points of the trajectories for various ΔTs initialized from the four local minima.

Then, we apply "the chaotic annealing" method, especially modefied for the discrete inertial gradient dynamics. The parameters of it is defined by $\Delta T_0 = 0.08$, $d = \Delta T_0/400$, $C_1 = 10$, $C_2 = 20$ and $T = 1.0$. For the inertial gradient model, in order to assure the convergence to local optima, reset of $\boldsymbol{v}$, $\boldsymbol{v} = a\boldsymbol{u}$, is done just after each decreases of ΔT. The annealing diagram, we call in this paper for convenience, is shown in Figure 3. It plots all samples $\boldsymbol{x}(t)$ as the candidates for the global optimum, in relation with ΔT during the annealing. We can see that all trajectories goes to either of the local optima through the annealing, which itself is the great advance thanks to the idea of resetting the inner state $\boldsymbol{v}$ after each change of ΔT.

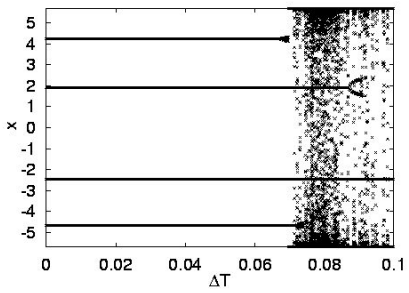

Figure 2: Bifurcation diagram of eq.(8) with eq.(17).

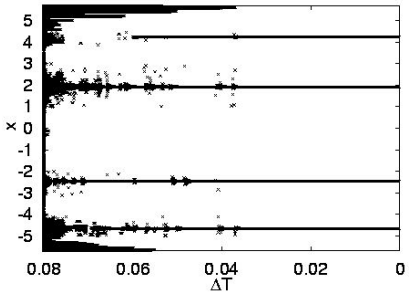

Figure 3: Annealing diagram of eq.(8) with eq.(17).

Besides, the convergence rate of the trajectories to either of the local optima, under the condition that whose initial points are given randomly can be used. The rate to the global optimum $x = 1.925$ from randomly defined initial points is 79.25%. The rates to local optima are, $x = -4.667$: 0.59%, $x = -4.667$: 19.68%, and $x = 4.243$: 0.06% respectively. 0.52% went to $x = -5.7$.

In the next, consider another minimization function named "Girewank",

$$E(\boldsymbol{x}) = -\cos(x_1)\cos(\frac{x_2}{\sqrt{2}}) + \frac{x_1^2 + x_2^2}{200}, \tag{18}$$

with bounds pair $x_1, x_2 \in (-25, 25)$, is considered. The structure of eq.(18) is shown in Figure 4, and the global optimum is $(x_1, x_2) = (0, 0)$. There also exist many local optima. The global bifurcation diagram for this problem, drawed by plotting the passing points of the trajectory initialized from the global minimum, is shown in Figure 5.

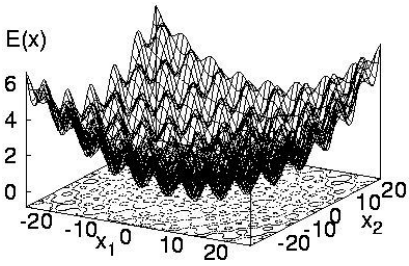

Figure 4: Structure of the minimization function eq.(18).

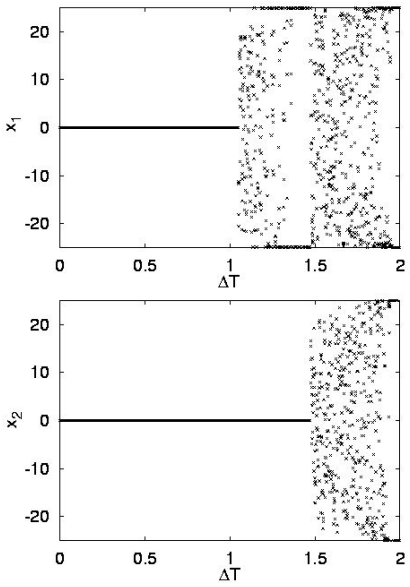

Figure 5: Bifurcation diagram of eq.(8) with eq.(18).

Under the configurations of parameters $\Delta T_0 = 1.8$, $d = \Delta T_0/600$, $C_1 = 10$, $C_2 = 20$ and $T = 0.001$, the bifurcation diagram is shown in Figure 6. In addition, the rate of convergence from randomly given initial points to the global optimum is 99.17%.

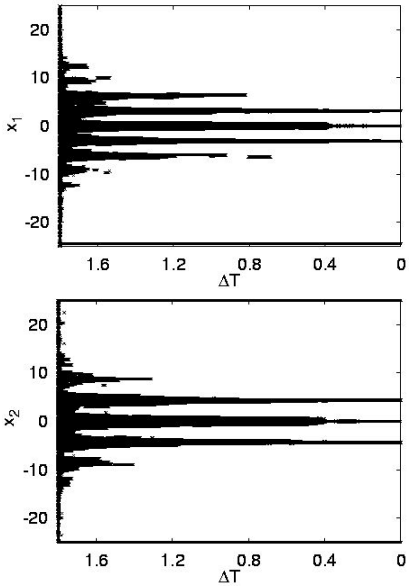

Figure 6: Annealing diagram of eq.(8) with eq.(18).

4.2 Result for a Problem on a Simplex

Consider a constrained problem on a simplex, whose minimization function E is named "Levy-Montalvo",

$$\min_{\boldsymbol{x}} \quad E(\boldsymbol{x}) = \frac{\pi}{n}\left\{B\sin^2(\pi y_1) + \sum_{i=1}^{n-1}(y_i - A)^2(1 + B\sin^2(\pi y_{i+1})) + (y_n - A)^2\right\}, \tag{19a}$$

$$y_i = 1.0 + 10.0(x_i - 0.25),$$

$$\text{subj.to} \quad \sum_{i=1}^{n} x_i = 1, \tag{19b}$$

$$x_i \geq 0, \quad i = 1, \cdots, n, \tag{19c}$$

where parameters are given by $A{=}1.0$, $B{=}5.0$.

Let $n = 4$ here. Though the structure of corresponding eq.(19a) is impossible to draw in a figure, it has so many local optima on a simplex. The global optimum is $(x_1, x_2, x_3, x_4) = (0.25, 0.25, 0.25, 0.25)$.

The global bifurcation diagram for this problem, with the trajectory initialized from the global minimum, is shown in Figure 8.

The annealing diagram for the problem eq.(19) is shown in Figure 8. The configurations of parameters are, $\Delta T_0 = 0.003$, $d = \Delta T_0/500$, $C_1 = 10$, $C_2 = 20$ and $T = 1.0$. But the convergence effect seems to be insufficient, and the convergence rate to the global optimum goes up to only 22.63%.

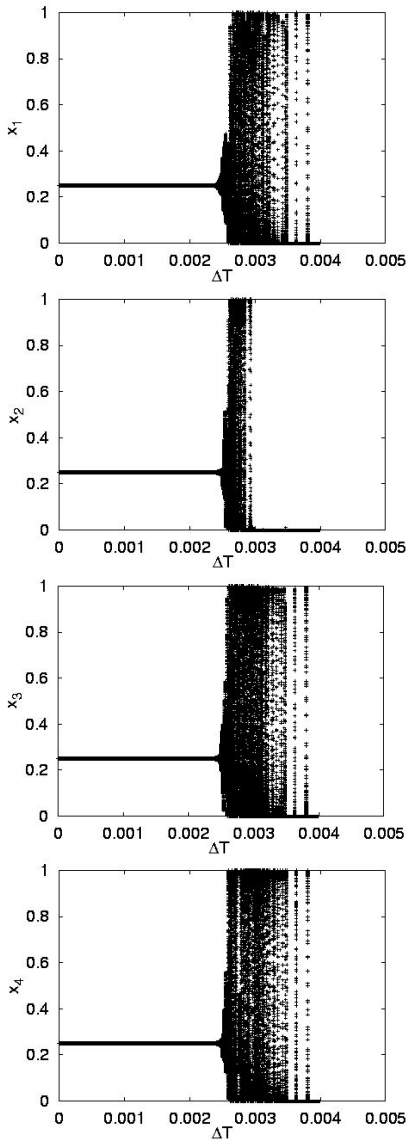

Figure 7: Bifurcation diagram of eq.(15) for the problem eq.(19).

5. Conclusion

In this paper, we proposed a new discrete inertial gradient dynamics with a little modefication of non-inertial dynamics. They are expected to to overcome the unwilling local optima during the global search. We also solved the the difficulty in the initial state configuration and the assurance of the convergence to local optima.

Then, we considered to avail the chaotic characteristic generated from the discrete model of it for global optimation. We used the chaotic annealing method, but we proposed a special modefication of the chaotic annealing in order to use it with the inertial gradient dynamics in this paper. It solved the problem we have found before that the decision variable doesn't converge to any of the local optima if the chaotic annealing is applied to the inertial gradient dynamics in the same way as to the non-inertial ones.

In the numerical simulations section, we demonstrated the distinguishing characteristics of the chaos generated from our proposing model. Such were the great progresses that the modefied chaotic annealing method applied to the discrete inertial gradient dynamics shown to be available for global optimization. On the other hand, judging from the comparison with our previous results in which the chaotic annealing was used with discrete non-inertial gradient dynamics[3, 4] or the convergence rate to the global optimum, we found it still difficult to argue in this paper that the use of chaos generated in the discrete inertial dynamics gives significant advantages.

In addition, we found a great difficulty in availing the chaos of the discrete inertial dynamics for global optimization on a simplex, which will be a new task for us to solve.

References

[1] H.Sugata and K.Shimizu, Global Optimization Using Chaos in a Quasi-Steepest Descent Method, IEICE Transactions (A), Vol. J79-A, No. 3, pp. 658–668, 1996 (in Japanese)

[2] I.Tokuda, K.Onodera, R.Tokunaga, K.Aihara and T.Nagashima, Global Bifurcation Scenario for Chaotic Annealing Dynamical System which Solves Optimization Problem and Analysis on Its Optimization Capability, IEICE Transactions (A), Vol. J80-A, No. 6, pp. 936–948, 1996 (in Japanese)

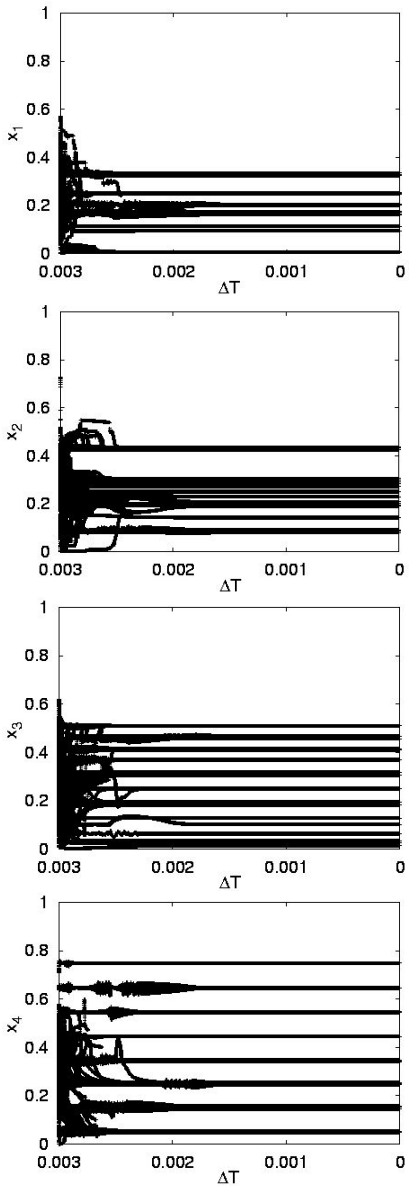

Figure 8: Annealing diagram of eq.(15) for the problem eq.(19).

[3] K.Masuda and E.Aiyoshi, Hybrid Type of Global Optimization Method with Discretized Chaos Mappings and Increase Accepting Methods, T.IEE Japan, Vol. 122-C, No. 5, pp. 892–899, 2002 (in Japanese)

[4] K.Masuda and E.Aiyoshi, Chaotic Dynamics on a Simplex and Global Optimization Method with Normalized Constraints, IEEJ. Trans. EIS., Vol. 123-C, No. 6 (in printing), 2003 (in Japanese)

[5] T.Okamoto, K.Masuda and E.Aiyoshi, Constrained Optimization by an Inertia Model on a Simplex – Optimization by Dissipative Chaos Dynamics on a Simplex –, SICE Symp. on Systems and Imformation, pp. 121–124, 2002 (in Japanese)

[6] M.Hata, Euler's Finite Difference Scheme and Chaos in R^n, Proc. Japan. Acad., vol. 58(A), pp. 178–181, 1982

[7] K.Masuda and E.Aiyoshi, Hybrid Type of Global Optimization Methods with Discretized Chaotic Mappings and Accepting Methods, SICE Symp. on Systems and Imformation, pp. 31–36, 2001 (in Japanese)