

Zero-Variance Principle for Monte Carlo Algorithms

Roland Assaraf* and Michel Caffarel†

CNRS-Laboratoire de Chimie Théorique Tour 22-23, Université Pierre et Marie Curie,
4 place Jussieu, 75252 Paris Cedex 05, France

(Received 7 June 1999)

We present a general approach to greatly increase at little cost the efficiency of Monte Carlo algorithms. To each observable to be computed we associate a renormalized observable (improved estimator) having the same average but a different variance. By writing down the zero-variance condition a fundamental equation determining the optimal choice for the renormalized observable is derived (zero-variance principle for each observable separately). We show, with several examples including classical and quantum Monte Carlo calculations, that the method can be very powerful.

PACS numbers: 02.70.Lq, 05.10.Ln, 05.50.+q, 75.40.Mg

Since the pioneering work of Metropolis *et al.* [1] Monte Carlo methods have been widely used in many areas of natural sciences. At the root of Monte Carlo methods lies a very efficient stochastic method for calculating many-dimensional integrals (or sums) written under the general form

$$\langle O \rangle \equiv \frac{\int_S dx \pi(x) O(x)}{\int_S dx \pi(x)}, \quad (1)$$

where $O(x)$ is some arbitrary observable (real-valued function) defined on the configuration space S (continuous or discrete) and $\pi(x)$ some probability distribution. In Monte Carlo methods the integrals are evaluated using a large but finite set of configurations $\{x^{(i)}\}_{i=1,N}$ distributed according to π and generated by a step-by-step stochastic procedure (Markov chain),

$$\langle O \rangle = \frac{1}{N} \sum_{i=1}^N O[x^{(i)}] + \delta O, \quad (2)$$

where δO is the statistical error associated with the finite statistics. For a large enough number N of Monte Carlo steps, standard statistical arguments lead to the following expression of the error:

$$\delta O = K \frac{\sigma(O)}{\sqrt{N}}, \quad (3)$$

where K is some positive constant proportional to the amount of correlation between configurations, and $\sigma(O)$ is a measure of the fluctuations of the observable,

$$\sigma(O) \equiv \sqrt{\langle O^2 \rangle - \langle O \rangle^2}. \quad (4)$$

In this Letter, it is shown that by introducing a suitably renormalized observable $\tilde{O}(x)$ the statistical error can be drastically reduced and even suppressed, thus defining a zero-variance principle for the Monte Carlo calculation of observables. To realize this, a trial operator H and a trial function $\psi(x)$ are introduced (a trial matrix and a trial vector in the discrete case). The operator H is supposed to be Hermitian (in all practical applications, real symmetric) and is chosen such that

$$\int dy H(x, y) \sqrt{\pi(y)} = 0. \quad (5)$$

On the other hand, the trial function $\psi(x)$ is a rather arbitrary function which is simply supposed to be integrable.

Now, the renormalized observable $\tilde{O}(x)$ associated with the observable $O(x)$ is defined as follows:

$$\tilde{O}(x) = O(x) + \frac{\int dy H(x, y) \psi(y)}{\sqrt{\pi(x)}}. \quad (6)$$

As a direct consequence of Eq. (1) and of the very definition of the Hermitian operator H , Eq. (5), we have the important property

$$\langle \tilde{O} \rangle = \langle O \rangle. \quad (7)$$

In other words, both quantities $O(x)$ and $\tilde{O}(x)$ can be used as estimators of the desired average. However, the statistical errors, which are controlled by $\sigma(O)$ and $\sigma(\tilde{O})$, can be very different. The optimal choice for (H, ψ) is obtained by imposing the renormalized function to be constant and equal to the exact average. This leads to the following fundamental equation:

$$\begin{aligned} \int dy H(x, y) \psi(y) &= -[O(x) - \langle O \rangle] \sqrt{\pi(x)} \\ \Leftrightarrow \sigma(\tilde{O}) &= 0. \end{aligned} \quad (8)$$

At this point it should be emphasized that the idea of using renormalized estimators for reducing the variance is not new. A number of applications have been performed using various “improved” estimators having a lower variance (see, e.g., [2,3]). The basic idea is to construct new estimators by integrating out some intermediate degrees of freedom and, therefore, removing the corresponding source of fluctuations. However, to the best of our knowledge, no general and systematic approach based on a zero-variance principle and valid for any type of Monte Carlo methods has been proposed so far.

In this work the following strategy is proposed. First, a Hermitian operator H verifying (5) is chosen. Second, some approximate solution of Eq. (8) is searched for. The various parameters entering ψ are then optimized by minimizing the fluctuations of the renormalized observable over a finite set of points distributed according to π and obtained from a short Monte Carlo calculation. Finally, a standard much longer Monte Carlo simulation is performed using $\tilde{O}(x)$ instead of $O(x)$ as estimator.

Choice of H .—Clearly, a large variety of choices are possible for the trial operator H . For Monte Carlo algorithms satisfying the detailed balance condition (in practice, the vast majority of MC schemes) a very natural choice is at our disposal. Denoting $p(x \rightarrow y)$ the transition probability distribution defining the Monte Carlo dynamics, the detailed balance condition is written as $\pi(x)p(x \rightarrow y) = \pi(y)p(y \rightarrow x)$ for all pairs (x, y) in configuration space. A most natural operator to consider is

$$H(x, y) = \sqrt{\frac{\pi(x)}{\pi(y)}} [p(x \rightarrow y) - \delta(x - y)]. \quad (9)$$

From the detailed balance condition it follows that the operator H is symmetric, $H(x, y) = H(y, x)$. The fundamental property (5) is verified since the sum-over-final states for a transition probability is equal to one. For continuous systems Schrödinger-type Hamiltonians can also be considered

$$H = -\frac{1}{2} \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} + V(x), \quad (10)$$

where $V(x)$ is some local potential constructed to fulfill condition (5):

$$V(x) = \frac{1}{2\sqrt{\pi(x)}} \sum_{i=1}^d \frac{\partial^2 \sqrt{\pi(x)}}{\partial x_i^2}, \quad (11)$$

where d is the number of degrees of freedom. Note that in Eq. (10), H is written using the standard quantum-mechanical notation for a *local* Hamiltonian in the x -space realization.

Choice of ψ .—Once the operator H has been chosen, the optimal choice for ψ is the exact solution of the fundamental equation. Of course, in practice only approximate solutions are available. What particular form to choose for ψ is very dependent on the problem at hand, on the type of observables considered, and also on the form chosen for the trial operator H . However, a most important point to be stressed is that the global normalization factor associated with ψ is a pertinent parameter of the trial function. Minimizing the fluctuations of the renormalized function $\sigma(\tilde{O})$ with respect to it, we get

$$\sigma(\tilde{O})^2 = \sigma(O)^2 - \frac{\left\langle \frac{O(x) \int dy H(x, y) \psi(y)}{\sqrt{\pi(x)}} \right\rangle^2}{\left\langle \left(\frac{\int dy H(x, y) \psi(y)}{\sqrt{\pi(x)}} \right)^2 \right\rangle}. \quad (12)$$

The correction to $\sigma(O)^2$ being negative, we obtain the important result that, whatever the choice made for the trial function (even the most unphysical one), the optimization of the multiplicative factor always leads to a reduction of the statistical error.

Our first application concerns the Monte Carlo calculation of the internal energy of the standard 2D Ising model at various temperatures and linear sizes $L = 5, 10, 20$, and 25. The observable considered is the en-

ergy function given by $E(S) = -\sum_{\langle i, j \rangle} S_i S_j$ (coupling constant $J = 1$, sum limited to nearest neighbors, and periodic boundary conditions). The probability distribution is $\pi(S) = \exp[-\beta E(S)]$ with $\beta = 1/k_B T$. Here, $S \equiv (S_1, \dots, S_N)$ with $S_i = \pm 1$, and $N = L \times L$ is the total number of spins. Simulations have been performed using a Swendsen-Wang-type algorithm [4] (nonlocal updates of clusters of spins). To construct the trial operator H we have chosen to use the transition probability distribution of Monte Carlo algorithms with local updates (“heat-bath”-type algorithms). The probability of flipping the spin $S_i = \pm 1$ at site i is given by

$$p(S_i \rightarrow \epsilon S_i) = \frac{e^{\beta \epsilon S_i \tilde{S}_i}}{e^{\beta S_i \tilde{S}_i} + e^{-\beta S_i \tilde{S}_i}}, \quad (13)$$

where $\epsilon = 1$ (no flip) or -1 (flip), and \tilde{S}_i is the sum of neighboring spin values. With this choice and using Eq. (9) the fundamental equation (8) can be rewritten under the form

$$\sum_{i=1}^N p(S \rightarrow T_i S) [Q(S) - Q(T_i S)] = E(S) - \langle E \rangle, \quad (14)$$

$$\psi(S) = Q(S) \sqrt{\pi(S)},$$

where the application T_i ($i = 1, \dots, N$) describes a flip at site i , and is defined by $T_i(S_1, \dots, S_i, \dots, S_N) = (S_1, \dots, -S_i, \dots, S_N)$. At $\beta = 0$ ($T = \infty$) the transition probability distribution becomes constant and the exact solution is easily found to be $Q(S) = E(S)/2$. For finite temperatures some approximate solution has to be found. Here we introduce for $Q(S)$ a polynomial expansion up to the fourth order in the variables $X = \sum_{i=1}^N S_i$ (magnetization) and $Y = \sum_{i=1}^N g(S_i \tilde{S}_i)$ [“generalized energy,” the usual energy being recovered for $g(x) = x$]. Precisely, we have chosen the form $Q(S) = e^{-Z} \sum_{n+m \leq 4} c_{nm} X^n Y^m$ where $Z = \sum_{i=1}^N h(S_i \tilde{S}_i)$. The set of variational parameters of ψ consists of all coefficients c_{nm} of the polynomial plus the ten possible values of functions g and h . All coefficients have been optimized by minimizing the fluctuations of the renormalized energy $\sigma(\tilde{E})$ defined by (4) and calculated from 2000 to 5000 different spin configurations $S^{(i)}$ drawn according to π . Finally, the last step consists in performing a long Monte Carlo simulation to compute accurately the various quantities. The number of clusters built varies from 10^6 (for the larger size) to 2×10^8 (for the smaller size). Results are presented in Table I. Three different temperatures have been considered. $T = 3$ corresponds to the low-temperature regime, $T = T_c = 4/\ln(\sqrt{2} + 1)$ is the critical temperature for the infinite lattice, and $T = 8$ is in the high-temperature regime of the model. At $T = 3$, our representation is extremely good whatever the size of the lattice considered. The variance associated with the renormalized energy is drastically reduced with respect to the bare value and the gain in computational effort can be as great as ~ 360 . Here, the gain in computational effort is defined as the ratio of the squared statistical errors $(\delta \tilde{E} / \delta E)^2$. In

TABLE I. Internal energies for the 2D Ising model at different temperatures. N is the number of sites. Statistical uncertainties on the last digit are indicated in parentheses.

Size	5×5	10×10	20×20	25×25	$\infty \times \infty$
$T = 3$					
$\sigma(E)^2/N$	1.789(1)	1.777(2)	1.78(1)	1.79(1)	
$\sigma(\tilde{E})^2/N$	0.012 5(4)	0.006 1(1)	0.006 0(2)	0.006 1(2)	
Ratio of variances	~ 143	~ 291	~ 297	~ 293	
$\langle E/N \rangle$	-3.902 044(31)	-3.902 200(55)	-3.902 17(21.4)	-3.902 42(29)	
$\langle \tilde{E}/N \rangle$	-3.902 020(2.4)	-3.902 229(3)	-3.902 25(1.2)	-3.902 22(1.5)	
Gain in computational effort ^a	~ 167	~ 336	~ 318	~ 360	
$\langle E/N \rangle$ Exact value	-3.902 021 4...				-3.902 233 1... ^b
$T = T_c = 4.538 37 \dots$					
$\sigma(E)^2/N$	18.581(4)	25.97(3)	33.1(2)	35.3(2)	
$\sigma(\tilde{E})^2/N$	0.215(2)	4.85(1)	16.5(1)	16.9(2)	
Ratio of variances	~ 86	~ 5.4	~ 2.0	~ 2.1	
$\langle E/N \rangle$	-3.073 34(13)	-2.952 14(33)	-2.890 2(12)	-2.880 0(14)	
$\langle \tilde{E}/N \rangle$	-3.073 45(1.3)	-2.952 36(13)	-2.890 8(7)	-2.878 8(8)	
Gain in computational effort ^a	~ 100	~ 6.4	~ 3	~ 3.1	
$\langle E/N \rangle$ Exact value	-3.073 439 6...				-2.828 427 1... ^b
$T = 8$					
$\sigma(E)^2/N$	13.17(1)	10.96	11.1(2)	10.9(3)	
$\sigma(\tilde{E})^2/N$	0.041	0.455	0.8(1)	0.9(1)	
Ratio of variances	~ 321.2	~ 24	~ 13.9	~ 12	
$\langle E/N \rangle$	-1.164 40(33)	-1.115 56(48)	-1.115 6(20)	-1.116 5(25.6)	
$\langle \tilde{E}/N \rangle$	-1.163 48(1.6)	-1.115 02(8.2)	-1.114 5(4.4)	-1.114 5(5.6)	
Gain in computational effort ^a	~ 425	~ 34	~ 20.7	~ 20.9	
$\langle E/N \rangle$ Exact value	-1.163 492 6...				-1.114 544 4... ^b

^aSee text for definition.^bReference [5].

other words, according to Eq. (3) it represents the factor by which it would be necessary to increase the number of Monte Carlo steps in the standard approach to get the same accuracy. Note that for $L = 5$ our Monte Carlo value coincides with the exact one (computed by exact numeration of the 2^N configurations) with an accuracy of less than 10^{-6} . Note also that our MC values converge as the size is increased to the exact infinite-lattice value as given by the Onsager solution [5]. At $T = 8$ (high-temperature regime) our representation is not as good, but still very satisfactory. As a function of the size, the gain in computational effort converges and a value of about 20 is gotten. At the infinite-lattice critical value the results are less spectacular but still of interest. A converged value of about 3 for the gain in efficiency is obtained. At this temperature the correlation length for the spin variables diverges and more accurate representations for the solution of Eq. (14) are needed. Starting from our basic equation built from a transition probability corresponding to *local* moves we need to resort to approximate solutions which contain in some way the collective spin excitations. Alternatively, we can change our fundamental equation by resorting to a nonlocal transition probability density, and then to a new operator H . A natural choice is, of course, the transition probability of the Swendsen-Wang

algorithm used here to generate configurations. Preliminary calculations show that statistical fluctuations are indeed strongly decreased. However, to sum up analytically all contributions corresponding to the different Swendsen-Wang clusters (action of H on ψ) is very time consuming, and the advantages of the method can be lost. Some approximate scheme is clearly called for; this is left for future development. Finally, a last important point is that the gain in computational effort is found to be systematically greater (by about 50%) than the corresponding ratio of variances. This result is a direct consequence of the fact that the integrated autocorrelation time known to control the amount of correlation between successive measurements (see, e.g., [3]) has been decreased when passing from the bare observable to the renormalized one. Note that a similar behavior has also been obtained in applications based on improved estimators [2,3]. Without entering into the details, it can be shown that this result is directly related to the fact that the fluctuations of the renormalized observable are much smaller than in the bare case.

The second application illustrates the method in the case of a continuous configuration space (calculation of multi-dimensional integrals). We have calculated a mean energy as it appears in the so-called variational Monte Carlo

TABLE II. Energy of the helium atom. All quantities are given in atomic units. Statistical uncertainties on the last digit are indicated in parentheses.

Variational Monte Carlo	
$\sigma(E_L)^2$	0.040 9(2)
$\sigma(\tilde{E}_L)^2$	0.006 88
Ratio of variances	~ 5.9
$\langle E_L \rangle$	-2.896 71(4.8)
$\langle \tilde{E}_L \rangle$	-2.896 74(1.6)
Gain in computational effort ^a	~ 9
Exact Green's function Monte Carlo ^b	
$\sigma(E_L)^2$	0.041 1(9)
$\sigma(\tilde{E}_L)^2$	0.008 55(8)
Ratio of variances	~ 4.9
$\langle E_L \rangle$	-2.903 745(99)
$\langle \tilde{E}_L \rangle$	-2.903 734(33)
Gain in computational effort ^a	~ 9
Exact energy	-2.903 724 377 ... ^c

^aSee text for definition.

^bReference [9].

^cReference [10].

(VMC) methods [6]. Starting from a quantum Hamiltonian H_Q (to be distinguished from our trial operator H) and a known trial-wave function ψ_T , our purpose is to compute the variational energy E_v associated with ψ_T . E_v can be easily rewritten as an average over the probability distribution ψ_T^2 , $E_v = \langle E_L \rangle$ where $E_L \equiv H_Q \psi_T / \psi_T$ is called the local energy. Here, we consider the case of the helium atom described by the Hamiltonian $H_Q = -1/2(\vec{\nabla}_1^2 + \vec{\nabla}_2^2) - 2/r_1 - 2/r_2 + 1/r_{12}$ (atomic units) with usual notations. As trial wave function a standard form has been chosen [7],

$$\psi_T(\vec{r}_1, \vec{r}_2) = \exp\left[\frac{ar_{12}}{1 + br_{12}} - c(r_1 + r_2)\right] 1s(r_1)1s(r_2), \quad (15)$$

where $1s(r)$ is the Hartree-Fock orbital as given by Clementi and Roetti [8], and the variational parameters have been chosen to be $a = 0.5$, $b = 0.522$, and $c = 0.0706$. As already remarked, a natural choice for the trial Hamiltonian H is a Schrödinger operator admitting ψ_T as ground state, Eqs. (10) and (11). Regarding ψ we have chosen a form similar to the trial wave function multiplied by some function of the potential energy. Configurations are generated using a standard Metropolis algorithm with local moves constructed using a Langevin equation [7]. Results are presented in Table II. It is seen that the introduction of the renormalized local energy increases the efficiency of the Monte Carlo calculation by about 1 order of magnitude.

In the last application it is shown that the method can even be used in exact (zero-temperature) quantum Monte Carlo (QMC) calculations. In QMC a combination of diffusion and branching process is used to construct a stationary density proportional to $\psi_T \psi_0$, where ψ_0 is the exact unknown ground-state wave function. By averaging the local energy over this distribution, an estimate of the exact energy E_0 is obtained [6]. Although the analytical form of the stationary density is no longer known, a renormalized function whose average is identical to that of the bare local energy can still be defined, $\tilde{E}_L = E_L + (H - E_0)\psi / \psi_T$, where H admits ψ_T as eigenvector, $H\psi_T = 0$. Calculations have been done using the exact Green's function Monte Carlo of Ceperley and Alder [9]. Results are presented in Table II. They are of a quality similar to that obtained in the variational case. About 1 order of magnitude in computer time has been gained.

To conclude, we have presented a simple and powerful method to greatly increase at little cost the efficiency of Monte Carlo calculations. The examples presented have been chosen to illustrate the great versatility of the method (discrete and continuous configuration spaces, classical or quantum Monte Carlo, local or nonlocal Monte Carlo updates). Although our examples have only been concerned with total energies, let us emphasize that the zero-variance principle is valid for any type of observable including important quantities such as local properties other than energy, differences of energies, spatial correlation functions, etc.

*Email address: ra@lct.jussieu.fr

†Email address: mc@lct.jussieu.fr

- [1] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1097 (1953).
- [2] M. Sweeny, *Phys. Rev. B* **27**, 4445 (1983); G. Parisi, R. Petronzio, and F. Rapuano, *Phys. Lett. B* **128**, 418 (1983); U. Wolff, *Nucl. Phys.* **B334**, 581 (1990).
- [3] B. Ammon, H. G. Evertz, N. Kawashima, M. Troyer, and B. Frischmuth, *Phys. Rev. B* **58**, 4304 (1998).
- [4] R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* **58**, 86 (1987).
- [5] L. Onsager, *Phys. Rev.* **65**, 117 (1944).
- [6] W. von der Linden, *Phys. Rep.* **220**, 53 (1992).
- [7] K. E. Schmidt and J. W. Moskowitz, *J. Chem. Phys.* **93**, 4172 (1990).
- [8] E. Clementi and C. Roetti, *At. Data Nucl. Data Tables* **14**, 177 (1974).
- [9] D. M. Ceperley and B. J. Alder, *J. Chem. Phys.* **64**, 5833 (1984).
- [10] K. Frankowski and C. L. Pekeris, *Phys. Rev.* **146**, 46 (1966).