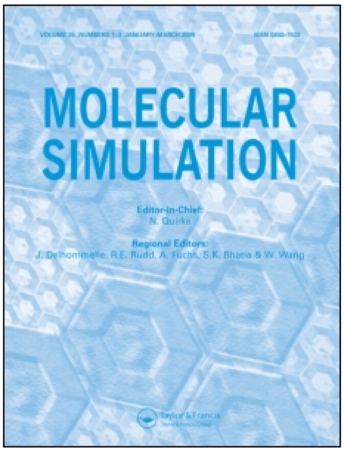

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Methods for Minimizing Errors in Linear Thermodynamic Integration

Jürgen Schlitter[a]
[a] Institut für Biophysik, Ruhr-Universität, Bochum, Germany

To cite this Article Schlitter, Jürgen(1991) 'Methods for Minimizing Errors in Linear Thermodynamic Integration', Molecular Simulation, 7: 1, 105 — 112

To link to this Article: DOI: 10.1080/08927029108022452

URL: http://dx.doi.org/10.1080/08927029108022452

Molecular Simulation, 1991, Vol. 7, pp. 105–112
Reprints available directly from the publisher
Photocopying permitted by license only

METHODS FOR MINIMIZING ERRORS IN LINEAR THERMODYNAMIC INTEGRATION

JÜRGEN SCHLITTER

Institut für Biophysik, Ruhr-Universität, D-W-4630 Bochum, Germany

(*Received September 1990, accepted October 1990*)

Thermodynamic integration (TI) as introduced by Kirkwood is a well established technique for calculating free energy differences in physical and chemical systems. The integrand is an average of a difference energy function usually being estimated by means of simulation methods like Monte Carlo or molecular dynamics. Inherent statistical errors and the singular behaviour in interesting applications can make TI difficult and costly. In this work a unique transformation is derived which minimizes the statistical error in linear TI. It is shown that also singularities and integration errors are completely removed in this way for an important class of applications. A numerical example is used to demonstrate the applicability of three different methods derived from the general rule.

KEY WORDS: Thermodynamic integration, perturbation methods, free energy.

1. INTRODUCTION

The computation of experimentally relevant free energy or enthalpy differences by the method of thermodynamic integration (TI) originating from Kirwood [1] has recently found wide application due to the development of simulation software for microscopic systems in physics and chemistry. For a review, see Mezei and Beveridge [2]. TI is applicable whenever a thermodynamic change of state can be described by transition from an initial energy function E_0 to a final E_1. The corresponding free energy change is simply

$$\Delta A = \int_0^1 \langle \partial E/\partial \lambda \rangle_\lambda d\lambda \tag{1.1}$$

where $E(\lambda)$ is an arbitrary continuous function with $E(0)$ and $E(1) = E_1$ and the average is taken over an ensemble governed by $E(\lambda)$. Linear TI denotes applications where $E(\lambda)$ is chosen to be linear in λ. There are three problems arising in practical calculations: (i) singularities of the integrand making the numerical integration difficult and costly, (ii) statistical errors with the same effect, and (iii) long relaxation times inducing systematic errors in the evaluation of mean values. This paper is dealing with the first two problems.

The problem of singularities arises e.g. when a van der Waals potential with a divergent repulsive core is switched on/off or changed in simulations of particle creation/anhilation or substitution and has attracted theorists since a long time [2]. Some autors truncated the repulsive core above a few kT [cited in 2]. If the functional form of the integrand is known near the singularity, which is indeed the case sometimes [3,2], the crucial part of the integral may be computed from only one value in the neighborhood [2]. Nonlinear TI resting on a linear change the parameters of a Lennard-Jones potential was used by Cross [4]. By a coordinate transformation condensing the grid points at the singularity a nonsingular integrand could be generated by Squire and Hoover [3] and Mruzik et al. [5]. The problem of statistical errors seems not have been considered in connection with the integration problem.

The present work starts from the observation that the behaviour of the statistical error near a singularity is even worse than the one of the above integrand. Hence minimization of the total statistical error is expected also to solve the problem of numerical integration. Apart from an analytical prescription an algorithmic approach is developed aiming at practical applications.

2. BASIC RELATIONS

Although the method is applicable to more general problems, we shall confine the consideration to a canonical ensemble of classical systems which undergo a change of state by transition from an initial potential function V_0 to a final V_1. Using linear TI one obtains the corresponding free energy change by the integral

$$\Delta A = \int_0^1 f(\lambda)\, d\lambda \tag{2.1}$$

where $f(\lambda) = \langle \Delta V \rangle_\lambda$ is the mean value of the difference $\Delta V = V_1 - V_0$ taken for an ensemble with an intermediate potential $V(\lambda) = (1 - \lambda)\, V_0 + \lambda V_1 = V_0 + \lambda \Delta V$,

$$f(\lambda) = \int \Delta V \exp(-\beta V(\lambda))\, dq / \int \exp(-\beta V(\lambda))\, dq \tag{2.2}$$

The variable q stands for the complete set of cartesian space variables defining the configuration of the system.

The entropy difference ΔS is most easily calculated from ΔA and the energy difference $\Delta U = \langle V_1 \rangle_1 - \langle V_0 \rangle_0$ as $\Delta S = \beta(\Delta U - \Delta A)$, but there is an alternative form [2]

$$T\Delta S = -\beta \int_0^1 (\langle V(\lambda)\Delta V \rangle_\lambda - \langle V(\lambda) \rangle_\lambda \langle \Delta V \rangle_\lambda)\, d\lambda \tag{2.3}$$

which offers an independent check of the previous method.

In all interesting cases the integrals (2.1) and (2.3) have to be calculated by numerical integration over averages obtained by some kind of computer simulations like Monte Carlo (MC) or molecular dynamics (MD), which are contaminated by statistical errors. It is a special advantage of linear TI that the variance of ΔV is readily expressed as

$$\sigma^2_{\Delta V} = -f'(\lambda)/\beta \tag{2.4}$$

where the prime denotes the derivative with respect to λ. Therefore, if f is calculated as the average of M independent values of ΔV, its variance becomes

$$\sigma_f^2 = -f'(\lambda)/(\beta M) \tag{2.5}$$

As the variance cannot be negative, this implies a monotonous decrease of $f(\lambda)$ and gives a hint to where statistical errors become important. One finds quite generally that divergent potential terms contained in ΔV can lead to divergent functions f and f'. In particular, if one has a potential V_1 proportional to r^{-m}, where r is the distance between particles, it turns out that $f \propto \lambda^{-(1-d/m)}$ at small λ, d being the dimension of the one-particle space [3,2]. For instance, van der Waals repulsion in three dimensions with $m = 12$ makes f diverge like $\lambda^{-3/4}$ and $\sigma_f^2 \propto f$ like $\lambda^{-7/4}$. The divergence of f, which is still an integrable function, will in general induce considerable integration errors. In addition large contributions to the variance are produced near $\lambda = 0$.

By variety of coordinate transformations, one can generate a regular smoothly variing integrand as shown e.g. by Mruzik *et al.* [5]. Among these transformations, however, the best choice will be the one which gives the minimum total error by minimizing the statistical error. The even stronger divergence of σ_f^2 in the example is suggesting an approach which starts directly by minimizing the statistical instead of the integration error. This proceeding is expected to yield a unique prescription and to solve both problems obviously being closely related.

3. NUMERICAL INTEGRATION

3.1 Optimal grid points

For the numerical integration of (2.1) we first introduce a new variable $n(\lambda)$ denoting the number of intervals between 0 and λ. For the sake of convenience $n(\lambda)$ is taken as a continuous monotonously increasing function of λ with

$$n(0) = 0 \text{ and } n(1) = N, \tag{3.1}$$

where N is the number of integration steps. Substituting $\lambda = \lambda(n)$ in (2.1) one obtains

$$\Delta A = \int_0^N dn f(\lambda)/n'(\lambda) \tag{3.2}$$

The variance of ΔA can be written as

$$\sigma_{\Delta A}^2 = \int_0^N dn\, \sigma_f^2/(n'(\lambda))^2 \tag{3.3}$$

or, by inserting (2.5), as

$$\sigma_{\Delta A}^2 = -(1/\beta M) \int_0^N dn f'(\lambda)/(n'(\lambda))^2 \tag{3.4}$$

The validity of (3.3) is easily shown by rewriting the integrals as sums, see (3.8) below. In order to minimize the statistical error for a given function f, one has to apply the variational condition $\delta\sigma_{\Delta A}^2 = 0$ with respect to the density of intervals, $n'(\lambda)$ with the constraints (3.1), and obtains by standard methods

$$n'(\lambda)(f'(\lambda))^{1/2} \propto \sigma_{\Delta V} \propto \sigma_f \tag{3.5}$$

Any of the relations (3.5) can be used, depending on which of the functions is known. Inserting of (3.5) in (3.3) or (3.4) shows that the condition is equivalent to choosing integration intervals in such a way that each will contribute the same amount to the total error.

Let us now consider the integration error of the transformed perturbation integral (3.2) for divisions chosen according to the condition (3.5) which so far was only shown to minimize the statistical error. One finds that the integrand f/n' vanishes at $n = 0$ and is just a linear function of n, $f/n' = \text{const}\cdot n$, whenever $f(\lambda)$ is an integrable power of λ, $f \propto \lambda^m$ with $m \neq 0$. This follows from (3.5) since $n' \propto \lambda^{(m-1)/2}$, $f/n' \propto \lambda^{(m+1)/2}$, and $n \propto \lambda^{(m+1)/2}$. In this very important case the numerical integration of (3.2) by means of the simple trapezoidal rule is exact without any error remaining by numerical integration. In all practical cases the latter is hence expected to yield reliable results. The perturbation integral (3.2) is expressed as

$$\Delta A = \sum_{n=1}^{N-1} f_n/n'_n + (f_0/n'_0 + f_N/n'_N)/2 \qquad \text{(Method A)} \qquad (3.6)$$

where Δn was set equal to 1, $\lambda_n = \lambda(n), f_n = f(\lambda_n)$ and so on.

We now give the optimal abscissas for the above van der Waals problem. As $f \propto \lambda^{-3/4}$, (3.5) together with (3.1) yields

$$\lambda(n) = \lambda_n = (n/N)^8 \text{ for } n = 0, 1 \ldots N \qquad (3.7)$$

Accuracy is achieved by the high density of grid points near $\lambda = 0$.

The particular choice (3.7) is of course appropriate to all cases where van der Waals potentials are contained in the perturbation, but does not give minimum errors regarding other contributions like binding potentials. As the functional form of f is not known then, it is impossible to find the best choice for λ_n *a priori*. Methods of minimizing the statistical error *a posteriori*, i.e. by using information obtained in the course of simulation, are presented in the next section.

3.2 Iterative improvement

To simplify the discussion we first use the relation $d\lambda = dn/n'(\lambda)$ and rewrite the above integrals in the less accurate form

$$\Delta A = \sum_{n=1}^{N-1} f_n h_n \qquad \text{and} \qquad \sigma^2_{\Delta A} = \sum_{n=1}^{N-1} \sigma^2_{f,n} h^2_n \qquad (3.8)$$

where $h_n = \lambda_n - \lambda_{n-1}$ and the distribution of grid points is given by (3.7) or some other approximate prescription. Consider now the contibution of one interval, $\sigma^2 h^2$, to the total variance. If σ^2 is roughly constant over the interval one may calculate an additional value of f in the middle of the interval. As their contributions are proportional to the square of the step size, each subinterval gives only one fourth of the original value, and the total variance is diminished by $\sigma^2 h^2/2$. A better rule which takes into account that σ^2 can vary over the interval and the stepsize should, according to (3.5), be taken proportional to σ^{-1} is used in the following scheme.

Method B

(i) start with an appropriate distribution $\lambda_0 \ldots \lambda_N$
(ii) pick out the interval n giving the largest error
(iii) insert a new $\lambda = \lambda_{n-1} + h_n \sigma_{f,n}/(\sigma_{f,n-1} + \sigma_{f,n})$
(iv) remove the original contributions of the interval and instead insert the values of the new subintervals in the sums (3.8)
(v) go to (i)

In this way ΔA is calculated with increasing accuracy and the statistical error is diminished, the contributions of all becoming approximately equal as claimed by the general rule.

The shortcoming of the previous method is the fact that the continuous transition from the initial to the final state is disturbed, which leads to unfavorable starting configurations for sampling. The following method is designed to avoid this problem.

Method C

(i) start with a few equidistant λ_n
(ii) estimate the standard deviation σ_f for each interval from the values calculated at both ends
(iii) divide the interval by a number of divisions proportional to σ_f
(iv) calculate ΔA for increasing values of λ

The method approximately satisfies the condition (3.5) and circumvents the problem of method B with a little larger amount of computational work. Clearly the methods B and C cannot replace method A in intervalls where the variance diverges, but help to reduce errors anywhere else. In the last step ΔA should always be calculated with the trapeziodal rule instead of (3.8).

4. NUMERICAL EXAMPLE

The methods developed above were tested at a one-dimensional example modelling the rupture of a chemical bond. As shown in Figure 1, the initial state is given by a

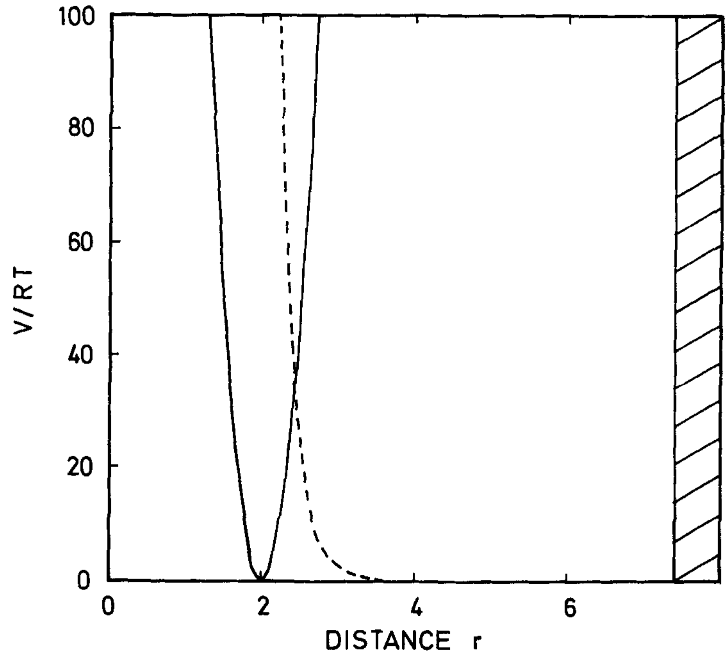

Figure 1 The figure shows the initial binding potential V_0 (full line) and the final repulsive van der Waals potential V_1 (dashed) in units of RT. The hatched area is excluded by the square potential. The energy scale was extended to large values to give an idea of the immense variation of $\Delta V = V_1 - V_0$ for which averages have to be calculated. In the initial state the system is near $r = 2$ with large values of V_1 in the final state large values of $-V_0$ are sampled at $r > 3$.

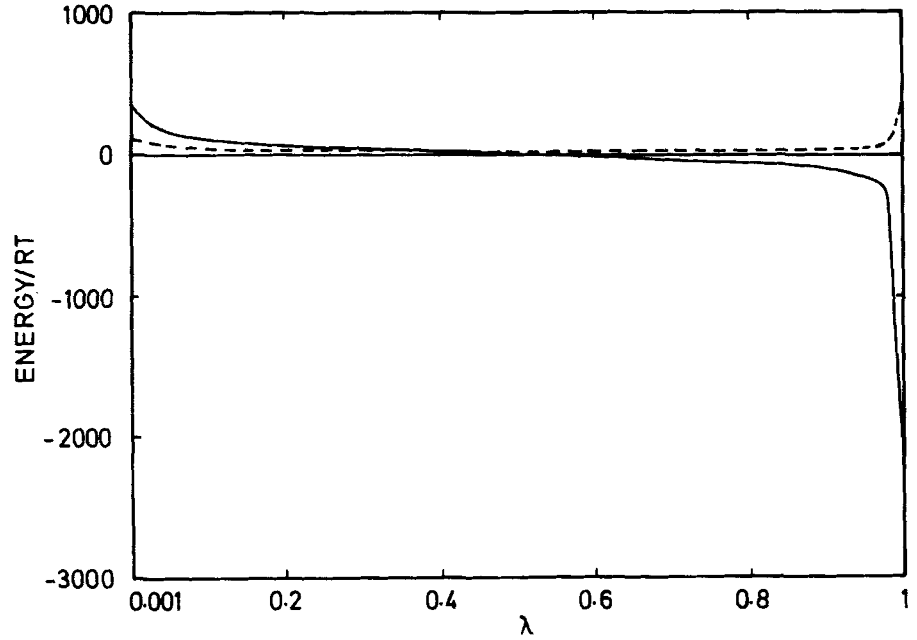

Figure 2 The function $f(\lambda) = \langle\Delta V\rangle_\lambda$ (full line) and its standard deviation $\sigma_{\Delta V}$ (dashed) are shown for $\lambda \geqslant 0.001$, i.e. outside the singularity which occurs at $\lambda = 0$. Note that at both ends values much larger than RT are found. The final ΔA is only a few RT due to compensation of larger positive and negative contributions, which makes integration sensitive against any kind of errors.

harmonic binding potential allowing fluctuations of ± 0.05 of the distance variable r about a mean value of 2, whereas in the final state there is a repulsive van der Waals potential, with $V_0/RT = 200\,r^2$ and $V_1/RT = 1.36 \times 10^6/r^{12}$. The system is contained in a potential well with infinitely high walls confining the distance to $0 < r < 7.3$. Mean values of ΔV were calculated with the MC technique [6,7]. As preliminary calculations had yielded a correlation length of less than 10 steps for ΔV all averages were calculated from 100 independent values out of a sample of 1000 MC steps.

Figure 2 shows $f(\lambda) = <\Delta V>_\lambda$ and the standard deviation $\sigma_{\Delta V}$ as computed from (2.2) as (2.4). f tends to infinity as $\lambda \to 0$ where the repulsive potential arises and becomes again large at $\lambda = 1$ where the binding potential is finally vanishing. A similar behaviour is shown by the error. We hence decided (i) to take method A for $0 < \lambda \leqslant 0.001$ and (ii) to use iterative improvement for the remaining interval. The figure also demonstrates that at both ends considerable integration errors can occur due to the curvature of the graph.

In the one-dimensional case one has $f(\lambda) \propto \lambda^{-11/12}$ which is a stronger divergence than in three dimensions and requires grid points $\lambda_n \propto n^{24}$ near $\lambda = 0$ according to (3.5). We set $\lambda_n = 0.001\,(n/N)^{24}$ with $N = 3$ and 5, and found excellent results with method A as shown in the table. It is clear from Figure 2 that extension of the interval up to $\lambda \sim 0.8$ would have been possible without loss of accuracy.

For testing the other methods we used instead 50 grid points for $0.001 \leqslant \lambda \leqslant 1$ chosen in different ways. The largest error occurs at equidistant points, see Table 1.

Table 1 Free energy differences calculated by MC simulation and different methods of thermodynamic integration for the example described in the text.

interval	*method applied*	$\Delta A_{calculated}/RT$
$0 < \lambda \leqslant 0.001$	method A, 3 grid points	1.371 ± 0.037
	5 grid points	1.383 ± 0.038
$0.001 \leqslant \lambda \leqslant 1$	equidistant grid points	-4.231 ± 2.180
	method C	-4.639 ± 1.129
	method B	-4.557 ± 0.810
	exact	-4.762

Method B and C were started with 10 equidistant points being completed by further 40 points according to the respective prescription. In both cases most additional points are inserted near $\lambda = 1$ and less at the left end. The smallest error is obtained by method B which iteratively makes use of the increasing information and puts more points at the right end. For comparison the table also gives the exact value for the interval as calculated directly from (2.2) and (2.1) by Simpson integration.

5. SUMMARY

Apart from its simple implementation in numerical codes, linear TI has the advantage to exhibit a simple relation between the integrand $f(\lambda)$ of the perturbation integral and its statistical error. This allows one to find a prescription as to how the statistical error is minimized by either using the analytical form of f or the "experimentally" accessible variance of a sample average. The first method (A) is essential for the treatment of singularities arising e.g. at particle creation/annihilation. The second approach (methods B and C) will be successful in all other cases.

The methods are based on a transformation yielding optimally distributed grid points, which is the only freedom one has in the framework of linear TI. They are chosen in such a way that each interval $n = 1 \ldots N$ yields the same contribution to the total statistical error. In this sense the error is a constant function of n. At least for an important class of problems the integrand was shown to be linear in n, which removes completely the numerical integration error. This is a hint that also the more complicated nonlinear TI probably cannot further reduce the total error.

For the three-dimensional van der Waals problem our method A yields a transformation similar to the one of Squire and Hoover [3] which used a power of 4 (instead of 8) in order to generate a constant integrand. This approach obviously aimes at merely minimizing the integration error, but their satisfactory results, see also [5], show that condensation of grid points near a singularity is the right strategy for minimizing the total error.

Acknowledgment

The author would like to thank D. Husmeier for doing the numerical calculations and C. Bassaris for preparing the figures.

References

[1] J.G. Kirkwood, "Statistical mechanics of fluid mixtures", *J. Chem. Phys.*, **3**, 300 (1935).

[2] M. Mezei and D.L. Beveridge, "Free energy simulations", *Ann. N.Y. Acad. Sci.*, **482**, 1 (1986).

[3] D.R. Squire and W.G. Hoover, "Monte Carlo simulation of vacancies in rare-gas crystals", *J. Chem. Phys.*, **50**, 701 (1969).

[4] A. Cross, "Influence of hamiltonian parametrization on convergence of Kirkwood free energy calculations", *Chem. Phys. Lett.*, **128**, 198 (1986).

[5] M.R. Mruzik, F.F. Abraham, D.E. Schreiber, and G.M. Pound, "A Monte Carlo study of ion-water clusters", *J. Chem. Phys.*, **64** 481 (1976).

[6] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, "Equation of state calculations by fast computing machines", *J. Chem. Phys.*, **21**, 1087 (1953).

[7] K. Binder, ed, Monte Carlo Methods in Statistical Physics, Springer, Berlin, 1979.