



Hybrid parallel tempering and simulated annealing method

Yaohang Li^{a,*}, Vladimir A. Protopopescu^b, Nikita Arnold^c, Xinyu Zhang^d, Andrey Gorin^e

^a Department of Computer Science, North Carolina A&T State University, Greensboro, NC 27411, USA

^b Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

^c Department of Soft Matter Physics, Institute of Experimental Physics, Johannes Kepler Universität Linz, Altenbergerstraße 69, A-4040 Linz, Austria

^d Department of Mechanical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

^e Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

ARTICLE INFO

Keywords:

Markov chain Monte Carlo
Simulated annealing
Parallel tempering

ABSTRACT

In this paper, we propose a new hybrid scheme of parallel tempering and simulated annealing (hybrid PT/SA). Within the hybrid PT/SA scheme, a composite system with multiple conformations is evolving in parallel on a temperature ladder with various transition step sizes. The simulated annealing (SA) process uses a cooling scheme to decrease the temperature values in the temperature ladder to the target temperature. The parallel tempering (PT) scheme is employed to reduce the equilibration relaxation time of the composite system at a particular temperature ladder configuration in the SA process. The hybrid PT/SA method reduces the waiting time in deep local minima and thus leads to a more efficient sampling capability on high-dimensional complicated objective function landscapes. Compared to the approaches PT and parallel SA with the same temperature ladder, transition step sizes, and cooling scheme (parallel SA) configurations, our preliminary results obtained with the hybrid PT/SA method confirm the expected improvements in simulations of several test objective functions, including the Rosenbrock's function and the "rugged" funnel-like function, and several instantiations of the traveling salesman problem. The hybrid PT/SA may have slower convergence than genetic algorithms (GA) with good cross-over heuristics, but it has the advantage of tolerating "bad" initial values and displaying robust sampling capability, even in the absence of additional information. Moreover, the hybrid PT/SA has natural parallelization potential.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Markov chain Monte Carlo (MCMC) methods have long been recognized as effective tools in difficult statistical sampling and global optimization problems arising from a wide range of applications, including physics, biology, medical science, chemistry, material science, computer science, and economics. In the original Metropolis–Hastings-based MCMC [1,2], a Markov process is built to sample a target probability distribution:

$$\pi(x) = Z^{-1} e^{-E(x)},$$

where $E(x)$ refers to the objective function and Z denotes the normalization constant (partition function). In this scheme, a new state y is generated from the current state x of the Markov process by drawing y from a proposal transition function $q(x,y)$. The new state y is accepted with the probability $\min(1,r)$, where r is the Metropolis–Hastings ratio:

* Corresponding author.

E-mail address: yaohang@ncat.edu (Y. Li).

$$r = \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}.$$

One of the outstanding problems of MCMC is the “waiting time dilemma” [3] which in general arises in systems having local minima separated by high barriers. In high-dimensional global optimization problems with difficult multi-mode objective functions in high dimensions, the waiting time problem becomes even more conspicuous. Theoretically, in the generalized Metropolis–Hastings algorithm, the Markov process converges to the target distribution using any positive transition function $q(x,y)$ and any initial configuration. Nevertheless, in practice, the Markov process can be easily trapped into a deep local minimum from where it cannot escape in reasonable time.

Many techniques have been proposed to address this waiting time problem, including simulated annealing (SA) [4], simulated tempering (ST) [5], parallel tempering [6], dynamic-weighting Monte Carlo [7], and various versions of the jump-walking algorithm [8–10]. These techniques have been applied to a variety of sampling and optimization problems and have been shown to alleviate to various extents the slow barrier-crossing problem. However, in difficult, high-dimensional optimization problems such as protein folding, large-scale traveling salesman’s problem, and spin glasses, a more efficient sampling approach is needed to allow more effective barrier-crossing and thereby ensure ergodic sampling.

In this paper, we present a hybrid scheme of simulated annealing and parallel tempering (hybrid PT/SA) with an efficient barrier-crossing capability. Within this scheme, Markov processes with various step sizes are carried out at different temperature levels, namely larger transition step sizes are allowed at higher temperatures. The PT is used to reduce the relaxation time of the conformation system in the SA process, which can significantly reduce the waiting time in deep local minima and thus lead to a better chance to discover the global minimum. Applications to complex trial functions and high-dimensional traveling salesman problems are presented.

2. Simulated annealing and parallel tempering

Simulated annealing (SA) [4] is a popular approach to address the “waiting time dilemma” by introducing a variable temperature T into the objective function of the target distribution,

$$\pi(x) = Z^{-1} e^{-E(x)/T}.$$

This is based on the fact that it is easier for a Markov process to escape from a local energy minimum at a higher temperature. The main idea behind SA is to interpret the objective function $E(x)$ as the energy of a thermodynamic system. In order to anneal the system to the ground state with the global energy minimum, the system is initially “melt” at a high temperature T , and then T is slowly decreased, allowing the system to be close to thermodynamic equilibrium at each stage by implementing the Metropolis–Hastings transition rules [1,2]. The SA achieves an improved barrier-crossing probability at high temperature. However, as the temperature decreases, the waiting time dilemma resurfaces as the problem of choosing the proper scheme for the cooling rate.

Consider a system of physical particles, such as the ideal gas or glass, which is originally treated in the work of Metropolis et al. [1]. In a rugged energy landscape, if the system is still trapped in a deep local minimum, as the temperature is lowered below a certain level, the probability that the system escapes from that deep local minimum becomes lower and lower. For complicated and topologically intricate energy landscapes, the Markov process may remain trapped in a local minimum unless the temperature is decreased impractically slowly so that the global thermodynamic equilibrium state could be reached. The key to tackle this problem in the SA method is to accelerate the relaxation speed of the system to thermodynamic equilibrium at one temperature before the temperature is further reduced. In practice, multiple SA processes are usually carried out in parallel with different initial values and random trajectories to locate conformations at or near the global minimum.

The simulated tempering (ST) [6] belongs to the Monte Carlo accelerated algorithms, where the probability to cross high barriers is artificially enhanced by adding a random walk in the temperature space to the sampling of the objective function at different temperature. The parallel tempering (PT) (replica exchange) method (also known as the multiple Markov chain or replica exchange method) [6,11] is an enhanced version of ST that was developed to achieve good sampling of systems having a complex objective function landscape with many local minima by exchanging system replicas at various temperature levels. In a typical PT algorithm, a composite system is constructed with one subsystem configuration per temperature level. All temperature levels form a temperature ladder. The subsystems at highest temperature levels easily climb over all barriers in the energy landscape while the low temperature subsystems mainly explore the local energy minima. The principle of PT is to induce a Monte Carlo trial transition by switching the subsystems at two neighboring temperature levels. Therefore, from a single subsystem point of view, *unlike in the SA scheme*, the temperature of a single subsystem may now either increase or decrease. As compared to SA, the subsystem exchange improves the mixing rate of MCMC samplers at different temperatures [12]. As a result, the PT scheme reduces the relaxation time of the overall composite system to the global minimum. Successful applications of PT include the simulation of biomolecules [13], spin glass modeling [14–16], determination of X-ray structures [17], and structure prediction of small proteins [18,19]. The efficient sampling property of PT lead us to combine the advantages of the PT and SA schemes in a hybrid PT/SA scheme that would reduce the waiting time in deep local minima and therefore achieve a faster convergence to the global minimum.

3. Implementation of hybrid PT/SA method

Similarly to the PT scheme, in the hybrid PT/SA algorithm, we track N sets of the system configurations, each at a different temperature level, T_i . The different temperature levels form a temperature ladder and each configuration forms a subsystem at each level on the temperature ladder. A state of the composite system is specified by $X = \{x_1, x_2, \dots, x_N\}$, where x_i is the subsystem at temperature level i . The equilibrium distribution of the “macrostate” X is,

$$P(X) = \prod_{i=1}^N \frac{e^{-\beta_i E(x_i)}}{Z(T_i)},$$

where $\beta_i = 1/T_i$, $E(x_i)$ is the objective (energy) function, and $Z(T_i) = \int e^{-\beta_i E(x_i)} dx_i$ is the partition function of the subsystem at temperature T_i .

The hybrid PT/SA method is composed of a PT part and a SA part. The goal of the PT part is to reduce the relaxation time of subsystems towards equilibrium state within the composite system. On the other hand, in the SA part, the temperature at each temperature level is gradually reduced and thus the overall composite system is gradually guided towards the target temperature while always remaining in a state close to thermodynamic equilibrium.

3.1. PT scheme

In the PT implementation of the hybrid PT/SA method, at each iteration step t , the Markov chains can be realized with two types of transitions – the Metropolis transition and the replica transition:

1. *Metropolis transition:* The Metropolis transition is employed as local Monte Carlo moves for the conformation at each temperature level. The transition probability depends only on the change of the objective function, $E(X_i^t)$, where X_i^t is the conformation at temperature level T_i . The Metropolis–Hastings ratio at temperature level T_i is calculated as:

$$w_{Local}(x_i^t \rightarrow x_i^{t+1}) = e^{-\beta_i \Delta_i E} = e^{-\beta_i (E(x_i^{t+1}) - E(x_i^t))}.$$

The new state is accepted with the probability $\min(1, w_{Local}(x_i^t \rightarrow x_i^{t+1}))$.

2. *Replica transition:* The replica transition takes place with the probability θ and is used to exchange configurations at two neighboring temperature levels, i and $i + 1$:

$$x_i^{t+1} \leftrightarrow x_{i+1}^{t+1}.$$

The exchange is accepted according to the Metropolis–Hastings criterion with probability $\min(1, e^{-\beta_i E(x_{i+1}^{t+1}) - \beta_{i+1} E(x_i^{t+1}) + \beta_{i+1} E(x_{i+1}^{t+1}) + \beta_i E(x_i^{t+1})})$). The relaxation rate [21] can be characterized by the ergodic measure, the so called fluctuation metric,

$$\Omega(X^t) = \sum_{j=1}^N [\beta_j E(x_j^t) - \overline{\beta E(x^t)}]^2 / N,$$

where $\overline{\beta E(x^t)} = \sum_{k=1}^N \beta_k E(x_k^t) / N$ is the ergodic average at iteration step t . The replica transitions lead to an improvement of the relaxation rate of the overall simulation of the composite system. The goal of the PT part of the hybrid PT/SA method is to quickly lead the composite system to an equilibrium state.

3.2. SA scheme

The main purpose of the SA part of the hybrid PT/SA method is to allow the temperature at each temperature level to decrease gradually toward a target temperature T^{target} . The cooling scheme is a key element of SA and there have been many cooling schemes discussed in literature [20]. Here we use a proportional cooling schedule,

$$T_i^{new} \leftarrow T^{target} + (T_i^{old} - T^{target})C,$$

where C is the cooling factor with $0 < C < 1$, which is selected to be very close to 1, and T^{target} is the target temperature. Similar to the pure PT method, the temperature of a single subsystem in hybrid PT/SA may either increase or decrease. However, following the annealing schedule of the SA scheme, the amplitude of the temperature fluctuation for each subsystem will become smaller and smaller as the temperatures converge to the target temperatures T^{target} . As a result, the SA cooling scheme drives the composite system out of its current equilibrium to a new equilibrium state. Eventually, all temperatures T_i s converge to T^{target} . Actually, it is not necessary to cool the temperatures at all temperature levels to T^{target} . However targeting the same T^{target} will increase the production rate of the hybrid PT/SA scheme, since in many Monte Carlo simulation applications, only conformations at T^{target} are of interest.

Theoretically, the convergence rate of the Markov process to its stationary distribution depends on the second largest eigenvalue of the Markov chain transition matrix. Practically, one can compute the integrated autocorrelation time (IAT)

[21] to evaluate the relaxation rate, $\Omega(X^t)$ of the Markov chain to equilibrium at iteration step t . However, the computation of IAT requires the record of generated samples in all previous steps, which is computation and memory intensive and becomes impractical for long-running chains. For the composite system used in the hybrid PT/SA scheme, the probability of finding a subsystem in the “microstate” i with objective function $E(x_i)$ is given by the Boltzmann distribution,

$$P(E(x_i)) = \frac{\exp(-E(x_i)/T_i)}{\sum_j \exp(-E(x_j)/T_j)}$$

By examining the objective function distribution of each subsystem at a certain iteration step in the hybrid PT/SA evolution, one could determine whether the composite system has reached its equilibrium state by using statistical tests such as the Kolmogorov–Smirnov (KS) test [22]. The SA cooling scheme will take effect to decrease the temperature values on the temperature ladder only when the composite system has reached the equilibrium state.

3.3. Various transition step sizes at different temperature levels

In the hybrid PT/SA scheme, multiple Markov chains are employed at various levels of the temperature ladder. The fundamental idea of employing a temperature ladder with various temperature values is to allow the Markov processes at different temperature levels to evolve somewhat independently, by ignoring different levels of local complexity [23]. The role of the high temperature levels is to give the Markov process a better chance to escape from the current local minimum. A large transition step employs a proposal transition function $q(x,y)$ that can generate a new state y that deviates largely from the old state x (or have a higher probability to generate a largely deviated new state). A large transition step usually helps the system to escape from a deep local minimum, but may miss “local details”. However, since only samples at the lowest level are relevant for the final result, “local details” at high temperature can be overlooked without serious consequences. This suggested the idea of using different step sizes for different temperature levels in setting up the transition step sizes within the hybrid PT/SA scheme. For lower temperature levels, we use smaller transition steps (or small transition steps with higher probability), which enable us to explore the “local details” of the objective function landscape. On the other hand, for higher temperature levels, larger transition steps (or large transition steps with higher probability) are appropriate, since they reduce the computational time and facilitate escape from the local minima.

3.4. Acceptance ratio and exchange ratio

To optimize performance of the hybrid PT/SA scheme, determining the temperature value of each level on the temperature ladder is of great importance. There is an optimal relationship between the temperature and transition step size at each temperature level, which is dictated by the experimentally established acceptance ratio. In Rathore et al.’s case study in PT simulations [24], an acceptance ratio of 20% for Metropolis sampling at a given temperature yielded the best possible performance. By analyzing how the temperature interval affects the performance of the system, Kone and Kofke also suggested that the acceptance probability of 23% is optimal in PT [25]. Although the 23% acceptance probability has been claimed to be optimal in [25], 20–30% acceptance probability is regarded as good temperature heuristic in MCMC practice. In the hybrid PT/SA scheme, we adopted a 20–30% acceptance ratio in each Markov chain at each temperature level. To achieve this, we employed an adaptive approach by adjusting the temperature at each level according to the acceptance ratio. Initially, the temperature in each subsystem is adjusted to allow its acceptance ratio to be relatively high at 20–50%. As the subsystem is cooling down in the SA scheme, the acceptance ratio is gradually adjusted to 15–25%. An alternative, and probably more appropriate, way to control the temperature levels is to adjust only the lowest temperature, T_1 , and the highest temperature, T_N , according to the desired acceptance ratio in the SA scheme and then adopt the feedback-optimized approach [26] to set up the temperatures T_2, \dots, T_{N-1} in the rest of the temp to minimize the round-trip times between T_1, \dots, T_N .

Another important question in hybrid PT/SA scheme is to determine how frequently the replica transition should take place. Schug and Wenzel [27] found empirically that an exchange ratio θ ranging from 0.5% to 2% in PT scheme is good. In the hybrid PT/SA, we adopt this exchange empirical range and introduce an adjustable variable T into the computation of replica exchange acceptance probability,

$$\min \left(1, e^{\frac{-\beta_i E(x_{i+1}^{t+1}) - \beta_{i+1} E(x_i^{t+1}) + \beta_{i+1} E(x_{i+1}^{t+1}) + \beta_i E(x_i^{t+1})}{T}} \right),$$

where T behaves like a “global temperature” of the system. By adjusting T , the replica exchange ratio can be kept in the range from 0.5% to 2% during the overall optimization process.

3.5. Hybrid PT/SA pseudocode

Fig. 1 shows the flowchart of the hybrid PT/SA algorithm. We summarize the hybrid PT/SA by the following sequence of steps:

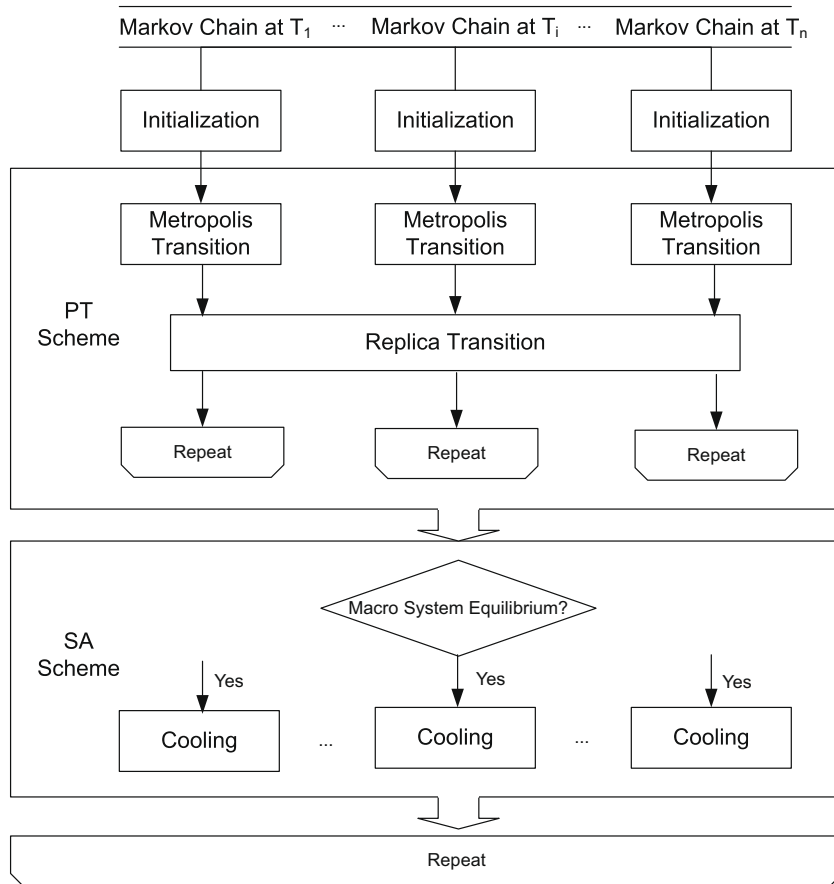


Fig. 1. Flowchart of hybrid PT/SA algorithm.

- The temperature ladder is initialized. Each Markov chain, i , starts at a temperature level T_i , configuration x_i , and a proposal transition function $q_i(\cdot)$.
- Repeat
 - Repeat in parallel:
 - Each Markov chain i evolves according to the Metropolis transition.
 - Each neighbor pair, i and $i + 1$, of the Markov chain performs replica transition with probability θ .
 - Until a certain number of steps has completed.
 - Adjust temperatures at the temperature ladder using the SA cooling scheme.
- Until a certain number of steps has been completed or the desired relaxation rate has been achieved.

4. Experimental results

4.1. Rosenbrock's function

The generalized n -dimensional Rosenbrock's function is defined as

$$f(x_1, \dots, x_n) = \sum_{i=2}^n (100(x_i - x_{i-1}^2)^2 + (1 - x_{i-1})^2),$$

where the global minimum is attained at $(1.0, 1.0, \dots, 1.0)$. The Rosenbrock's function is a notorious benchmark function in optimization because of its slow convergence for most optimization methods. Due to the long narrow valley present in this function, gradient-based methods may have to spend a large number of iterations before the global minimum is reached. Fig. 2 shows the two-dimensional Rosenbrock's function. The Rosenbrock's function is also interesting for simulations of the protein folding process, because it is commonly believed that for native proteins there exists a definite folding pathway formed by high energy barriers [28]. The Rosenbrock's function landscape has no local minima, which in principle should make the SA an efficient method.

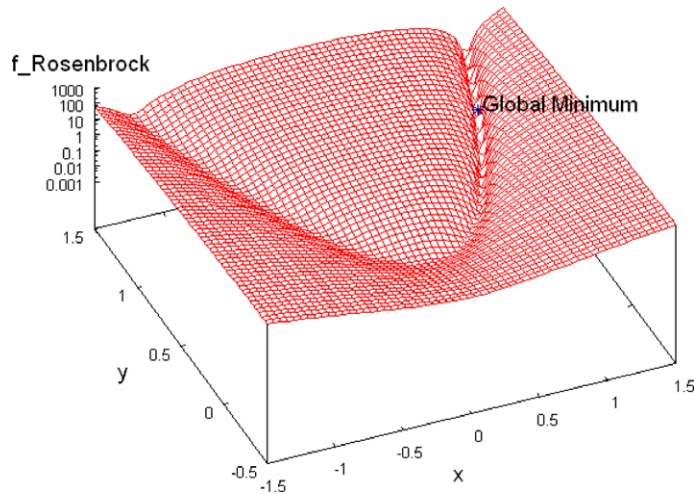


Fig. 2. Two-dimensional Rosenbrock's function. Note the logarithmic scale of the function's axis.

Fig. 3 shows the comparison of searching the global minimum in the 100-dimensional Rosenbrock's function using hybrid PT/SA, PT, and SA schemes with 10^7 iterations. The hybrid PT/SA scheme employs 10 Markov chains, where at the lowest level the temperature is 2.0 and the transition step size is 10^{-5} . For each upper levels, the values of temperature and step size are increased by the constant factor of 2.0 and 1.6, respectively. The initial location is at $(10, 10, \dots, 10)$. The PT scheme starts with the same configuration setup without gradually reducing the temperatures at different levels. The parallel SA scheme uses 10 independent Markov chains, the same configuration as those in hybrid PT/SA scheme, and records the lowest function values among these chains. During the minimization process, the temperature at each level is gradually reduced by a factor of 0.99 at every 10^4 iterations unless the current acceptance rate is too low ($<18\%$) or the current objective function distribution fails the statistical test. As a result, the acceptance ratio at each temperature level is gradually adjusted to the range of 20–25%. As shown in Fig. 2, the hybrid PT/SA scheme yields a faster convergence to the global minimum at value 0 in the Rosenbrock's function than simple parallel SA or PT schemes using multiple subsystems.

4.2. "Rugged" funnel-like function

One possible application of the hybrid PT/SA method is protein structure prediction. Proteins typically adopt a single structure, corresponding to the global minimum of the free energy under physiological conditions. Various studies [29,30]

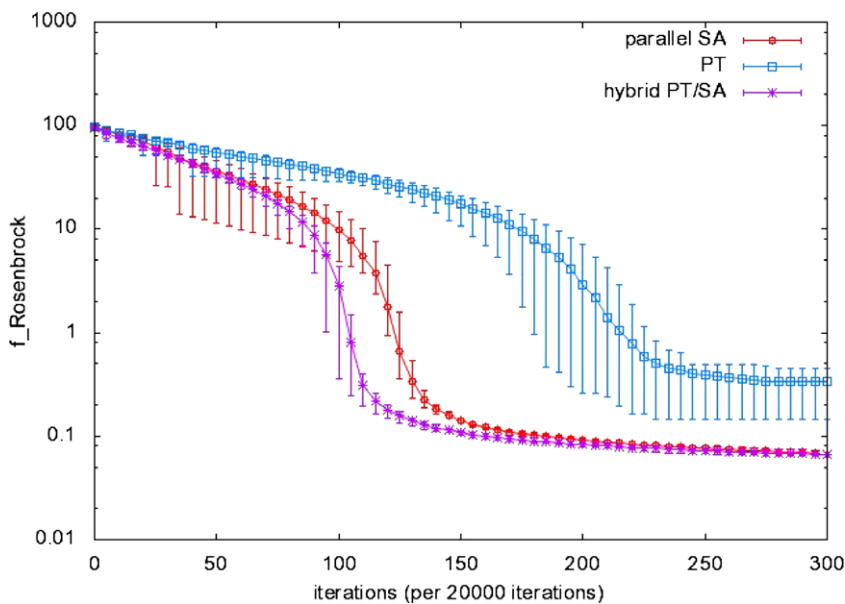


Fig. 3. Comparison of SA, PT, and hybrid SA/PT in searching for global minimum in 100-dimensional Rosenbrock's Function.

suggest that the energy landscape in the protein folding problem generally resembles a “rough” funnel with local minima where the protein can transiently reside. In this experiment, we construct a “rugged” funnel-like function:

$$E(x_1, \dots, x_n) = -\frac{c}{2n} \sum_{k=1}^n \left\{ 1 + A \sum_{i=0}^m a^i \cos(b^i x_k) \right\},$$

where $A = (a - 1)/(a^{m+1} - 1)$, n is the dimension, and $m, a, b,$ and c are some tunable constants to determine the depth of the funnel and the number of local minima along it. The only global minimum is located at $(0, \dots, 0)$ and is equal to $-c$. Fig. 4 shows the a “rugged” funnel-like function where $c = 1.0$.

In our computational experiments, to ensure deep local minima disposed in the high-dimensional objective function, we select $c = 10^6$ and $n = 100$. We intend to use this function to simulate the protein folding energy landscape and study the behavior of the hybrid PT/SA method. Fig. 5 shows the comparison of searching for the global minimum in the 100-dimensional “rugged” funnel-like function using hybrid the PT/SA scheme, the PT scheme, and each individual subsystem using a SA scheme. A similar cooling scheme in the computational experiment in the Rosenbrock’s function presented in Section 4.1

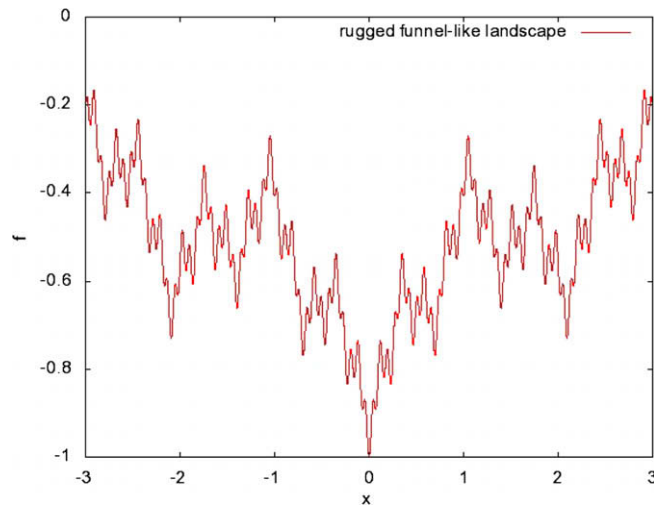


Fig. 4. A “rugged” funnel-like function ($m = 4, a = 0.7, b = 3.0, c = 1.0$).

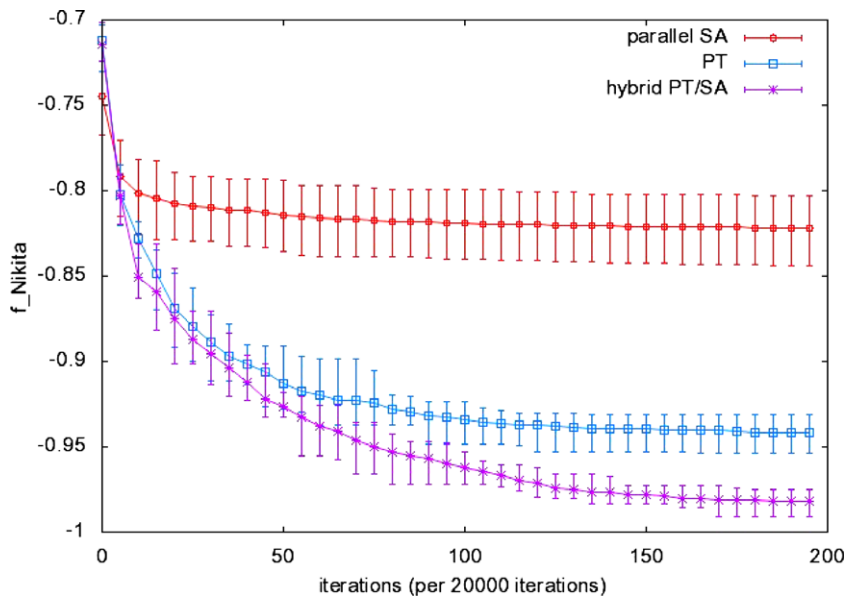


Fig. 5. Comparison of SA, PT, and hybrid SA/PT in searching for global minimum in 100-dimensional rugged funnel-like function with 2000 cycles, each cycle containing 10^4 iterations.

is also employed here. High acceptance ratio (40–50%) is allowed initially and is then gradually adjusted to 18–25%. The hybrid PT/SA has the same cooling scheme as the parallel SA and the same replica exchange criteria as PT. Similar to the observation of the experiments in Rosenbrock’s function, the hybrid PT/SA scheme exhibits a faster barrier-crossing and shows a faster approaching to the global minimum in this rugged funnel-like function than that of the SA or PT schemes.

4.3. Traveling salesman problem

The traveling salesman problem (TSP) asks for the shortest tour to visit N cities of specific locations exactly once. The TSP is a well-known NP-complete problem. A number of large TSPs are collected in TSPLIB [31] to serve as challenging tests for newly proposed global optimization algorithms.

To apply hybrid PT/SA to the TSP, we use two types of proposal transition moves as suggested in [32]: (1) permutation, whereby a section of path is removed and then replaced with the same cities running in a permuted order; and (2) replacement, whereby a section of path is removed and then replaced between a randomly chosen city pair. Multiple subsystems are running on the multiple-level temperature ladder – the one at the higher temperature levels takes more aggressive moves with longer sections in permutation or replacement. The temperature at each level is adjusted to maintain the 20–40% acceptance rate. Table 1 shows the mean, best, and worst optimal results found in 100 runs of hybrid PT/SA, PT, and parallel SA in TSPLIB maps of att532 and ptr1002 using 10^9 iterations. Fig. 6 shows the tour with minimum path cost found in att532 of TSPLIB at every 10^6 iterations using hybrid PT/SA, PT, and SA approaches in one run. In this computational experiment, the PT and parallel SA approaches use the same temperature and transition move configurations as the hybrid PT/SA; however, the PT approach does not use the SA cooling scheme while the parallel SA approach does not use the replica transitions. Also, similar to those computational results in the high-dimensional Rosenbrock’s function and rugged funnel-like function, the hybrid PT/SA scheme finds faster tours close to the optimal one. Figs. 7 and 8 show the solutions of att532 and ptr1002 from TSPLIB obtained by the hybrid PT/SA scheme, which are very close to the known optimal solutions. Fig. 9 shows the hybrid PT/SA solution of a 40×40 grid, which is the exact optimal solution. Unlike att532 and ptr1002 with a single global minimum, the 40×40 grid is regular so that there exist many solutions (global minima); thus, it is easier for a global optimization algorithm to locate one of them. Fig. 10a shows the att532 computational time comparison of the implementation hybrid PT/SA and parallel SA using MPI and Fig. 10b shows the progress to global minimum of hybrid PT/SA and

Table 1
Comparison of hybrid PT/SA, PT, and parallel SA in TSPLIB maps att532 and pr1002 in 100 runs.

	Hybrid PT/SA			Parallel tempering			Parallel SA		
	Best	Mean	Worse	Best	Mean	Worse	Best	Mean	Worse
att532	27,765	28,247	29,195	28,555	29,481	32,406	33,369	36,064	39,775
pr1002	261,993	263,533	264,696	267,484	274,221	296,298	263,192	265,451	267,465

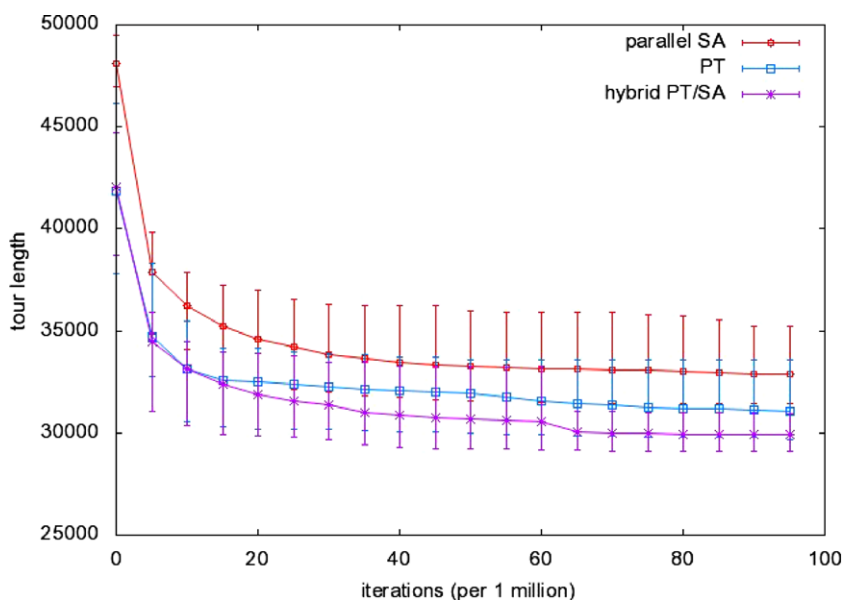


Fig. 6. Comparison of parallel SA, PT, and hybrid PT/SA in search for minimum tour in map att532 of TSPLIB.

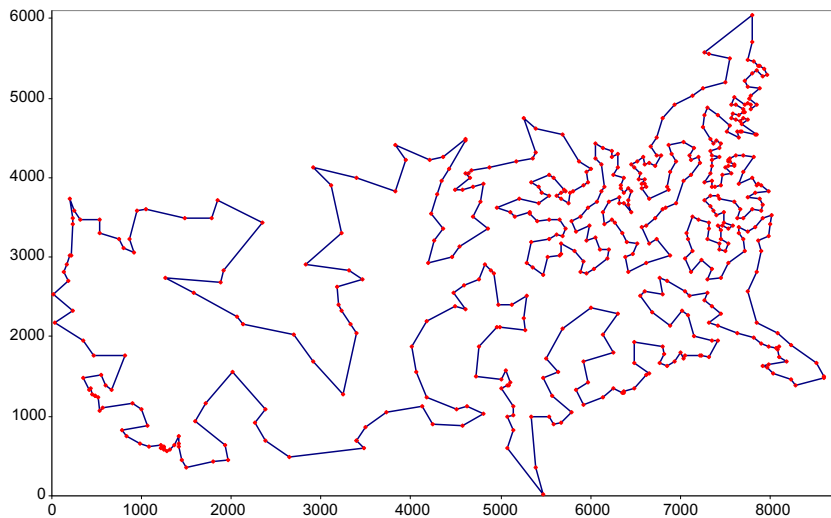


Fig. 7. Hybrid PT/SA solution at att532 (532 cities, best known optimal solution = 27,686) hybrid PT/SA solution = 27,765, 0.285% longer than the optimal solution.

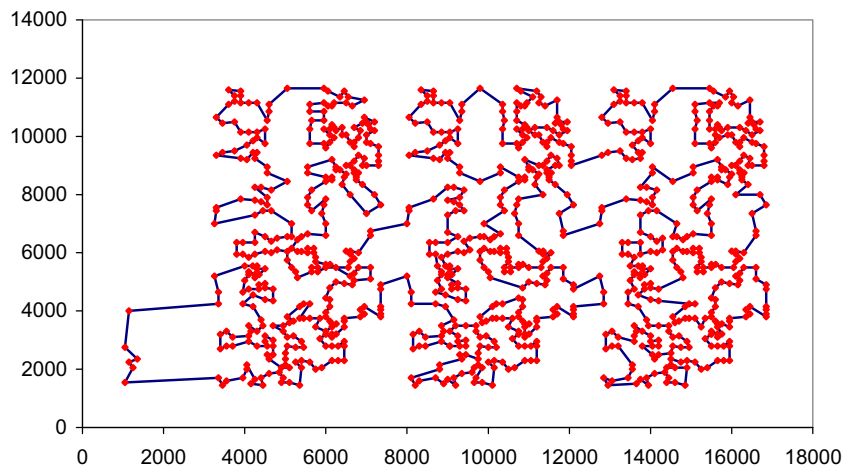


Fig. 8. Hybrid PT/SA solution at ptr1002 (1002 cities, optimal solution = 259,045) hybrid PT/SA solution = 261,993, 1.13% longer than the optimal solution.

parallel SA as a function of time. The computations are conducted on an IBM xSeries Cluster with 16 2.0 GHz Xeon nodes using LAM-MPI – each node carries out a Markov process in one level within the temperature ladder. Since the exchange rate is controlled in a small range (0.5–2.0%) in hybrid PT/SA scheme, the interprocessor communication does not pose a significant overhead in the overall global optimization process (see Fig. 11).

Usually, unlike those using the TSP-dependent heuristic like Lin–Kernighan [33], approaches using general problem-independent heuristics such as SA and genetic algorithms (GA) perform quite poorly in large-scale TSP problems. This is possibly due to the large number of transition choices and local minima [34]. In our experiment, using the hybrid PT/SA with 10^9 iterations using the problem-independent heuristics simply led by the objective function gradient and scrambling subsequence of the TSP tour, tours within 5% of the optimal solution in large-scale TSP problems in TSPLIB can be found as shown in Table 2.

4.4. Comparison with genetic algorithms

If good crossover heuristics exist, GA usually outperforms conventional Monte Carlo by exhibiting convergence to global minimum with fewer objective function evaluations [35]. Table 3 shows the average number of function evaluations to reach “global minimum valley” using simple GA implemented by GALib, differential evolution (DE) [36], and hybrid PT/SA on sev-

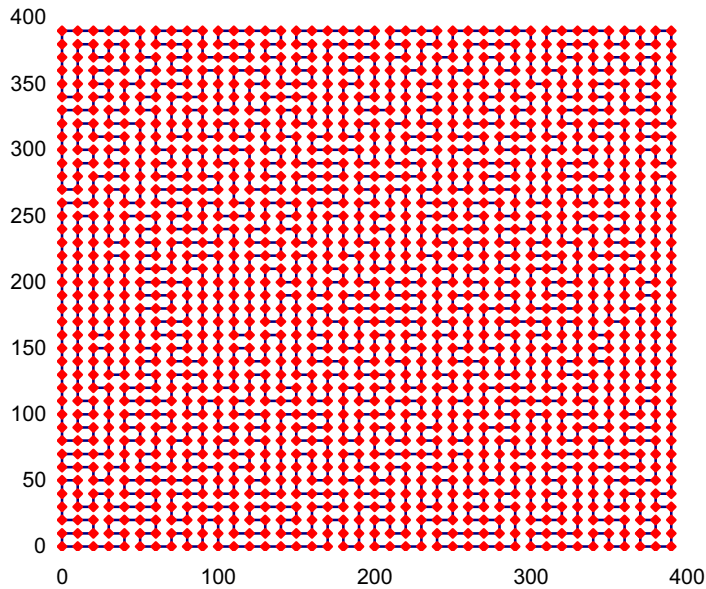


Fig. 9. Hybrid PT/SA solution on a 40×40 grid (exact optimal solution).

eral well-known benchmark functions [37] for global optimization. Hybrid PT/SA yields better performance than simple GA but DE outperforms both hybrid PT/SA and simple GA with its good heuristic of vector difference in continuous functions.

Nevertheless, compared to GA, the hybrid PT/SA method has the following advantages, which are inherited from MCMC:

1. Hybrid PT/SA can perform reasonably well even in the absence of additional information and can tolerate “bad” initial values. We tested the hybrid PT/SA method and the genetic algorithm with differential evolution using the following 10-dimensional simple multi-modal objective function:

$$f = -3 * e^{-\sum (x_i - 5)^2 / 2} - 2 * e^{-\sum (x_i)^2 / 2} - 4 * e^{-\sum (x_i + 4)^2 / 2}.$$

Fig. 10 shows the simple multi-modal objective function in 2D. The objective function has three “witch-hats” with local minima of -2 and -3 and global minimum of -4 . Table 3 shows the best solution found using standard GA (population = 100, $P_{\text{crossover}} = 1.0$, $P_{\text{mutation}} = 0.1$) developed by MIT GALib, differential evolution (population = 100, $F = 0.8$ and $P_{\text{crossover}} = 0.9$ as recommended), and hybrid PT/SA using in 100 runs with random initial values. The standard GA was trapped in local minima in 71 runs and reached global minimum in only 29 runs. Differential evolution was also trapped in local minima 19 times. In contrast, hybrid PT/SA was trapped in local minima only once and by increasing the number of iterations to 10^7 , all 100 runs reached the global minimum, -4 .

2. Hybrid PT/SA is more suitable for sampling applications. In many physical or biological systems, we are not only interested in obtaining the global minimum, but also the deep local minima. Based on the same multi-modal objective function in Fig. 10, Table 4 shows the distribution of the terminated solutions using standard GA, differential evolution, and hybrid PT/SA. Hybrid PT/SA has 18% of the terminated solutions (i.e., solutions when all temperatures in different levels are reduced to 0.0) at local minimum -2 , 28% at local minimum -3 , and 54% at global minimum -4 , which is approximately proportional to the size of the three “witch-hats”, where standard GA and differential evolution do not show this correct proportion with respect to the multi-modal function (see Table 5).
3. The hybrid PT/SA can be parallelized more efficiently. In hybrid PT/SA, the replica transition requires interprocessor communication, which can be implemented by exchanging of temperature – a floating point number instead of the replica configuration. This leads to a small interprocessor communication overhead.
4. The hybrid PT/SA can handle optimization/sampling in situations where GA crossover is not feasible. However, GA and MCMC could be combined to form hybrid algorithms to achieve better optimization or sampling capability. For example, an approach combining DE and MCMC has reported improved sampling efficiency [38].

5. Summary

In this paper, we proposed a new global optimization scheme, the so-called the hybrid PT/SA scheme, by combining the SA and PT schemes. We achieve a better barrier-crossing capability, which leads to a more aggressive sampling over com-

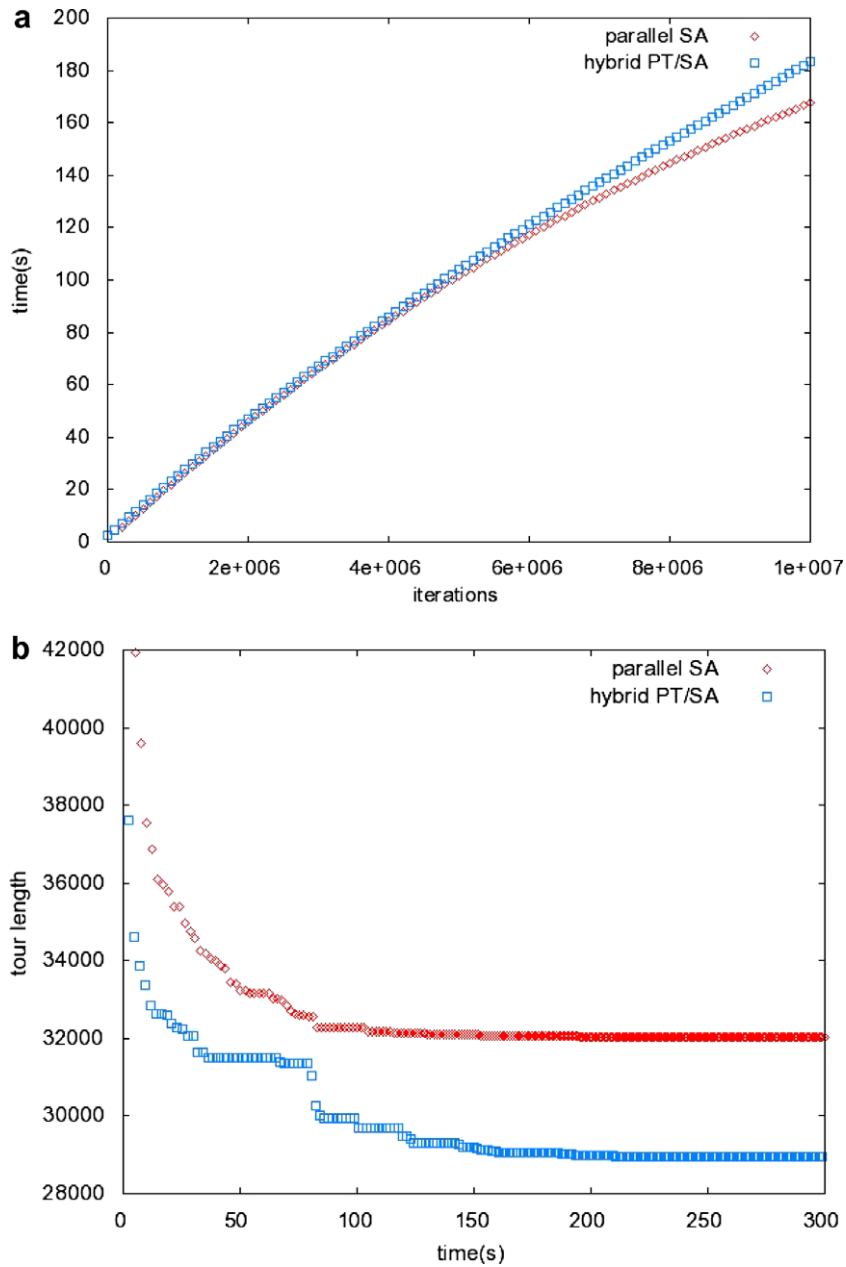


Fig. 10. Computational time comparison of parallel implementation of hybrid PT/SA and parallel SA using MPI on TSPLIB map att532. Hybrid PT/SA and parallel SA employ the similar temperature ladder configuration while hybrid PT/SA performs additional interprocessor replica exchange operations, which poses insignificant intercommunication overhead but accelerate the search process.

plicated high-dimensional landscapes. The hybrid PT/SA scheme uses a composite system composed of multiple subsystems with various transition step sizes at different temperature levels on the temperature ladder. Subsystems at higher temperature employ larger transition step sizes. The PT scheme reduces the composite system's relaxation time at a temperature configuration in the SA process. Replica transition in the PT scheme helps to reduce the waiting time of Markov chains running at low temperatures when being trapped in deep local minima. Overall, the hybrid PT/SA approach exhibits improved barrier-crossing capability compared to both the PT and the parallel SA approaches. The computational results from high-dimensional objective functions including the Rosenbrock's function, the "rugged" funnel-like function, and the applications of the traveling salesman problem confirm the expected improvement. The hybrid PT/SA also has advantages of: (i) tolerating "bad" initial values, (ii) more efficient sampling, and (iii) easy parallelization, without using additional information.

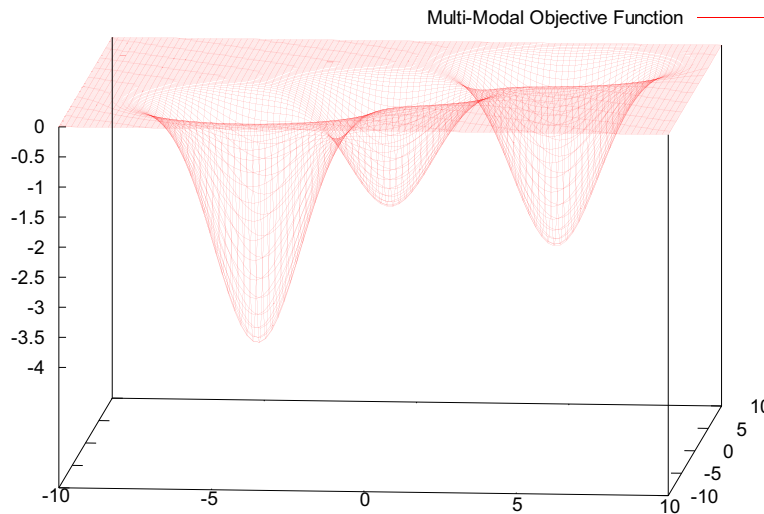


Fig. 11. Simple multi-modal objective function in 2D.

Table 2

Optimal tour found in hybrid PT/SA in compared with the known optimal solution in TSPLIB library.

	# of cities	Optimal tour found	Optimal solution	% exceed
att532	532	27,765	27,686	0.285
pr1002	1002	261,993	259,045	1.13
r15934	5934	577,667	556,045	3.89
usa13509	13,509	20,509,831	19,982,859	2.63
d15112	15,112	1,631,832	1,573,084	3.73
swe24978	24,978	891,584	855,597	4.21

Table 3

Average number of objective function evaluations (10^6) in 10 runs to reach the global minimum valley (<0.1 root mean square from global minimum location) using simple GA (population = 100, $P_{crossover} = 1.0$, $P_{mutation} = 0.1$), DE (population = 100, $F = 0.8$, $P_{crossover} = 0.9$), and hybrid PT/SA ($P_{replica\ exchange} = 0.1$) in various benchmark objective functions in 100 dimensions.

Objective functions	Definition	Sim. GA	DE	Hybrid PT/SA
Ackley	$f_{Ack}(x) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^n \cos(2\pi x_i)}$ $x_i \in [-30, 30]$, $x^* = (0, 0, \dots, 0)$, $f_{Ack}(x^*) = 0$	2.8	2.0	2.4
Rosenbrock	$f_{Ros}(x) = \sum_{i=1}^{n-1} (100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$ $x_i \in [-2.048, 2.048]$, $x^* = (1, 1, \dots, 1)$, $f_{Ros}(x^*) = 0$	98.2	10.7	5.5
Schwefel	$f_{Sch}(x) = 418.9829n + \sum_{i=1}^n x_i \sin(\sqrt{ x_i })$ $x_i \in [-500, 500]$, $x^* = (-420.9687, \dots, -420.9687)$, $f_{Sch}(x^*) = 0$	720.1	3.2	260.8
Rastrigin	$f_{Ras}(x) = 10n + \sum_{i=1}^n (x_i^2 - 10 \cos(2\pi x_i))$ $x_i \in [-5.12, 5.12]$, $x^* = (0, 0, \dots, 0)$, $f_{Ras}(x^*) = 0$	>1000	0.46	52.5
Griewank	$f_{Gri}(x) = 1 + \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos(x_i/\sqrt{i})$ $x_i \in [-600, 600]$, $x^* = (0, 0, \dots, 0)$, $f_{Gri}(x^*) = 0$	>1000	0.2	0.3

Table 4

Best solution found using standard GA, differential evolution, and hybrid PT/SA in 100 runs with random initial values.

	Function evaluations	Local minimum -2	Local minimum -3	Global minimum -4
Standard GA	10^6	46	25	29
Differential evolution	10^6	19	0	81
Hybrid PT/SA	10^6	0	1	99
Hybrid PT/SA	10^7	0	0	100

Table 5

Distribution of the terminated solutions using standard GA, differential evolution, and hybrid PT/SA.

	Function evaluations	Local minimum -2 (%)	Local minimum -3 (%)	Global minimum -4 (%)
Standard GA	10^6	46	25	29
Differential evolution	10^6	19	0	81
Hybrid PT/SA	10^6	18	28	54

Acknowledgements

The work is partially supported by NSF under grant number CCF-0829382 to Y. Li, 2006 ORAU/ORNL Summer Faculty Participation Program to Y. Li, and LDRD Program of the Oak Ridge National Laboratory managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725, to A. Gorin.

References

- [1] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* 21 (1953) 1087–1092.
- [2] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* 57 (1970) 97–109.
- [3] W.H. Wong, F. Liang, Dynamic weighting in Monte Carlo and optimization, *Proc. Natl. Acad. Sci. USA* 94 (1997) 14220–14224.
- [4] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Optimization by simulated annealing, *Science* 220 (1983) 671–680.
- [5] E. Marinari, G. Parisi, Simulated tempering: a new Monte Carlo scheme, *Europhys. Lett.* 19 (1992) 451–458.
- [6] C.J. Geyer, E.A. Thompson, Annealing Markov chain Monte Carlo with applications to ancestral inference, *J. Am. Stat. Assoc.* 90 (1995) 909–920.
- [7] J.S. Liu, F. Liang, W.H. Wong, A theory for dynamic weighting in Monte Carlo computation, *J. Am. Stat. Assoc.* 96 (2001) 561–573.
- [8] S. Brown, T. Head-Gordon, Cool walking: a new Markov chain Monte Carlo sampling method, *J. Comput. Chem* 24 (2003) 68–76.
- [9] C.J. Tsai, K.D. Jordan, Use of the histogram and jump-walking methods for overcoming slow barrier crossing behavior in Monte Carlo simulations: applications to the phase transitions in the $(Ar)_{13}$ and $(H_2O)_8$ clusters, *J. Chem. Phys.* 99 (9) (1993) 6957–6970.
- [10] R. Zhou, B.J. Berne, Smart walking: a new method for Boltzmann sampling of protein conformations, *J. Chem. Phys.* 107 (21) (1996) 9185–9196.
- [11] J.J. Moreno, H.G. Katzgraber, A.K. Hartmann, Finding low-temperature states with parallel tempering, simulated annealing and simple Monte Carlo, *Int. J. Mod. Phys. C* 14 (3) (2003) 285–299.
- [12] J.S. Liu, Markov Chain Monte Carlo and Related Topics, Technical Report, Stanford University, 1999.
- [13] U. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, *Chem. Phys. Lett.* 281 (1997) 140–150.
- [14] H.G. Katzgraber, M. Palassini, A.P. Young, Monte Carlo simulation of spin glasses at low temperatures, *Phys. Rev. B* 63 (2001) 184422.
- [15] H.G. Ballesteros, A. Cruz, L.A. Fernandez, V. Martin-Mayor, J. Pech, J.J. Ruiz-Lorenzo, A. Tarancon, P. Tellez, C.L. Ullod, C. Ungil, Critical behavior of the three-dimensionalising spin glass, *Phys. Rev. B* 62 (21) (2000) 14237–14245.
- [16] K. Hukushima, K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, *J. Phys. Soc. (JPN)* 65 (1996) 1604–1608.
- [17] M. Falconi, M.W. Deem, A biased Monte Carlo scheme for zeolite structure solution, *J. Chem. Phys.* 110 (1999) 1754–1766.
- [18] A. Schug, T. Herges, A. Verma, W. Wenzel, Investigation of the parallel tempering method for protein folding, *J. Phys.: Condens. Matter* 17 (2005) 1641–1650.
- [19] A. Schug, W. Wenzel, Predictive *in silico* all atom folding of a four helix protein with a free energy model, *J. Am. Chem. Soc.* 126 (2004) 16737.
- [20] Y. Nourani, B. Andresen, A comparison of simulated annealing cooling strategies, *J. Phys. A: Math. Gen.* 31 (1998) 8373–8385.
- [21] B.A. Berg, Markov Chain Monte Carlo Simulations and Their Statistical Analysis, World Scientific, 2004. p. 197.
- [22] D.E. Knuth, The Art of Computer Programming, third ed., Addison-Wesley Publishing Company, 1997.
- [23] Y. Li, V.A. Protopopescu, A. Gorin, Accelerated simulated tempering, *Phys. Lett. A* 328 (4) (2004) 274–283.
- [24] N. Rathore, M. Chopra, J.J. de Pablo, Optimal allocation of replicas in parallel tempering simulations, *J. Chem. Phys.* 122 (2) (2005) 024111.
- [25] A. Kone, D.A. Kofke, Selection of temperature intervals for parallel tempering simulations, *J. Chem. Phys.* 122 (20) (2005) 206101.
- [26] H.G. Katzgraber, S. Trebst, D.A. Huse, M. Troyer, Feedback-optimized parallel tempering Monte Carlo, *J. Stat. Mech.: Theor. Exp.* (2006) P03018.
- [27] A. Schug, W. Wenzel, All-atom folding of the trp-cage protein with an adaptive parallel tempering method, *Europhys. Lett.* 67 (2) (2004) 307–313.
- [28] D. Baker, A surprising simplicity to protein folding, *Nature* 405 (2000) 39–42.
- [29] C.L. Brooks, M. Gruebele, J.N. Onuchic, P.G. Wolynes, Chemical physics of protein folding, *Proc. Natl. Acad. Sci. USA* 95 (19) (1998) 11037–11038.
- [30] J.N. Onuchic, Contacting the protein folding funnel with NMR, *Proc. Natl. Acad. Sci. USA* 94 (1997) 7129–7131.
- [31] M. Junger, G. Reinelt, G. Rinaldi, The Traveling Salesman Problem: A Bibliography, Annotated Bibliography in Combinatorial Optimization, Wiley, 1997.
- [32] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C++: The Art of Scientific Computing, Cambridge University Press, 2002.
- [33] S. Lin, B.W. Kernighan, An effective heuristic algorithm for the traveling-salesman problem, *Oper. Res.* 21 (1973) 498–516.
- [34] R. Baraglia, H.I. Hidalgo, R. Perego, A hybrid heuristic for the traveling salesman problem, *IEEE Trans. Evol. Comput.* 5 (6) (2001) 613–622.
- [35] R. Unger, J. Moulton, Genetic algorithm for protein folding simulation, *J. Mol. Biol.* 231 (1) (1993) 75–81.
- [36] R. Storn, K. Price, Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces, *J. Global Optim.* 11 (1997) 341–359.
- [37] D. Ortiz-Boyer, C. Hervas-Martinez, N. Garcia-Pedrajas, CIXL2: a crossover operator for evolutionary algorithms based on population features, *J. Artif. Intell. Res.* 24 (2005) 1–48.
- [38] C.J.F. Ter Braak, A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces, *Stat. Comput.* 16 (2006) 239–249.