

Single Temperature for Monte Carlo Optimization on Complex Landscapes

Denis Tolkunov^{1,2,*} and Alexandre V. Morozov^{1,2,†}

¹*Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA*

²*BioMaPS Institute for Quantitative Biology, Rutgers University, Piscataway, New Jersey 08854, USA*

(Received 1 February 2012; revised manuscript received 30 March 2012; published 20 June 2012)

We propose a new strategy for Monte Carlo (MC) optimization on rugged multidimensional landscapes. The strategy is based on querying the statistical properties of the landscape in order to find the temperature at which the mean first passage time across the current region of the landscape is minimized. Thus, in contrast to other algorithms such as simulated annealing, we explicitly match the temperature schedule to the statistics of landscape irregularities. In cases where these statistics are approximately the same over the entire landscape or where nonlocal moves couple distant parts of the landscape, a single-temperature MC scheme outperforms any other MC algorithm with the same move set. We also find that in strongly anisotropic Coulomb spin glass and traveling salesman problems, the only relevant statistics (which we use to assign a single MC temperature) are those of irregularities in low-energy funnels. Our results may explain why protein folding is efficient at constant temperature.

DOI: [10.1103/PhysRevLett.108.250602](https://doi.org/10.1103/PhysRevLett.108.250602)

PACS numbers: 05.10.Ln, 02.50.Ey, 02.60.Pn, 02.70.Tt

Numerous problems in science and technology such as protein structure prediction, evolution on fitness landscapes, stochastic dynamics of complex systems, and machine learning require efficient global optimization of multivariate objective functions or “energies.” The objective function $U(\vec{x})$ can be viewed as a multidimensional landscape in which a certain quantity (potential energy, free energy, cost function, likelihood of a model) is assigned to every configuration \vec{x} of an arbitrary number \mathcal{D} of discrete or continuous state variables. The optimization task is then to find a global minimum (or maximum) on arbitrary landscapes as efficiently as possible. Since exact global optimization methods are not available, various empirical approaches have been devised. A popular class of algorithms is based on the Metropolis Monte Carlo (MC) scheme [1]. This class includes the simulated annealing (SA) algorithm [2], as well as simulated tempering [3], parallel tempering [4], replica exchange [5], ensemble MC schemes [6,7], and multicanonical MC schemes [8]. Non-Metropolis schemes for global optimization, such as random cost [9] and genetic algorithms [10], have also been developed. Another class of algorithms enables more efficient exploration of the novel regions of the configuration space by making adaptive changes to the landscape [11–14].

Unfortunately, the empirical nature of these algorithms makes it impossible to predict which approach would perform best on a given problem. In addition, most algorithms depend on *ad hoc* adjustable parameters such as the SA cooling schedule. Here, we address these concerns by proposing a universal guiding principle for analyzing global optimization problems. Our interest is not only in developing efficient, parameter-free global optimization schemes, but also in understanding how stochastic simulations run by nature (such as protein folding driven by

thermal fluctuations at constant temperature) appear to be so much simpler than corresponding human-designed algorithms.

Our intuition is based on the notion of the global (negative) gradient that leads towards good solutions [Fig. 1(a), upper panel]. Landscapes without such a gradient are of the golf-course type or even the “misleading” type in which the negative gradient leads away from the global minimum [Fig. 1(a), middle and lower panels]. In the golf-course and misleading scenarios it is necessary to sample $O(N^{\mathcal{D}})$ possible states, where N is the number of distinct values adopted by (discretized) state variables. In contrast, global gradients define funnels on the landscape that can in principle be traversed in $O(N)$ steps, making efficient optimization possible. A famous problem of this kind is protein folding [15], but any landscape in which gradual improvements lead toward a good solution will have the funnel

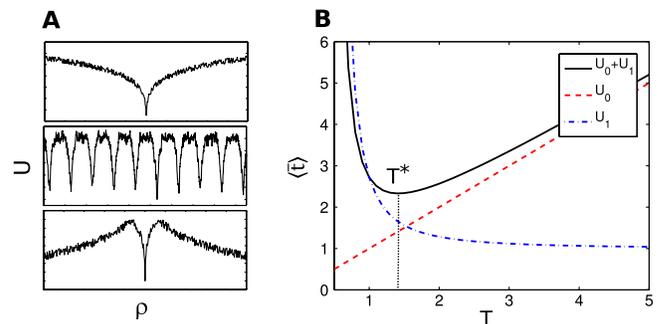


FIG. 1 (color online). (a) Top to bottom: Examples of funnel, golf-course, and misleading landscapes. (b) $\langle \bar{t} \rangle$ as a function of T for $U_0(x) = x$ and $U_1 = \mathcal{N}(0, 1)$. Black (solid) curve: diffusion in $U_0 + U_1$, $T^* = \sqrt{2}$. Red (dashed) curve: diffusion in U_0 , $T^* = 0$. Blue (dash-dotted) curve: diffusion in U_1 , $T^* = \infty$. Note that we have set $b - x = DF$, $\bar{t}_0(x, 0) = 1$ for simplicity.

structure. However, in realistic problems the global gradient will be weak and obscured by the local “noise” or irregularities in the objective function (after all, in the absence of such noise any local optimizer would be successful). As a result, the global gradient will be invisible at the smallest scale of a single MC step and can only be detected from the average over a macroscopic local region. If this region is still relatively small, the global gradient will be approximately constant over it. Furthermore, in strongly anisotropic problems the gradient may not be present everywhere but only in narrow low-energy valleys, whereas the high plateaus surrounding the valleys will be of the golf-course or misleading type.

Thus the global optimization problem can be formulated as diffusion (i.e., Metropolis MC sampling) in a potential which consists of random fluctuations with arbitrary magnitude and correlation length superimposed onto a weak, approximately linear term. Note that global optimization is different from computing thermodynamic properties, which requires at least partial equilibration and detailed balance. In contrast, global optimization is a strongly nonequilibrium process, with diffusion at any point in the simulation affected only by the landscape features in its immediate neighborhood.

In the absence of random fluctuations, diffusion in a local region \mathbf{L} subject to the constant force $\vec{\mathbf{F}} = -\partial U/\partial \vec{\mathbf{x}}$ is described by the Fokker-Planck (FP) equation in \mathcal{D} dimensions

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial \vec{\mathbf{x}}^2} - \vec{\mathbf{v}} \frac{\partial \rho}{\partial \vec{\mathbf{x}}}, \quad (1)$$

where $\rho(\vec{\mathbf{x}}, t)$ is the probability distribution, D is the diffusion constant, and $\vec{\mathbf{v}} = D\beta\vec{\mathbf{F}}$ is the drift velocity ($\beta = 1/T$ is the inverse temperature). We choose a coordinate system in which one of the axes (x) is parallel to $\vec{\mathbf{F}}$. In this system, Eq. (1) factorizes into a one-dimensional (1D) FP equation with a linear potential and $\mathcal{D} - 1$ FP equations describing free diffusion. We place an absorbing boundary $\perp \vec{\mathbf{F}}$ and focus on the 1D FP equation with the linear potential.

The speed of propagation along $\vec{\mathbf{F}}$ is characterized by the mean first passage time (MFPT) $\bar{t}(x, \beta)$, defined as the mean time required for a particle starting at x to reach the absorbing boundary b for the first time. With a reflecting boundary at a ($a < x < b$), the MFPT is given by [16,17]

$$\bar{t}(x, \beta) = \frac{1}{D} \int_x^b dx' e^{\beta U(x')} \int_a^{x'} dx'' e^{-\beta U(x'')}, \quad (2)$$

where $U(x)$ is the 1D potential (objective function) in the direction of $\vec{\mathbf{F}}$. For a linear potential in the absence of noise, $U_0(x) = F(b - x)$, where $F \equiv |\vec{\mathbf{F}}| > 0$, we obtain

$$\bar{t}_0(x, \beta) = \frac{1}{D\beta F} \left[(b - x) + \frac{1}{\beta F} (e^{-\beta F(b-a)} - e^{-\beta F(x-a)}) \right]. \quad (3)$$

In the $\beta \rightarrow 0$ (or $F \rightarrow 0$) limit the free-diffusion expression is recovered [16]: $\bar{t}_0(x, 0) = (1/2D)[(b - a)^2 + (x - a)^2]$. However, for finite β and F we can always set a so that the exponential terms on the right-hand side of Eq. (3) are vanishingly small, eliminating the effect of the reflecting boundary on the diffusion process (formally, we take the $a \rightarrow -\infty$ limit)

$$\bar{t}_0(x, \beta) = \frac{b - x}{v_x}, \quad (4)$$

where $v_x = D\beta F$. Now, assuming that the potential consists of the regular part and the irregular part, $U(x) = U_0(x) + U_1(x)$, and that the characteristic length scale of U_1 , l_c , is much smaller than the size \mathcal{L} of the 1D region over which U_0 is approximately linear, we take a spatial average over the irregular part [18]

$$\begin{aligned} \langle \bar{t}(x, \beta) \rangle &= \frac{1}{D} \int_x^b dx' \int_{-\infty}^{x'} dx'' e^{\beta F(x'' - x')} \\ &\times \int_{x' - L/2}^{x' + L/2} dx_1 e^{\beta U_1(x_1)} \int_{x'' - L/2}^{x'' + L/2} dx_2 e^{-\beta U_1(x_2)}, \end{aligned} \quad (5)$$

where $l_c \ll L \ll \mathcal{L}$. Under these conditions and the additional assumption that U_1 statistics does not change over \mathcal{L} , the two spatial averages are independent of each other and of x', x'' , yielding

$$\langle \bar{t}(x, \beta) \rangle = H(\beta) \bar{t}_0(x, \beta), \quad (6)$$

where $H(\beta) = \int_{-\infty}^{\infty} dU_1' P(U_1') e^{\beta U_1'} \int_{-\infty}^{\infty} dU_1'' P(U_1'') e^{-\beta U_1''}$ [we have switched from x_1 and x_2 to $U_1' \equiv U_1(x_1)$ and $U_1'' \equiv U_1(x_2)$ in the spatial averages]. Furthermore, $H(\beta) = \int_{-\infty}^{\infty} d\Delta P(\Delta) e^{\beta \Delta}$, where $P(\Delta)$ is the distribution of $\Delta = U_1' - U_1''$ for x_1, x_2 constrained by $|x_1 - x_2| \gg l_c$. The last condition guarantees that $P(\Delta)$ is independent of $|x_1 - x_2|$.

Clearly, if MFPT along $\vec{\mathbf{F}}$ is minimized for all local regions \mathbf{L} with the constant gradient, the total time to reach a good solution will also be minimized. The inverse temperature β^* that minimizes MFPT is given by

$$\left. \frac{dH(\beta)}{d\beta} \right|_{\beta^*} = \frac{H(\beta^*)}{\beta^*}. \quad (7)$$

Note that β^* is independent of F . Equation (7) can be used to find β^* numerically for arbitrary $P(\Delta)$. If $P(U_1) = \mathcal{N}(0, \sigma^2)$, $P(\Delta) = \mathcal{N}(0, 2\sigma^2)$ and $H(\beta) = e^{\beta^2 \sigma^2}$, yielding $T^* = \sqrt{2}\sigma$. With $T \ll T^*$ the diffusing particle gets stuck in local minima ($\langle \bar{t} \rangle \sim T e^{\sigma^2/T^2}$), while for $T \gg T^*$ the diffusion is no longer optimally along the gradient of U_0 and $\langle \bar{t} \rangle \sim T$ [Fig. 1(b)]. $\langle \bar{t}(x, \beta^*) \rangle \sim \sigma/F$, indicating

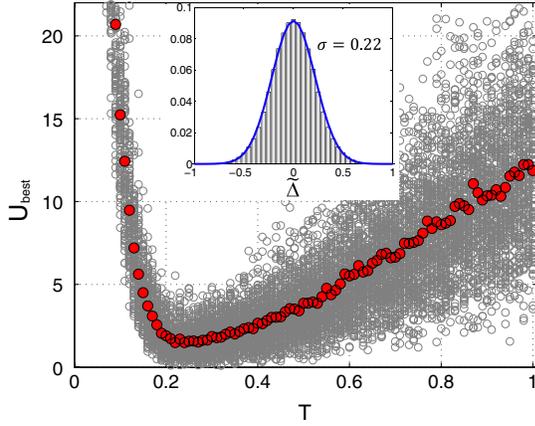


FIG. 2 (color online). Distribution of best predicted energies as a function of temperature for the 4D Griewank function (Table I). For each T , N_{trials} independent trajectories with N_{iter} MC steps each (both accepted and rejected) were created by Metropolis MC sampling. The lowest energy U_{best} from each trajectory is shown as a gray open circle. Red filled circles are the average of U_{best} at a given T . Inset: Histogram of energy differences from a random sample of the landscape with a Gaussian fit (blue solid curve).

that diffusion is impeded by noise and aided by the gradient. If $U_1 = 0$ everywhere, $P(\Delta) = \delta(\Delta)$ and $T^* = 0$. Thus, as expected, the optimal solution in the absence of noise is to roll down the potential at zero temperature. However, $\bar{t}_0(x, \infty) = 0$ since Eq. (2) does not accurately describe ballistic dynamics or strong forces. Finally, if $F \rightarrow 0$, $\langle \bar{t} \rangle \sim e^{\sigma^2/T^2}$ then $T^* = \infty$ [although any $T > \sqrt{2}\sigma$ will work almost as well, Fig. 1(b)], yielding $\langle \bar{t}(x, 0) \rangle = \bar{t}_0(x, 0)$.

In order to find the best MC temperature T^* , we need to estimate $P(\Delta)$ by sampling. Let us consider $P(\tilde{\Delta})$, where $\tilde{\Delta} = U(\tilde{\mathbf{x}} + \Delta\tilde{\mathbf{x}}) - U(\tilde{\mathbf{x}})$ and $\Delta\tilde{\mathbf{x}}$ is a single MC step of constant length. $|\Delta\tilde{\mathbf{x}}|$ can be made so small that the contribution from U_0 is negligible. With uncorrelated noise ($|\Delta\tilde{\mathbf{x}}| > l_c$), $P(\tilde{\Delta}) = P(\Delta)$ and Eq. (7) can be applied immediately. However, if $|\Delta\tilde{\mathbf{x}}| < l_c$, U_1 is smooth at the scale of a single MC step, and $P(\tilde{\Delta})$ and T^* will depend on

the move set. Indeed, $T^* \sim |\Delta\tilde{\mathbf{x}}|$ if the MC steps are so fine that $U_1(\tilde{\mathbf{x}} + \Delta\tilde{\mathbf{x}}) - U_1(\tilde{\mathbf{x}})$ is approximately linear. Nonetheless, we find that for complex landscapes where $U(\tilde{\mathbf{x}})$ is a sum of many independent terms, $P(\tilde{\Delta})$ quickly adopts a Gaussian shape if the sampling is over a region $\gg l_c$. Since MC walks are memoryless, $T^* = \sqrt{2}\sigma$ still holds but now σ depends on the move set. As $|\Delta\tilde{\mathbf{x}}|$ increases beyond l_c , σ converges to a universal value.

We have tested our approach on a set of standard functions often used to check performance of global optimization algorithms [19] (Fig. 2, Table I). To estimate $P(\tilde{\Delta})$, we use a single trajectory with 10^3 (G), 5×10^4 (R), and 5×10^2 (A) random steps (all steps are accepted). These parameters ensure that $P(\tilde{\Delta})$ is close to a Gaussian, with the rate of convergence dependent on the long-range order in the landscape and on the complexity of the potential function. The histograms yield accurate predictions of T^* (Table I) despite the fact that the landscapes are correlated and anisotropic and the gradient is not guaranteed to be weak.

To compare our approach with SA, we have carried out two additional sets of runs with $N_{\text{trials}} = 5 \times 10^2$ and $N_{\text{iter}} = 5 \times 10^4$, 1.5×10^4 , 5×10^3 for G , R , and A functions, respectively. In one set the MC temperature was fixed at T_{comp}^* , while in the other SA was employed with the initial temperature $T_i = 2.0, 8.0, 2.5$ (G, R, A) and the final temperature $T_f = 0$. T_i was chosen as the difference between the maximum and the minimum values of the noise term in each test function. If the global minimum is reached N_{suc} times in a given set of runs, $\alpha \equiv N_{\text{trials}}N_{\text{iter}}/N_{\text{suc}}$ gives the average number of iterations per success (if $N_{\text{suc}} = 0$, a lower bound of $N_{\text{trials}}N_{\text{iter}}$ is obtained). We find that T^* runs are superior to SA in all cases (Table I).

Next, we turn to two more challenging global optimization problems: the Coulomb spin glass (CSG) [6] and the traveling salesman (TS) problem [2]. With CSG, we place $N = 50$ charges randomly within a 3D unit cube: $U(\tilde{\mathbf{s}}) = \sum_{i=1, i \neq j}^N \sum_{j=1}^N s_i s_j / |\tilde{\mathbf{r}}_i - \tilde{\mathbf{r}}_j|$, where $s_i = \pm 1$ and the charge positions are fixed. A move involves flipping all signs in a randomly chosen subset of charges. The CSG

TABLE I. Predicted and computed optimal temperatures for standard test functions [19]: the 4D Griewank (G) function ($U(\tilde{\mathbf{x}}) = 1 + \frac{1}{4000} \sum_{i=1}^4 x_i^2 - \prod_{i=1}^4 \cos(\frac{x_i}{\sqrt{i}})$, $x_i \in [-600, 600]$, $\forall i$), the 4D Rastrigin (R) function ($U(\tilde{\mathbf{x}}) = 4 + \sum_{i=1}^4 [x_i^2 - \cos(18x_i)]$, $x_i \in [-5, 5]$, $\forall i$), and the 4D Ackley (A) function ($U(\tilde{\mathbf{x}}) = 20 + e - 20 \exp(-0.2 \sqrt{\frac{1}{4} \sum_{i=1}^4 x_i^2}) - \exp[\frac{1}{4} \sum_{i=1}^4 \cos(2\pi x_i)]$, $x_i \in [-32.8, 32.8]$, $\forall i$). All three functions have multiple local minima and a single global minimum located at $\tilde{\mathbf{x}} = 0$ [$U(0) = 0$]. Each MC step is taken in a random direction and has a constant length of 1.0 (G), 0.05 (R), and 0.3 (A). T_{pred}^* is based on a Gaussian fit to the histogram of $\tilde{\Delta}$. T_{comp}^* is the temperature at which the average U_{best} is at minimum. U_{best} distribution at each T is based on N_{trials} Metropolis MC runs with N_{iter} steps each. $\alpha(T^*)$ and $\alpha(\text{SA})$ are the average number of iterations required to find the global minimum using MC sampling with T_{comp}^* and SA, respectively.

| Function | N_{iter} | N_{trials} | T_{pred}^* | T_{comp}^* | $\alpha(T^*)$ | $\alpha(\text{SA})$ |
|----------|-------------------|---------------------|---------------------|---------------------|-------------------|---------------------|
| G | 1.5×10^4 | 10^2 | 0.22 | 0.22 | 4.2×10^6 | $> 2.5 \times 10^7$ |
| R | 5×10^3 | 10^2 | 0.85 | 0.90 | 3.1×10^5 | 2.5×10^6 |
| A | 5×10^3 | 10^2 | 0.40 | 0.45 | 6.1×10^3 | 1.9×10^4 |

problem is characterized by the separation of scales: $P(\tilde{\Delta})$ estimated using unconstrained random walks (as was done for the test functions) yields a very high temperature since most of the landscape consists of high-energy plateaus that are either flat or have gradients pointing in random directions [Fig. 3(a), upper panel]. MC runs at this temperature would not be able to utilize the global gradient information, which is restricted to low-energy funnels. We therefore focus on the funnels to estimate $P(\tilde{\Delta})$: from the current position with U_{cur} , up to N_m (2×10^4 for CSG) random moves are attempted. If the new state is found with $U_{\text{new}} \leq U_{\text{cur}}$, the loop terminates and the new state becomes the current state. Otherwise, the lowest energy among N_m new energies is chosen. For CSG, $P(\tilde{\Delta})$ is based on 10^3 accepted states, which required $\approx 1.3 \times 10^7$ iterations. The resulting histogram [Fig. 3(a), lower panel] correctly predicts the optimal temperature seen in Fig. 3(b), even

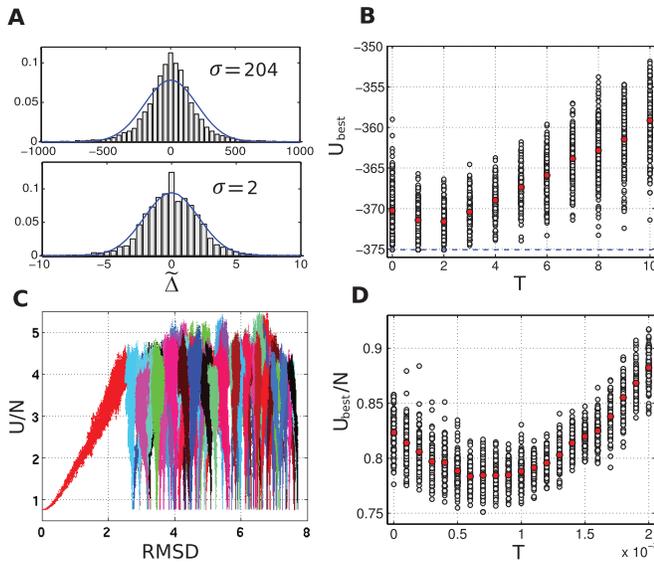


FIG. 3 (color). (a) $P(\tilde{\Delta})$ estimated with unconstrained random walks (10^2 trials with 10^4 steps each) (upper panel) and with the funnel-sampling algorithm (lower panel). (b) Distribution of U_{best} (gray open circles) as a function of temperature for CSG ($N_{\text{trials}} = 2 \times 10^2$, $N_{\text{iter}} = 5 \times 10^5$). Red filled circles are the average of U_{best} at each T . The dashed horizontal line is the lowest energy found in all runs. This energy has been reached two, six, and five times at $T = 0, 1, 2$, respectively, yielding $\alpha(T^* = 2) = 2 \times 10^7$ iterations. SA runs ($T_i = 200$, $T_f = 0$) with $N_{\text{trials}} = 2 \times 10^2$ and $N_{\text{iter}} = 1 \times 10^6$ have failed to find the lowest energy, yielding $\alpha^{\text{SA}} > 2 \times 10^8$ iterations. (c) Multifunnel structure of the TS landscape. U/N is the average distance between neighboring cities in a given trajectory. One hundred best minima were chosen from (d), and for each minimum the funnel was mapped out using 10 random walks with 2×10^4 local steps each and plotted in a distinct color. Local steps involve exchanging two randomly picked neighboring cities. The rms deviation (RMSD) is computed with respect to the best solution in (d). (d) Distribution of U_{best}/N as a function of temperature for the TS problem ($N_{\text{trials}} = 10^2$, $N_{\text{iter}} = 2.5 \times 10^5$).

though the moves are no longer of constant length. Its Gaussian shape suggests that the $\tilde{\Delta}$ distribution is isotropic in the funnels. Surprisingly, even $T = 0$ simulations yield reasonable results, indicating that some deep funnels are smooth.

In the TS problem, one is given a list of cities and their locations, and the goal is to find the shortest possible tour that visits each city exactly once. We considered $N = 180$ cities randomly distributed within a $N^{1/2} \times N^{1/2}$ square so that the average distance between neighboring cities is independent of N [2]. We use Euclidean distances to compute U/N and employ nonlocal moves in which a segment of the trajectory is chosen at random and the direction in which all cities within that segment are traversed is inverted [2]. To reduce the degeneracy of low-scoring solutions, we start all trajectories from the same city. The TS landscape has a complex multifunnel structure [Fig. 3(c)] with high plateaus that dominate the landscape so that only $\tilde{\Delta}$ statistics within the funnels are relevant. As in CSG, we use the funnel-sampling algorithm (with $N_m = 2 \times 10^3$ and 10^4 accepted states) to obtain $\sigma = 7.6 \times 10^{-4}$. This value is consistent with Fig. 3(d). Additional fixed- T sampling ($T^* = 7 \times 10^{-4}$, $N_{\text{trials}} = 10^3$, $N_{\text{iter}} = 5 \times 10^5$) has found 32 solutions with $U_{\text{best}}/N \leq 0.755$. In contrast, SA runs ($T_i = 1/\sqrt{N}$, $T_f = 0$) [2] with the same N_{trials} and N_{iter} have not found any solutions with $U_{\text{best}}/N \leq 0.9$.

Throughout this Letter, we have focused on minimizing MFPT. Instead, one may want to maximize the fraction of runs with U_{best} below a certain cutoff. The tail of the U_{best} distribution at a given T is affected by both its mean and standard deviation σ' , making it possible that the temperature with the best mean is not the same as the temperature optimized for yielding extremely low-energy solutions. However, Figs. 2, 3(b), and 3(d) show that σ' varies with T rather slowly. As a result, the MFPT-based T^* remains valid, although in some cases the interplay between the mean and σ' may make temperatures in a small range around T^* equally acceptable.

If U_1 statistics is approximately constant and isotropic either throughout the landscape or in the low-energy funnels, there is a unique MC temperature T^* for the most efficient optimization of the objective function (the anisotropic case will be presented elsewhere). However, if the nature of irregularities changes across the landscape, two scenarios are possible: First, if $P(U_1)$ stays approximately the same in regions $\gg L$, the best temperature can be found for each such region but needs to be updated periodically as the landscape is traversed. Mixing statistics from multiple regions will yield a single T^* that will not be the absolute best solution but may still be a good approximate one. Second, it is possible that different scales are mixed in a region $\ll L$, e.g., due to anisotropy. In this case $P(\tilde{\Delta})$ will be non-Gaussian but Eq. (7) still applies, yielding a single T^* . Using a single temperature works

especially well with nonlocal steps such as those employed in the TS problem, which can traverse a sizable part of the landscape in a single leap. Indeed, we find that even a multiscale TS problem, in which cities are clustered rather than randomly distributed [2], is characterized by a unique best temperature with nonlocal steps.

Since MFPT has a unique global minimum at $T = T^*$ [Fig. 1(b)], all other schemes such as SA will yield sub-optimal performance, as confirmed by our comparisons between SA and fixed- T runs. In fact, if the amplitude of U_1 gradually increases with decreasing U_0 , our prescription calls for increasing the temperature as the simulation progresses—the exact opposite of the SA cooling schedule. The SA algorithm assumes *a priori* that all scales are present in the problem. Querying some of the landscape statistics allows us to substantially improve on this “one size fits all” SA technique by matching a given landscape and move set to the appropriate temperature(s). We look forward to applying our approach to protein structure prediction and other global optimization challenges.

This research was supported by National Institutes of Health (HG 004708) and by an Alfred P. Sloan Research Fellowship to A. V. M.

*Present address: The Cancer Institute of New Jersey, UMDNJ, New Brunswick, NJ 08903, USA.

†Corresponding author.
morozov@physics.rutgers.edu

- [1] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [2] S. Kirkpatrick, C. Gelatt, Jr., and M. Vecchi, *Science* **220**, 671 (1983).
- [3] E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
- [4] U. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- [5] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [6] F.-M. Dittes, *Phys. Rev. Lett.* **76**, 4651 (1996).
- [7] B. Hesselbo and R. B. Stinchcombe, *Phys. Rev. Lett.* **74**, 2151 (1995).
- [8] B. Berg and T. Neuhaus, *Phys. Lett. B* **267**, 249 (1991).
- [9] B. Berg, *Nature (London)* **361**, 708 (1993).
- [10] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison Wesley, Reading, 1989).
- [11] J. Barhen, V. Protopopescu, and D. Reister, *Science* **276**, 1094 (1997).
- [12] W. Wenzel and K. Hamacher, *Phys. Rev. Lett.* **82**, 3003 (1999).
- [13] K. Hamacher, *Europhys. Lett.* **74**, 944 (2006).
- [14] D. Cvijović and J. Klinowski, *Science* **267**, 664 (1995).
- [15] J. Bryngelson, J. Onuchic, N. Socci, and P. Wolynes, *Proteins: Struct. Func. Genet.* **21**, 167 (1995).
- [16] S. Lifson and J. Jackson, *J. Chem. Phys.* **36**, 2410 (1962).
- [17] G. Weiss, *Adv. Chem. Phys.* **13**, 1 (1967).
- [18] R. Zwanzig, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2029 (1988).
- [19] A. Törn and A. Žilinskas, *Global Optimization* (Springer-Verlag, Berlin, 1989).