Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physics Letters A 328 (2004) 274–283
PHYSICS LETTERS A
www.elsevier.com/locate/pla

Accelerated simulated tempering

Yaohang Li [a], Vladimir A. Protopopescu [b], Andrey Gorin [b,*]

[a] *Department of Computer Science, North Carolina A&T State University, Greensboro, NC 27411, USA*
[b] *Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

Received 23 January 2004; received in revised form 18 May 2004; accepted 18 May 2004

Available online 26 June 2004

Communicated by C.R. Doering

Abstract

We propose a new stochastic global optimization method by accelerating the simulated tempering scheme with random walks executed on a temperature ladder with various transition step sizes. By suitably choosing the length of the transition steps, the accelerated scheme enables the search process to execute large jumps and escape entrapment in local minima, while retaining the capability to explore local details, whenever warranted. Our simulations confirm the expected improvements and show that the accelerated simulated tempering scheme has a much faster convergence to the target distribution than Geyer and Thompson's simulated tempering algorithm and exhibits accuracy comparable to the simulated annealing method.

1. Introduction

In a seminal paper, Metropolis et al. [1] proposed the by now famous and widely used algorithm of simulating a Markov process to efficiently and accurately sample a target distribution $\pi(x) = Z^{-1}e^{-h(x)}$, where $h(x)$ refers to the energy function and Z denotes the normalization constant. In the original version of the algorithm, a new state y is generated from the current state x of the Markov process by drawing y from a symmetric proposal transition function $q(x, y)$. Later, Hastings [2] generalized the algorithm by extending $q(x, y)$ to functions that are not necessarily symmetric. Either case, the Metropolis–Hastings ratio is computed by

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

and the new state y is accepted with probability $\min(1, r)$. The central limit theorems for ergodic averages are valid for the Metropolis–Hastings method. The convergence rate to the required expectation of the Metropolis–Hastings method is $O(N^{-1/2})$, where N is the number of samples [21]. This convergence rate is independent of the dimensionality of the problem, thereby avoiding in a certain sense the "curse of dimensionality" and yielding a remarkably good performance in high-dimensional problems. As a result, this method has been adopted in many scientific computing applications, including biology [3], medical science [10], chemistry, material sciences, computer sciences, economics, and physics.

* Corresponding author.
E-mail address: agor@ornl.gov (A. Gorin).

0375-9601/$ – see front matter
doi:10.1016/j.physleta.2004.05.067

One important application of the Metropolis–Hastings method is protein folding [11]. Proteins typically adopt a single structure, corresponding to the global minimum of the free energy under physiological conditions. Levinthal showed that protein folding could not occur via a systematic enumeration of all possible configurations compatible with the very large number of degree of freedom of the amino acids, which is popularly known as the "Levinthal paradox" [12]. Various studies suggest that the energy landscape in the protein folding problem generally resembles a "rough" funnel with local minima where the protein can transiently reside. This motivates the use of Metropolis–Hastings based simulations to speed up the exploration of the conformational space of amino acid sequences, to identify the global minimum energy configuration.

Theoretically, the Markov process defined by the Metropolis–Hastings method converges to the target distribution for any trial transition function $q(x, y) > 0$ and any starting point [15]. Nevertheless, in practice, the actual implementation of the Metropolis–Hastings method is very sensitive to the length of the transition steps. If the transition steps are too small, the Markov process can be easily trapped into a deep local minimum from where it cannot escape in practical time. On the other hand, if the transition steps are too large, the acceptance rate tends to be low. More importantly, long transition steps may ignore some "local details" of $\pi(x)$ and then lead to an incorrect equilibrium distribution. This quandary is referred to as the "waiting time dilemma" [4]. Unfortunately, in most real-life applications that resort to the Metropolis–Hastings method, there are no general guidelines to select the appropriate length of the transition steps.

One approach to avoid the "waiting time dilemma" is the simulated annealing (SA) method [5]. This method is based on the fact that it is easier for a system to escape from a local minimum energy at a significantly higher temperature. Thus, within the SA scheme, the temperature is initially raised to a high value. The system evolves according to the Metropolis–Hastings criterion, while the temperature is very slowly reduced during the simulation procedure. However, unless the temperature is decreased impractically slowly, the Markov process may still be trapped in another local minimum. This is almost certainly the case in multi-minima systems when the temperature has reached a sufficiently low value. This problem can be addressed by the recent cool walking (CW) method [19] which propagates two Markov chains in tandem at both high and low temperatures and generates jumps using SA. We note that like the Metropolis–Hastings method, the SA method or CW method are sensitive to the length of transition steps. Indeed, when the local minimum is too deep and/or the transition steps too small, even raising the temperature to a high level might not be able to help the system escape from that local minimum.

Several methods have been proposed to deal with this drawback by adjusting the length of transition steps in the Markov process. The dynamic importance-weighting scheme [4] introduces weights into the dynamic Monte Carlo process to allow the system to make large transitions, which would not be permitted by the standard Metropolis transition rules. This scheme yields good results for many optimization problems. However, since the weight variable increases as the Markov process evolves, the dynamic weighting scheme loses efficiency in a long chain [6]. Finally, the multiple-try Metropolis (MTM) [7] enables a sampler to make large step-size jumps. In practice, MTM is still prone to be trapped by steep energy minima.

2. Simulated tempering

In the 1990s, Marinari and Parisi [8] and, a few years later, Geyer and Thompson [9] proposed the simulated tempering (ST) algorithm which seemed to offer the key to a satisfactory solution. The new idea in ST is to treat the temperature, t, as an additional dynamic variable. As a result, in the ST scheme, the system is defined on an augmented space $X \times I$, where X is the original phase space and I denotes the one-dimensional temperature space. By varying the temperature ladder t_i, $i = 1, \ldots, m$, m stationary distributions $\pi_i(x)$ are constructed, where $\pi_1(x) = \pi(x)$ and $1 = t_1 < t_2 < \cdots < t_m$. Then, a random walk is performed on this temperature ladder.

We outline the ST algorithm as presented by Geyer and Thompson [9]:

(1) At the current temperature level, i, update x using a Metropolis–Hastings update for $\pi_i(x)$.

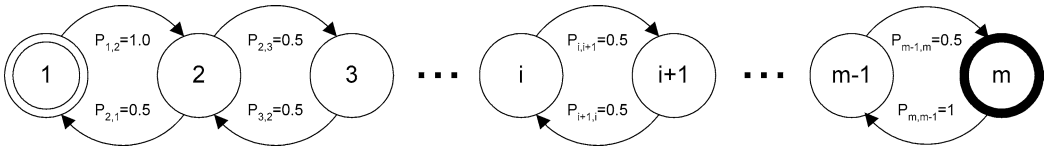

Fig. 1. State diagram of Geyer and Thompson's algorithm *with* m temperature levels.

(2) Set $j = i \pm 1$ according to probabilities $p_{i,j}$, where $p_{1,2} = p_{m,m-1} = 1.0$ and $p_{i,i+1} = p_{i,i-1} = 0.5$ if $1 < i < m$.
(3) Calculate the equivalent of the Metropolis–Hastings ratio for the ST method,

$$r_{\mathrm{ST}} = \frac{c(j)\pi_j(x)p_{j,i}}{c(i)\pi_i(x)p_{i,j}},$$

where $c(i)$ are tunable constants and accept the transition from i to j with probability $\min(r_{\mathrm{ST}}, 1)$. The distribution π is called the pseudo-prior because the function $c_i(x)\pi_i(x)$ resembles formally the product of likelihood and prior.

The equilibrium distribution of the ST Markov chain is a joint distribution for the pair of variables $(x, i) \in X \times I$, where x is a random realization of the state variable and i is a random realization of the "temperature" level. We note that unlike the SA method, where each time the temperature is changed the system is driven out of equilibrium, the ST maintains the system in equilibrium. As suggested by Marinari and Parisi, it is as if instead of finding the minimum of the energy, one finds the minimum of the free energy. This does not affect the final result, since for the interesting level, corresponding to the zero temperature the energy and free energy coincide. The ST method has been applied successfully to the simulations of random field Ising models [8] and models of ancestral inference in the Hutterites [9].

Compared to the SA method, the ST method enables the system to escape from local minima and to locate multiple minima by allowing the system to move up and down the temperature ladder, according to the Metropolis–Hastings rule. The tradeoff here is that the system needs to spend more computational time at high temperature levels.

In practice, the ST algorithm generates samples much more slowly than the original Metropolis–Hastings method or the SA method. The reason is that, during an execution of the ST program, the Markov chain is frequently pacing up and down between the intermediate levels of the temperature ladder, which accounts for most of the computational time. However, what we are usually interested in are only those samples generated in level 1, with the coolest temperature. Fig. 1 shows the state diagram of Geyer and Thompson's algorithm with m different temperature levels, where the transition probabilities from temperature level 1 to 2, $p_{1,2}$, and m to $m-1$, $p_{m,m-1}$, are 1 and all the other transition probabilities $p_{i,i+1}$ and $p_{i,i-1}$ are 0.5. In Geyer and Thompson's algorithm, based on the state diagram shown in Fig. 1, the probability that the system visits level 1 during the evolution of the Markov process is roughly $1/2(m-1)$, which indicates that the algorithm is approximately $2m-1$ times slower than a simple Metropolis–Hastings method. Also, for a particular system, if an inappropriate length of transition step is chosen, the ST algorithm may become ineffective.

A computationally efficient variant of ST is the parallel tempering (PT) [9,20] sampling scheme, which can be viewed as "multi-copy" ST. The difference is that the PT method has m Markov chains, one at each temperature level, and randomly swaps a pair of walkers at neighboring temperatures. The PT method does not need to evaluate the constants $c(i)$; however, as a tradeoff, PT has to keep track of m Markov chains. Provided that the constants $c(i)$ are computed appropriately, the ST, PT, and ST methods should yield equivalent behaviors. This reasonable expectation is indeed supported by numerical simulations (see results section).

3. Accelerated simulated tempering

Based on the analysis of Geyer and Thompson's algorithm, we propose a new idea embodied in three complementary technical improvements to accelerate convergence to the target distribution. The inclusion

of these techniques into the ST algorithm resulted in an accelerated simulated tempering (AST) algorithm that we will present next. Within the AST algorithm, the system continues to be in equilibrium (like in the ST) and the detailed balance condition of the Markov process remains valid; hence, AST converges also to the target distribution. Moreover, we will show that the AST scheme is insensitive to the length of the transition step and can provide an effective solution for a wide variety of energy landscapes. The three new, complementary techniques ideas designed to accelerate the ST algorithm are discussed next.

3.1. Temperature ladder with various step sizes

In the ST scheme, the fundamental idea of building a temperature ladder with different levels is to allow the system to evolve, while ignoring different levels of local complexity. The role of the higher-level temperature is to allow the system to escape from the current local minimum. As noted before, a longer transition step may help the system to escape from such a local minimum more easily but may miss "local details." However, since only samples at the lowest level are relevant, "local details" at high temperature can be overlooked without serious consequences. This suggested *the idea of using different step sizes for different temperature levels*. For lower temperatures, we use smaller transition steps, which enable us to explore the "local details" of the system. On the other hand, for higher temperatures, longer transition steps are appropriate, since they reduce the computational time and render easier escape from local minima.

To implement this idea and accelerate convergence to the target distribution, we propose a random walk on a *modified temperature ladder*, with variable step sizes. More precisely, we construct an alternative ladder that combines temperature and step size, (t_i, s_i), where $1 = t_1 < t_2 < \cdots < t_m$, and $s_1 < s_2 < \cdots < s_m$.

3.2. "Lean the ladder"

Since the main purpose of higher temperatures is to enable the Markov process to escape from possible local minima, one does not need during the simulation to visit these temperature levels too often. Indeed, for a given system, at high temperatures the sampler encounters bigger features, larger basins, and "rougher" relief than it would at low temperatures. Thus, exploring the energy landscapes at high temperatures may not require as many walks as at low temperature. To account for this reality we increase the number of walks at low temperatures and reduce their number at high temperatures. As a result, in AST, we *lean the temperature ladder* by setting $p_{i,i+1} < p_{i,i-1}$, i.e., by letting the system have higher probabilities to visit the lower temperatures. The "leaned temperature ladder" is expected to improve the acceptance rate of the system by generating more walks at low temperatures than at high temperatures. Fig. 2 shows the state diagram of ST on a leaned temperature ladder.

Then, the approximate probability that, at the end of the process, the system finds itself at the lowest temperature level is

$$\pi_1 = \frac{\prod_{j=2}^{m-1}(1 - p_{j,j+1})}{\sum_{j=2}^{m-1}\left(\prod_{s=2}^{j-1} p_{s,s+1}\prod_{t=j}^{m-1} p_{t,t+1}\right)}.$$

We can actually control the probability of visiting each temperature level by manipulating the corresponding transition probabilities. We note that "leaning the ladder" is accompanied by a suitable calculation of the constants $c(i)$, which ensures that the populations on each temperature level are compatible with the equilibrium requirements. The net effect of leaning the ladder is that, while maintaining the system at

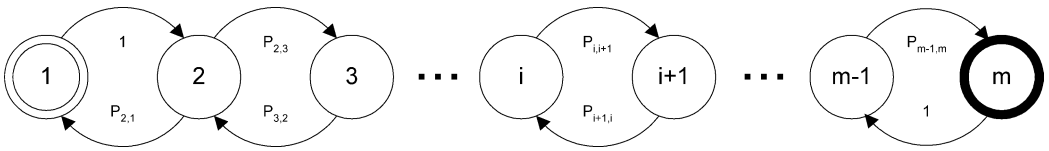

Fig. 2. State diagram of simulated tempering on a leaned temperature ladder: $p_{i,i+1} < p_{i,i-1}$, $i = 1, \ldots, m-1$.

equilibrium, the fundamental level becomes more favored.

3.3. Efficient constant evaluation

Like in the original ST algorithm, the constants $c(i)$ are tunable and are usually selected as the normalizing constants of the distributions $\pi_i(x)$ [9]. The evaluation of these normalizing constants is no easier than in the original problem of simulating from $\pi_i(x)$ [16]. Fortunately, neither in the ST nor in the AST algorithms, it is necessary to calculate the constants $c(i)$ with very high accuracy. Therefore, we propose to adopt a crude Monte Carlo method to obtain very fast, but good approximation of the normalizing constants. Indeed, by using a small number of independent random samples, $\xi_1, \ldots, \xi_n$, from a normal Gaussian distribution, G, the constants $c(i)$ are approximately given by:

$$c(i) \approx \frac{1}{n} \sum_{k=1}^{n} \frac{\pi_i(\xi_k)}{G(\xi_k)}.$$

3.4. AST algorithm

We incorporated the new idea and three technical improvements outlined above into the ST algorithm, to obtain the following AST algorithm:

1. Initialization: use a crude Monte Carlo method with a few random samples (e.g., 10 000) to quickly evaluate a reasonable approximation of the integral
$$Z_i = \int_{-\infty}^{+\infty} \pi_i(x)\,dx.$$
Estimate the constants $c(i)$ with $c(i) = 1/Z_i$.
2. Simulated tempering on the new temperature ladder with variable step sizes, (t_i, s_i):
   2.1. At the current temperature level, i, update x using a Metropolis–Hastings update for $\pi_i(x)$, with step s_i.
   2.2. Perform random walk on the temperature level.
      2.2.1. Set the proposed next temperature level, $j = i \pm 1$, according to probabilities $p_{i,j}$, where $p_{1,2} = p_{m,m-1} = 1$ and $p_{i,i+1} > p_{i,i-1}$, if $1 < i < m$.
      2.2.2. Calculate the ratio
$$r_{\mathrm{ST}} = \frac{c(j)\pi_j(x)p_{j,i}}{c(i)\pi_i(x)p_{i,j}}$$
and accept the transition from i to j with probability $\min(r_{\mathrm{ST}}, 1)$.
   2.3. Back to step 2.1, until all the samples are generated.
   2.4. Stop.

The main difference between the ST and the AST schemes is the random walk on the "double" temperature-ladder of transition steps as opposed to the "single" temperature ladder. This enables the system to jump across energy barriers in considerably shorter times. Equally importantly, due to the presence of transition steps of various lengths at each temperature level, the AST algorithm is not sensitive to a particular requirement of transition step length. This enables the algorithm to perform well for problems with different transition step specifications.

The basic idea of the ST method consists of changing the temperature while remaining at equilibrium—which is in contrast with the SA method. It is easy to show that the random walk on the temperature ladder with various step sizes does not change the equilibrium status of the system. Indeed, there are two types of transitions that may take place within the AST algorithm, namely, the transitions within a temperature level and transitions between different temperature levels. Due to the Metropolis–Hastings rule, the transitions within a temperature level maintain the system in equilibrium, regardless of the length of the transition step. We note that the transitions between different temperature levels do not change the equilibrium status of the system either, because the construction of pseudo-prior in the ST scheme is specifically designed to maintain the system at equilibrium when transitioning between different temperatures. Thus, the AST maintains the system in equilibrium and eventually converges to the target distribution.

4. Results

We illustrate the performance of the new AST algorithm by comparing it with the original Metropolis–Hastings, SA, ST, PT, and CW algorithms on a few

typical examples. Applications to larger scale, directly applied problems will be published elsewhere [17].

4.1. Widely distributed minima

We tested the AST method using the following target distribution density function in 10 dimensions:

$$f = -0.6N(-6,1) - 0.1N(0,1) - 0.3N(5,1),$$

where $N(a, b^2)$ generically denotes a 10-dimensional normalized standard distribution with mean a and standard deviation b in each of the ten variables, $x = (x_1, x_2, \ldots, x_{10})$. The function f has three "witch-hat" minima that are widely distributed across the energy landscape. The global minimum of f is at -6 while 0 and 5 are two local minima. This function is a prototype for energy landscapes occurring, for instance, in biomolecular simulations, where the energy may depend very sensitively on the distances between atoms (e.g., the Van-der-Walls potential energy is of the form $A/r^{12} - B/r^6$, where r is distance between a given pair of atoms and A and B are positive constants).

We applied the Metropolis–Hastings, SA, ST, PT, CW, and AST algorithms to generate 10^7 random samples, respectively. Each computation starts at the origin. Fig. 3 shows the distribution of these samples projected on one dimension, compared to the actual target distribution density of f. The simple Metropolis–Hastings algorithm is trapped at local minimum of 1. In the SA algorithm, the initial temperature is raised to be 100 with an annealing rate 0.99. Starting at a sufficiently high initial temperature, the system can escape from the local minimum at 0, but then gets trapped by the local minimum at 5 and misses the global minimum. We implemented Geyer and Thompson's ST method using a temperature ladder with 10 levels, where the highest temperature level is 100. We also implemented the CW sampling with the high temperature level at 100, CW jump probability of 0.2, annealing rate of 0.99, and CW complete probability of 0.03. In our experiments, both Geyer and Thompson's ST and CW methods missed the global minimum, even though they can escape from the local minimum at 1.

The proposed AST uses the same ladder with various step sizes, which can find rapidly all the local minima and properly reconstitute the shape of the target distribution. The original PT method with fixed transition step size is similar to the ST method. Their similar behaviors are shown in Fig. 4. However, if PT uses a modified temperature ladder with various step sizes, then its behavior becomes similar to the ASTs. Fig. 4 also shows that PT with various step sizes at different temperature levels converges to the target distribution as AST does. Our experiments show that simply manipulating the temperature may not be sufficient to avoid the "waiting time dilemma." Strategy with larger step sizes at higher temperature and smaller step sizes at lower temperatures is key to escaping from deep local minima.

Also, in this experiment, AST yields a much better computational performance than the ST algorithm. Table 1 shows the performance comparison among SA, ST, PT, CW, and AST. Our experimental results show that AST method is quite effective, displaying fast convergence to the stationary distribution. As mentioned before, the reason for this good performance

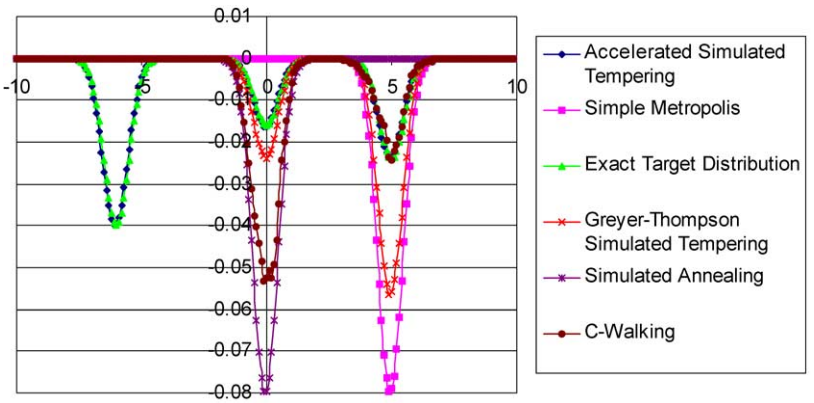

Fig. 3. Comparison of simple Metropolis–Hastings method, SA, ST, CW and AST after generating 10 000 000 samples.

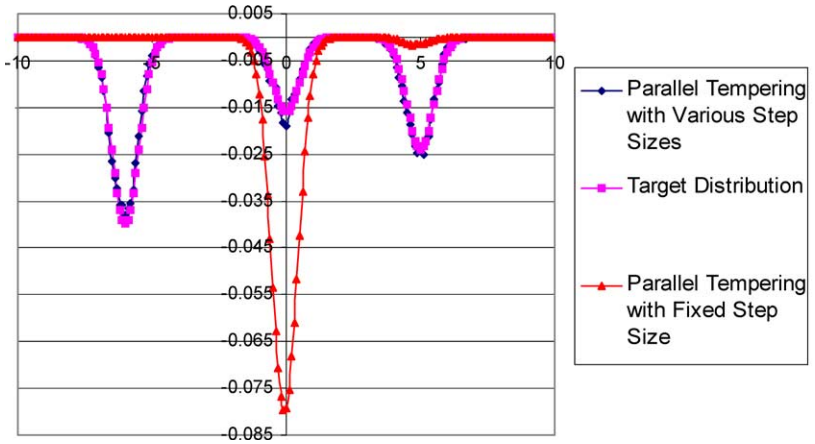

Fig. 4. PT with various step sizes vs. PT with fixed step size at different temperature levels.

Table 1
Performance comparison of SA, CW, ST, PT, and AST[1]

# of samples	Time (s)				
	Simulated annealing	Cool-walking	Simulated tempering	Parallel tempering	Accelerated simulated tempering
10^7	302	12480	6746	7209	904
10^8	2946	113781	71532	76809	9002

[1] All experiments have been performed on a DEC Alpha DS10 6/466 with 256M memory.

is that by using larger steps at high temperatures, the system can easily escape from one minimum into another region of the energy landscape. The performance of the AST in this experiment is much better than the original ST algorithm and much closer to the performance of SA. Indeed, while the overall efficiency of the algorithms on a specific problem depends on the energy landscape, one can safely say that *in general*, SA has a faster computational speed compared to the ST, but often gets trapped by a local minimum when its temperature is reduced to a certain level. On the other hand, ST has a much lower chance to be trapped by the local minimum. AST inherits this property, but, at the same time, has a much better convergence rate compared to ST. The CW jump is costly, which makes the CW algorithm very time consuming. PT with various step sizes at different temperature levels yields computational results similar to AST's. While PT does not need to evaluate the normalization constants, it has to keep track though of the Markov chains at all temperature level. AST has to compute the fast constant evaluation at the beginning, which is a one time overhead. After that, the sample producing speed is faster than PT because it only needs to evolve with a single Markov chain.

4.2. Periodic energy landscape

Many real life systems have periodic energy landscapes, in which the minima are periodically distributed. An example of such system would be an extended polymer chain, whereupon angle of rotation around each bond energy of the whole chain goes through 3 local minima. As a result the energy landscape has at least 3N minima of approximately equal depth.

In such a system, the simple Metropolis–Hastings method or the SA methods are very sensitive to the step size. Figs. 5 and 6 show the sampling of an energy landscape with periodic minima using SA with different time steps. We also notice that the SA method is very sensitive to the transition step size. More precisely, upon using small time steps, the system is easily trapped in one of the local minima

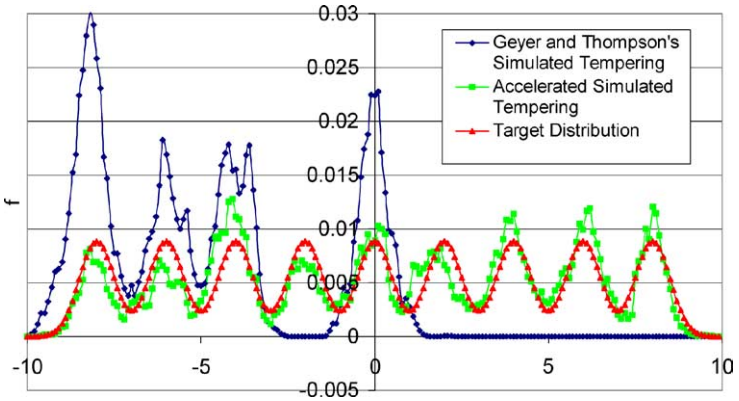

Fig. 5. Performance of the AST algorithm for a periodic energy landscape.

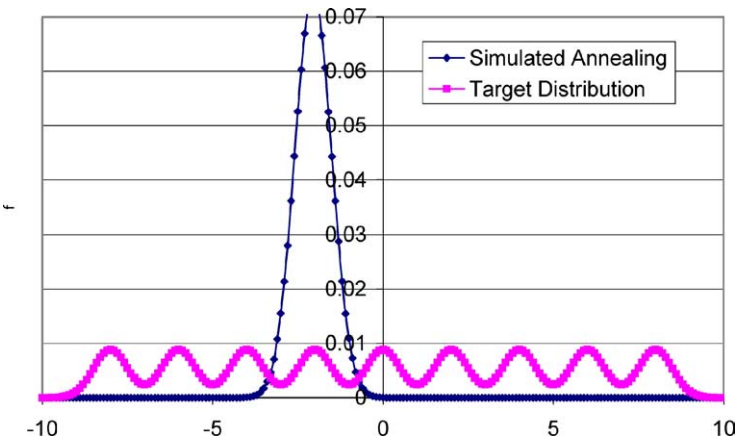

Fig. 6. Performance of the SA algorithm for a periodic energy landscape with small transition step size.

(Fig. 5), while upon using larger time steps, the system's acceptance rate is low and the system is essentially unable to escape the local minimum when the temperature is decreased below a certain low level (Fig. 6). On the contrary, AST has a fast convergence in the periodic energy landscape (Fig. 7).

4.3. Funnel-like energy landscape

The energy landscape of protein folding roughly resembles a funnel, with hierarchically disposed local minima, where the protein can transiently reside [13, 14,18]. Within such an energy landscape, the simple Metropolis–Hastings method or SA method can be extremely easily trapped in a local minimum.

In our experiment, we construct a target distribution with a rough funnel shape. Fig. 8 shows that the AST method converges fast to the target distribution, without getting trapped in a local minimum, in contradistinction, the SA algorithm is easily trapped, as illustrated in Fig. 9.

5. Conclusions

We presented a new stochastic algorithm for global optimization, namely the Accelerated Simulated Tempering (AST). This algorithm builds upon Geyer and Thompson's ST algorithm by incorporating a new idea and three complementary technical improvements based thereupon. The new, quite natural idea is to use different types and numbers of transition steps at different temperature levels. This idea is warranted by the fact that exploration at high temperatures could

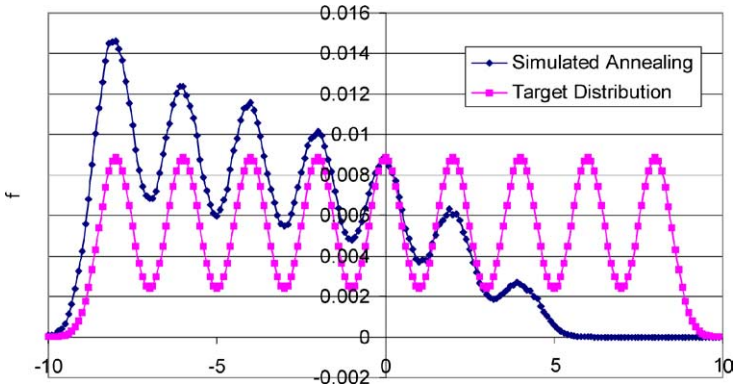

Fig. 7. Performance of the SA algorithm for a periodic energy landscape with a large transition step size.

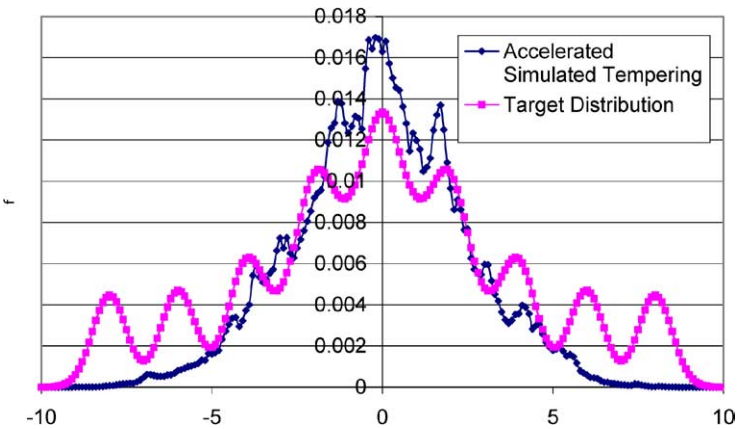

Fig. 8. Performance of the AST algorithm for a funnel-like energy landscape with 10^7 samples.

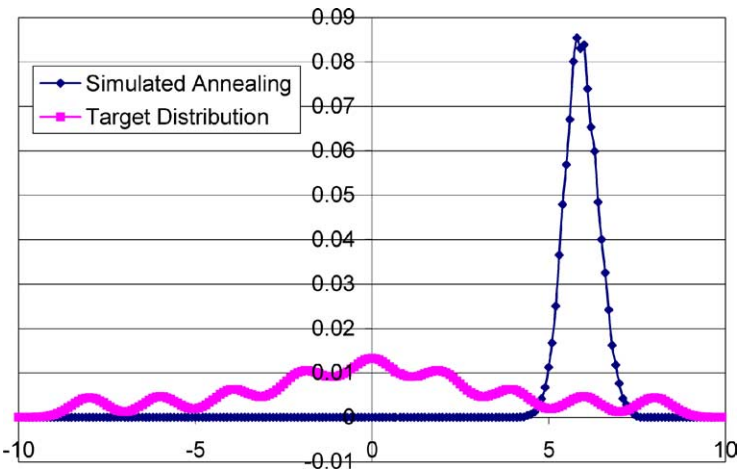

Fig. 9. Performance of the SA algorithm for a rough funnel-like energy landscape with 10^7 samples (note the difference between the vertical scales in Figs. 8 and 9: the target distribution is the same in both figures).

proceed fast, by ignoring local details, while at low temperatures a more accurate exploration is needed, which requires a larger number of shorter steps. The three technical improvements based on this idea are: (i) the temperature ladder with various temperature and transition steps; (ii) the leaned temperature ladder, which, during the transitions, favors the lower temperatures over the higher ones; and (iii) a fast recipe for computing the normalization constants. Our results in the experiments show that simply manipulating the temperature may not be sufficient to avoid the "waiting time dilemma." Strategy with larger step sizes at higher temperature and smaller step sizes at lower temperatures is key to escaping from deep local minima. Employing the techniques of random walks executed on a temperature ladder with various transition step sizes, application of AST to a few typical examples shows that the new algorithm consistently outperforms the Metropolis–Hastings, SA, ST, PT, and CW algorithms.

Acknowledgements

We acknowledge partial support from the US Department of Energy's Genomics: GTL program (www.doegenomestolife.org) under the project, "Carbon Sequestration in Synechococcus Sp.: From Molecular Machines to Hierarchical Modeling"; Laboratory Directed Research and Development Program of the Oak Ridge National Laboratory; and the Division of Materials Science and Engineering, US Department of Energy. O.R.N.L. is managed by UT-Battelle, LLC, under Contract DE-AC05-00OR22725.

References

[1] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.
[2] W.K. Hastings, Biometrika 57 (1970) 97.
[3] Z. Li, H.A. Scheraga, Proc. Natl. Acad. Sci. USA 84 (19) (1987) 6611.
[4] W.H. Wong, F. Liang, Proc. Natl. Acad. Sci. USA 94 (1997) 14220.
[5] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Science 220 (1983) 671.
[6] J.S. Liu, F. Liang, W.H. Wong, J. Am. Stat. Assoc. 96 (2001) 561.
[7] J.S. Liu, F. Liang, W.H. Wong, The use of multiple-try method and local optimization in Metropolis sampling, Technical report, Department of Statistics, Stanford University, 1998.
[8] E. Marinari, G. Parisi, Europhys. Lett. 19 (1992) 451.
[9] C.J. Geyer, E.A. Thompson, J. Am. Stat. Assoc. 90 (1995) 909.
[10] C. Berzuini, Medical monitoring, in: Markov Chain Monte Carlo in Practice, Chapman & Hall/CRC, 1996, Chapter 18.
[11] J.N. Onuchic, Z. Luthery-Schulten, P. Wolynes, Annu. Rev. Phys. Chem. 48 (1997) 545.
[12] C. Levinthal, Mössbauer spectroscopy in biological systems, in: Proceedings of a Meeting held at Allerton House, Monticello, IL, University of Illinois Press, 1969, p. 22.
[13] J.D. Bryngelson, J.N. Onuchic, N.D. Socci, P.G. Wolynes, Proteins: Struct. Funct. Gen. 21 (1995) 167.
[14] L. Piela, J. Kostrowicki, H.A. Scheraga, J. Phys. Chem. 93 (1989) 3339.
[15] J.S. Liu, Markov chain Monte Carlo and related topics, Technical report, Department of Statistics, Stanford University, 1999.
[16] J.S. Liu, Metropolis Algorithm and Beyond, in: Monte Carlo Strategies in Scientific Computing, Springer-Verlag, 2001.
[17] Y. Li, V. Protopopescu, A. Gorin, in preparation.
[18] J.N. Onuchic, Z. Luthey-Schulten, P. Wolynes, Annu. Rev. Phys. Chem. 48 (1997) 545.
[19] S. Brown, T. Head-Gordon, J. Comput. Chem. 24 (2003) 68.
[20] M. Falcioni, M.W. Deem, J. Chem. Phys. 110 (1999) 1754.
[21] J.M. Hammersley, D.C. Handscomb, Monte Carlo Methods, Spottiswoode, Ballantyne, 1964.