

## TRANSFORMATIONS FOR ACCELERATING MCMC SIMULATIONS WITH BROKEN ERGODICITY

Mark Fleischer

ULTRANETX, LLC.  
5308 Nightshade Ct.  
Columbia, MD 21045, U.S.A.

### ABSTRACT

A new approach for overcoming broken ergodicity in Markov Chain Monte Carlo (MCMC) simulations of complex systems is described. The problem of broken ergodicity is often present in complex systems due to the presence of deep “energy wells” in the energy landscape. These energy wells inhibit the efficient sampling of system states by the Metropolis Algorithm thereby making estimation of the Boltzmann Partition Function (BPF) more difficult. The approach described here uses transformation functions to create a family of modified or smoothed energy landscapes. This permits the Metropolis Algorithm and the MCMC approach to sample system states in a way that leads to accurate estimates of a *modified* BPF (mBPF). Theoretical results show how it is then possible to extrapolate from this mBPF to the BPF value associated with the original landscape with a small absolute error. Computational examples are provided.

### 1 INTRODUCTION

A recurring problem in statistical mechanics and simulations of complex systems is the computation of the Boltzmann Partition Function (BPF) (Technically, this is more accurately denoted as the *canonical partition function* where the exponential involves the Boltzmann constant.)  $Z(t) = \sum_{i=1}^N e^{-f_i/t}$  ( $N$  is the number of system states), a value related to many quantities of interest in the study of large ensembles of interacting entities. Among other things, this value serves as the normalizing constant in expressions involving the stationary probability of the system in question being in state  $i$  with corresponding energy value  $f_i$ . It is often useful to estimate this stationary probability using Markov Chain Monte Carlo (MCMC) simulation techniques in conjunction with the Metropolis Algorithm (Metropolis et al. 1953). The Metropolis Algorithm governs the transition probabilities of the simulation and hence, how it moves from one state to another. These transition probabilities are affected by the “energy” levels of the states. This seminal work,

developed over 50 years ago, has provided a mathematically sound basis for estimating the stationary probability of any particular state.

Such estimates, if they are accurate, provide valuable insights regarding the frequency of certain interesting or rare events. In many problems however, obtaining accurate estimates of the stationary probabilities (or related values) is impaired because of the nature of the energy landscape. Sometimes there are deep pits or “energy wells” in the energy landscape. Once the state of the system reaches these deep energy wells, the probability of escaping them (transitioning to other sets of states) becomes exceedingly small resulting in very biased estimates of the state probabilities. Under these circumstances, repetitions of such a simulation with different initial conditions often yield vastly different estimates of the stationary probabilities. This situation is often referred to as the *broken ergodicity problem*.

One straightforward approach for addressing this problem is to compensate for the small escape probabilities by requiring very long simulations. Long runs provide a sufficient number of opportunities for the simulation to escape from the energy wells and thereby yield output statistics more in line with the theoretical probabilities.

Many other methods exist however that overcome this problem *without* such inordinately long simulation runs. The basic approaches usually involve modifying the MCMC technique or the Metropolis Algorithm in some way for “problem” energy landscapes. In this article, the broken ergodicity problem is addressed differently—the approach described here utilizes transformations of the energy landscape itself while leaving the essential features of the MCMC technique and the Metropolis Algorithm unchanged. These transformation functions *smooth out* the energy landscape and thereby enable the MCMC technique to avoid or reduce the possibility of “getting stuck” in deep energy wells. Section 2 provides the necessary background material on the BPF, the MCMC approach, and the broken ergodicity problem. Section 3 then describes the motivation and mathematical foundations of the energy transformation functions. Of course, smoothing out the landscape changes the stationary

probabilities from what the original landscape would yield in an MCMC simulation. These transformed energy landscapes yield a *modified* BPF (mBPF) value. Consequently, some way must be developed that relates the mBPF values to the BPF values associated with the original landscape. Section 4.1 describes some theoretical results pertaining to such relationships. Section 5 describes computational experiments that provide examples of the techniques described here. Finally, Section 6 provides some concluding remarks and directions of future research.

## 2 BACKGROUND

The development of the Metropolis Algorithm spawned a great deal of progress in many diverse fields—from condensed matter physics to optimization techniques such as simulated annealing (Fleischer 1995)—by providing a simple mechanism for the computer simulation of many complex phenomena (Conference: 2003). The basic machinery of the Metropolis Algorithm is the *Metropolis Acceptance Criterion* that governs movement from state  $i$  with value  $f_i$  to a ‘candidate’ state  $j$  with value  $f_j$ :

$$\Pr\{\text{Accept state } j \mid \text{Current state} = i\} = \begin{cases} e^{-\Delta f_{ji}/t} & \text{if } \Delta f_{ji} = f_j - f_i > 0 \\ 1 & \text{if } f_j - f_i \leq 0 \end{cases} \quad (1)$$

which leads to stationary probabilities

$$\pi_i(t) = \frac{e^{-f_i/t}}{Z(t)} \quad (2)$$

where  $Z(t) = \sum_i e^{-f_i/t}$  is the BPF and in (2) serves as the normalizing constant (Metropolis et al. 1953, Conference: 2003).

The BPF has a great deal of significance in statistical mechanics and “...contains all of the essential information about the system under consideration” (Landau and Binder 2000, p.7). It is therefore not surprising that a great deal of research has explored various ways for *accurately* estimating this value. Naturally, the Metropolis Algorithm has played a significant role in these efforts. If, for example, one can estimate the value in (2) for some state  $i$  with known energy  $f_i$ , then one can obtain an estimate of  $Z(t)$ . The accuracy of these estimates obviously depends on the degree to which the simulation state frequencies comports to the actual state probabilities. As noted above, for certain landscapes, the broken ergodicity problem makes the accurate estimation of  $Z(t)$  more computationally expensive. Figure 1 illustrates an energy landscape that might exhibit the broken ergodicity problem.

The point highlighted in the graphic is in such a deep energy well with neighboring solutions that have relatively very high energy values. Thus, the probability of moving

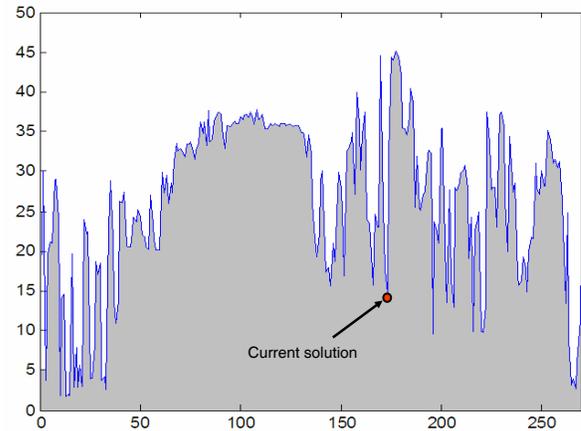


Figure 1: Illustration of the broken ergodicity. Problem.

to another state is very low and simulation based strictly on the Metropolis Algorithm and this landscape requires a very large number of iterations to accurately estimate  $Z(t)$ .

This problem is exacerbated when the temperature is lowered as the acceptance probability becomes exponentially smaller. Unfortunately for researchers in many disciplines that have use for such simulations, many of the more interesting *critical phenomena* such as *phase transitions* occur under just such low temperature conditions. As Straub and Andricioaei (2001) point out “[o]vercoming broken ergodicity is a non-trivial problem...” (p.193) and describe two general approaches for addressing it: 1) using tailored methods that utilize specific features of the particular system involved; and 2) more generalized methods that are more broadly effective. Several examples of the latter use the so-called *Wang-Landau Sampling* technique (Landau 2003). Topper et al. (2003) further describes other techniques to overcome what they describe as an “insidious problem” in referring to the “quasi-ergodicity” problem. They describe methods that involve modifications of the basic MCMC approach such as “Mag-Walking”, “Subspace Sampling”, the “Jump Between Wells Method”, the “J-Walking” method and several others. Importance sampling methods have also been widely used (Straub and Andricioaei 2001, Landau and Binder 2000).

All these generalized approaches involve significant changes to the MCMC approach or involve significant ancillary computation of other quantities. Rather than using some specific modification of the MCMC technique, the approach described here is general and involves modifications of the entire energy landscape. To be technically correct, the approach described here is really a hybrid as it involves the basic MCMC machinery but, in effect, modifies the acceptance probabilities by virtue of the energy transformation functions. Other forms of energy landscape modifications are given by Tsallis (1988) which demonstrate power-law distributions. It is also worth noting that the approach described here is rooted in non-extensive statistical

mechanics (Fleischer 2005a). The next section describes an extrapolation concept that lies at the heart of the approach used here.

### 2.1 An Old “Smoothing” Technique—Increasing the Temperature

One approach for solving the broken ergodicity problem is to run simulations at high temperatures, where the landscape is effectively smoother, and using results thus obtained to estimate the BPF at low temperatures. This possibility is suggested in Figure 2 and illustrates the characteristic ‘S’ shaped curve for  $Z(t)$  plotted as a function of  $\log(t)$  (Landau and Binder 2000).

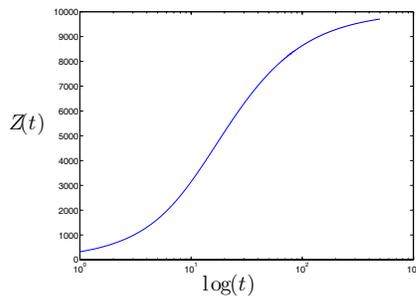


Figure 2: The Boltzmann partition function.

These curves typically exhibit regions of convexity and concavity, and hence the possibility of *extrapolating* to the value of  $Z(t)$  at a low temperature from a corresponding value at a high temperature where more accurate sampling is possible. This requires estimation of the derivative of  $Z(t)$ , a well-known quantity in statistical mechanics (Landau and Binder 2000).

This ‘extrapolation’ idea is illustrated in Figure 3 which highlights the relatively high temperature values at which the derivative must be estimated. Clearly, the only way for this extrapolation idea to work is if the correct temperature at which to estimate the derivative of  $Z(t)$  can be identified. Unfortunately, not all landscapes are alike—each landscape can have its own derivative point. Indeed, even when landscapes have the same size and a similar distribution of values, the derivative points will likely be different. Figure 3 illustrates this problem for two randomly generated landscapes (both have 10,000 energy values with the indicated ranges). The landscape associated with Figure 3a indicates that the log of this critical temperature is a bit less than 200 while for the landscape associated with Figure 3b the log of the critical temperature is something greater than 200. Obviously, this *Derivative Point Identification Problem* (DPIP) must be overcome for this idea to work. An approach for addressing the DPIP is presented in the following sections.

### 3 TRANSFORMATION FUNCTION PRINCIPLES

An important aspect in generating transformation functions is the capacity for creating a *family* of smoothed landscapes each of which is related to the original landscape. As will be shown, two such curves are required to provide an effective solution to the DPIP. The following three principles therefore provide the basis for formulating a *transformation function*.

**Principle 1—Parameterization:** Three parameters are defined: a smoothing parameter  $a \geq 0$ , a ‘family’ parameter  $m \geq 1$  and the temperature  $t > 0$ . Define a new function,  $\hat{f}_i(a, m, t)$  based on  $f_i$  and using the parameters of interest. This will often be denoted more simply by  $\hat{f}_i \equiv \hat{f}_i(a, m, t)$  or  $\hat{f}_i(a)$  where the functional dependencies are understood.

**Principle 2—Boundary Conditions:** A boundary condition is required that recovers the original value of  $f_i$ . Thus, for all  $i$ ,  $m \geq 1$ , and  $t > 0$  with  $a = 0$ ,  $\hat{f}_i(0, m, t) = f_i$ .

**Principle 3—Smoothing:** The third consideration is to establish the mechanisms for smoothing out the landscape. This obviously requires that the difference between any  $\hat{f}_j(a)$  and  $\hat{f}_i(a)$  should be decreased depending on the smoothing parameter  $a$ . It would also be useful if the high energy values were decreased to a greater degree than lower energy values. This would make escaping from very deep energy wells more likely.

With these considerations in mind, a relationship must be defined between the *rate change* of transformed energy values  $\hat{f}_i(a, m, t)$  with respect to  $a$  and the *magnitude* of  $\hat{f}_i(a, m, t)$ . The rate of decrease in  $\hat{f}_i(a, m, t)$  as  $a$  gets larger is negatively proportional to the size of  $\hat{f}_i(a, m, t)$ . To produce a *family* of relationships, the following differential equation is proposed:

$$\frac{\partial \hat{f}_i/t}{\partial a} = - \left( \frac{\hat{f}_i}{t} \right)^m. \quad (3)$$

#### 3.1 Defining the Transformation Function

Eq. (3) is solved by rearranging and integrating and keeping in mind the boundary condition  $\hat{f}_i(0, m, t) = f_i$ . Letting  $g_i(a) \equiv \hat{f}_i(a, m, t)/t$  to simplify notation and keeping just the parameter  $a$  as it is involved in the integration to follow, we solve the differential equation

$$\frac{\partial g_i(a)}{\partial a} = -(g_i(a))^m \Rightarrow \int_0^a \frac{\partial g_i(s)}{-g_i(s)^m} = \int_0^a \partial s = a. \quad (4)$$

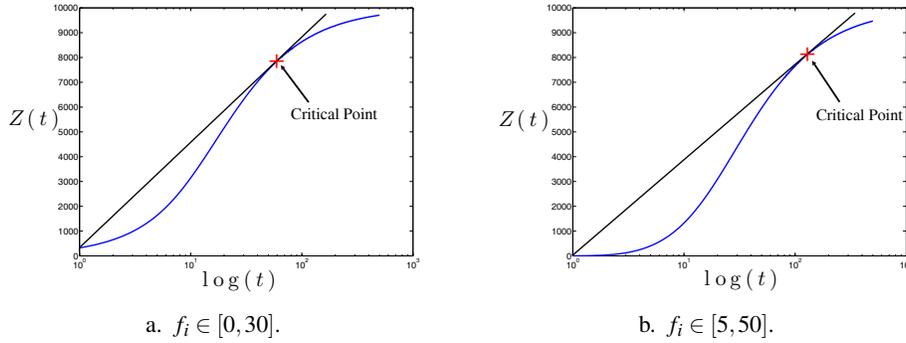


Figure 3: Illustration of the derivative point identification problem.

Thus,  $\frac{g_i(s)^{-m+1}}{m-1} \Big|_{s=0}^{s=a} = a$  and hence

$$\frac{1}{(m-1)g_i(a)^{m-1}} - \frac{1}{(m-1)g_i(0)^{m-1}} = a. \quad (5)$$

Simplifying (5), and substituting back  $\hat{f}_i(a, m, t)/t$  for  $g(a)$  and keeping in mind the boundary conditions (i.e.,  $g_i(0) = f_i$ ), we get the general energy landscape transformation function for all states  $i$ :

$$\hat{f}_i(a, m, t) = \frac{f_i}{[1 + a(m-1)(f_i/t)^{m-1}]^{1/(m-1)}}. \quad (6)$$

**Case  $m = 1$ :** The integrals in (4) yield the logarithmic form:  $-\ln g_i(s) \Big|_{s=0}^{s=a} = a$  and hence  $-\ln g_i(a) + \ln g_i(0) = a$ . This leads to

$$g_i(a) = e^{-a} g_i(0) \quad (7)$$

and therefore

$$\hat{f}_i(a, 1, t) = e^{-a} f_i. \quad (8)$$

Notice that (8) does not involve  $t$  as they cancel each other out on both sides of (7) whereas for  $m > 1$ , (6) does involve  $t$ . This same result in (8) can also be obtained from (6) and taking limits; i.e.,  $\lim_{m \rightarrow 1} \hat{f}_i(a, m, t) = e^{-a} f_i$ .

Insofar as computing the value of the BPF, observe that for  $\hat{f}_i(a, 1, t)$  denoted here as  $\hat{f}_i$ ,

$$\sum_{i=1}^N e^{-\hat{f}_i/t} = \sum_{i=1}^N e^{-c f_i/t} = \sum_{i=1}^N e^{-f_i/t'} \quad (9)$$

where  $c = e^{-a}$  and hence the effective temperature  $t' = t/c = t e^a$ . Thus, for  $m = 1$  the BPF value associated with the transformed landscape is equivalent to one obtained by raising the temperature a factor of  $e^a$  using the original landscape.

**Case  $m = 2$ :** In this case, (6) becomes

$$f_i(a, 2, t) = \frac{f_i}{1 + a f_i/t}. \quad (10)$$

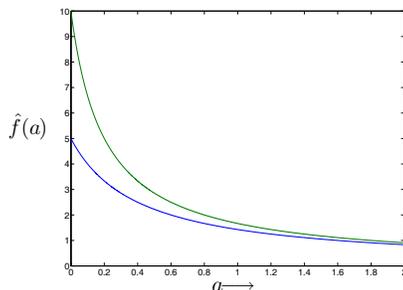
See Fleischer (2005b) where (10) forms the basis of a scale invariance structure in non-extensive systems.

Figures 4a and b illustrate the effects of these transformations and certain important monotonicity properties. Figure 4a highlights the smoothing effect using two energy values  $f = 5$  and  $f = 10$ . Notice that the two transformed energy values  $\hat{f}$  become closer in magnitude as  $a$  increases and intercept the  $y$ -axis at the untransformed values  $f = 5$  and  $f = 10$  per the boundary conditions.

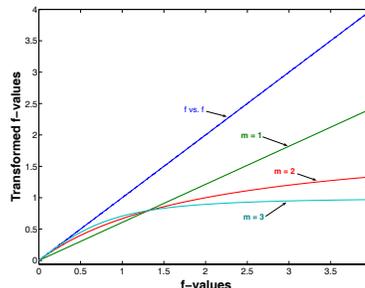
Figure 4b depicts how the transformed energy values relate to the original energy values. The topmost curve represents the case where  $\hat{f}_i(a, m, t) = f_i$  ( $a = 0$ ). This curve is basically the  $y = x$  line. The three curves below it correspond to transformed values with  $a = 0.5$  and  $m = 1, 2$  and  $3$ , respectively.

Figure 4b also provides a great deal of insight into the monotonicity properties of these transformations. Note that for  $m = 1$  the transformation function changes the energy landscape by a linear relation as shown in the top two curves in Figure 4b. For any state  $i$ , plotting  $\hat{f}_i(a, 1, t)$  versus  $f_i$  results in a line with a lower slope than the curve  $\hat{f}_i(0, m, t) = f_i$  versus  $f_i$ . These transformations thus preserve the rank-order of energy values, smooth-out the energy landscape, and have certain boundedness properties. (Proofs of these assertions are available from the author.)

Figures 5a and b illustrate how these energy transformations modify the relief of the landscape. For  $a > 0$ , the original landscape (the black curves in both graphics) is transformed to a set of ‘smoothed-out’ landscapes for  $m = 1, 2$  and  $3$ . Higher values of  $a$  produce a smoother landscape for a given value of  $m$  (compare the corresponding curves in Figures 5a and b). Note that each curve in Figure 5a that is a transformation of the original landscape (the top-most landscape in black) majorizes the corresponding curve (one with the same value of  $m$ ) in Figure 5b per the monotonicity properties. The linear transformation of the



a.  $\hat{f}(a)$  vs.  $a$  for two values  $f_i = 5, 10$ .



b. Transformations with different values  $m$ .

Figure 4: The effects of the transformation function.

landscape when  $m = 1$  is readily apparent—compare the top two curves of each graphic. The non-linear effects for  $m \geq 2$  curves are also apparent. Finally, observe that the rank order of all points is also preserved.

These smoothed-out landscapes allow the MCMC approach to accurately estimate the BPF. (On the other hand, if  $a$  is very large and the landscape almost flat (too smooth) we have another type of ‘broken ergodicity’ problem in that now all states can be visited with almost uniform probability. Under these circumstances, the very slight (yet definite) undulations in the landscape require a very large number of MCMC iterations to obtain statistically significant differences in energy values. This problem could rightly be referred to as the *super-ergodicity* problem.) The next section explains how this is accomplished.

#### 4 THE MODIFIED BOLTZMANN PARTITION FUNCTION

As stated earlier, the approach taken here does not modify the basic Metropolis Algorithm and instead modifies the energy values. Thus, (1) is only changed by substituting the transformed values  $\hat{f}_i$  for every  $f_i$  resulting in the stationary probabilities

$$\hat{\pi}_i(t) = \frac{e^{-\hat{f}_i/t}}{\sum_j e^{-\hat{f}_j/t}} = \frac{e^{-\hat{f}_i/t}}{\hat{Z}(a, m, t)} \quad (11)$$

where  $\hat{Z}(a, m, t) = \sum_j e^{-\hat{f}_j/t}$  is the mBPF associated with a smoother landscape. It is easy to see that  $Z(0, m, t) = Z(t)$  for any  $m$  and  $t$  from the boundary conditions. It is also worth noting that for  $m > 1$ , (11) is asymptotically power-law distributed (Fleischer 2005b). Assuming that this smoothed landscape permits a more accurate estimate of the mBPF, the question arises as to how to use mBPF values to estimate the BPF at low temperature. The following section describes how this can be done. And now, for some magic!

#### 4.1 Relating the mBPF to the BPF

Now that a set of transformations functions associated with the same underlying landscape exists, the relationships between the curves generated by  $\hat{Z}(a, m, t)$  as functions of  $a$  can be examined. An almost magical relationship becomes readily apparent: The functions  $\hat{Z}(a, 1, t)$  and  $\hat{Z}(a, 2, t)$  intersect at a point *very near* the point at which the derivative of  $\hat{Z}(a, 1, t)$  allows extrapolation to the value of  $\hat{Z}(0, 1, t) = Z(t)$ , hence effectively solves the DPIP!

Figure 6 illustrates this relationship using the same landscapes from Figure 3. Note how the tangent lines intersect the  $y$ -axis at the points denoted here by  $\tilde{Z}(t) \approx \hat{Z}(0, m, t)$ , *i.e.*, the extrapolated values based on the tangent line to  $\hat{Z}(a, 1, t)$  are very close to the mBPF evaluated at  $a = 0$ , hence, very close to the BPF  $Z(t)$ . As shown below, this relationship holds for any finite landscape and any  $t > 0$ .

#### 4.2 The Derivative of $\hat{f}_i$ With Respect to $a$

Taking the derivative of  $\hat{Z}(a, m, t)$  with respect to  $a$ :

$$\begin{aligned} \frac{\partial \hat{Z}(a, m, t)}{\partial a} &= \frac{\partial}{\partial a} \sum_{i=1}^N e^{-\hat{f}_i/t} = \sum_{i=1}^N \frac{\partial e^{-\hat{f}_i/t}}{\partial a} \\ &= \sum_{i=1}^N e^{-\hat{f}_i/t} \frac{\partial (-\hat{f}_i/t)}{\partial a} = \sum_{i=1}^N e^{-\hat{f}_i/t} \left( \frac{\hat{f}_i}{t} \right)^m \end{aligned} \quad (12)$$

where (12) is obtained by substituting in (3). For  $m = 1$ , and ignoring arguments for notational clarity, (12) becomes  $\frac{\partial \hat{Z}}{\partial a} = \langle \frac{\hat{f}}{t} \rangle$ , the expectation of  $\hat{f}(a, 1, t)/t$ .

#### 4.3 The Derivative Point and Term-Wise Analysis

As shown in Figure 6, solving the DPIP requires determining the value  $a^*$  such that  $\hat{Z}(a^*, 1, t) = \hat{Z}(a^*, 2, t)$ , *i.e.*, where the curves intersect. An important observation however is that one can produce very similar graphs as in Figure 6 for just the *single* exponential term associated with the lowest value  $f_i$  in the mBPF. Moreover, it is much simpler and mathematically tractable to examine the important relationships for single

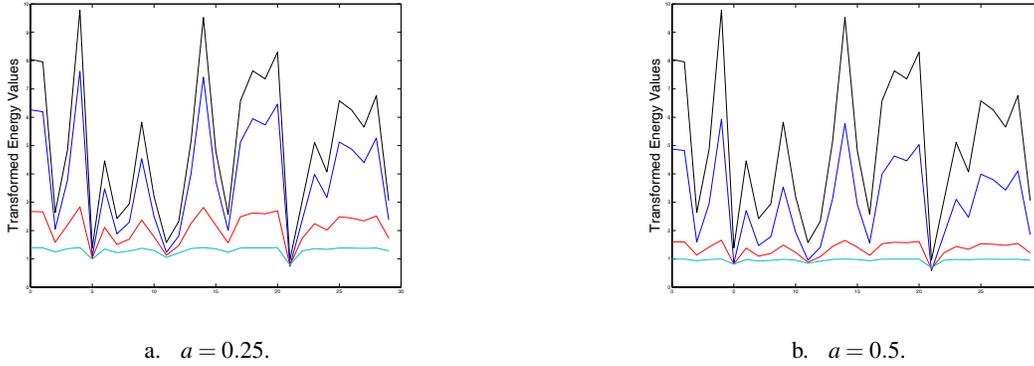


Figure 5: Plots of a random energy landscape (curve in black) and its associated transformed with  $m = 1, 2,$  and  $3$ . The transformations change the relief of the landscapes.

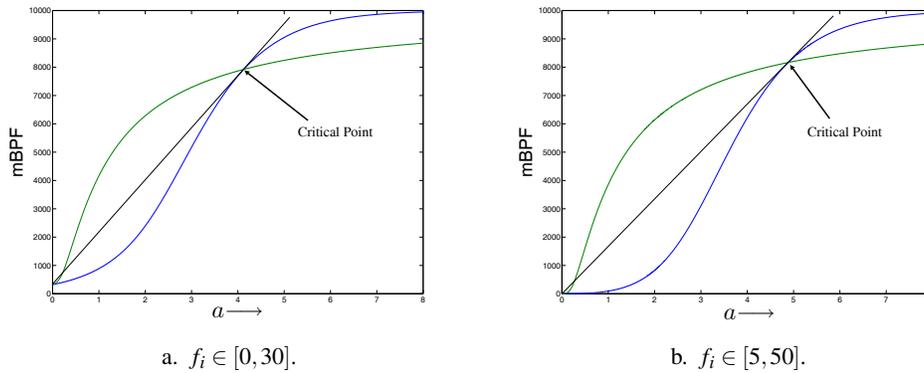


Figure 6: Illustration of the solution to the derivative point identification problem.

terms of the mBPFs and further provides a number of important insights for error analysis. For  $f_i \geq 0$  define

$$\begin{aligned} \hat{Z}_i(a, 1, t) &\equiv \exp(-e^{-a} f_i/t) \\ \hat{Z}_i(a, 2, t) &\equiv \exp\left(\frac{-f_i}{1+a(f_i/t)} / t\right) \end{aligned} \quad (13)$$

with  $Z_i(t) = e^{-f_i/t}$  and  $a_i \geq 0$  such that

$$\hat{Z}_i(a_i, 1, t) = \hat{Z}_i(a_i, 2, t). \quad (14)$$

Simplifying (14) yields

$$a_i = \ln(1 + a_i(f_i/t)) \Leftrightarrow \frac{e^{a_i} - 1}{a_i} = \frac{f_i}{t}. \quad (15)$$

It is easy to show that for every  $f_i \geq 0$ , there exists a unique  $a_i > 0$  such that (14) holds. Also, for any given  $f_i \geq 0$ , the fixed point  $a_i$  on the left in (14) can easily be calculated based on the Fixedpoint Theorem. (Note that [Havil \(2003\)](#) describes an intriguing connection between (15) and the Riemann Zeta Function.) Using Figure 6 as a guide, define

the following estimate of a BPF term:

$$\tilde{Z}_i(t) = \hat{Z}_i(a_i, 1, t) - a_i \left( \frac{\partial \hat{Z}_i(a, 1, t)}{\partial a} \Big|_{a=a_i} \right) \quad (16)$$

where  $f_i \geq 0$  and  $a_i$  satisfies (15). The following theorem shows that the absolute error of the extrapolated value for the exponential term is less than 4.8%.

**Theorem 1** Using the definitions in (14) and (16), define the error (the difference between the extrapolated estimate based on (16) and the actual value)

$$E_i(t) = \tilde{Z}_i(t) - Z_i(t). \quad (17)$$

Then the following are true for all  $f_i \geq 0$  and  $t > 0$ :

- A:**  $\lim_{f_i \rightarrow \infty} E_i(t) = 0,$
- B:**  $E_i(t) \geq 0,$  and
- C:**  $\max_{f_i \geq 0} E_i(t) < 0.04792.$

*Proof:*

**Statement A:** Re-writing the expression in (16) by substi-

tuting in the expressions in (12) and (13) yields

$$\begin{aligned} \tilde{Z}_i(t) &= \exp\{-e^{-a_i} f_i/t\} - a_i [e^{-a_i} (f_i/t) \exp\{-e^{-a_i} f_i/t\}] \\ &= \exp\{-e^{-a_i} f_i/t\} \left[1 - \frac{a_i e^{-a_i} f_i}{t}\right] \end{aligned} \quad (18)$$

Using the fact that  $a_i = \ln(1 + a_i(f_i/t)) \Rightarrow e^{a_i} = 1 + a_i(f_i/t)$  and simplifying (18),

$$\begin{aligned} \tilde{Z}_i(t) &= \exp\left(\frac{-f_i}{1 + a_i(f_i/t)} / t\right) \left[1 - \frac{a_i(f_i/t)}{1 + a_i(f_i/t)}\right] \\ &= \frac{\exp\left(\frac{-f_i}{1 + a_i(f_i/t)} / t\right)}{1 + a_i(f_i/t)} = \frac{\exp(-e^{-a} f_i/t)}{1 + a_i(f_i/t)}. \end{aligned}$$

Consequently,

$$\begin{aligned} E_i(t) &= \tilde{Z}_i(t) - Z_i(t) \\ &= \frac{\exp\left(\frac{-f_i}{1 + a_i(f_i/t)} / t\right)}{1 + a_i(f_i/t)} - e^{-f_i/t}. \end{aligned} \quad (19)$$

Note that  $\lim_{f \rightarrow \infty} \frac{f}{1 + af/t} = \frac{t}{a}$ , hence the first term of (19) converges to  $e^{-1/a_i}$ . From (15),  $a_i$  increases as  $f_i$  increases, it follows that  $e^{-1/a_i} \rightarrow 1$  as  $f_i \rightarrow \infty$  while the denominator  $1 + a_i(f_i/t) \rightarrow \infty$ . Thus, the first term converges to 0 as does the second term hence  $E_i(t) \rightarrow 0$  as  $f_i \rightarrow \infty$ .

**Statement B:** To see that  $E_i(t) \geq 0$  for all  $f_i \geq 0$ , consider the following cases:

**Case 1:** For all  $f_i \in [0, 1]$ ,  $a_i = 0$  and it is obvious from (19) that the error  $E_i(t) = 0$ .

**Case 2:** For all  $f_i > 1$ , it follows from (15) that  $a_i > 0$ . Note that for all integers  $k \geq 3$ ,  $k < 2^{k-1}$ , hence for all  $k \geq 3$  and  $a_i > 0$ ,  $\frac{a_i^k}{k!} > \frac{a_i^k}{k!} \left(\frac{k}{2^{k-1}}\right)$ . Thus, for infinite sequences,

$$\begin{aligned} 1 + a_i + \frac{a_i^2}{2!} + \sum_{k=3}^{\infty} \frac{a_i^k}{k!} &> 1 + a_i + \frac{a_i^2}{2!} + \sum_{k=3}^{\infty} \frac{a_i^k}{k!} \left(\frac{k}{2^{k-1}}\right) \\ &= 1 + a_i + \frac{a_i^2}{2!} + \sum_{k=3}^{\infty} \frac{a_i a_i^{k-1}}{(k-1)!} \left(\frac{1}{2^{k-1}}\right) \\ &= 1 + a_i \left(1 + (a_i/2) + \sum_{k=3}^{\infty} \frac{(a_i/2)^{k-1}}{(k-1)!}\right) \\ &= 1 + a_i \left(1 + (a_i/2) + \sum_{k=2}^{\infty} \frac{(a_i/2)^k}{k!}\right). \end{aligned} \quad (20)$$

Noting the forms on both sides of (20) we obtain the inequality  $e^{a_i} > 1 + a_i e^{a_i/2}$  and

$$\frac{e^{a_i} - 1}{a} > e^{a_i/2}. \quad (21)$$

Recall that the left-hand side of (21) is  $f_i/t$ , hence  $f_i/t > e^{a_i/2}$  and therefore  $2 \ln(f_i/t) > a_i = \ln(1 + a_i(f_i/t))$ . Exponentiating both sides yields  $(f_i/t)^2 > 1 + a_i(f_i/t)$ . Rearranging,  $0 > 1 + a_i(f_i/t) - (f_i/t)^2$  and multiplying by  $a_i > 0$  yields

$$0 = a_i + a_i^2 (f_i/t) - a_i (f_i/t)^2. \quad (22)$$

Subtracting  $f_i/t$  from both sides of (22) and factoring the right-hand side yields  $\frac{-f_i}{t} > \left(a_i - \frac{f_i}{t}\right) \left(1 + \frac{a_i f_i}{t}\right)$  and therefore

$$\frac{-f_i/t}{1 + a_i(f_i/t)} > a_i - (f_i/t) = \ln(1 + a_i(f_i/t)) - (f_i/t).$$

After exponentiating both sides then

$$\exp\left\{\frac{-f_i}{1 + a_i(f_i/t)} / t\right\} > (1 + a_i(f_i/t)) \exp\{-f_i/t\}$$

and upon rearranging, it immediately follows from (19) that for all  $f_i > 1$ ,  $E_i(t) > 0$ .

**Statement C:** Numerical methods are necessary for calculating the maximum error and the corresponding critical values of  $f$  and  $a$ . This is facilitated by modifying the error function (19) by substituting in  $\frac{e^{a_i} - 1}{a_i}$  per (15) for every  $f_i/t$  thereby eliminating every appearance of  $f_i$  and  $t$  (in the first term). The error function thus becomes an expression involving only  $a$ , hence

$$E(a) = e^{-\frac{a^2 + e^{-a} - 1}{a}} - e^{-\frac{1 - e^a}{a}}. \quad (23)$$

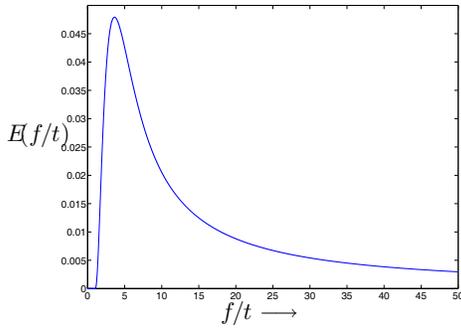
Taking the derivative of  $E_i(a)$  with respect to  $a$  and without showing all the calculations,

$$\begin{aligned} \frac{dE(a)}{da} &= \frac{\exp\left(\frac{-a^2 + e^{-a} - 1}{a}\right)}{a^2} [1 - a^2 - (a + 1)e^{-a}] \\ &\quad + \frac{\exp\left(\frac{1 - e^a}{a}\right)}{a^2} [1 + (a - 1)e^a]. \end{aligned} \quad (24)$$

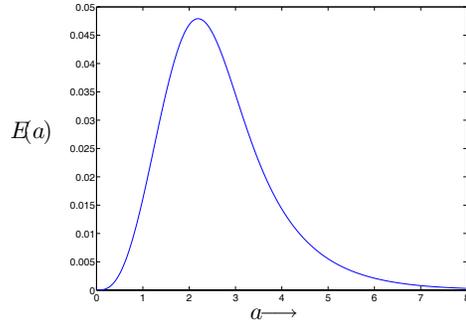
Figure 8 depicts the curve  $dE/da$  vs.  $a$  and a unique, non-trivial zero. Setting (24) equal to zero and simplifying yields the following function for which the zero value corresponds to the zero in (24).

$$g(a) = a^2 - \cosh(a) - \frac{a}{2} \ln(1 + a) + 1 = 0. \quad (25)$$

Application of the bisection method on (25) yields the following values associated with the maximum error  $E_i(t)$  and the corresponding values for  $a$  and  $f$  (denoted here



a. Error vs.  $f/t$ .



b. Error vs.  $a$ .

Figure 7: Plot of the error  $E_i$  as functions of  $f/t$  and  $a$ .

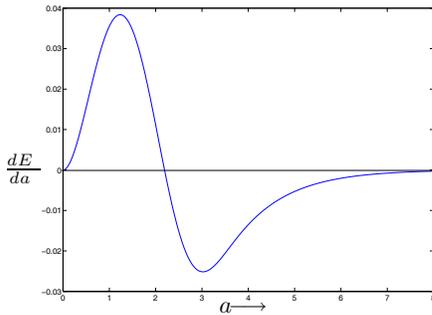


Figure 8: Plot of the derivative of  $E$  with respect to  $a$ .

with the \*).

$$\begin{aligned} a^* &= 2.19241298854854 \\ f^* &= 3.62924310192440 \\ \max_{f \geq 0} E(f) &= 0.04791525323212. \end{aligned} \quad (26)$$

Figures 7a and b show the magnitude of error plotted against  $f$  and  $a$  respectively at  $t = 1$  and illustrate the asymptotic convergence of the error function. ■

Because of the existence of a closed form expression for the error, the following identity is proved.

**Corollary to Theorem 1:** For all  $f_i \geq 0$ ,

$$\begin{aligned} Z_i(t) &= \hat{Z}_i(a_i, 1, t) \\ -a_i \left( \frac{\hat{f}_i(a_i, 1, t)}{t} \right) \hat{Z}_i(a_i, 1, t) &- E_i(a_i) \end{aligned} \quad (27)$$

*Proof:*

This expression is obtained by recalling that  $\partial \hat{Z}_i(a, 1, t) / \partial a = (\hat{f}_i/t) \hat{Z}_i(a, 1, t)$ , substituting into (16) and applying the error term in the theorem. ■

The next section shows one possible general approach for using these results to produce estimates of the BPF.

## 5 COMPUTATIONAL EXAMPLES FOR ESTIMATING THE BPF

Estimating the sum of exponentials without brute-force calculation has long been a challenging task—there are no easy formulas that convert a sum of exponentials into simpler expressions. In the end, only the actual summation of these terms provides the exact answer. There are however very well known techniques that attempt to mitigate this problem and can produce reasonable approximations. Perhaps the most famous of these is the widely used *Laplace Method* that takes advantage of the fact that the largest contribution in a sum of exponentials (Laplace Integrals) comes from the terms with the lowest values of  $f_i$  since the influence of the other terms on the sum decreases exponentially and so can be discounted somewhat or ignored. The approach described below also uses this idea along with the theoretical results presented here to suggest a general approach for solving the broken ergodicity problem.

Estimate  $Z(t) = \sum_{i=1}^N e^{-f_i/t}$  by

1. first estimating the critical value  $a^*$  by running MCMC simulations using two transformed landscapes with  $m = 1$  and  $m = 2$ . Modify the simulations by changing  $a$  until estimates are obtained where  $\hat{Z}(a, 1, t) \approx \hat{Z}(a, 2, t)$ .
2. Calculate  $\tilde{Z}(t)$  as an estimate of  $Z(t)$  using the following equation similar in structure to (27):

$$\tilde{Z}(t) = \hat{Z}(a^*, 1, t) - a^* \left\langle \frac{\hat{f}}{t} \right\rangle - E(a^*) \quad (28)$$

where the second term corresponds to  $a^*$  times the derivative of  $\hat{Z}(a, 1, t)$  from (12), and the third term is based on the term-wise error in (23).

The following analytical computations illustrate this approach.

In Table 1 experiment #1 was based on uniformly and randomly generated values for  $f_i$  in the range 0 to 50.

Experiment #2 was similarly based on values for  $f_i$  ranging from 10 to 100.

Table 1: Numerical experiments.

Exp.	# Terms	$Z(t)$	$\tilde{Z}$	% Error
1	10,000	193.40	402.53	2.09
2	100,000	0.0495	1779.73	1.77

## 6 CONCLUSION

This article highlights a transformation function that smoothes an energy landscape. MCMC simulations on such smoothed-out landscapes can provide a means to more accurately estimate a *modified* BPF (mBPF). Mathematical relationships are described that allows one to estimate the original BPF based on estimates of the corresponding mBPFs. Error analysis and computational examples show a small absolute error in the estimated BPFs. This mathematical relationship can thus be utilized to estimate the BPF in MCMC simulations that exhibit broken ergodicity. Future research will examine the best manner in which to estimate certain critical parameters such as  $\alpha^*$  under actual MCMC simulations and other issues relating to upper and lower bounds on the resulting estimates of the BPF and values of  $\alpha$  that yield the most efficient statistics in a simulation.

## ACKNOWLEDGMENTS

Much of the foregoing research took place while the author was a member of the Senior Professional Staff at the Johns Hopkins University Applied Physics Laboratory. The author wishes to thank Dr. Alan Weiss and three anonymous referees for their very helpful comments and suggestions.

## REFERENCES

- Conference: 2003, June 9-11. The monte carlo method in the physical sciences: Celebrating the 50th anniversary of the metropolis algorithm. J. Robert Oppenheimer Study Center, Los Alamos, NM: Los Alamos National Laboratory.
- Fleischer, M. 1995. Simulated annealing: past, present, and future. In *Proceedings of the 27th conference on Winter simulation*, 155–161: ACM Press.
- Fleischer, M. 2005a. A generalization of tsallis' non-extensive entropy and energy landscape transformation functions. <http://arxiv.org/abs/cond-mat/0505665/>.
- Fleischer, M. 2005b. Scale invariance and symmetry relationships in non-extensive statistical mechanics. <http://arxiv.org/abs/cond-mat/0501293/>.
- Havil, J. 2003. *Gamma: Exploring euler's constant*. Princeton, New Jersey: Princeton University Press. According to Havil (p.84), the function  $f(x) = \frac{x}{e^x - 1}$  has some history to it and that  $\int_0^\infty \left(\frac{x}{e^x - 1}\right) dx = \Gamma(x)\zeta(x)$  where  $\zeta(x)$  is the Riemann Zeta Function.
- Landau, D. P. 2003, November 25, 2003. The metropolis monte carlo method in statistical physics. In *The Monte Carlo Method in the Physical Sciences: Celebrating the 50th Anniversary of the Metropolis Algorithm*, ed. J. E. Gubernatis, Volume 690 of *AIP Conference Proceedings*, pp. 134–146. Melville, NY: Los Alamos National Laboratory: American Institute of Physics.
- Landau, D. P., and K. Binder. 2000. *A guide to monte carlo simulations in statistical physics*. Cambridge, United Kingdom.: Cambridge University Press.
- Metropolis, N. A., A. T. A. Rosenbluth, M. Rosenbluth, and E. Teller. 1953. Equation of state calculations by fast computing machines. *J. of Chemical Physics* 21:1087–1092.
- Straub, J., and I. Andricioaei. 2001. *Nonextensive statistical mechanics and its application*, Chapter IV: Computational methods for the simulation of classical and quantum many body systems sprung from non-extensive thermostatics, pp.195–235. Lecture Notes in Physics. Berlin, Germany.: Springer-Verlag.
- Topper, R., D. Freeman, D.Bergin, and K. LaMarche. 2003. Computational techniques and strategies for monte carlo thermodynamic calculations, with applications to nanoclusters. *Reviews in Computational Chemistry* 19:pp.1–42.
- Tsallis, C. 1988. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics* 52 (1-2): 479–487.

## AUTHOR BIOGRAPHY

**MARK A. FLEISCHER** is the president and owner of ULTRANETX, LLC., a scientific consulting company specializing in trade-off analysis and multi-objective optimization software. He has held research and academic positions at Towson University, Johns Hopkins University, the Institute for Systems Research at the University of Maryland, College Park and at Old Dominion University in Norfolk, VA. He received his Ph.D. in operations research from Case Western Reserve University in 1994 and his B.Sc from the Massachusetts Institute of Technology in 1978. Prior to his academic and research career, he practiced law in the State of Ohio. He enjoys playing baseball, mushroom hunting, hiking and camping with his family. His company's website is located at [www.ultranetx.com](http://www.ultranetx.com). He can be reached via email at [MarkFleischer@alum.mit.edu](mailto:MarkFleischer@alum.mit.edu).