Journal of Computational Physics 228 (2009) 6119–6129

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Optimized convergence for multiple histogram analysis

Tristan Bereau *, Robert H. Swendsen

Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, United States

ARTICLE INFO

Article history:
Received 22 January 2009
Received in revised form 1 May 2009
Accepted 5 May 2009
Available online 18 May 2009

PACS:
05.70.Ce
05.10.−a

Keywords:
WHAM
Free-energy calculation
Multiple histogram analysis

ABSTRACT

We propose a new algorithm for solving the weighted histogram analysis method (WHAM) equations to estimate free energies out of a set of Monte Carlo (MC) or molecular dynamics (MD) simulations. The algorithm, based on free-energy differences, provides a more natural way of approaching the problem and improves convergence compared to the widely used direct iteration method. We also study how parameters (temperature, pressure, etc.) of the independent simulations should be chosen to optimize the accuracy of the set of free energies.

1. Introduction

The multiple histogram method, also known as the weighted histogram analysis method (WHAM), is a technique for combining data from multiple Monte Carlo (MC) or molecular dynamics (MD) computer simulations to calculate thermodynamic properties as continuous functions over a wide range of values of temperature, pressure, etc. [1–3]. WHAM has proven efficient and accurate in many applications, but some questions remain as to their optimal implementation.

The questions we are concerned with in this paper are:

1. What is the most efficient method for computing the self-consistent set of free-energy parameters from the histogram data?
2. How should parameters (temperature, pressure, etc.) of the independent computer simulations be chosen to optimize the accuracy of the final results?

In Section 2, we recall the WHAM equations [1–3], and discuss the convergence of the direct iteration (DI) algorithm for solving these equations. We will then develop a more efficient approach based on the relationship of the WHAM equation to an earlier equation derived by Bennett [4]. The efficiency of these approaches turns out to be surprisingly sensitive to apparently harmless variations in the details of their implementation. We present numerical comparisons of the efficiency of the different variations and provide recommendations of the best way to proceed, depending of the number and degree of overlap of the histograms from the independent simulations.

* Corresponding author. Tel.: +1 412 268 8367; fax: +1 412 681 0648.
E-mail addresses: bereau@cmu.edu (T. Bereau), swendsen@cmu.edu (R.H. Swendsen).

0021-9991/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2009.05.011

In Section 3 we address the question of the optimum choice of overlap between neighboring histograms. Although it might seem that maximizing the overlap would be optimal, we will show that neighboring histograms should be separated by somewhat more than double their *rms* width.

We conclude the paper in Section 4 with recommendations for an efficient, nearly automatic implementation of the choice of simulation parameters, together with an efficient solution of the WHAM equations.

2. The WHAM equations and the free-energy parameters

For simplicity of notation we will only discuss the WHAM equations for the case of varying the temperature and measuring the energy in the simulations. Although WHAM can be applied to a wide variety of physical models, simulation parameters, and thermodynamic properties, all other cases can be obtained from our equations by simple analogies.

The partition function is given as a sum over the energy,

$$Z = \sum_E W(E) \exp[-\beta E], \tag{1}$$

where $\beta = 1/k_B T$ and $W(E)$ is the density of states. The free-energy is then given by

$$F = -\beta^{-1} \ln Z. \tag{2}$$

We will consider a set of R computer simulations that have each produced a set of histograms $\{H_n(E)|n = 1,\ldots,R\}$ for the distributions of energies at the inverse temperatures $\beta_n = 1/k_B T_n$. The WHAM equations give the best estimate for the unnormalized probability of the energy E at inverse temperature β as

$$P(E,\beta) = \frac{\sum_{n=1}^{R} g_n^{-1} H_n(E) \exp[-\beta E]}{\sum_{m=1}^{R} N_m g_m^{-1} \exp[-\beta_m E - f_m]}, \tag{3}$$

where $g_n = 1 + 2\tau_n$ is a measure of the independence of the values of the energy for the nth MC simulation [5,6], τ_n is the correlation time, and the free-energy parameters $f_n = -\beta_n F(\beta_n)$ are determined by the equations

$$e^{f_n} = \sum_E P(E,\beta). \tag{4}$$

Some features of these equations will prove important in applications.

It can be immediately seen that if $g_n = g_1$ for all n, the g_n's cancel out of Eq. (3). Furthermore, as the statistical accuracy of the data in the histograms improves, the values of the g_n become less important. We can see this by considering the expectation values of the histograms

$$\langle H_n(E) \rangle = N_n W(E) \exp(-\beta_n E)/Z_n, \tag{5}$$

where $W(E)$ is the density of states and $Z_n = \sum_E \exp(-\beta_n E)$ is the partition function at the inverse temperature β_n. If we substitute Eq. (5) in Eq. (3), we find that the equation is satisfied for all values of g_n. In the following discussion, we will keep the g_n's in the equations for those cases in which they are important.

Another property of the equations is that the values of the free-energy parameters are only determined up to an arbitrary additive constant. This means that there are only $R - 1$ independent parameters.

Naturally, to obtain the absolute value of the free-energy from the WHAM equations, it is necessary to know the free-energy at some reference point, such as infinite or zero temperature.

2.1. Solution by direct iteration (DI)

The first numerical task in calculating the free-energy from these equations is to compute a consistent set of free-energy parameters as solutions to Eqs. (3) and (4). The obvious and most widely used choice is to begin with a guess for the values of the free-energy parameters and to iterate these equations.

The first piece of good news is that the guess can be very crude: to begin by setting all free-energy parameters to zero is almost always acceptable, if not terribly efficient.

The second piece of good news is that direct iteration is almost always sufficient when there are only a small number of histograms.

The third piece of good news is that the convergence is always exponentially rapid with the number of iterations, even for a large number of histograms.

The bad news is that even though the convergence is exponential, it is often still very slow, sometimes requiring many thousands of iterations and a great deal of computer time.

Fig. 1 shows an example of such convergence. The graph plots the absolute error on the whole set of free energies with respect to time. The absolute error δ is defined as the sum of squared differences of the calculated free-energy f_i compared to its converged value f_i^c (defined in a previous simulation when the relative error between successive free-energy differences was less than 10^{-12})

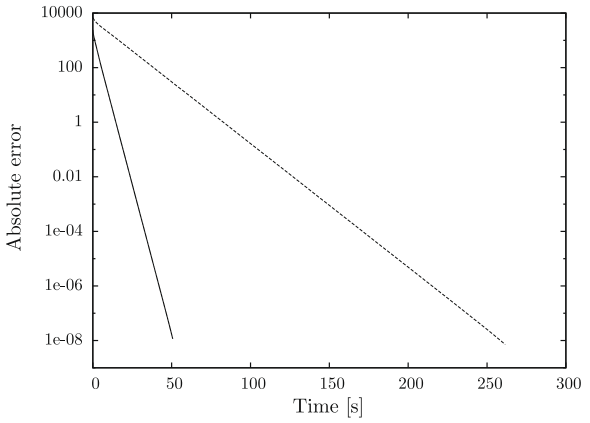

Fig. 1. Convergence of the absolute error made on the set of free energies, using the DI method introduced in [1]. The graph shows the convergence of free energies for a 2D Ising model (lattice size $L = 32$) with 14 simulations which cover the whole temperature range (solid line) and 40 simulations in the same range (dashed line). In the 14-temperature system, temperatures were separated by an amount of 2.5σ, where σ is the second moment of the energy distribution of a simulation. This corresponds to the optimal separation between histograms proposed in this paper (see Section 3).

$$\delta^2 = \sum_i (f_i - f_i^c)^2. \tag{6}$$

A simple 2D Ising model with lattice size of $L = 32$ took around 1 min on a single laptop processor to converge the absolute error below 10^{-8} for a set of 14 temperatures that spanned the whole temperature range. When a different set of 40 temperatures was used to cover the same range, convergence took more than 4 min.

As we will show in this paper, the overlap between histograms has a strong influence on the error made on the calculation of free energies.

Various approaches to speeding up the direct iteration method have been suggested. In particular, Alan Ferrenberg has developed some interesting methods in his thesis [7]. Unfortunately, even with these techniques convergence remains slow in many problems of interest.

2.2. Nearest-neighbor histograms

To make further progress in developing faster convergence techniques, we will focus our attention on the differences in the free-energy parameters of neighboring histograms, rather than the parameters themselves. Once these differences are found, the full set of free energies is obtained by simple addition, up to the usual arbitrary additive constant.

To develop a general iterative method for solving for the free-energy parameter differences, we first consider the simple case of just two neighboring histograms at inverse temperatures β_i and β_{i+1}. Combining Eqs. (3) and (4), we find

$$e^{f_i} = \sum_E \frac{\left[g_i^{-1} H_i(E) + g_{i+1}^{-1} H_{i+1}(E)\right] \exp[-\beta_i E]}{N_i g_i^{-1} \exp[-\beta_i E - f_i] + N_{i+1} g_{i+1}^{-1} \exp[-\beta_{i+1} E - f_{i+1}]}, \tag{7}$$

which can be re-written as

$$1 = \sum_E \frac{\left[g_i^{-1} H_i(E) + g_{i+1}^{-1} H_{i+1}(E)\right]}{N_i g_i^{-1} + N_{i+1} g_{i+1}^{-1} \exp\left[-\Delta\beta_i E - \Delta f_i\right]}, \tag{8}$$

where $\Delta\beta_i = \beta_{i+1} - \beta_i$ and $\Delta f_i = f_{i+1} - f_i$, as before. This equation was originally derived by Bennett [4] although he did not include the g_i factors. We will refer to Eq. (8) and its variations as Bennett equations.

There is an essential asymmetry in this equation that turns out to be quite important. The origin of this asymmetry is that we could equally well exchange i and $i + 1$ to obtain

$$1 = \sum_E \frac{\left[g_i^{-1} H_i(E) + g_{i+1}^{-1} H_{i+1}(E)\right]}{N_i g_i^{-1} \exp\left[\Delta\beta_i E + \Delta f_i\right] + N_{i+1} g_{i+1}^{-1}}. \tag{9}$$

While Eqs. (8) and (9) lead to the same solutions, the numerical behavior of the generalized Bennett equations derived below is different. A hint of why this might be so can be seen by considering the limit of $N_i \to 0$. Eq. (8) gives an estimate for Δf_i, while Eq. (9) reduces to an uninformative identity.

A numerical solution for Δf_i can always be found for either Eq. (8) or Eq. (9). Using the false position method between upper and lower limits tends to be much more reliable than Newton–Raphson. It is also important to solve this equation and the equations given below to high accuracy, since errors can propagate.

If we include more simulations, we can still use Eq. (8) or Eq. (9) as a starting point, which we will call the Bennett approximation. It turns out to be very good, but we can improve on it by including more distant neighboring histograms systematically.

2.2.1. Combining the Bennett equation with the DI method

As stated before, the DI method is extremely efficient when converging the set of free energies of small systems. We have seen that giving accurate starting points for the DI method yields very good results, even for reasonable system sizes. In this respect, it has proven efficient to feed in the results of the Bennett approximation as a guess to the DI method. Convergence is exponentially fast, stable, and rather straightforward to implement.

2.3. Generalizing the Bennett equation to include more distant neighboring histograms

We can rewrite the basic WHAM Eqs. (3) and (4) to make the inclusion of q neighboring histograms explicit, and then recast the result to emphasize the similarity with Bennett's equation (8). This will lead us to an efficient iterative solution for the full set of equations including all histograms.

Although Bennett's equation provides an estimate for each Δf_n separately, without any influence from other free-energy differences, the inclusion of more distant neighbors means that the set of values $\{\Delta f_n | n = 1,\ldots,R-1\}$ must be obtained iteratively.

Let us introduce the notation $\Delta f_i^{(n,q)}$ to refer to the nth level of iteration for a calculation that includes q neighbors on each side of the original pair. For the lowest level, there are no extra neighbors, so $q = 0$, and the solution of the Bennett equation does not require iteration, so $n = 0$. The Bennett approximation is therefore denoted as $\Delta f_i^{(0,0)}$, and is given by the solution to Eq. (8).

We can rewrite Eqs. (3) and (4) for f_i when q neighbors are included in the calculation as

$$e^{f_i} = \sum_E \sum_{n=i-q}^{i+1+q} g_n^{-1} H_n(E) \exp\left[-\beta_i E\right] \Big/ z_i(E), \tag{10}$$

where

$$z_i(E) = \sum_{m=i-q}^{i+1+q} N_m g_m^{-1} \exp[-\beta_m E - f_m]. \tag{11}$$

To avoid cluttering up the notation, it is understood that the sum is to be truncated if either $i - q < 1$ or $i + 1 + q > R$.

To recast these equations in a form similar to that of Eq. (8), divide both sides of Eq. (10) by the factor e^{f_i} and write it as

$$1 = \sum_E \frac{\sum_{n=i-q}^{i+1+q} g_n^{-1} H_n(E)}{z_i(E) \exp[\beta_i E + f_i]}. \tag{12}$$

The denominator in Eq. (12) can be found by multiplying Eq. (11) by the factor $\exp[\beta_i E + f_i]$ and writing it as

$$z_i(E) \exp[\beta_i E + f_i] = \sum_{m=i-q}^{i+1+q} N_m g_m^{-1} \exp\left[(\beta_i - \beta_m)E + f_i - f_m\right]. \tag{13}$$

Combining Eqs. (12) and (13), we find

$$1 = \sum_E \frac{\sum_{n=i-q}^{i+1+q} g_n^{-1} H_n(E)}{\sum_{m=i-q}^{i+1+q} N_m g_m^{-1} \exp\left[(\beta_i - \beta_m)E + f_i - f_m\right]}. \tag{14}$$

It will prove more convenient to write this as two equations:

$$1 = \sum_E \frac{\sum_{n=i-q}^{i+1+q} g_n^{-1} H_n(E)}{\Omega_i(E)}, \tag{15}$$

and

$$\Omega_i(E) = \sum_{m=i-q}^{i+1+q} N_m g_m^{-1} \exp\left[(\beta_i - \beta_m)E + f_i - f_m\right]. \tag{16}$$

With this notation, we can recast the Eqs. (15) and (16) in a convenient form that can be solved for Δf_i^{new}.

Fix all free-energy parameter differences with $m \neq i$ at their old values $\{\Delta f_m^{\text{old}} | m \neq i\}$. For $m < i, f_i - f_m = \sum_{k=m}^{i-1} \Delta f_k^{\text{old}}$, and for $m > i+1, f_{i+1} - f_m = -\sum_{k=i+1}^{m-1} \Delta f_k^{\text{old}}$. For $m = i + 1$, we have the parameter we wish to determine, $f_{i+1} - f_i = \Delta f_i^{\text{new}}$. Eq. (16) becomes

$$\Omega_i(E) = N_i g_i^{-1} + N_{i+1} g_{i+1}^{-1} \exp\left[-\Delta\beta_i E - \Delta f_i^{\text{new}}\right] + \sum_{m=i-q}^{i-1} N_m g_m^{-1} \exp\left[(\beta_i - \beta_m)E + \sum_{k=m}^{i-1} \Delta f_k^{\text{old}}\right] + \sum_{m=i+2}^{i+1+q} N_m g_m^{-1} \exp\left[(\beta_i - \beta_m)E - \Delta f_i^{\text{new}} - \sum_{k=i+1}^{m-1} \Delta f_k^{\text{old}}\right], \tag{17}$$

or

$$\Omega_i(E) = N_i g_i^{-1} + N_{i+1} g_{i+1}^{-1} \exp\left[-\Delta\beta_i E - \Delta f_i^{\text{new}}\right] + \sum_{m=i-q}^{i-1} N_m g_m^{-1} \exp\left[(\beta_i - \beta_m)E + \sum_{k=m}^{i-1} \Delta f_k^{\text{old}}\right] + \exp\left[-\Delta f_i^{\text{new}}\right] \times \sum_{m=i+2}^{i+1+q} N_m g_m^{-1} \exp\left[(\beta_i - \beta_m)E - \sum_{k=i+1}^{m-1} \Delta f_k^{\text{old}}\right]. \tag{18}$$

The pair of Eqs. (15) and (18) contains only a single unknown Δf_i^{new}, so it is easy to solve numerically. However, note that Δf_i^{new} occurs in two places in (18).

The asymmetry that we saw in the Bennett equation also propagates when adding more neighbors. If we had started with the equation for f_{i+1} instead of that for f_i, we would have found

$$1 = \sum_E \frac{\sum_{n=i-q}^{i+1+q} g_n^{-1} H_n(E)}{\sum_{m=i-q}^{i+1+q} N_m g_m^{-1} \exp\left[(\beta_{i+1} - \beta_m)E + f_{i+1} - f_m\right]}. \tag{19}$$

The limits on the sum in the denominator are determined by the neighbors of the central pair, which doesn't change when we shift from f_i to f_{i+1}. It is straightforward to rewrite Eqs. (15) and (16) for this new set of indices.

2.4. Solving the generalized Bennett equations with q neighboring histograms

From (15) and (16), there are different ways one can initialize and converge a set of free energies. Here, we describe the two algorithms that were most efficient for our test case: the 2D Ising model.

2.4.1. Successive iteration over neighboring histograms (SINH)

The SINH algorithm is based on the decreasing weight of more distant histograms.

To solve the generalized equations, we start with the Bennett approximation in Eq. (8), which gives us the set $\{f_n^{(0,0)} | n = 1, \ldots, R-1\}$, using only information from nearest-neighbor histograms ($q = 0$).

For the next step, we will combine the nearest neighbors on each side of each pair of histograms. This corresponds to $q = 1$. Unlike the $q = 0$ case, when $q \geqslant 1$ the set of equations must be iterated to find a fully consistent set of free-energy differences.

After obtaining a converged set of free-energy differences with $q = 1$, we find a new set for $q = 2$ neighbors, and so on until the sets for q and $q + 1$ neighbors agree. At this point we do not need to include more neighbors, since all overlapping histograms have been taken care of. This is an important point, as the number of relevant neighboring histograms is usually much smaller than the total number of simulations. The final converged set can then be substituted into the full set of equations as a check.

At each iteration, we will be going from an "old" set of free-energy differences, $\{\Delta f_i^{\text{old}}\}$, to a "new" set $\{\Delta f_i^{\text{new}}\}$. For the lowest level of iteration when the number of neighboring histograms is incremented from $q-1$ to q, $\{\Delta f_i^{\text{old}}\} = \{\Delta f_i^{(n_c,q-1)}\}$, where n_c indicates the last (converged) iteration at level $q-1$, so that $\{\Delta f_i^{(n_c,q-1)}\} = \{\Delta f_i^{(n_c-1,q-1)}\}$. The new estimate for the free-energy difference is $\{\Delta f_i^{\text{new}}\} = \{\Delta f_i^{(0,q)}\}$. For subsequent iterations at the level q, $\{\Delta f_i^{\text{old}}\} = \{\Delta f_i^{(j,q)}\}$, and $\{\Delta f_i^{\text{new}}\} = \{\Delta f_i^{(j+1,q)}\}$.

2.4.2. Convergence by iteration over all neighbors (CIAN)

When histograms are close to each other, the SINH algorithm needs to converge over many levels of approximation q, until all overlapping histograms have been considered. The idea behind the CIAN algorithm is to give an accurate first guess on the free-energy differences, and then converge the whole set at once.

For the initialization procedure, better accuracy will be given by including as many overlapping neighbors as possible in Eqs. (15) and (18). One has to be careful though, stability is often an issue when solving these equations, they require good guesses on the set of unknowns. One way around this is the following: when computing Δf_i, include all histograms H_n for all n such that $n \leqslant i + 1$. The first free-energy difference Δf_0 will only include its two constituents: H_0 and H_1. As one loops over the index i, more histograms will be included, all of them initialized using at least two histograms. This step is extremely fast since no iteration is involved. We have seen that errors on initial free energies propagate rapidly throughout this process, and may lead to unsolvable equations because of instability.

The second step is to sweep through the entire set of free-energy differences by using Eqs. (15) and (18), now including all neighbors. The advantage is that there is only one convergence run. The disadvantage is that the algorithm includes all neighbors, even non-overlapping ones, which might slow down convergence when including many simulations. To deal with

this problem, it might be useful to analyze histograms beforehand by looking at their overlap, and then setting an interaction matrix M. The matrix element M_{ij} would determine whether histograms i and j overlap, and therefore if their pair should be considered when converging free-energy differences.

2.4.3. Technical details

As mentioned earlier, all equations were solved using the false position method. This algorithm requires two initial conditions. These are best estimated by coming back to the two-simulation case. We recall Eqs. (8) and (9) which are two equivalent ways of calculating the free-energy difference between two simulations. For each equation, there is a contribution coming from each simulation. The free-energy difference comes in the equation multiplied by a pre-factor belonging to one of the simulations. One can solve analytically a simplified (and therefore approximate) version of these equations by keeping terms that belong to this simulation only. The following equations can be written

$$\Delta f_i^a \simeq -\ln \sum_E \frac{H_{i+1}(E)}{N_{i+1}\exp(-\Delta\beta_i E)}, \tag{20}$$

$$\Delta f_i^b \simeq \ln \sum_E \frac{H_i(E)}{N_i \exp(\Delta\beta_i E)}, \tag{21}$$

where the superscripts a and b refer to the different approximations. By this mean we take advantage of the asymmetry in the equations. Although the exact equations lead to identical results, the path used to solve for the unknown is different.

The tolerance required for the convergence of free-energy differences should be much weaker than the one used for solving the equations (a factor of $\sim$1000 between the two has proven to be sufficient). We expect convergence to be rapid since the effective weights of the neighboring histogram should only have a small adjustment at each iteration. However, experience shows that iteration procedures do not always converge. We have found both limit cycles and even divergences. Averaging successive iterations was usually enough to overcome cyclic behaviors and instabilities.

2.5. Effect of overlap and efficiency of the algorithms

Out of the two algorithms proposed here, SINH seems to be the fastest and most stable algorithm regarding numerous tests we have done.

Fig. 1 shows the result of two simulations of a 2D Ising model at $L = 32$ using the DI method. For the first case, we selected the temperatures of our simulations by taking a separation of 2.5σ between histograms (see Section 3), starting at infinite temperature. The second test used three times as many simulations in the same range to study the effect of large overlap on the different algorithms.

It should be noted that although the absolute error of the free energies is plotted, the program was set as to converge the relative difference between successive sets of free energies.

The results of the 14-temperature simulation using the SINH algorithm is shown on Fig. 2. The many curves show the different levels of approximation q. The reason why they seem to level off again is because we plot the absolute error. This seems reasonable as a fully converged set of free energies that do not include all relevant neighbors will contain an absolute error.

The main results of the 14-and 40-temperature simulations for the DI, SINH, and CIAN algorithms are given in Table 1. For each simulation, the time necessary for convergence ($\delta \leqslant 10^{-12}$) of the whole set of free energies is reported.

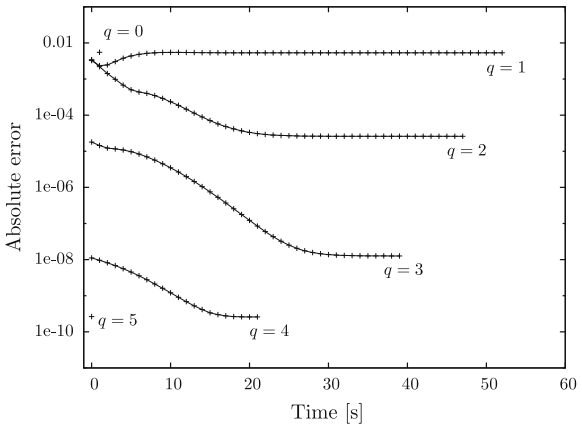

Fig. 2. Absolute error with respect to time for the set of free energies coming from the 14-temperature simulation with the SINH algorithm. As the number of neighbors taken into account is increased (denoted by q), convergence improves.

Table 1
Time of convergence (in seconds) of the 14-and 40-temperature simulations for different algorithms.

	DI	SINH	CIAN
14-Temperature simulation	66	12	16
40-Temperature simulation	440	180	Failed

When stacking up 40 simulations in the same energy interval as the optimal 14-temperature range, the DI algorithm was $\simeq$7 times longer than before, which scales with the square of number of simulations added. The SINH algorithm was 15 times longer for the 40-temperature simulations compared to the 14-temperature case, though it was still more than twice as fast as the DI method. The reason why the SINH algorithm is becoming less advantageous is due to the high overlap between simulations. Moreover, the DI method is efficient at converging simulations that are close to each other. The CIAN algorithm was not able to successfully initialize all its free-energy differences in the 40-temperature case, owing to the large overlap between histograms. This is inherent to the initialization process of the free energies in the algorithm: only previous neighbors are considered, and stability becomes an issue when the overlap is important.

The efficiency of the equations proposed in this work strongly depend on the different number of histograms and their overlaps. Unlike the DI method which is very efficient for free energies that are close to each other, our algorithm relies on the overlap between neighboring histograms. The more overlap, the longer it will take to converge. This is actually good news, since less overlap means a larger parameter space span for fixed computer time. The following section explains how one should choose the optimal separation between neighboring simulations.

3. Optimized choice of simulation parameters

So far, we have addressed the question of how to optimize the solution of the WHAM equations for the free-energy parameters, $\{f_i\}$ from the histograms for a given set of computer simulations. However, since the ease of finding the free-energy parameters depends on how much overlap there is between histograms that are not nearest neighbors, we are naturally led to the question of what the optimal degree of overlap might be. For ease of solution of the equations for the free-energy parameters, we would like to have overlap only between nearest-neighbor histograms, but the primary criterion must remain the overall accuracy of the calculation.

At first sight, it might be thought that the final errors would be minimized by having a great deal of overlap to get more information about the free-energy parameters. This would suggest making the temperature differences between simulations as small as possible.

However, there is a compensating effect. To cover a given range of temperatures with a fixed amount of computer time, narrowing the spacing between histograms means that more simulations are needed. The simulations must be correspondingly shorter, and the statistical errors in each simulation corresponding larger.

The statistical errors in WHAM calculations for a given spacing between the simulations can be regarded as arising from two related sources. The first concerns the determination of the free-energy parameters $f_i = -\beta_i F(\beta_i)$ for each of the R simulations. The second source of error concerns the error in the free-energy as a function of temperature given the free-energy parameters.

To investigate the dependence of these errors on the separation between simulations, we use the nearest-neighbor ($q = 0$) approximation. This is a good approximation when the overlap between more distant neighbors is not too large, which turns out to be the case for the optimal separation.

3.1. Error in the free-energy parameters

Since we are using the nearest-neighbor approximation to estimate the errors in WHAM calculations, we need only consider the original Bennett Eq. (8). To simplify the equations, we ignore the relatively unimportant g_n factors. To further simplify the notation, we label the two simulations as 1 and 2, rather than i and $i+1$. This leaves us with

$$1 = \sum_E \frac{H_1(E) + H_2(E)}{N_1 + N_2 \exp\left[-\Delta\beta_1 E - \Delta f_1\right]}. \tag{22}$$

We begin by varying the values of the histograms and Δf_1 in Eq. (22),

$$0 = \sum_E \frac{\delta H_1(E) + \delta H_2(E)}{N_1 + N_2 \exp\left[-\Delta\beta_1 E - \Delta f_1\right]} + \sum_E \frac{H_1(E) + H_2(E)}{\left[N_1 + N_2 \exp\left[-\Delta\beta_1 E - \Delta f_1\right]\right]^2} N_2 \exp\left[-\Delta\beta_1 E - \Delta f_1\right] \delta\Delta f_1, \tag{23}$$

and averaging over all possible simulation results.

Since separate simulations are statistically independent, we have $\langle \delta H_1(E) \delta H_2(E') \rangle = 0$. We will also assume that the correlation time is zero, so that

$$\langle \delta H_1(E) \delta H_1(E') \rangle = \delta^2 H_1(E) \delta(E, E'), \tag{24}$$

where $\delta(E,E')$ is the Kronecker delta function. This approximation should only affect the magnitude of the error and not its dependence on overlap.

Moving the second term in Eq. (23) to the left side of the equation, squaring, and averaging over all possible simulations, we find

$$\delta^2\Delta f_1\left[\sum_E \frac{H_1(E)+H_2(E)}{[N_1+N_2\exp[-\Delta\beta_1 E-\Delta f_1]]^2}\exp[-\Delta\beta_1 E-\Delta f_1]\right]^2 N_2^2 = \sum_E \frac{\delta^2 H_1(E)+\delta^2 H_2(E)}{[N_1+N_2\exp[-\Delta\beta_1 E-\Delta f_1]]^2}. \tag{25}$$

Again using the assumption of zero correlation time to insert $\delta^2 H_i(E) = H_i(E)$ in Eq. (25), this becomes

$$\delta^2\Delta f_1\left[\sum_E \frac{H_1(E)+H_2(E)}{[N_1+N_2\exp[-\Delta\beta_1 E-\Delta f_1]]^2}\exp[-\Delta\beta_1 E-\Delta f_1]\right]^2 N_2^2 = \sum_E \frac{H_1(E)+H_2(E)}{[N_1+N_2\exp[-\Delta\beta_1 E-\Delta f_1]]^2} \tag{26}$$

Solving for $\delta^2\Delta f_1$, we find

$$\delta^2\Delta f_1 = N_2^{-2}\frac{\left[\sum_E \frac{H_1(E)+H_2(E)}{[N_1+N_2\exp[-\Delta\beta_1 E-\Delta f_1]]^2}\right]}{\left[\sum_E \frac{H_1(E)+H_2(E)}{[N_1+N_2\exp[-\Delta\beta_1 E-\Delta f_1]]^2}\exp[-\Delta\beta_1 E-\Delta f_1]\right]^2} \tag{27}$$

This expression for the error simplifies if the simulations are the same length ($N_1 = N_2 = N$), which should be sufficient for our purposes.

$$\delta^2\Delta f_1 = \frac{\left[\sum_E \frac{H_1(E)+H_2(E)}{[1+\exp[-\Delta\beta_1 E-\Delta f_1]]^2}\right]}{\left[\sum_E \frac{H_1(E)+H_2(E)}{[1+\exp[-\Delta\beta_1 E-\Delta f_1]]^2}\exp[-\Delta\beta_1 E-\Delta f_1]\right]^2} \tag{28}$$

This expression also explicitly displays the expected dependence of the error on the length of the simulation.

3.1.1. Approximation of the histograms by two Gaussians

To evaluate the dependence of the error in the free-energy parameters in Eq. (28), consider the case in which the probability distributions at both β_1 and β_2 have the same Gaussian form, centered about $\pm s\sigma/2$. To simplify the notation, we also define $x \equiv \Delta\beta E$.

$$H_1(x) = N(2\pi\sigma^2)^{-1/2}\exp\left[-\frac{(x+s\sigma/2)^2}{2\sigma^2}\right], \tag{29}$$

and

$$H_2(x) = N(2\pi\sigma^2)^{-1/2}\exp\left[-\frac{(x-s\sigma/2)^2}{2\sigma^2}\right]. \tag{30}$$

By symmetry, the exact solution for the free-energy parameters are $f_1 = f_2$, which gives $\Delta f_1 = 0$. Inserting this into Eq. (28) and changing the sums to integrals, we find

$$\delta^2\Delta f_1 = \frac{(2\pi\sigma^2)^{1/2}}{N}\frac{\left[\int dx \frac{\exp\left[-\frac{(x+s\sigma/2)^2}{2}\right]+\exp\left[-\frac{(x-s\sigma/2)^2}{2}\right]}{[1+\exp[-x]]^2}\right]}{\left[\int dx \frac{\exp\left[-\frac{(x+s\sigma/2)^2}{2}\right]+\exp\left[-\frac{(x-s\sigma/2)^2}{2}\right]}{[1+\exp[-x]]^2}\exp[-x]\right]^2}, \tag{31}$$

or

$$\delta^2\Delta f_1 = \frac{(2\pi\sigma^2)^{1/2}}{N}\frac{\left[\int dx \frac{\exp\left[-\frac{(x+s\sigma/2)^2}{2}\right]+\exp\left[-\frac{(x-s\sigma/2)^2}{2}\right]}{[1+\exp[-x]]^2}\right]}{\left[\int dx \frac{\exp\left[-\frac{(x+s\sigma/2)^2}{2}\right]+\exp\left[-\frac{(x-s\sigma/2)^2}{2}\right]}{[1+\exp[-x]]^2}\exp[-x]\right]^2}. \tag{32}$$

This expression can then be evaluated numerically, which will be done later in Section 3.3.1. It is shown in Appendix A that the error for large separations grows exponentially.

3.2. Error in the free-energy as a function of temperature, given the free-energy parameters

The error in the partition function Z can be estimated using the assumption that the errors in $W(E)$ are independent.

$$\delta^2 Z = \sum_E \delta^2 W(E)\exp[-2\beta E]. \tag{33}$$

The standard relationship for the error in the density of states $W(E)$ for the multiple histogram method (WHAM) [2] is

$$\delta W(E) = W(E)\left[\sum_i H_i(E)\right]^{-1/2}. \tag{34}$$

Combining the equations, this gives us

$$\delta^2 Z = \sum_E W(E)^2 \exp[-2\beta E]\left[\sum_i H_i(E)\right]^{-1}. \tag{35}$$

From this we can see that the error in the free-energy F is related to the error in Z by

$$\delta F = -\beta^{-1}\delta Z/Z. \tag{36}$$

Eq. (35) then becomes

$$\delta^2 F = Z^{-2}\sum_E W(E)^2 \exp[-2\beta E]\left[\sum_i H_i(E)\right]^{-1}. \tag{37}$$

3.2.1. Gaussian approximation for the error in the free-energy

Let us now approximate thermal distributions by Gaussians.

$$W(E)\exp[-\beta E]Z^{-1} \approx [2\pi\sigma^2]^{-1/2}\exp\left[-\frac{\beta(E-\langle E\rangle)^2}{2\sigma^2}\right]. \tag{38}$$

As Subsection 3.1, we will again ignore the variation of the width of the thermal distribution on temperature.

We are most interested in minimizing the maximum error in determining $F(\beta)$, given the free-energy parameters $\{f_n\}$. The maximum error will occur when the denominator in the sum is minimum, which is when $\langle E\rangle$ is located half way between the histograms on either side of it, or $\langle E\rangle = 0$. The maximum error in the free-energy as a function of temperature can then be found by evaluating Eq. (37) for $\langle E\rangle = 0$.

$$\delta^2 F_{max} = \beta^{-2}[2\pi\sigma^2]^{-1}\sum_E \exp\left[-\frac{\beta E^2}{\sigma^2}\right]\left[\sum_i H_i(E)\right]^{-1}. \tag{39}$$

As in Subsection 3.1, we denote the separation between the histograms by $\Delta E = s\sigma$. Inserting the approximations for the histograms in Eqs. (29) and (30), we find

$$\delta^2 F_{max} = N^{-1}\beta^{-2}[2\pi\sigma^2]^{-1/2}\times\sum_E \exp\left[-\frac{\beta E^2}{\sigma^2}\right]\left[\exp\left[-\frac{(x+s\sigma/2)^2}{2\sigma^2}\right]+\exp\left[-\frac{(x-s\sigma/2)^2}{2\sigma^2}\right]\right]^{-1}. \tag{40}$$

The error calculated from Eq. (40) will clearly grow rapidly as s increases, since the denominator in the sum has a minimum at $\beta E = 0$ where the numerator has a maximum.

3.3. Total error

The error in the free-energy parameters is given as $\delta^2\Delta f_1$ in Eq. (32), and the maximum error in the free-energy, given the free-energy parameters, is $\delta^2 F_{max}$ in Eq. (40).

Finally, for a fixed total number of MC configurations N_{total}, the number N devoted to each simulation must be reduced when the spacing between simulations is reduced and the number of simulations increases.

$$N = \frac{N_{total}}{M}. \tag{41}$$

Since s is inversely proportional to $M-1$, when M is large,

$$N = \frac{N_{total}}{M} \approx sN_{total}. \tag{42}$$

The square of the total error must then be proportional to

$$\delta^2 F_{total} \approx \beta^{-2}\frac{1}{sN_{total}}\left[\beta^{-2}\delta^2\Delta f_1 + \delta^2 F_{max}\right]. \tag{43}$$

By minimizing this quantity with respect to s, we should be able to determine the optimum separation between histograms.

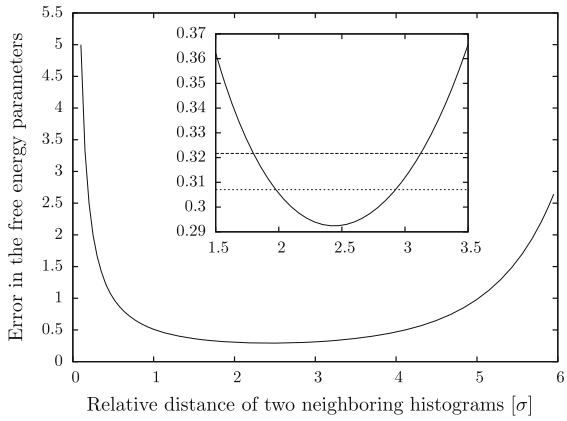

Fig. 3. Error in the free-energy parameters as a function of the relative distance between two histograms in units of σ, the width of the histogram (see Section 3.1.1). The inset shows the same plot at a finer scale, plus the five and ten percentile errors.

3.3.1. Numerical determination of the optimal separation between histograms

This result is shown on Fig. 3 where the error in the free-energy is plotted as a function of neighboring histograms distance. The minimum is located at 2.45σ, and the five and ten percentile errors are plotted as well. This graph disproves the common belief that close simulations will yield better results.

The advantage of using the optimized separation of 2.45σ between neighboring simulations is twofold: first we need fewer simulations for a given parameter space, which implies better statistics on each of them for fixed amount of computer time. Second, such a separation between histograms will yield fast convergence of our optimized algorithm.

This optimized spacing of histograms needs special attention when coming close to a phase transition. The width of a histogram grows with the specific heat ($\sigma^2 \propto C_V/\beta^2$), which means that a single histogram will likely represent an important part of the energy spectrum close to the phase transition. However, this simulation is most likely to experience critical slowing-down and suffer from important correlation times. This will be taken into account when converging free energies by properly evaluating the terms g_n in Eqs. (15) and (16).

The complete optimization of a set of simulations will depend on additional considerations. As we optimized the spacing between histograms based solely on free-energy convergence, it is unlikely that this will be identical for example with choices of temperatures that optimize parallel tempering [8]. Various optimization schemes have been proposed ranging from a constant acceptance probability of attempted swaps independent of temperature [9], to the minimization of round-trip times between the lowest and highest temperatures [10].

4. Summary

In this paper, we propose a new algorithm for solving the WHAM equations to estimate free energies out of a set of MC or MD simulations (see algorithm in Section 2.4.1). The algorithm uses free-energy differences, which provide a more natural way of approaching the problem. Fast convergence relies on histograms being far from each other (less overlap between histograms). We quantify this distance by providing an optimized choice of separation between histograms which minimizes the error made on the free energies. We showed for our test case – the 2D Ising model – that this separation favors the algorithm by successive iteration over neighboring histograms (SINH), based on overlap between histograms.

Even though the second part of this paper argues for a rather wide separation between histograms, there are various scenarios where the SINH algorithm is most likely a better candidate in terms of efficiency than the conventional DI method. It will turn out useful in the analysis of preexisting data, where histograms may not have been optimally spaced. Also, memory requirements can become important when the DI method is used on a large number of simulations because it relies on the complete data set of histograms in order to estimate each free-energy, rather than its relevant neighborhood.

Acknowledgments

One of us (RHS) would like to thank Thomas Neuhaus for stimulating discussions that led to the work described in this paper. TB would like to thank Markus Deserno for financial support.

Appendix A. Errors for large separations

For $s\sigma \gg 1$, we can approximate the factor $\exp[-x]$ for each Gaussian in Eq. (32) by $\exp[\pm s\sigma]$. This gives us

$$\delta^2\Delta f_1 \simeq \frac{(2\pi\sigma^2)^{1/2}}{N}\frac{\left[\int dx \frac{\exp\left[-\frac{(x+s\sigma/2)^2}{2}\right]}{[1+\exp[s\sigma]]^2}\right] + \left[\int dx \frac{\exp\left[-\frac{(x-s\sigma/2)^2}{2}\right]}{[1+\exp[-s\sigma]]^2}\right]}{\left[\int dx \frac{\exp\left[-\frac{(x+s\sigma/2)^2}{2}\right]}{[1+\exp[s\sigma]]^2}\exp[s\sigma] + \int dx \frac{\exp\left[-\frac{(x-s\sigma/2)^2}{2}\right]}{[1+\exp[-s\sigma]]^2}\exp[-s\sigma]\right]^2}, \tag{44}$$

leading to

$$\delta^2\Delta f_1 \simeq \frac{1}{N}\frac{\left[[1+\exp[s\sigma]]^{-2} + [1+\exp[-s\sigma]]^{-2}\right]}{\left[[1+\exp[s\sigma]]^{-2}\exp[s\sigma] + [1+\exp[-s\sigma]]^{-2}\exp[-s\sigma]\right]^2}. \tag{45}$$

Again using $s\sigma \gg 1$, we can write

$$\delta^2\Delta f_1 \simeq \frac{1}{N}\frac{[\exp[2s\sigma]+1]}{[\exp[-s\sigma]+\exp[-s\sigma]]^2} \simeq \frac{1}{4N}\exp[4s\sigma]. \tag{46}$$

This is an exponentially large error.

References

[1] A.M. Ferrenberg, R.H. Swendsen, Optimized Monte Carlo data analysis, Phys. Rev. Lett. 63 (1989) 1195–1198.
[2] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules, J. Comp. Chem. 13 (1992) 1011–1021.
[3] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, Multidimensional free-energy calculations using the weighted histogram analysis method, J. Comp. Chem. 16 (1995) 1339–1350.
[4] C.H. Bennett, Efficient estimation of free-energy differences from Monte Carlo data, J. Comp. Phys. 22 (1976) 245–268.
[5] H. Müller-Krumbhaar, K. Binder, Dynamic properties of the Monte Carlo method in statistical mechanics, J. Stat. Phys. 8 (1973) 1–24.
[6] A.M. Ferrenberg, D.P. Landau, R.H. Swendsen, Statistical errors in histogram reweighting, Phys. Rev. E 51 (1995) 5092–5100.
[7] A.M. Ferrenberg, Efficient use of Monte Carlo simulation data, Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, 1989.
[8] R.H. Swendsen, J.S. Wang, Replica Monte Carlo simulation of spin-glasses, Phys. Rev. Lett. 57 (1986) 2607–2609.
[9] A. Kone, D.A. Kofke, Selection of temperature intervals for parallel-tempering simulations, J. Chem. Phys. 122 (2005).
[10] H.G. Katzgraber, S. Trebst, D.A. Huse, M. Troyer, Feedback-optimized parallel-tempering Monte Carlo, J. Stat. Mech. (2006).