Evaluation of the physical annealing strategy for simulated annealing: A function-based analysis in the landscape paradigm

M. Hasegawa
Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba 305-8573, Japan
(Received 31 January 2012; published 22 May 2012)

The effectiveness of the actual annealing strategy in finite-time optimization by simulated annealing (SA) is analyzed by focusing on the search function of the relaxation dynamics observed in the multimodal landscape of the cost function. The rate-cycling experiment, which was introduced in the previous study [M. Hasegawa, Phys. Rev. E **83**, 036708 (2011)] to examine the role of the relaxation dynamics in optimization, and the temperature-cycling experiment, which was developed for a laboratory experiment on relaxation-related phenomena, are conducted on two types of random traveling salesman problems (TSPs). In each experiment, the SA search starting from a quenched solution is performed systematically under a nonmonotonic temperature control used in the actual heat treatment of metals and glasses. The results show that, as in the previous monotonic cooling from a random solution, the optimizing ability is enhanced by allocating a lot of time to the search performed near an effective intermediate temperature irrespective of the annealing technique. In this productive phase, the relaxation dynamics successfully function as an optimizer and the relevant characteristics analogous to the stabilization phenomenon and the acceleration of relaxation, which are observed in glass-forming materials, play favorable roles in the present optimization. This nonmonotonic approach also has the advantage of a wider operation range of the effective relaxation dynamics, and in conclusion, the actual annealing strategy is useful and more workable than the conventional slow-cooling strategy, at least for the present TSPs. Further discussion is given of an illuminating aspect of computational physics analysis in the optimization algorithm research.

DOI: 10.1103/PhysRevE.85.056704 PACS number(s): 02.60.Pn, 05.10.−a, 89.20.−a, 81.40.Ef

I. INTRODUCTION

Various research topics related to both physics and computer science have attracted many researchers during the last few decades. Optimization is one of these interdisciplinary topics, and several subjects have been presented and studied according to their individual interests. As examples, we can cite the analysis of the physically relevant properties of combinatorial optimization problems [1–3], the development of heuristic optimization algorithms inspired by physical concepts and methods [4–10], and the application of general-purpose optimization methods to the solution of problems of physical interest [11,12]. These kinds of studies have evolved into an attempt to create a mutually beneficial relationship between the two communities [11] or to unify the understanding of the related complex systems on an equal footing [13–15]. Whereas these challenges have stimulated interdisciplinary interest, their progress is gradual; therefore, we still need further efforts to acquire precise knowledge of the relationship between complex physical systems and hard optimization problems.

In a recent study [16], we reconsidered the physical analogy of simulated annealing (SA) [17,18], which is one of the heuristic optimization algorithms and was proposed in the early days of this growing interdisciplinary field. An optimization problem is usually formulated as a task of finding a solution minimizing a given cost function, and in the application of SA, the Metropolis algorithm (MA) [19] for equilibrium sampling used in the Monte Carlo method is employed with a slow decrease in temperature by replacing the feasible solution and its cost with the microscopic state and its energy, respectively. The previous study was motivated by the fact that a correct understanding of the function of SA appeared to be hampered by a misguided physical analogy. (For a summary of the relevant background, see the introduction in Ref. [16].) To find a precise connection between optimization dynamics and physical phenomena, the search trajectory was analyzed in a multimodal landscape of the cost function and its behavior was compared to the dynamics of glass-forming molecular systems in a rugged landscape of the energy surface. The analysis was performed for the solution of the traveling salesman problem (TSP) [20,21], which is one of the hard optimization problems and has been well studied in both communities. The results showed that, contrary to the conventional scenario, the effectiveness of SA comes not from equilibrium sampling at low temperatures but from the search function of the relaxation dynamics occurring before equilibrium. These dynamics effectively work as an optimizer by slow cooling crossing an intermediate temperature analogous to the glass transition temperature, and therefore the promising design of the temperature schedule can be learned from analogy to vitrification phenomena, at least for the random TSPs considered there. These findings imply that the physical analogy still serves the purpose of optimization and that the actual annealing strategy used in the heat treatment of metals [22,23] and glasses [24] is expected to be useful also in the finite-time implementation of SA. From the present interdisciplinary interest, it should be necessary to verify this unnoticed consistency between simulated and actual annealing.

We attempt here to do just that: The effectiveness of nonmonotonic temperature schedules taking advantage of an effective temperature is examined for the solution of the same optimization problems. Although similar approaches to optimization have been discussed in some earlier papers [25,26], we seek to obtain well-founded knowledge of the optimization mechanism from analogy to the behavior of

glass-forming systems. For this purpose, the rate-cycling experiment, which was introduced in the previous study [16], and the temperature-cycling experiment, which has been utilized in glass research [27], are conducted numerically in a coordinated manner. The reason we adopt these experiments, the details of which are described in the next section, is that each framework of the temperature schedule is adjustable to fit an actual annealing scheme and is convenient for systematic investigation. Another reason for the application of the temperature-cycling experiment is that since it has been widely used in studies on liquid and glass [27–30], complementary studies on hard optimization problems will also be helpful in fertilizing the integrated research field [31].

As in the previous study, we examine the hidden search dynamics (the transition process among the basins of attraction in the solution space) as well as the optimization performance because the former reflects the relaxation dynamics and directly affects the latter. From the results of both experiments, we evaluate the effectiveness of the actual annealing strategy and elucidate how each annealing technique works in the implementation of the algorithm. As shown in Sec. III, the performance depends primarily on the search performed near an effective intermediate temperature irrespective of the annealing technique. In addition, various behaviors analogous to relaxation-related phenomena are clearly observed in the hidden search dynamics and some of them appear to work profitably in finite-time optimization by SA. A discussion of these and other findings gives us a fresh insight into the optimization algorithm research in the landscape paradigm.

For the purpose here, both experiments were designed for a better understanding of the optimization characteristics without a direct attempt to propose new improvements of the algorithm. For this reason, the algorithm was sometimes employed under conditions unsuited for practical use. The remainder of this paper is organized as follows: In Sec. II, we give the definition of the TSPs used in the present study and explain the methods of the two experiments in detail. In Sec. III, the results of these experiments are presented in an organized way. In Sec. IV, we discuss the effectiveness of the actual annealing strategy in finite-time optimization by SA and further the potential of the function-based analysis of the search algorithm. In Sec. V, we conclude our study with a brief remark on the practical implementation.

II. EXPERIMENTAL METHODS

The TSP [20,21] is a combinatorial optimization problem to find the shortest tour that passes through each of the given cities once and returns to the starting city. We consider the two types of random problems used as test beds in the previous study [16]: (i) the random Euclidean TSP (RE-TSP), where the locations of N cities are sampled uniformly in the unit square region and the intercity distance is computed under the Euclidean metric; and (ii) the random distance matrix TSP (RD-TSP), where the intercity distances between N cities are sampled uniformly in the unit interval. Even in the latter type, we consider only the symmetric case, namely, that the intercity distance does not depend on the order of two cities.

We introduce here some notation and terminology, which are the same as used in Ref. [16]. Let x be a tour (a feasible solution), $f(x)$ be the cost function defined by the tour length, and $\mathcal{N}(x)$ be the neighborhood function. Throughout the study, we use a 2-opt neighborhood [21], which is defined as a set of tours constructed by any change of two intercity paths from the tour x. We write x_n $(n = 0,1,2,\ldots)$ to represent the (actual) search history and x^* to represent the incumbent solution (the best-so-far solution). A locally optimal solution is a solution that does not have any improved solution in its neighborhood. A basin is defined as a set of solutions attracted to the same locally optimal solution by a simple local search, that is, a repetition of the move to the best solution in the neighborhood. Let $y(x)$ be the locally optimal solution inside the basin including the solution x and we refer to its cost $f(y(x))$ as the cost of the basin. [Obviously, $y(x)$ can be found by a simple local search starting from the solution x; this identification method is analogous to the mapping-onto-minima approach used in studies on liquid and glass [32–34]. The temperature is denoted by T; however, we often use the logarithmic temperature Θ $(=\log_{10} T)$.

Next, the methods of the two experiments are described in detail. In each experiment, the SA algorithm is implemented with a nonmonotonic temperature schedule.

A. Rate-cycling experiment

In the rate-cycling experiment, the effect of the search performed near a specified temperature is examined by changing the rate of temperature variation cyclically. To emulate the strategy used in the actual heat treatment, the previous framework of a monotonic temperature schedule is extended so that we can replicate the annealing process beginning with a heating stage starting from a quenched solution. In what follows, the rate-cycling process beginning with a heating stage is referred to simply as h-annealing and that beginning with a cooling stage (starting from a random solution) as c-annealing, and both annealing procedures are considered here. The initial quenched solution for h-annealing is taken to be a locally optimal solution generated by a simple local search from a random solution. Although h-annealing starting from a quenched solution is more analogous to another heat treatment known as tempering, we use the former name throughout the paper to avoid excessive terminology.

The framework of the temperature schedule is constructed in the following manner. First, an initial logarithmic temperature Θ_s, a final logarithmic temperature Θ_e, and a total search time t_e (or a total number of search steps, n_e) are selected. Then the whole process is divided into three stages so that the cooling in the intended logarithmic temperature range, $\Theta_c + \frac{1}{2}\Delta\Theta > \Theta > \Theta_c - \frac{1}{2}\Delta\Theta$, is performed as its second stage. We refer to the middle temperature Θ_c as the target (logarithmic) temperature and the above logarithmic temperature range simply as the target range. In c-annealing, as in the previous study, the cooling rate on the logarithmic temperature scale is selected and fixed at each stage so that the rates in the first and third stages are equal and are λ times the rate in the second stage. In h-annealing, the construction is the same as in the above c-annealing except that the heating rate is selected and fixed in the first stage and is set equal to the cooling rate in the third stage. In this framework, the six parameters, Θ_s, Θ_e, Θ_c, $\Delta\Theta$ (the logarithmic temperature

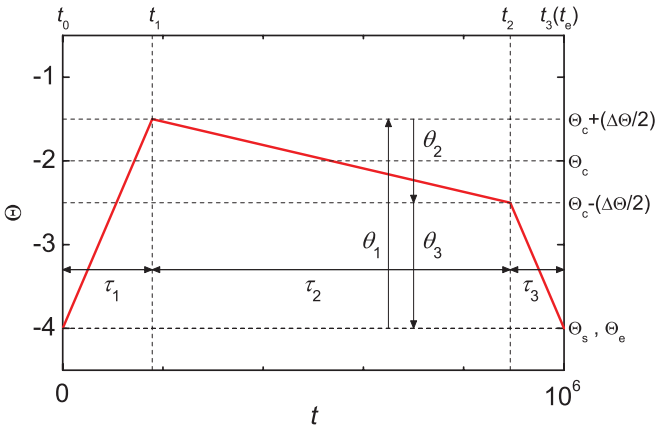

FIG. 1. (Color online) An example of the temperature schedule $\Theta(t)$ used in the rate-cycling experiment; depicted is the case for h-annealing ($\Theta_s = \Theta_e = -4$, $\Theta_c = -2.0$, $\Delta\Theta = 1.0$, $\lambda = 10$, and $t_e = 10^6$). See text for symbols.

width of the target range), λ (the rate ratio), and t_e ($=n_e$), uniquely determine the temperature schedule irrespective of the annealing procedure.

For experimental convenience, we use the following expression of the logarithmic temperature schedule in the kth stage:

$$\Theta(t) = \Theta(t_{k-1}) + \frac{t - t_{k-1}}{\tau_k}\,\theta_k \quad (t_{k-1} \leqslant t \leqslant t_k), \qquad (1)$$

where t_{k-1} and t_k are the beginning and the ending time of this stage, respectively, and θ_k and τ_k are the variation in logarithmic temperature ($\theta_k > 0$ at the heating stage and $\theta_k < 0$ at the cooling stage) and the time allocated in this stage, respectively; hence t_k equals the accumulated time of τ_j's up to $j = k$ ($t_0 = 0$). An example of the temperature schedule is depicted in Fig. 1 for h-annealing. The values of θ_k's are determined by Θ_s, Θ_e, Θ_c, and $\Delta\Theta$, and those of τ_k's are determined by θ_k's, λ, and t_e by the following equations: $\tau_1 = (s\theta_1/\bar{\theta})t_e$, $\tau_2 = (-\lambda\theta_2/\bar{\theta})t_e$, and $\tau_3 = (-\theta_3/\bar{\theta})t_e$, where $s = \mathrm{sgn}(\theta_1)$ and $\bar{\theta} = s\theta_1 - \lambda\theta_2 - \theta_3$. [Note that in c-annealing ($s = -1$), the present θ_k and $\bar{\theta}$ are identical to the previous $-h_k$ and H [16], respectively.]

In both annealing procedures, the time allocated in the second stage increases with the value of λ. Note that the temperature schedule of c-annealing with $\lambda = 1$ is equivalent to the geometric cooling schedule, which has generally been adopted in the baseline implementation of SA. In the present study, we use this unbiased c-annealing as a reference and the optimization characteristics of each annealing procedure are evaluated in comparison with those of this reference case.

The algorithm of this experiment is summarized as follows.

[RC1] Determine the logarithmic temperature schedule $\Theta(t)$; that is, select the values of the six parameters, Θ_s, Θ_e, Θ_c, $\Delta\Theta$, λ, and t_e ($=n_e$).

[RC2] Set $n := 0$ and generate an initial solution x_0 (a quenched solution for h-annealing and a random solution for c-annealing). Set $\Theta := \Theta(0)$ and $x^* := x_0$.

[RC3] Implement a single step of the MA and increment n by 1. (At this point, the current solution is x_n and the current incumbent solution is x^*.)

[RC4] If the termination condition, $n = n_e$, is satisfied, output x^* and stop. Otherwise, set $\Theta := \Theta(n)$ and return to [RC3].

In a single step of the MA in [RC3], the following [MA1] is implemented (the logarithmic temperature Θ has to be converted to the temperature T in advance).

[MA1] Choose a trial solution $x'_n \in \mathcal{N}(x_n)$ randomly and set $\Delta := f(x'_n) - f(x_n)$. If $\Delta < 0$, accept the trial solution and set $x_{n+1} := x'_n$; furthermore, if $f(x_{n+1}) < f(x^*)$, renew the incumbent solution, $x^* := x_{n+1}$. If $\Delta \geqslant 0$, accept the trial solution with the probability $\exp(-\Delta/T)$ and set $x_{n+1} := x'_n$; with the complementary probability, reject the trial solution and set $x_{n+1} := x_n$.

The experimental conditions were determined by taking into account the characteristic temperatures found in the previous study [16]. The parameter values were selected in the following way: In h-annealing, the values of Θ_s and Θ_e were equally fixed at -4 and the values of Θ_c were selected to cover the whole interesting temperature range found previously. The value of $\Delta\Theta$ was selected from the range from 0.2 to 1.8 and the value of λ was selected from $\{1, 10, 100\}$. Several values of t_e (or n_e) were chosen from a wide range to evaluate the scalability of the algorithm. In c-annealing, the conditions were the same as those for the above h-annealing except that the value of Θ_s was fixed at 1 instead of -4.

B. Temperature-cycling experiment

The same effect as that considered in the rate-cycling experiment is also examined in the temperature-cycling experiment, which is performed by changing the temperature itself cyclically. As in the h-annealing in the former experiment, a quenched solution is used for the initial solution.

The framework of the temperature schedule is constructed in the following manner. First, the target logarithmic temperature Θ_c, the logarithmic temperature width of the target range $\Delta\Theta$, and the total number of search steps, n_e, are selected. Then the length of one cycle, $2L$, and the number of cycles, M, are selected so as to satisfy the inequality $2ML < n_e$. The logarithmic temperature is fixed at $\Theta_c - \frac{1}{2}\Delta\Theta$ during the first half of the cycle and at $\Theta_c + \frac{1}{2}\Delta\Theta$ during the last half, and the cycle starts at the beginning of the search process and repeats M times. After the final cycle, the logarithmic temperature is returned to and kept fixed at $\Theta_c - \frac{1}{2}\Delta\Theta$ during the rest of the process. The temperature schedule in this experiment is thus determined by the five parameters, Θ_c, $\Delta\Theta$, L, M, and n_e. An example of the schedule is depicted in Fig. 2. To determine the influence of the presence or absence of the cycling, we also consider the process without cycling ($M = 0$), which is nothing other than the MA search performed at a lower temperature ($\Theta_c - \frac{1}{2}\Delta\Theta$) starting from a quenched solution.

The algorithm of this experiment is summarized as follows.

[TC1] Select the values of the five parameters, Θ_c, $\Delta\Theta$, L, M, and n_e.

[TC2] Set $n := 0$ and generate an initial solution x_0 (a quenched solution). Set $x^* := x_0$. If $M = 0$, go to [TC4]. Otherwise, set $m := 0$.

[TC3] Repeat the following lines until the condition $m = M$ is satisfied: Set $\Theta := \Theta_c - \frac{1}{2}\Delta\Theta$ and repeat [TC5] until the condition $n = (2m + 1)L$ is satisfied; set $\Theta := \Theta_c + \frac{1}{2}\Delta\Theta$

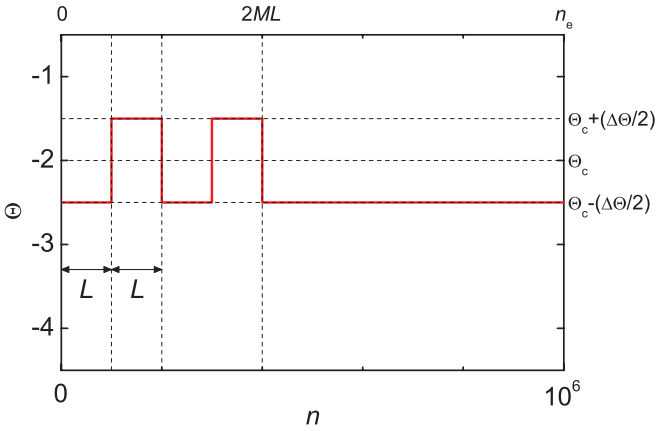

FIG. 2. (Color online) An example of the temperature schedule used in the temperature-cycling experiment ($\Theta_c = -2.0$, $\Delta\Theta = 1.0$, $M = 2$, and $n_e = 10L = 10^6$). See text for symbols.

and repeat [TC5] until the condition $n = (2m + 2)L$ is satisfied; increment m by 1.

[TC4] Set $\Theta := \Theta_c - \frac{1}{2}\Delta\Theta$ and repeat [TC5] until the termination condition, $n = n_e$, is satisfied. Output x^* and stop.

[TC5] (Before repeating this line, the logarithmic temperature Θ has to be converted to the temperature T.) Implement a single step of the MA, namely, [MA1] and increment n by 1. (At this point, the current solution is x_n and the current incumbent solution is x^*.)

The experimental conditions were chosen according to those adopted in the above rate-cycling experiment: The values of Θ_c, $\Delta\Theta$, and n_e were the same as in the rate-cycling experiment. The value of L was taken to be $n_e/10$, which was always an integer number in the present experiment, and hence the value of M was selected from non-negative integers not exceeding 4.

C. Additional remarks

In both experiments, the optimization performance evaluated by $f(y(x^*))$ and the hidden search dynamics described by $f(y(x_n))$ were investigated for various combinations of the parameter values. The cost variation along the actual search history, $f(x_n)$, which here is called the actual search dynamics, is sometimes referenced in the next section. The experiments were conducted in the following manner. First, each experiment was performed for a single instance for each type of problem, namely, RE-TSP and RD-TSP; three system sizes, $N = 100$ $(=10^2)$, $N = 316$ $(\approx 10^{2.5})$, and $N = 1000$ $(=10^3)$, were considered here. The average behavior was observed over I independent runs, where the value of I was taken to be 2^4 for $N = 100$, 2^3 for $N = 316$, and 2^2 for $N = 1000$. After surveying all the results and characterizing them, each experiment was repeated only for $N = 316$ to confirm the first findings. This confirmation experiment was done for two different instances for each type of problem with a larger I $(=2^5)$.

III. RESULTS

In this section, the results of the two experiments are presented in an organized way. Although the results are shown only for the first instance of the RE-TSP in the confirmation experiment ($N = 316$), the characteristic features described below were similarly observed in the other cases (which include the cases of RD-TSP). Throughout this confirmation experiment, the values of Θ_c were taken from -2.8 to -1.2 in increments of 0.1, and the values of $\Delta\Theta$ and n_e were selected from $\{0.2, 0.6, 1.0, 1.4, 1.8\}$ and $\{10^5, 10^6, 10^7\}$, respectively. All results are averages over I $(=2^5)$ independent runs, which are not described each time in the text below; it was confirmed that all I initial solutions differ from one another even if they are prepared by quenching.

A. Rate-cycling experiment

We begin with the results of the rate-cycling experiment. In the previous study [16], we found that the standardized SA with the geometric cooling schedule can be improved within the framework of monotonic cooling by taking the search function of the relaxation dynamics appropriately into account. As shown just below, h-annealing in the present experiment can also enhance the optimizing ability of SA through the same mechanism. Figure 3 shows the optimization performance of h-annealing on a long time scale ($n_e = 10^7$) for the case $\Delta\Theta = 1.0$. We find that the performance is improved by slow cooling crossing an intermediate temperature. The most effective temperature is close to that found in c-annealing and the resulting performance is comparable to that of the latter cooling-only process. (These are reported in Table I.) In Fig. 4, the hidden search dynamics for various Θ_c's are plotted for the case $\lambda = 100$ in Fig. 3. The performance is maximized when the relaxation dynamics observed as downward interbasin transitions work most successfully on the experimental time scale.

In this relaxation process, as predicted previously [16], a behavior analogous to the stabilization phenomenon observed in glass-forming materials [24] occurs over a wide range of parameter settings. Figure 5 shows the actual and hidden search dynamics for the best parameter setting for h-annealing. In this figure, the result for c-annealing with the same values of Θ_c, $\Delta\Theta$, and λ and that for the reference SA are also depicted for

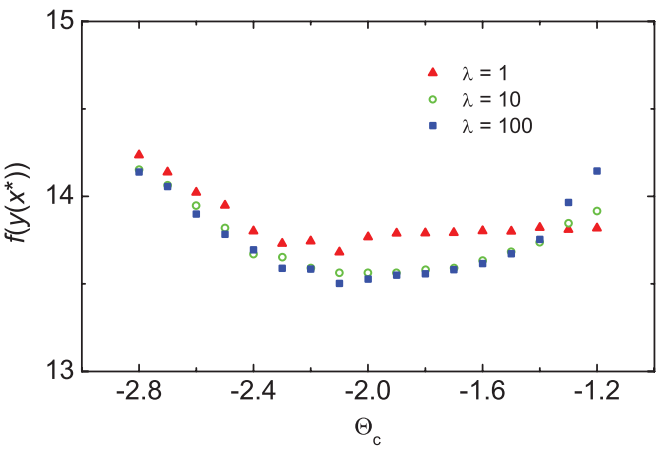

FIG. 3. (Color online) Optimization performance $f(y(x^*))$ (RE-TSP with $N = 316$; $\Theta_s = \Theta_e = -4$, $\Delta\Theta = 1.0$, $n_e = 10^7$, and $I = 2^5$). Results are averaged over I search processes and plotted against the target logarithmic temperature Θ_c.

TABLE I. Optimization performance $f(y(x^*))$ (RE-TSP with $N = 316$; $I = 2^5$). The best average result among the 15 cases (see text for details) is listed for each target logarithmic temperature Θ_c on each experimental time scale n_e. The corresponding values of $\Delta\Theta$ and λ are also listed in parentheses in this order. The very best result among all Θ_c's (ranging from -2.8 to -1.2 in increments of 0.1) on each n_e is listed in boldface. For comparative purposes, the average result for the reference SA (c-annealing with $\lambda = 1$; $\Theta_s = 1$ and $\Theta_e = -4$) is listed in the bottom row. All results are averages over I search processes.

	n_e		
Θ_c	10^5	10^6	10^7
	h-annealing ($\Theta_s = \Theta_e = -4$)		
−1.4	14.519	14.154	13.629
	(0.2,1)	(1.8,100)	(1.8,100)
−1.7	14.507	13.978	13.536
	(0.2,1)	(0.6,100)	(1.4,100)
−1.9	14.484	13.872	13.521
	(0.2,10)	(0.6,100)	(0.6,100)
−2.0	14.471	**13.866**	13.513
	(0.6,1)	(0.6,100)	(0.6,100)
−2.1	14.443	13.908	**13.502**
	(1.0,100)	(1.0,10)	(1.0,100)
−2.2	**14.407**	13.952	13.554
	(0.6,100)	(1.4,100)	(1.4,10)
−2.4	14.408	14.009	13.564
	(1.0,100)	(1.4,100)	(1.8,100)
−2.7	14.440	14.097	13.663
	(1.8,100)	(1.8,10)	(1.8,100)
	c-annealing ($\Theta_s = 1$ and $\Theta_e = -4$)		
−1.4	**14.492**	14.144	13.593
	(1.0,100)	(1.4,100)	(1.8,100)
−1.7	14.586	13.954	13.554
	(0.2,100)	(0.6,100)	(1.4,100)
−1.9	14.620	**13.902**	13.509
	(1.0,100)	(0.6,100)	(0.6,100)
−2.0	14.641	13.907	**13.501**
	(1.0,100)	(0.6,100)	(0.6,100)
−2.1	14.658	14.013	13.531
	(0.6,10)	(1.4,100)	(1.0,100)
−2.2	14.678	14.038	13.542
	(1.8,100)	(1.4,100)	(1.4,100)
−2.4	14.668	14.118	13.582
	(0.2,10)	(1.8,100)	(1.8,100)
−2.7	14.707	14.305	13.684
	(0.2,10)	(1.8,100)	(1.8,100)
Reference SA	14.749	14.481	13.726

comparative purposes, and all the results are plotted against the logarithmic temperature Θ instead of the time step n. As clearly shown in Fig. 5(b), the search trajectory starting from the initial quenched solution does not return in the heating stage to the original random solution, which should be found in the same basin, but the trajectory progresses into low-lying basins in the final phase of the heating stage and the subsequent slow-cooling stage. It should be noted that for this parameter setting, the progress of the stabilization-like behavior, which seems to be beneficial to the present optimization, is first recognized in detail in the hidden search dynamics [cf. the actual search dynamics depicted in Fig. 5(a)]. On this time scale, overall, the search dynamics of h-annealing and that of the corresponding c-annealing were close in the cooling process; therefore, there appeared to be no significant difference in the optimization performance of these two annealing procedures regardless of the value of Θ_c.

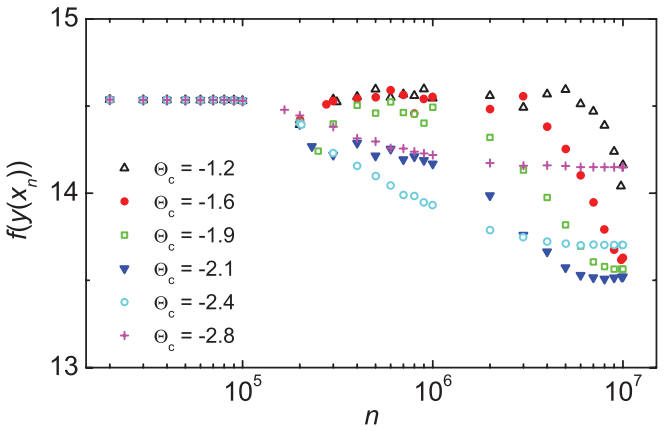

FIG. 4. (Color online) Hidden search dynamics $f(y(x_n))$ (RE-TSP with $N = 316$; $\Theta_s = \Theta_e = -4$, $\Delta\Theta = 1.0$, $\lambda = 100$, $n_e = 10^7$, and $I = 2^5$). Points are the results at some representative time steps. All results are averages over I search processes.

These features were commonly observed if the experimental time was long enough. Looking over the results for various time scales, however, we see that the situation varies with the total search time. The results of the performance on different time scales are listed in Table I for both annealing procedures. This table was constructed in the following manner. First, for each annealing procedure, the best average result among all possible combinations of five values of $\Delta\Theta$ and three values of λ was chosen for each Θ_c on each n_e. Then the chosen result was listed with the corresponding values of $\Delta\Theta$ and λ. For comparative purposes, the average result of the reference SA is listed in the bottom row. We see from this table that, as mentioned above, the performance is almost the same in the two procedures on a long time scale ($n_e = 10^7$). On shorter time scales, however, a relative superiority of h-annealing is found at low target temperatures. The search dynamics of such a case are shown in Fig. 6 for $n_e = 10^6$. We find a separation between the trajectories of the two procedures not only in the actual search dynamics [Fig. 6(a)] but also in the hidden search dynamics [Fig. 6(b)], which is probably due to an incomplete system response to the temperature variation. Similar behaviors are observed over a wider range of parameter settings on a much shorter time scale. As an example, the result for the best parameter setting for h-annealing for $n_e = 10^5$ is shown in Fig. 7. The two trajectories deviate significantly from each other in the actual search dynamics [Fig. 7(a)], and curiously, roughly upward transitions appear in the hidden search dynamics for both c-annealing and the reference SA [Fig. 7(b)]. In these c-annealings, the incumbent solution x^* is still renewed in the final phase of the search process; therefore, the optimization performance evaluated by $f(y(x^*))$ deteriorates beyond that of the simple local search. [Note that the average performance of the simple local search is shown in Figs. 5–7 as the value of $f(y(x_0))$'s, which is also equal to

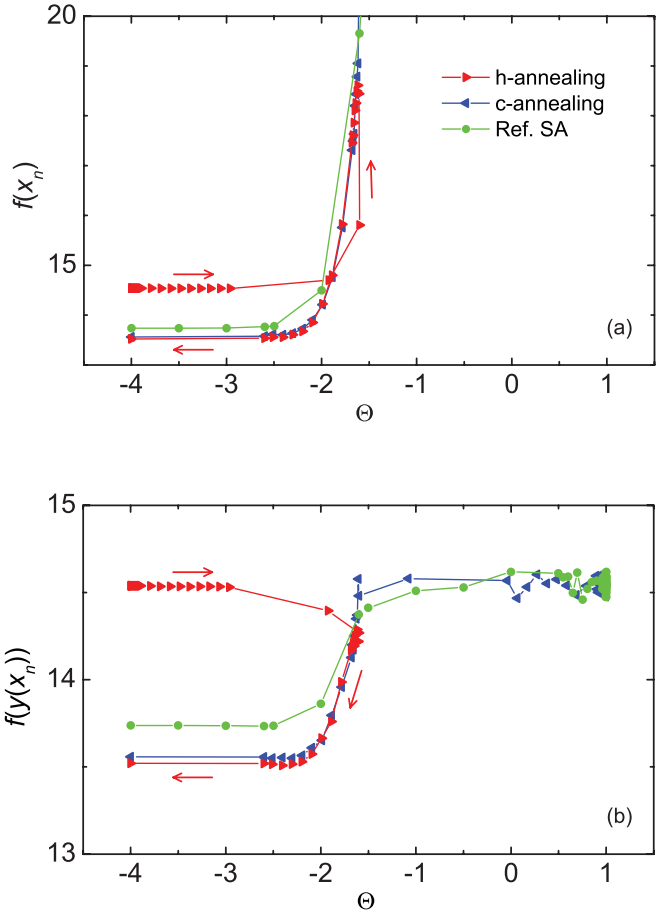

FIG. 5. (Color online) Actual [(a) $f(x_n)$] and hidden [(b) $f(y(x_n))$] search dynamics (RE-TSP with $N = 316$; $\Theta_c = -2.1$, $\Delta\Theta = 1.0$, $\lambda = 100$, $n_e = 10^7$, and $I = 2^5$). Depicted are the results for h-annealing ($\Theta_s = \Theta_e = -4$), c-annealing ($\Theta_s = 1$ and $\Theta_e = -4$), and the reference SA (c-annealing with $\lambda = 1$; $\Theta_s = 1$ and $\Theta_e = -4$). Note that the abscissa represents the logarithmic temperature Θ instead of the time step n. Points are the results at some representative time steps and are connected by line segments along the search trajectory. Arrows indicate the direction of the time evolution for h-annealing. All results are averages over I search processes.

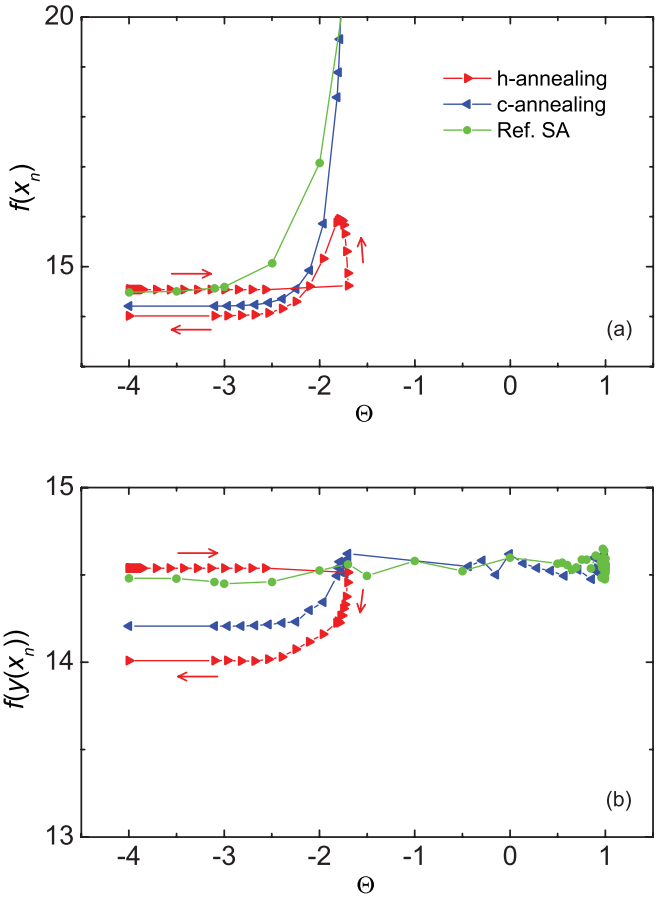

FIG. 6. (Color online) Actual [(a) $f(x_n)$] and hidden [(b) $f(y(x_n))$] search dynamics (RE-TSP with $N = 316$; $\Theta_c = -2.4$, $\Delta\Theta = 1.4$, $\lambda = 100$, $n_e = 10^6$, and $I = 2^5$). Other details are the same as in Fig. 5.

that of $f(x_0)$ in h-annealing.] In contrast, we can still receive benefit from downward interbasin dynamics in h-annealing, and consequently, h-annealing outperforms c-annealing. From the optimization point of view, this superiority of h-annealing is not striking: The improvement in the basin's cost during the whole process is far less than that achieved in the extension of the search time. This advantage of h-annealing, however, can have practical importance in the application to large-size instances in a given search time. The reason is that in c-annealing, the critical time scale for the onset of the above malfunction of SA became longer with increasing system size. This functional limitation of c-annealing is mentioned again in Sec. IV from the scalability point of view.

B. Temperature-cycling experiment

Next we turn to the results of the temperature-cycling experiment. The optimizing ability can also be enhanced by this nonmonotonic technique. Figure 8 shows the optimization performance for the case $M = 2$ for $n_e = 10^7$. The performance is successfully improved in cases where the whole search process is performed near the effective intermediate temperature found in the above rate-cycling experiment. However, even if the target temperature deviates from this effective point, a larger $\Delta\Theta$ can compensate for the deterioration to a certain extent. To overview the results for various parameter settings, the results of the performance are listed in Table II in a manner similar to that used in Table I. Here, the best average result among all possible combinations of five values of $\Delta\Theta$ and four positive values of M is chosen and listed with the corresponding values of $\Delta\Theta$ and M('s). From the table we find that the average result plotted in Fig. 8 for $n_e = 10^7$ (and $M = 2$) can be improved further for a high or low value of Θ_c. In general, as in this case, a smaller number of cycles was better for high Θ_c's, whereas a larger number was better for low Θ_c's. In every situation, however, the optimization performance was improved to, at most, a level comparable to that of the narrow cycling near the effective temperature within the present study. Table II also shows that the very best performance among all Θ_c's on this time scale is at the same level as that of the two annealings in the rate-cycling experiment.

Intriguingly, as shown just below, a variety of behaviors analogous to relaxation-related phenomena is observed in the

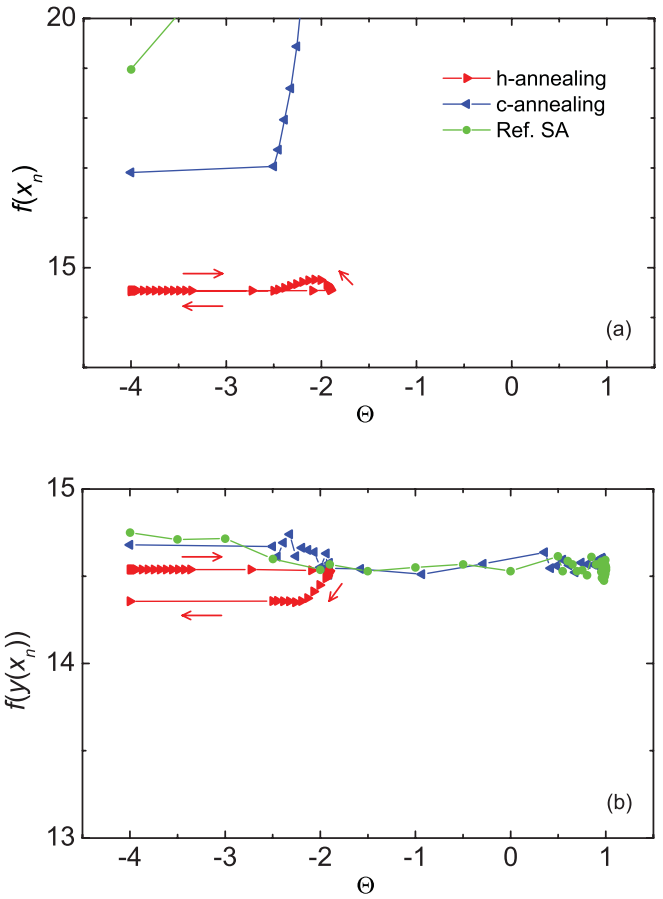

FIG. 7. (Color online) Actual [(a) $f(x_n)$] and hidden [(b) $f(y(x_n))$] search dynamics (RE-TSP with $N = 316$; $\Theta_c = -2.2$, $\Delta\Theta = 0.6$, $\lambda = 100$, $n_e = 10^5$, and $I = 2^5$). Other details are the same as in Fig. 5.

search dynamics and can be related to the above optimization characteristics. Typical results of the hidden search dynamics on a long time scale ($n_e = 10^7$) are shown in Figs. 9(a)–9(d). Except for the third one [Fig. 9(c)], the results are for the best parameter setting for the selected target logarithmic temperature, and in all panels, the result for the no-cycling case is superimposed to illustrate the effect of the cycling. In the best case for $\Theta_c = -1.4$ [Fig. 9(a)], relaxation occurs and proceeds at the lower temperature ($\Theta = -1.9$); however, the temperature cycling appears to ruin this productive process. This is because the solution is sufficiently randomized during the search at the upper temperature ($\Theta = -0.9$), and these observations remind us of the rejuvenation phenomenon found in spin glasses [27]. In such a situation, the effective relaxation dynamics contribute only after the end of the cycling process; in short, the temperature cycling plays no role in this optimization, and hence a smaller M is better. In the best case for $\Theta_c = -1.9$, [the most successful case in this experiment; Fig. 9(b)], the enhancement of the optimizing ability derives from the relaxation occurring near the effective temperature and from a slight acceleration of relaxation, caused in this case mainly by the second cycle; a similar acceleration phenomenon is known to be observed in a polymer glass [28]. At this target temperature, the optimization performance is maximized for the smallest $\Delta\Theta(= 0.2)$; however, even at the

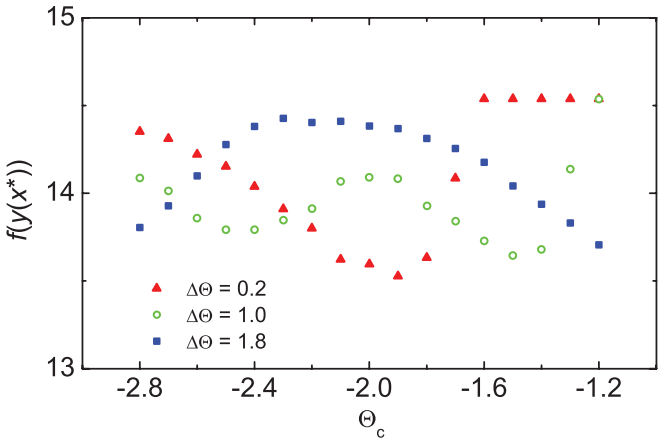

FIG. 8. (Color online) Optimization performance $f(y(x^*))$ (RE-TSP with $N = 316$; $M = 2$, $n_e = 10L = 10^7$, and $I = 2^5$). Results are averaged over I search processes and plotted against the target logarithmic temperature Θ_c.

same temperature, unfavorable upward interbasin dynamics similar to those found in the rate-cycling experiment are observed for a larger $\Delta\Theta$. To illustrate the latter situation, the result for the largest $\Delta\Theta(= 1.8)$ for the same Θ_c is shown in Fig. 9(c). We find that the search trajectory goes into high-lying basins after sudden cooling and that, in this case, the average cost of the final basin is still higher than that of the initial quenched solution. Such upward transitions occurred when cooling was performed rapidly over the effective target range, in which the relaxation dynamics successfully function as an optimizer if the system is cooled down slowly on a

TABLE II. Optimization performance $f(y(x^*))$ (RE-TSP with $N = 316$; $I = 2^5$). The best average result among the 20 cases (see text for details) is listed for each target logarithmic temperature Θ_c on each experimental time scale n_e. The corresponding values of $\Delta\Theta$ and M('s) are also listed in parentheses in this order. The very best result among all Θ_c's (ranging from -2.8 to -1.2 in increments of 0.1) on each n_e is listed in boldface. All results are averages over I search processes.

Θ_c	Temperature cycling: n_e (=10L)		
	10^5	10^6	10^7
−1.4	14.533	14.182	13.640
	(1.0,1–4)	(1.4,2)	(1.0,1)
−1.7	14.533	14.098	13.613
	(1.4,1–4)	(0.6,1)	(0.6,1)
−1.9	14.460	13.920	**13.527**
	(0.6,1)	(0.2,1)	(0.2,2)
−2.0	14.466	**13.897**	13.573
	(0.6,1)	(0.2,4)	(0.2,4)
−2.1	**14.405**	13.989	13.597
	(0.2,1)	(0.2,4)	(0.2,3)
−2.2	14.416	13.997	13.661
	(0.6,2)	(0.6,4)	(0.6,4)
−2.4	14.449	14.032	13.701
	(0.6,3)	(1.0,4)	(1.0,4)
−2.7	14.456	14.104	13.727
	(1.8,2)	(1.4,4)	(1.4,4)

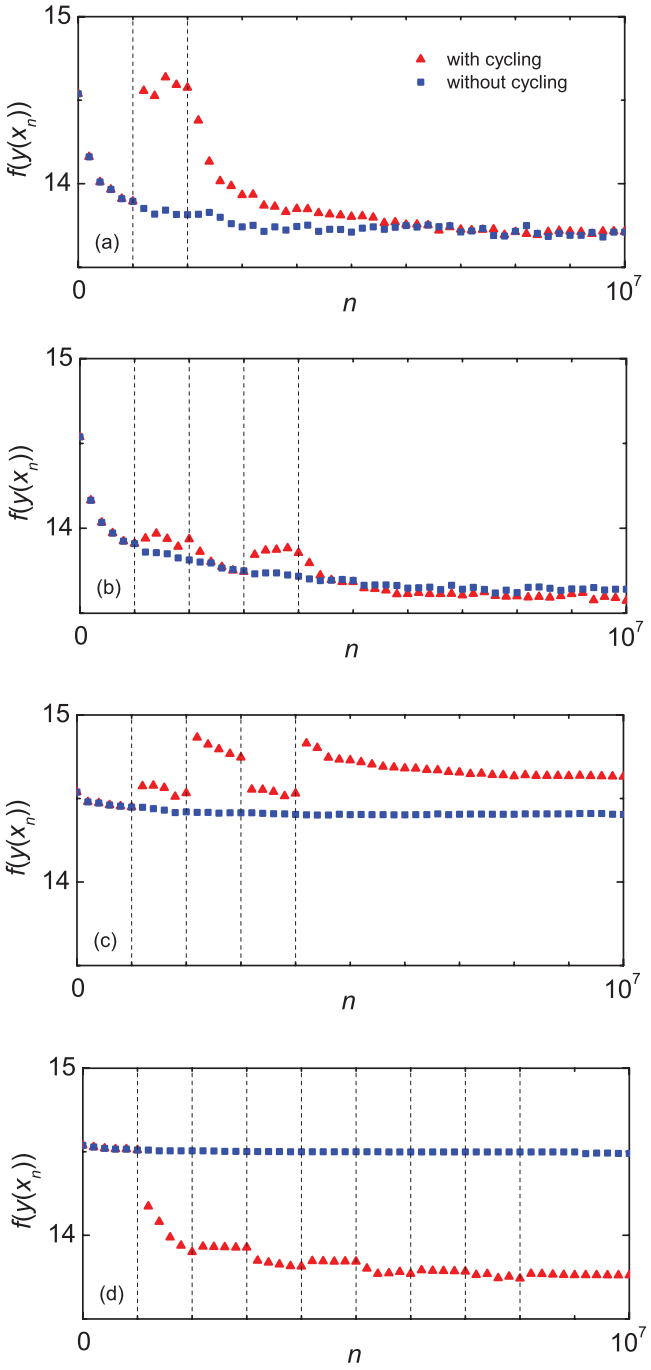

FIG. 9. (Color online) Hidden search dynamics $f(y(x_n))$ (RE-TSP with $N = 316$; $n_e = 10L = 10^7$ and $I = 2^5$). The result for the no-cycling case at the lower temperature ($\Theta_c - \frac{1}{2}\Delta\Theta$) is superimposed; the two behaviors are the same in the first $L(=10^6)$ steps. (a) $\Theta_c = -1.4$, $\Delta\Theta = 1.0$, and $M = 1$; (b) $\Theta_c = -1.9$, $\Delta\Theta = 0.2$, and $M = 2$; (c) $\Theta_c = -1.9$, $\Delta\Theta = 1.8$, and $M = 2$; (d) $\Theta_c = -2.7$, $\Delta\Theta = 1.4$, and $M = 4$. All results are averages over I search processes.

long time scale. In the best case for $\Theta_c = -2.7$ [Fig. 9(d)], acceleration of relaxation occurs every cycle, and in this case it always starts immediately after sudden heating. The system is almost frozen at the lower temperature ($\Theta = -3.4$) and therefore the relaxation proceeds intermittently only at the upper temperature ($\Theta = -2.0$), which recalls the memory effect observed in different glass-forming materials [27–29]. In such a situation, the incumbent solution was improved cycle by cycle, and hence a larger M is better.

As the time scale becomes shorter, the system response to the temperature change becomes insufficient, and as a result, the effect of the temperature cycling is suppressed. Nevertheless, we still receive the benefit of the relaxation dynamics: Even on the shortest time scale ($n_e = 10^5$), as reported in Table II, the highest performance is at the same level as that of h-annealing in the rate-cycling experiment.

IV. DISCUSSION

In both experiments under nonmonotonic temperature control, we found that there exists an effective intermediate temperature as in the previous case of monotonic cooling. If the experimental time is sufficiently long, the optimization performance is sensitive to cooling near this temperature. It is, however, relatively insensitive to preprocessing (heating from a quenched solution or cooling from a random solution) and to the schedule in the target range (monotonic cooling or cyclic temperature change). These characteristics result in no substantial difference in optimizing ability between the two nonmonotonic techniques and these techniques achieve the same degree of improvement as in the slow-cooling process. In addition, on a short time scale, whereas the effective relaxation dynamics disappear in c-annealing, the advantage of the effective temperature still remains in h-annealing and temperature cycling. Thus the present nonmonotonic approach has a wider operation range than the cooling-only process. From these observations, we conclude that the actual annealing strategy is effective also in finite-time optimization by SA and more workable than the conventional slow-cooling strategy.

What should be emphasized here is that the optimization mechanism under nonmonotonic temperature control can still be explained by analogy to physical relaxation-related phenomena. We found that the characteristic behaviors analogous to the stabilization phenomenon and the acceleration of relaxation can be observed in detail in the hidden search dynamics and these behaviors play favorable roles in optimization. In addition to the other relaxation-related behaviors known in laboratory experiments on liquid and glass, unfavorable search dynamics observed as upward interbasin transitions occur when the temperature is rapidly decreased over the effective intermediate range. In the optimization context, this alerts us to the fact that reheating and subsequent cooling with a mere hope of escape from a basin can be counterproductive. As reported in Tables I and II, in cases where the idea of annealing is implemented successfully, the optimum $\Delta\Theta$ is relatively small for target temperatures near the effective intermediate temperature and larger $\Delta\Theta$'s are preferable for high or low target temperatures. These show the critical need for the search performed near the effective temperature. This finding applies whether or not the temperature schedule is monotonic; therefore, it is considered that, as discussed also in the previous study [16], the utilization of relaxation dynamics occurring near the effective temperature should be a primary factor in the successful implementation of finite-time SA, at

least for the present TSPs. (Note that the results depend on the landscape structure determined by the cost and neighborhood functions.) In this way, the similarities found between the optimization dynamics and the relaxation process provide a unified understanding of the optimization mechanism of SA on the basis of the physical analogy.

A note is added on another relevant observation in the rate-cycling experiment. In the present study, the optimization performance was evaluated by the cost of the basin including the incumbent solution, $f(y(x^*))$. Considering the primary role of the relaxation dynamics, we can recognize that this commonly used criterion is inappropriate in the following situation. In h-annealing on a short time scale, it can happen that after the system is activated in the heating stage, the actual search dynamics do not return to the average cost level of the initial quenched solutions in the cooling stages, in which, however, downward transitions to low-lying basins appear in the hidden search dynamics. In this case, the present evaluation cannot reflect the influence of the relaxation dynamics occurring in the late steps because the incumbent solution found in an early step is not improved afterward. This disadvantage of h-annealing was noticeable in the solution of the RD-TSP; more specifically, c-annealing in the confirmation experiment ($N = 316$) with a high target temperature for $n_e = 10^6$ clearly outperformed the corresponding h-annealing in the comparison as in Table I. It was to be expected that in this c-annealing, the effect of the relaxation dynamics was properly assessed by $f(y(x^*))$ because downward interbasin transitions still occurred on this time scale and the incumbent solution was improved throughout the search process. When the performance was evaluated by the final basin's cost $f(y(x_{n_e}))$, which was substantially better than $f(y(x^*))$ in the h-annealing discussed here, the result did not contradict the above conclusion.

Finally, we review how computational physics analysis worked in the previous and present optimization studies from the interdisciplinary point of view; this gives fresh insight into the optimization algorithm research in the landscape paradigm. The experimental analysis in these studies, specifically that using the mapping-onto-minima approach, enabled us to identify the search function of the SA algorithm. In this respect, the analysis of the hidden search dynamics does not merely supply information unavailable in the conventional procedure-based algorithm research; rather, it creates a meaningful complementary approach, a function-based approach, to the studies on heuristic optimization algorithms. The significance of this approach can be highlighted by the following fact: The present comprehensive understanding of the optimization characteristics of various annealing schemes can never be acquired without reunifying the algorithms by their common search function: they have conventionally been distinguished from one another by their procedure detail. This understanding will probably be extensively applicable to the other threshold algorithms [35] (for specific examples, see, e.g., Refs. [5,6,36]), because the relaxation dynamics play a positive role also in their optimization process [37]. Another possible task created from the present study is a function-based evaluation of the scalability of the algorithm. The present rate-cycling experiment showed that in c-annealing from a random solution for a given system size, there exists a lower bound of the search time for downward interbasin dynamics to occur. Because this critical time becomes longer with increasing system size, there exists an upper bound on the system size for the effective relaxation dynamics to work on a given time scale. This means that if one hopes to evaluate properly the scalability of this algorithm, it must be done with clarification of the operation range of the intrinsic optimization function. As shown by these illustrations, we can make more careful and comprehensive considerations of the algorithm by focusing on the search function analyzed in the landscape paradigm. This should be stressed as an illuminating aspect of computational physics analysis in the optimization algorithm research.

V. CONCLUSIONS

The effectiveness of the actual annealing strategy in finite-time optimization by SA was examined in the hope of drawing well-founded knowledge of its optimization mechanism from the physical analogy. The rate-cycling and the temperature-cycling experiments were conducted systematically for the solution of two types of random TSPs. From the results, we confirmed for the present instances that the actual annealing strategy is effective and more workable than the conventional slow-cooling strategy. In the course of these experiments, a variety of behaviors analogous to those characteristic of glass-forming materials was observed in detail in the hidden search dynamics. From these observations, it was shown that the optimization mechanism of the present SA algorithms can be explained in a unified manner by analogy to relaxation and its related phenomena irrespective of the annealing technique. Furthermore, considering how computational physics analysis contributed in the previous and present studies, we have discussed the significance of the function-based approach to the optimization algorithm research. Using this approach, it is hoped that our knowledge of heuristic optimization algorithms will be refined in this interdisciplinary field.

Before closing, a brief remark is made on the practical application of the physical annealing strategy. In the present evaluation, the initial solution was generated by quenching from a random solution and yet we paid no attention to the time for this preparation of the locally optimal solution. This time is not necessarily negligible for large-size instances and therefore this fact undermines the merit of the present nonmonotonic approach. However, if one only takes advantage of the effective intermediate temperature, some easily generated short tour will probably serve as a substitute, and in fact we can find a similar idea in the existing method. In the approach of low-temperature starts [21], a short initial tour is generated in some constructive manner and the conventional SA is implemented with a low initial temperature. From the functionality point of view, it is considered that this approach preserves the optimizing ability of SA by maintaining the advantage of the effective relaxation dynamics in a limited search time.

ACKNOWLEDGMENT

This study was supported in part by JSPS KAKENHI (Grant No.20656017).

[1] S. Kirkpatrick and G. Toulouse, J. Phys. (France) **46**, 1277 (1985).
[2] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987), part 2.
[3] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, Oxford, 2001), chap. 9.
[4] P. Salamon, J. D. Nulton, J. R. Harland, J. Pedersen, G. Ruppeiner, and L. Liao, Comput. Phys. Commun. **49**, 423 (1988).
[5] C. Tsallis and D. A. Stariolo, Physica A **233**, 395 (1996).
[6] T. J. P. Penna, Phys. Rev. E **51**, R1 (1995).
[7] U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993).
[8] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Chem. Phys. Lett. **219**, 343 (1994).
[9] S. Boettcher and A. G. Percus, Phys. Rev. Lett. **86**, 5211 (2001).
[10] G. Zaránd, F. Pázmándi, K. F. Pál, and G. T. Zimányi, Phys. Rev. Lett. **89**, 150201 (2002).
[11] A. K. Hartmann and H. Rieger, *Optimization Algorithms in Physics* (Wiley-VCH, Berlin, 2002).
[12] F. Baletto and R. Ferrando, Rev. Mod. Phys. **77**, 371 (2005).
[13] P. Salamon, P. Sibani, and R. Frost, *Facts, Conjectures, and Improvements for Simulated Annealing* (Society for Industrial and Applied Mathematics, Philadelphia, 2002).
[14] K. H. Hoffmann and J. C. Schön, Found. Phys. Lett. **18**, 171 (2005).
[15] J. C. Schön, J. Phys. A **30**, 2367 (1997).
[16] M. Hasegawa, Phys. Rev. E **83**, 036708 (2011).
[17] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).
[18] V. Černý, J. Optim. Theory Appl. **45**, 41 (1985).
[19] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[20] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys (eds.), *The Traveling Salesman Problem, A Guided Tour of Combinatorial Optimization* (Wiley, Chichester, UK, 1985).
[21] D. S. Johnson and L. A. McGeoch, in *Local Search in Combinatorial Optimization*, edited by E. H. L. Aarts and J. K. Lenstra (Wiley, Chichester, UK, 1997), p. 215.
[22] R. K. Singh, D. Kumar, and B. Chen, in *Encyclopedia of Applied Physics*, edited by G. L. Trigg, Vol. 21 (Wiley-VCH, Weinheim, 1997), p. 67.
[23] H. Ohtani, in *Materials Science and Technology*, edited by R. W. Cahn, P. Haasen, and E. J. Kramer, Vol. 7 (VCH, Weinheim, 1992), p. 147.
[24] G. W. Scherer, in *Materials Science and Technology*, edited by R. W. Cahn, P. Haasen, and E. J. Kramer, Vol. 9 (VCH, Weinheim, 1991), p. 119.
[25] A. Möbius, A. Neklioudov, A. Díaz-Sánchez, K. H. Hoffmann, A. Fachat, and M. Schreiber, Phys. Rev. Lett. **79**, 4297 (1997).
[26] J. Schneider, I. Morgenstern, and J. M. Singer, Phys. Rev. E **58**, 5085 (1998).
[27] E. Vincent, J. P. Bouchaud, J. Hammann, and F. Lefloch, Philos. Mag. B **71**, 489 (1995).
[28] L. Bellon, S. Ciliberto, and C. Laroche, e-print arXiv:cond-mat/9905160v2.
[29] R. L. Leheny and S. R. Nagel, Phys. Rev. B **57**, 5154 (1998).
[30] F. Alberici, P. Doussineau, and A. Levelut, Europhys. Lett. **39**, 329 (1997).
[31] M. Hasegawa, AIP Conf. Proc. **832**, 578 (2006).
[32] F. H. Stillinger and T. A. Weber, Phys. Rev. A **25**, 978 (1982).
[33] F. H. Stillinger and T. A. Weber, Phys. Rev. A **28**, 2408 (1983).
[34] F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).
[35] E. H. L. Aarts, J. H. M. Korst, and P. J. M. van Laarhoven, in *Local Search in Combinatorial Optimization*, edited by E. H. L. Aarts and J. K. Lenstra (Wiley, Chichester, UK, 1997), p. 91.
[36] G. Dueck and T. Scheuer, J. Comput. Phys. **90**, 161 (1990).
[37] M. Hasegawa, Comput. Phys. Commun. **182**, 229 (2011).