Quadratic diffusion Monte Carlo algorithms for solving atomic many-body problems

Siu A. Chin*
Center for Theoretical Physics, Department of Physics, Texas A&M University, College Station, Texas 77843
(Received 10 November 1989; revised manuscript received 6 July 1990)

The diffusion Monte Carlo algorithm with and without importance sampling is analyzed in terms of the algorithm's underlying transfer matrix. The crucial role played by the Langevin algorithm in the importance-sampling process is made explicit and emphasized. The failure of existing second-order algorithms to converge quadratically for atomic many-body problems is shown to be caused by nonperturbative convergence errors due to the intrinsic inability of the Langevin algorithm to sample Slater orbitals. This failure can be simply circumvented by enforcing attractive cusp conditions on the trial function. Various new second-order diffusion Monte Carlo algorithms are systematically derived and their quadratic convergence numerically verified in cases of He and H_2.

I. INTRODUCTION

In the past decade, Monte Carlo methods for solving the many-body Schrödinger equation[1-3] have been widely applied to the study of quantum systems as diverse as liquid and solid helium,[4-6] electron gas,[7] small molecules,[8-11] few-nucleon bound states,[12-14] model nuclear systems,[15-17] and Hamiltonian lattice gauge theory.[18-23] With the advent of supercomputers, this trend will undoubtedly continue and fundamental improvements in Monte Carlo algorithms can significantly impact many areas of research simultaneously. Currently, there are two basic ensemble algorithms for solving the many-body Schrödinger equation. The first is Kalos's Green's-function Monte Carlo (GFMC) method,[1,2] which samples the exact ground state by iterating the resolvent operator $(E-z)/(H-z)$. (Here, E and z are just some convenient constants.) Computationally, this algorithm is complicated in that one can only sample the exact resolvent operator by a further iteration on a simpler, approximate resolvent. In effect, "a Monte Carlo sampling within a Monte Carlo sampling" must be performed. The other method is the diffusion Monte Carlo (DMC) algorithm[3,7-11] which projects out the exact ground state by iterating the imaginary-time evolution operator $e^{-\Delta t(H-E)}$. In this case, one can easily sample an approximate evolution operator which is exact in the limit of $\Delta t \to 0$. The disadvantage here is that current DMC algorithms with importance sampling[7-11,13-17,19-22] only converge linearly in Δt and one must repeat the calculation several times using rather small values of Δt to extrapolate to the $\Delta t = 0$ limit. Previous attempts[24-26] of applying second-order DMC algorithms to atomic problems were met with failures; theoretically second-order algorithms were found to converge only linearly in practice. On the other hand, DMC algorithms without step-size error[10,11] requires intermediate iterations to sample the exact density matrix, which is not unlike that of GFMC in spirit and in complexity.

In this paper, I show how the underlying transfer matrix of the DMC algorithm can be analyzed perturbatively to yield second-order algorithms systematically. I made explicit the crucial connection between the importance-sampling process and the Langevin algorithm. The failure of second-order DMC algorithms to converge quadratically is shown to be caused by nonperturbative convergent errors due to the intrinsic inability of the Langevin algorithm to sample Slater orbitals, and has nothing to do with overshooting.[24-26] I further show that, when computing the energy, this convergence failure can be averted by enforcing all attractive cusp conditions on the trial function. That this avoidance procedure works as proposed is numerically demonstrated in the case of the helium atom and the hydrogen molecule.

The paper is organized as follows. The perturbative analysis of the transfer matrix is outlined in Sec. II. In Sec. III, second-order Langevin algorithms are systematically derived. These are then used in Sec. IV to generate corresponding DMC algorithms. In Sec. V, the failure of second-order DMC algorithms to converge quadratically is traced to the convergence failure of their underlying Langevin algorithms. After enforcing cusp conditions, quadratic convergences of the ground-state energies of He and H_2 are demonstrated. In Sec. VI, the convergence behavior of the Langevin algorithm is studied in detail by Monte Carlo methods and by directly iterating the algorithms' integral equation. The fundamental mismatch between the Gaussian character of the algorithm and the exponential nature of the Slater orbital is shown to result in fractional power convergence errors that degrade the algorithm. Conclusions are summarized in Sec. VII.

The present DMC calculations confirm the expectation that larger step sizes can be used with second-order algorithms. Moreover, they suggest the possibility that calculations can be carried out at a reasonably large step size without the need of extrapolation. The efficiency of these new algorithms has been further verified in model nuclei[27,28] and helium droplets [29] calculations. Obviously, these second-order DMC algorithms have important implications for atomic calculations as well.

II. PERTURBATIVE ANALYSIS

In order to analyze the systematic errors due to the finite step size Δt, it is useful to view Monte Carlo

methods for solving the many-body Schrödinger equation as stochastic means of iterating a *transfer matrix* T such that for n sufficiently large, T^n converges as $T^n \to \text{const.}\ \lambda^n|\rho\rangle$, where $|\rho\rangle$ is its largest right eigenstate and λ is its largest eigenvalue. To illustrate this basic point of view, consider

$$T = e^{-\Delta t(H-E)}, \tag{1}$$

where $H = K + V$ is the $3N$-dimensional many-body Hamiltonian with kinetic energy $K = \frac{1}{2}\mathbf{p}^2$ and potential energy $V = V(\mathbf{x})$. The iteration of T will converge to its largest eigenstate $|\Psi_0\rangle$, which is the ground state of H. Its corresponding largest eigenvalue is $\lambda = e^{-\Delta t(E_0-E)}$, from which the ground-state energy E_0 of H can be extracted. If one were able to sample T directly, then the iteration will project out the ground state $|\Psi_0\rangle$ irrespective of the value of Δt. In practice, one can only iterate an approximate transfer matrix T'. For example, the approximate matrix

$$T' = e^{-\Delta t K} e^{-\Delta t(V-E)} \tag{2}$$

can be iterated by evolving an ensemble, or population, of configurations $\{\mathbf{x}^k\}$ according to its matrix element (C is an irrelevant normalization constant)

$$\langle \mathbf{x}'|T'|\mathbf{x}\rangle = C\, e^{-(\mathbf{x}'-\mathbf{x})^2/(2\Delta t)} e^{-\Delta t[V(\mathbf{x})-E]}$$

by replicating (or branching) each configuration $\mathbf{x}^k$ on the average $e^{-\Delta t[V(\mathbf{x}^k)-E]}$ times and random walk each configuration via $x_i'^k = x_i^k + \sqrt{\Delta t}\,\xi_i$, where ξ_i are independent Gaussian random variables with zero mean and unit variance. At each time step, the population will change by a factor $\lambda = e^{-\Delta t(E_0-E)}$ and can be stabilized by choosing $E = E_0$. This way of directly determining E_0 defines the growth or normalization energy.

A key point in my analysis is the recognition that, by virtue of the Campbell-Baker-Hausdorff (CBH) formula

$$e^A e^B = \exp\{A + B + \tfrac{1}{2}[A,B] + \tfrac{1}{12}[(A-B),[A,B]] + \cdots\},$$

the approximate transfer matrix T' can be reconstituted in the form $T' = e^{-\Delta t(H'-E)}$ with approximate Hamiltonian

$$H' = H - \tfrac{1}{2}\Delta t[H,V] + \tfrac{1}{12}\Delta t^2[(H-2V),[H,V]] + \cdots. \tag{3}$$

Thus the convergence of T' is governed by H', which in term, can be analyzed by simple perturbation theory. In the present case, the iteration of (2) will converge to the ground state of H' given by

$$|\Psi_0'\rangle = |\Psi_0\rangle + \tfrac{1}{2}\Delta t V|\Psi_0\rangle + \cdots, \tag{4}$$

which deviates systematically from the exact ground state by a term linear in Δt. Note, however, that since the first-order term in (3) vanishes between eigenstates, the eigenvalues themselves converge quadratically; in particular,

$$E_0' = E_0 - \tfrac{1}{24}\Delta t^2\langle\Psi_0|[V,[H,V]]|\Psi_0\rangle + \cdots. \tag{5}$$

Hence, whereas the trial energy

$$E_T = \langle H\Phi_0|\Psi_0\rangle / \langle\Phi_0|\Psi_0\rangle$$

converges only linearly via

$$E_T = E_0 + \tfrac{1}{2}\Delta t \frac{\langle\Phi_0|[H,V]|\Psi_0\rangle}{\langle\Phi_0|\Psi_0\rangle}, \tag{6}$$

the normalization energy always converges quadratically from below according to (5). By noting the correction term in the CBH formula, second-order algorithms can be obtained by taking the product of operators at half the step size with one set of operators in reverse order. A well-known example is

$$T' = e^{-\Delta tV/2} e^{-\Delta tK/2} e^{-\Delta tK/2} e^{-\Delta tV/2} = e^{-\Delta tV/2} e^{-\Delta tK} e^{-\Delta tV/2}, \tag{7}$$

with approximate Hamiltonian

$$H' = H + \tfrac{1}{24}\Delta t^2[(2H-V),[H,V]] + \cdots. \tag{8}$$

In this case, the normalization energy converges identically as (5) and the trial energy now converges quadratically as

$$E_T = E_0 + \tfrac{1}{24}\Delta t^2\left[\langle\Psi_0|\Delta H|\Psi_0\rangle - \frac{\langle\Phi_0|\Delta H|\Psi_0\rangle}{\langle\Phi_0|\Psi_0\rangle}\right], \tag{9}$$

where $\Delta H \equiv [(2H-V),[H,V]]$. I have verified the validity of (5), (6), and (9) both analytically and numerically in the case of the one-dimensional harmonic oscillator.

A detailed calculation of the ground states of (3) and (8), corresponding to a special case of the Langevin algorithm discussed in Sec. VI, is given in the Appendix.

III. IMPORTANCE SAMPLING AND THE LANGEVIN ALGORITHM

In most physical applications, the rapidly varying potential V precludes the use of naive algorithms (2) and (7). As first shown by Kalos and collaborators,[4] importance sampling with a trial function is essential for practical Monte Carlo calculations. As noted elsewhere,[22,23] importance sampling can be understood in terms of the transfer matrix as follows: Instead of the original T, one considers an alternative transfer matrix $\tilde{T}$ "similarly transformed" by a trial function Φ_0,

$$\tilde{T} = \Phi_0 e^{-\Delta t(H-E)}\Phi_0^{-1} = e^{-\Delta t(\tilde{H}-E)}, \tag{10}$$

where

$$\tilde{H} \equiv \Phi_0 H \Phi_0^{-1} = \tfrac{1}{2}\mathbf{p}^2 + i\mathbf{p}\cdot\mathbf{G} + E_L(\mathbf{x}) \tag{11}$$

is the transformed Hamiltonian, $E_L(\mathbf{x}) \equiv \Phi_0^{-1}(\mathbf{x})H\Phi_0(\mathbf{x})$ is the local energy, $G_k(\mathbf{x}) = -\nabla_k S(\mathbf{x})$, and $\Phi_0(\mathbf{x}) = e^{-S(\mathbf{x})}$. Although $\tilde{H}$ is no longer Hermitian, it is simple to see that its left and right eigenstates are, respectively, $\langle\Psi_n|\Phi_0^{-1}$ and $\Phi_0|\Psi_n\rangle$ with unchanged real eigenvalues E_n. Thus with importance sampling, the iteration will converge to the product state $\Phi_0|\Psi_0\rangle$.

Iterating $\tilde{T}$ is advantageous in that the bare potential $V(\mathbf{x})$ is replaced by a generally smoother local energy $E_L(\mathbf{x})$. More importantly, the simple Gaussian random walk associated with the diffusion evolution operator

$e^{-\Delta tK}$ is replaced by $e^{-\Delta tL}\equiv e^{-\Delta t(\frac{1}{2}\mathbf{p}^2+i\mathbf{p}\cdot\mathbf{G})}$, which one recognizes as the evolution operator for the Fokker-Planck equation[30]

$$-\frac{\partial}{\partial t}\Psi(\mathbf{x},t)=-\tfrac{1}{2}\nabla^2\Psi(\mathbf{x},t)+\nabla_i[G_i(\mathbf{x})\Psi(\mathbf{x},t)]\ , \tag{12}$$

with stationary distribution $\Phi_0^2(\mathbf{x})$. The associated random walks are now described by the Langevin algorithm. Since the diffusion operator $e^{-\Delta tK}$ by itself will simply converge to a uniform distribution, heuristically, one may view naive algorithms (2) and (7) as trying to generate Ψ_0 primarily through branching. In contrast, with importance sampling, the Fokker-Planck evolution operator $e^{-\Delta tL}$ by itself will converge to Φ_0^2 and branching is only required to supply the remaining factor Ψ_0/Φ_0. As Φ_0 varies from 1 to Ψ_0, the primary burden for generating Ψ_0 shifts from branching to random walks. In the limit that the trial function becomes the exact ground state, Ψ_0^2 is entirely generated by Langevin random walks and branching only produces an overall constant. *In this case, it is clear that the order of convergence of the importance sampled DMC algorithm is dictated by that of the associated Langevin algorithm.* In any case, unless this embedded Langevin algorithm is correctly simulated to second order, no overall second-order DMC algorithm is possible.

As pointed out elsewhere,[31] Langevin algorithms for evolving the Fokker-Planck equation can also be derived by regarding $T_L=e^{-\Delta tL}$ as a transfer matrix. A first-order approximation is obviously

$$T_L'=e^{-\Delta tK}e^{-\Delta tD}\ , \tag{13}$$

where $D\equiv i\mathbf{p}\cdot\mathbf{G}$. The matrix element of $e^{-\Delta tD}$ is the Green's function of the evolution equation

$$-\frac{\partial}{\partial t}\Psi(\mathbf{x},t)=\nabla_i[G_i(\mathbf{x})\Psi(\mathbf{x},t)]\ . \tag{14}$$

If one regards Ψ as a particle density function, then one can naturally interpret this as a *continuity equation* with current density $\mathbf{J}=\mathbf{G}\Psi$ and velocity field $\mathbf{v}(\mathbf{x})=\mathbf{G}(\mathbf{x})$. The trajectory of the particles are therefore determined by

$$\frac{d\mathbf{x}(t)}{dt}=\mathbf{G}(\mathbf{x})\ , \tag{15}$$

and the Green's function given by

$$\langle\mathbf{x}'|e^{-\Delta tD}|\mathbf{x}\rangle=\delta(\mathbf{x}'-\mathbf{x}(\Delta t))\ , \tag{16}$$

where $\mathbf{x}(t)$ is the solution to (15) with $\mathbf{x}(0)=\mathbf{x}$. Thus the matrix element

$$\langle\mathbf{x}'|T_L'|\mathbf{x}\rangle=C\exp\left[-\frac{1}{2\Delta t}[\mathbf{x}'-\mathbf{x}(\Delta t)]^2\right]$$

can be sampled by setting

$$x_i'=x_i(\Delta t)+\sqrt{\Delta t}\,\xi_i\ . \tag{17}$$

The decomposition (13), and hence (17), has the virtue of separating out the *stochastic* part, which is the Gaussian random walk, from the *deterministic* part (15), which represents the bias of the trial function. Since (13) is only correct to first order, it is sufficient to solve (15) to the same order:

$$x_i'=x_i+\Delta tG_i(\mathbf{x})+\sqrt{\Delta t}\,\xi_i\ . \tag{18}$$

A second-order Langevin algorithm is produced by the decomposition

$$T_L'=e^{-\Delta tD/2}e^{-\Delta tK}e^{-\Delta tD/2} \tag{19}$$

with matrix element

$$\begin{aligned}\langle\mathbf{x}'|T_L'|\mathbf{x}\rangle&=\int d\mathbf{y}\langle\mathbf{x}'|e^{-\Delta tD/2}|\mathbf{y}\rangle\langle\mathbf{y}|e^{-\Delta tK}e^{-\Delta tD/2}|\mathbf{x}\rangle\\&=\int d\mathbf{y}\delta(\mathbf{x}'-\mathbf{y}(\tfrac{1}{2}\Delta t))C\exp\left[-\frac{1}{2\Delta t}[\mathbf{y}-\mathbf{x}(\tfrac{1}{2}\Delta t)]^2\right]\ .\end{aligned} \tag{20}$$

The above structure naturally suggests a two-stage sampling procedure:

$$\begin{aligned}y_i&=x_i(\tfrac{1}{2}\Delta t)+\sqrt{\Delta t}\,\xi_i\ ,\\x_i'&=y_i(\tfrac{1}{2}\Delta t)\ .\end{aligned} \tag{21}$$

Thus with (15) solved correctly to second order, one obtains a two-stage random walk:

$$\begin{aligned}y_i&=x_i+\tfrac{1}{2}\Delta tG_i(\mathbf{x})+\tfrac{1}{2}(\tfrac{1}{2}\Delta t)^2G_j(\mathbf{x})\nabla_jG_i(\mathbf{x})+\sqrt{\Delta t}\,\xi_i\ ,\\x_i'&=y_i+\tfrac{1}{2}\Delta tG_i(\mathbf{y})+\tfrac{1}{2}(\tfrac{1}{2}\Delta t)^2G_j(\mathbf{y})\nabla_jG_i(\mathbf{y})\ .\end{aligned} \tag{22}$$

To avoid evaluating derivatives, this can be interpreted as

$$\begin{aligned}y_i&=x_i+\tfrac{1}{2}\Delta tG_i(\mathbf{x}+\tfrac{1}{4}\Delta t\mathbf{G}(\mathbf{x}))+\sqrt{\Delta t}\,\xi_i\ ,\\x_i'&=y_i+\tfrac{1}{2}\Delta tG_i(\mathbf{y}+\tfrac{1}{4}\Delta t\mathbf{G}(\mathbf{y}))\ .\end{aligned} \tag{23}$$

This algorithm requires four evaluations of $\mathbf{G}(\mathbf{x})$ per updating step, which is computationally very expensive.

A second algorithm can be devised according to the alternate decomposition

$$T_L'=e^{-\Delta tK/2}e^{-\Delta tD}e^{-\Delta tK/2}\ . \tag{24}$$

Similar manipulations as before yield

$$y_i = x_i + \sqrt{\Delta t/2}\,\xi_i \ ,$$
$$x_i' = y_i + \Delta t G_i[\mathbf{y} + \tfrac{1}{2}\Delta t \mathbf{G}(\mathbf{y})] + \sqrt{\Delta t/2}\,\xi_i' \ , \tag{25}$$

which requires two independent Gaussian vectors per updating step but only two evaluations of $\mathbf{G}(\mathbf{x})$. These two Langevin algorithms, (23) and (25), are canonical in the sense that they correspond to obvious ways of decomposing the original transfer matrix to second order.

A noncanonical second-order algorithm can be obtained by directly substituting in the intermediate step y_i in (23) and retaining terms up to second order in Δt:

$$x_i' = x_i + \Delta t G_i(\mathbf{x}) + \sqrt{\Delta t}\,\xi_i + \tfrac{1}{2}\Delta t\sqrt{\Delta t}\,\xi_j \nabla_j G_i(\mathbf{x}) + \tfrac{1}{2}\Delta t^2 G_j(\mathbf{x})\nabla_j G_i(\mathbf{x}) + \tfrac{1}{4}\Delta t^2 \xi_j \xi_k \nabla_j \nabla_k G_i(\mathbf{x}) \ . \tag{26}$$

This in turn can be reinterpreted as

$$y_i = x_i + \Delta t G_i(\mathbf{x}) + \sqrt{\Delta t}\,\xi_i \ ,$$
$$x_i' = y_i + \tfrac{1}{2}\Delta t[G_i(\mathbf{y}) - G_i(\mathbf{x})] \ , \tag{27}$$

which one recognizes as a second-order Runge-Kutta Langevin algorithm.[32–34] This algorithm was also independently derived in Ref. 24, but in a rather opaque manner. Since this algorithm only requires one Gaussian vector and two evaluations of $\mathbf{G}(\mathbf{x})$ per update, it is clearly faster than (23) and (25). However, as noted in Ref. 31, and will be further demonstrated below, the range of quadratic convergence of this algorithm is usually much shorter than the two canonical algorithms. Following Ref. 31, we will refer to algorithms (18), (23), (25), (27) as LGV1, LGV2a, LGV2b, and LGV2c, respectively, where LGV is short for Langevin, and 1, 2a, 2b, and 2c mean first-order and three second-order algorithms, respectively.

One other second-order Langevin algorithm can be derived by evaluating $e^{-\Delta t L}$ directly to second order via normal ordering,[35] however, the resulting algorithm, which requires the evaluation of higher derivatives of $G_i(\mathbf{x})$ and the sampling of multivariate Gaussian distributions, does not appear to be competitive with algorithms outlined above.

It should be noted that these algorithms are derived on the basis of perturbative analysis, i.e., on the assumption that the convergence error can be expanded in integral powers of Δt. Unfortunately, this is not the complete picture. As will be demonstrated in Sec. VI, depending on the specific form of $G_i(\mathbf{x})$, there can exist nonperturbative, nonintegral power convergence errors due to the intrinsic limitation of the Langevin algorithm. These nonintegral power errors can be of lower order than that of the algorithm and can therefore degrade the algorithm to lower order.

IV. SECOND-ORDER DMC ALGORITHMS

Algorithms for iterating $\tilde{T}$ can now be devised and analyzed identically as in Sec. II. Analysis of the step-size error is more tedious. However, for the purpose of constructing second-order algorithms, it is unnecessary to know these errors explicitly. To obtain a first-order DMC algorithm one can approximate $\tilde{T}$ via

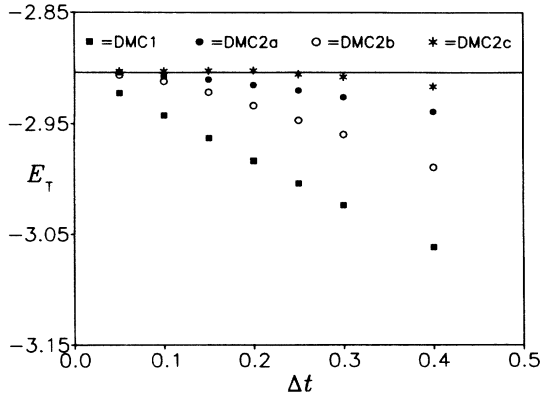

FIG. 1. The trial ground-state energy of helium as a function of Δt as computed by four DMC algorithms. The statistical errors are at most half the size of the plotting symbols. The horizontal line indicates the exact ground-state energy. The trial function used is (31) with $\zeta = 2$, $a = 0.5$, $b = 0.2$.

$$\tilde{T}' = e^{-\Delta t L} e^{-\Delta t (E_L - E)} \tag{28}$$

and sample $e^{-\Delta t L}$ to first order according to algorithm LGV1 (18). This algorithm will be designated correspondingly as DMC1. For a first-order algorithm, the computation of position-dependent Gaussian random walks[9,19–22] is unnecessary. If $e^{-\Delta t L}$ in (28) were sampled to second order using either LGV2a, LGV2b, or LGV2c, then, as discussed in Sec. II, the normalization energy would converge quadratically, despite that fact that the ground state of (28) only converges linearly. Genuine second-order algorithms can be obtained via

$$\tilde{T}' = e^{-\Delta t (E_L - E)/2} e^{-\Delta t L} e^{-\Delta t (E_L - E)/2} \ , \tag{29}$$

with $e^{-\Delta t L}$ sampled by second-order Langevin algorithms LGV2a, LGV2b, or LGV2c. These three DMC algorithms will be designated as DMC2a, DMC2b, and DMC2c, respectively. If, however, $e^{-\Delta t L}$ is sampled only to first order, then (29) can only converge linearly, irrespective of whether branching is symmetric, as in (29), or asymmetric, as in (28).

Again, in discussing the order of these algorithms, one should keep in mind the last paragraph of Sec. III.

V. APPLICATIONS TO He AND H_2

Because of continued interest in *ab initio* atomic calculations,[8–11,36] I first test these second-order algorithms on the nonrelativistic helium atom[37] with Hamiltonian

$$H = \tfrac{1}{2}\mathbf{p}_1^2 + \tfrac{1}{2}\mathbf{p}_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}} \ . \tag{30}$$

The trial function used is of the form

$$\Phi_0 = \exp(-\zeta r_1)\exp(-\zeta r_2)\exp[a r_{12}/(1 + b r_{12})] \ . \tag{31}$$

Figure 1 shows the trial energy obtained using various DMC algorithms with trial function $\zeta = 2$, $a = 0.5$, and $b = 0.2$. This trial function satisfies all the cusp conditions and yields a variational energy of $E_v = -2.878(2)$. For each algorithm, the time-step iteration is preceded by a variational calculation with configurations sampled

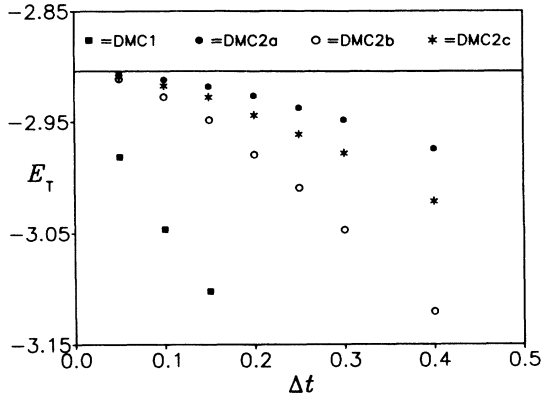

FIG. 2. Same as Fig. 1 but with $\zeta=2$, $a=0.0$, $b=0.0$.

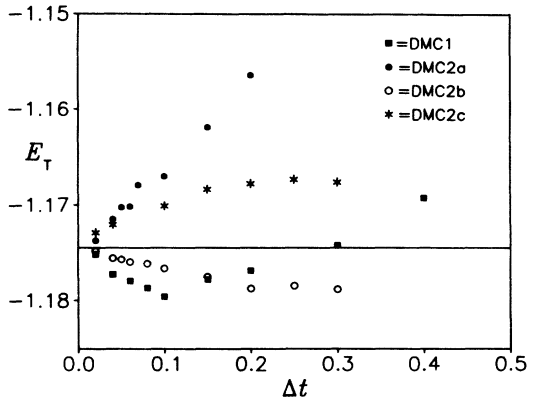

FIG. 3. The trial ground-state energy of H_2 as a function of Δt as computed by four DMC algorithms. The statistical errors are at most the size of the plotting symbols. The horizontal line indicates the exact ground-state energy. The trial function used is $\Phi_0^{\rm I}$ of (35).

from Φ_0^2 using the Metropolis method. The normalization energy is adjusted after each time step to maintain a population of about 200 configurations for $\Delta t=0.05$, 0.10, and 100 configurations for $\Delta t=0.15$, 0.20, 0.25, 0.30, and 0.40. At each value of Δt, after discarding the initial 1000–4000 generations, expectation values were computed using 100 block averages of 100 consecutive generations. The ratio of time used is roughly 1:2:1.7:1.5 for DMC1:DMC2a:DMC2b:DMC2c, with DMC1 consuming about 1.4 VAX2000 work-station hours per value of Δt with a population of 100 configurations. The linear convergence of DMC1 and the quadratic convergence of DMC2a and DMC2b are obvious. For DMC2c, its apparent rapid convergence at step size as large as $\Delta t=0.3$ is remarkable but deceiving. It actually overshoots the exact energy and seems to converge *from above* at $\Delta t \lesssim 0.2$.

The performance of these second-order algorithms as compared to the first-order one is even more striking when the trial function is poor. Figure 2 shows the identical calculation with trial function $\zeta=2$ and $a=b=0.0$. This simple product of two hydrogenic orbitals has $E_v=-2.75$ and does not satisfy the repulsive electron-electron cusp condition. It is clear from these two figures that the convergent behaviors of second-order algorithms are quadratic, and they are distinct from those of the first order.

The numerical values for the trial and normalization energy obtained by all four algorithms at the smallest time step size $\Delta t=0.05$ are compared in Table I. For both trial functions, all second-order results have essentially converged to the exact ground-state energy to within statistical error. This is to be contrasted with first-order results, where the systematic step-size error remains sizable. Moreover, as a corollary, second-order results at small Δt are less sensitive to the quality of the trial function.

Next, I apply these algorithms to compute the ground-state energy of the hydrogen molecule with Hamiltonian

$$H=\tfrac{1}{2}\mathbf{p}_1^2+\tfrac{1}{2}\mathbf{p}_2^2-\frac{1}{r_{1A}}-\frac{1}{r_{1B}}-\frac{1}{r_{2A}}-\frac{1}{r_{2B}}+\frac{1}{r_{AB}}+\frac{1}{r_{12}}\ , \tag{32}$$

where r_{1A}, r_{1B}, etc., are electron-nucleus separations and $r_{AB}=1.401$ is the known equilibrium internuclei distance. The trial wave function[8] is of the form

$$\Phi_0=\phi(r_{1A},r_{1B})\phi(r_{2A},r_{2B})\exp[ar_{12}/(1+br_{12})]\ , \tag{33}$$

with molecular orbital

$$\phi(r_A,r_B)=\exp(-\zeta r_A)+\exp(-\zeta r_B)\ . \tag{34}$$

The particular trial wave function[8] used is

$$\Phi_0^{\rm I}=\Phi_0(\zeta=1.285;a=0.28,b=0.05)\ , \tag{35}$$

with variational energy $E_v=-1.149\,80(35)$. The results

TABLE I. The trial (upper) and normalization (lower) energy for helium as computed by four DMC algorithms at $\Delta t=0.05$ using two different trial functions of the form (31). The exact energy (Ref. 37) is $E_0=-2.903\,724\ldots$.

Trial function	DMC1	DMC2a	MC2b	DMC2c
$\zeta=2$, $a=0.5$, $b=0.2$	−2.922 4(7)	−2.903 5(8)	−2.906 0(9)	−2.902 8(8)
	−2.922 3(17)	−2.903 1(20)	−2.902 8(19)	−2.903 6(18)
$\zeta=2$, $a=0.0$, $b=0.0$	−2.981 7(16)	−2.906 6(17)	−2.910 9(18)	−2.908 7(19)
	−2.967 8(25)	−2.904 6(29)	−2.909 7(29)	−2.902 8(27)

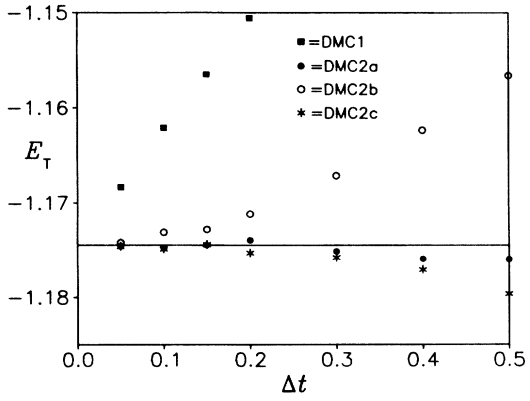

FIG. 4. Same as Fig. 3 but with trial function Φ_0^{II} of (36).

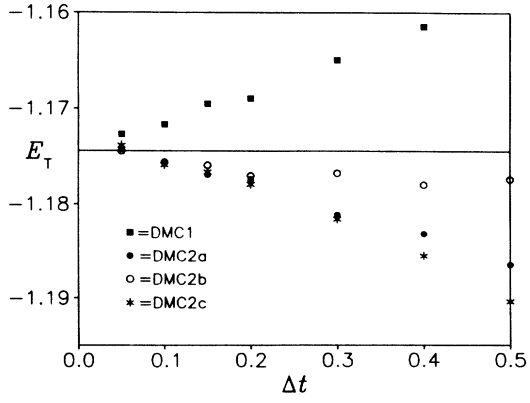

FIG. 5. Same as Fig. 3 but with trial function Φ_0^{III} of (37).

are shown in Fig. 3. (In this and subsequent calculations, I used symmetric branching even for DMC1.) The ensemble population was maintained at a target value of 200. Expectation values were computed using 100–300 block averages of 100 consecutive generations. Dependent on Δt, the total number of configurations sampled ranges from 2×10^6 to 6×10^6. The trial energy of all four algorithms appears to have converged correctly to the exact values[38] of $E_0=-1.174\,47\ldots$, but apparently with only a *linear* slope. Moreover, the range of the linear convergence is very small, with $\Delta t\lesssim 0.1$. (For DMC1, this range is in agreement with that of Ref. 8.) The failure of algorithm DMC2c to converge quadratically in the case of H_2 was previously observed by Vrbik and Rothstein,[25] using the identical trial function Φ_0^I. In a subsequent study by Rothstein *et al.*,[26] other algorithms similar (but not identical) to DMC2a and DMC2b were found to suffer the same fate.

This failure of second-order algorithms to converge quadratically in the case of H_2 is rather surprising in view of their success with He. In Ref. 25 Vrbik and Rothstein suggested that the problem may be due to the discontinuity of $\mathbf{G}(\mathbf{x})$ near each nucleus, which causes electrons to incorrectly overshoot past each nucleus. The fact that this cannot be the explanation is easily demonstrated by the counterexample of He; overshooting in that case clearly did not prevent the energy from converging quadratically. In Ref. 26, Rothstein *et al.* suggested that, somehow, the combination of branching and overshooting is to be blamed, but no details were given. In the next section, I will show that this convergence failure of second-order DMC algorithms is due to an intrinsic inability of their embedded Langevin algorithm to sample Slater orbitals correctly, and has nothing to do with overshooting. Fortunately, as will also be demonstrated, when computing the energy of the system, this problem can be simply circumvented by enforcing attractive cusp conditions on the trial function.

The fact this avoidance procedure indeed works as claimed is demonstrated in Fig. 4, where the same H_2 problem is solved with the trial function

$$\Phi_0^{II}=\Phi_0(\zeta=1.189\,032\,767;a=0.50,b=0.40)\ , \tag{36}$$

where the value for ζ is determined by the cusp equation $\zeta=1+e^{-\zeta r_{AB}}$. The corresponding variational energy is $E_v=-1.148\,36(25)$. The improvement is quite evident. The convergence behaviors of the first- and second-order algorithms are now distinct. Even DMC1 is improved in that its range of linear convergence now extends out to $\Delta t\approx 0.5$, five times its previous range with trial function Φ_0^I. The quadratic convergence of algorithm DMC2b is particularly convincing. The convergence of the other two is actually too good to tell, but as we shall see in the next section, there is no reason to doubt.

To test the robustness of these algorithms, a further calculation is performed with the trial function

$$\Phi_0^{III}=\Phi_0(\zeta=1.189\,032\,767;a=0.0,b=0.0)\ , \tag{37}$$

which again ignores the electron-electron cusp condition and their correlation. The variational energy is $E_v=-1.128\,68(52)$. The results are shown in Fig. 5. The quadratic convergences of second-order algorithms

TABLE II. The trial (upper) and normalization (lower) energy for H_2 as computed by four DMC algorithms at $\Delta t=0.05$ using two different trial functions (36) and (37). The exact energy (Ref. 38) is $E_0=-1.174\,474\ldots$.

Trial function	DMC1	DMC2a	DMC2b	DMC2c
Φ_0^{II}	−1.168 3(4)	−1.174 6(5)	−1.174 2(5)	−1.174 6(4)
	−1.166 8(9)	−1.172 0(11)	−1.173 6(11)	−1.175 0(11)
Φ_0^{III}	−1.172 7(6)	−1.174 3(6)	−1.174 5(6)	−1.173 9(6)
	−1.171 5(12)	−1.173 3(13)	−1.174 6(12)	−1.174 5(13)

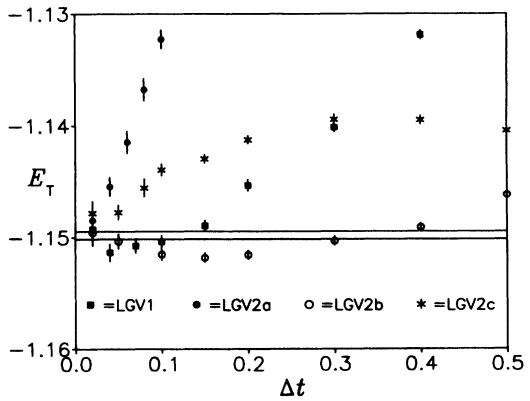

FIG. 6. The H_2 trial energy due to $\Phi_0^{\rm I}$ as obtained by four Langevin algorithms. At each value of Δt, expectation values are computed using 200–400 block averages of 5000 configurations. Horizontal lines bracket the variational energy computed via the Metropolis algorithm.

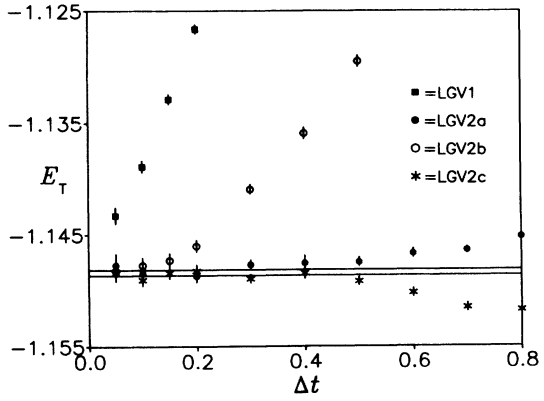

FIG. 7. Same as Fig. 6, but for trial function $\Phi_0^{\rm II}$ of (36).

remains evident. What is surprising is the fact that the linear convergence of DMC1 is improved by this supposedly poorer trial function. The trial and normalization energy at $\Delta t=0.05$ for $\Phi_0^{\rm II}$ and $\Phi_0^{\rm III}$ are compared in Table II. Again, all second-order results have essentially converged to the exact ground-state energy to within statistical errors and are not particularly sensitive to the trial function used.

VI. THE CONVERGENCE FAILURE AND ITS AVOIDANCE

Since the branching part of the DMC algorithm is straightforward, there is little doubt that it is correctly simulated. One must therefore scrutinize the Langevin part of the DMC algorithm for clues to its failure. Figure 6 shows the resulting trial energy for $\Phi_0^{\rm I}$ when branching is suppressed. In this case, one is simply running the Langevin algorithm to sample $(\Phi_0^{\rm I})^2$ and therefore the trial energy should reproduce the variational energy. This is indeed observed; however, as shown in Fig. 6, all four algorithms apparently only approach the variational energy *linearly*, and only when $\Delta t \lesssim 0.1$. By contrast, when the same calculation is carried out for $\Phi_0^{\rm II}$, as shown in Fig. 7, quadratic convergences of second-order Langevin algorithms are clearly evident. In particular, the convergence range of LGV2a and LGV2c is so wide that one is forced to extend the range of Δt in order to see their quadratic deviations. Thus failures of second-order DMC algorithms to converge quadratically can all be traced to failures of their embedded Langevin algorithms.

The failures of second-order Langevin algorithms to converge quadratically has nothing to do with overshooting. A simple argument by referees of this paper is sufficient to illustrate the point. Consider the case of sampling generic Slater atomic orbitals of the form $\Phi_0^2(r)=e^{-\zeta r}$, with velocity field $\mathbf{G}(\mathbf{r})=-\frac{1}{2}\zeta\hat{\mathbf{r}}$, which is discontinuous at $\mathbf{r}=\mathbf{0}$. As noted previously by Vrbik and Rothstein,[25] electrons will overshoot past the nucleus. In the limit of small Δt, this overshooting is dominated by the linear term. Hence, there is always a spherical region centered around the origin with radius $R\approx\zeta\Delta t$ in which the sampling is defective. However, in computing the energy, which is at most $1/r$ or less singular, this defective hole in the wave function can at most contribute an error,

$$\Delta E\approx\Phi_0^2(0)\int_0^R d^3r\frac{1}{r}\propto\Delta t^2\ , \tag{38}$$

which is second order in Δt and therefore cannot explain why the algorithm is apparently degraded to first order.

The real reason for the convergence failure is more subtle and is due to the intrinsic inability of the Langevin algorithm to sample exponential functions. This intrinsic failure give rises to fractional power convergence errors not accountable by the perturbative analysis of Sec. II. In computing specific expectation values, these errors can be of lower order than that of the algorithm and therefore can degrade the algorithm to lower order.

To investigate the detail convergence behavior of the Langevin algorithm in sampling Slater orbitals of the form $\rho_0(r)=\Phi_0^2(r)=N_0e^{-\zeta r}$, $(N_0=\zeta^3/8\pi)$, I directly iterate its corresponding integral equation. The drift step corresponding to applying the operator $e^{-\Delta tD}$ is

$$\rho(r,t+\Delta t)=(1+\Delta r/r)^2\rho(r+\Delta r,t)+\delta^3(\mathbf{r})\int_0^{\Delta r}d^3r\rho(r,t)\ , \tag{39}$$

where $\Delta r=|\mathbf{G}|\Delta t=\frac{1}{2}\zeta\Delta t$ is the distance moved in time Δt. Regarding $\rho(r,t)$ as an evolving particle density distribution, the first term on the right-hand side (rhs) simply corresponds to

$$4\pi r^2dr\rho(r,t+\Delta t)=4\pi(r+\Delta r)^2dr\rho(r+\Delta r,t)\ ,$$

reflecting the fact that particles now at r were originally from $r+\Delta r$. The second term expresses the fact that all particles that reach the origin in time Δt stay at the origin. This is thus exact with no overshooting. The random-walk step corresponding to applying $e^{-\Delta tK}$ is

$$\rho(r',t+\Delta t)=\frac{1}{Z_3}\int d^3r\rho(r,t)\exp\left[-\frac{1}{2\Delta t}(\mathbf{r}'-\mathbf{r})^2\right],$$

$$=\frac{1}{Z_1}\int_0^\infty dr\frac{r}{r'}\rho(r,t)\left[\exp\left[-\frac{1}{2\Delta t}(r'-r)^2\right]-\exp\left[-\frac{1}{2\Delta t}(r'+r)^2\right]\right], \tag{40}$$

where $Z_n=(2\pi\Delta t)^{n/2}$. The integral equation is only one dimensional in the present case of spherical symmetry. Note that after the random walk, the wave function is necessarily an even function of r', i.e., it is basically a Gaussian near the origin.

Substituting (39) into (40) gives the first-order algorithm LGV1,

$$\rho(r',t+\Delta t)=\frac{1}{Z_3}\int d^3r(1+\Delta r/r)^2\rho(r+\Delta r,t)\exp\left[-\frac{1}{2\Delta t}(\mathbf{r}'-\mathbf{r})^2\right]+\left[\int_0^{\Delta r}d^3r\rho(r,t)\right]\frac{1}{Z_3}\exp\left[-\frac{1}{2\Delta t}(\mathbf{r}')^2\right], \tag{41}$$

which can be directly solved by discretizing $\rho(r)$ and iterating. [In actual calculations, $\rho(r)$ is discretized into 400–1400 points depending on Δt.] Once (41) is solved, results of other algorithms can be obtained without much effort. For example, the related first-order algorithm corresponding to $e^{-\Delta tD}e^{-\Delta tK}$ can be obtained by applying $e^{-\Delta tD}$ to the stationary solution of (41), LGV2a can be obtained by applying $e^{-\Delta tD/2}$, and LGV2b can be obtained by applying $e^{-\Delta tK/2}$ in place of the last application of $e^{-\Delta tK}$.

By binning the density of particles in Monte Carlo calculations, one can compare the convergence of $\rho(r)$ from Monte Carlo sampling with that of directly iterating (41). For LGV1, this shown in Fig. 8. The agreement between the two is excellent. As discussed in Sec. II, the convergence behavior of an algorithm can also be studied via its underlying approximate Hamiltonian, which in the present case is

$$H'=H-\tfrac{1}{2}\Delta t[K,D]+\tfrac{1}{12}\Delta t^2[(K-D),[K,D]]+\cdots. \tag{42}$$

Plotted as a dashed line is the first-order perturbative ground state of H',

$$\rho_0^{(1)}(r)=N_1\exp\left[-\zeta r-\frac{\zeta\Delta t}{2r}\right]. \tag{43}$$

(For details, see the Appendix.) This first-order wave function is good at large r but fails catastrophically near the origin. As we shall see, this is simply due to the fact that, near the origin, the wave function in nonperturbative in powers of Δt. However, in computing expectation value of $\langle r^n\rangle$, for $n>-1$, the wave function is weighted with r^{n+2} and this failure is not very glaring. An even better wave function at large r is given by the second-order wave function (see the Appendix)

$$\rho_0^{(2)}(r)=N_2\exp\left[-\zeta r-\frac{\zeta\Delta t}{2r}-\frac{1}{2}\left[\frac{\zeta\Delta t}{2r}\right]^2\right]. \tag{44}$$

This is plotted as a dotted line.

Figure 9 shows the same comparison for LGV2b. Again, the agreement between Monte Carlo simulations and direct iteration is excellent. There is no perturbative wave-function comparison because there is no second-order correction to the exact wave function for Slater orbitals (see the Appendix). The convergence of LGV2b is obviously better than that of LGV1. Note, however, that

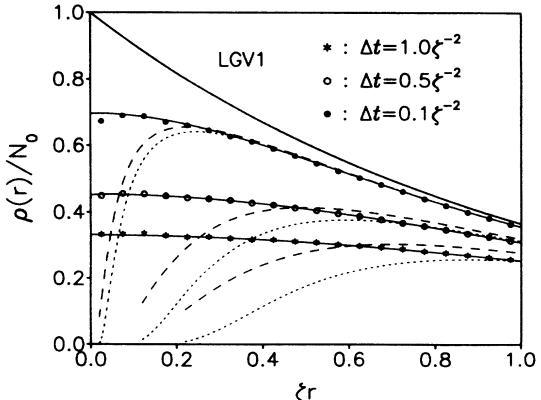

FIG. 8. The square of the wave function $\rho(r)=\Phi^2(r)$ as sampled by the first-order Langevin algorithm LGV1 at finite Δt. The topmost curve is the exact Slater orbital $e^{-\zeta r}$. Plotting symbols are binned Monte Carlo results. Error bars are mostly smaller than the size of the plotting symbols and are not shown. Solid lines are results obtained by directly iterating the integral equation corresponding to the algorithm. For comparison, the first- and the second-order perturbative wave functions (43) and (44) are drawn as dotted and dashed lines, respectively.

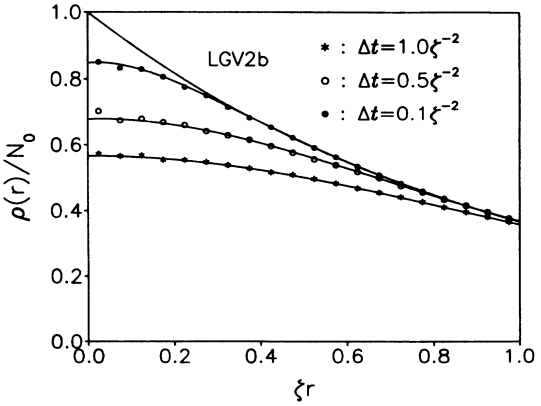

FIG. 9. Same as Fig. 8 for the second-order algorithm LGV2b. In this case, however, there is no second-order perturbative correction to the exact wave function (see the Appendix).

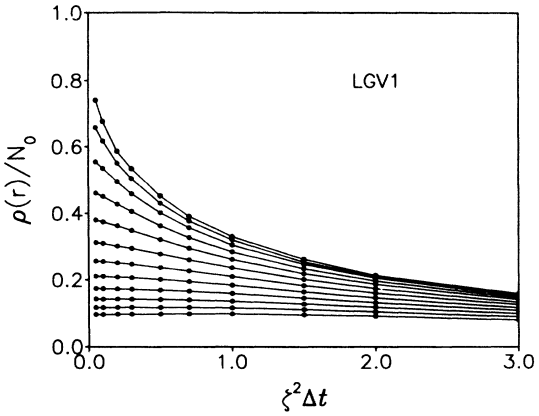

FIG. 10. The convergence of $\rho(r)/N_0$ at fixed r for LGV1. Data are from Monte Carlo calculations and are connected by straight lines to guide the eye. From top to bottom are equal distance values of the wave function from $\zeta r = 0.15$ to 2.35.

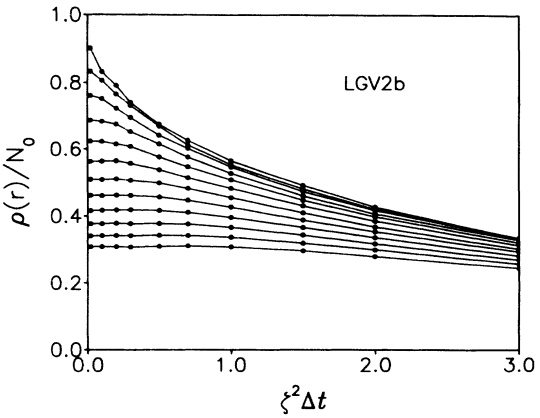

FIG. 11. Same as Fig. 10 for algorithm LGV2b.

in both cases, the fundamental mismatch between the Gaussian character of the evolving wave function and the exponential Slater orbital results in a persistent, triangular region of defect. This is shown even more clearly in Figs. 13 and 14.

In Figs. 10 and 11, we examine in detail how the wave function from Monte Carlo simulation converges as a function of r. Away from the origin, the convergence is correctly linear and quadratic. Near the origin, the convergence is clearly neither linear nor quadratic. Since statistics for binning the central density is poor in Monte Carlo simulations, we plot in Fig. 12, as a function of $\zeta\Delta t^{1/2}$, the convergence of $\rho(0)$ for both algorithms from direct iteration. The apparent $\zeta\Delta t^{1/2}$ convergence is supported by the following analytic estimate. From (41), the LGV1 central density is given by

$$\rho(0,t+\Delta t)=\frac{1}{Z_3}\int d^3r(1+\Delta r/r)^2\rho(r+\Delta r,t)\exp\left[-\frac{1}{2\Delta t}r^2\right]+\frac{1}{Z_3}\left[\int_0^{\Delta r}d^3r\rho(r,t)\right]. \tag{45}$$

As $\Delta t\to 0$, one can estimate the rhs by substituting in the exact wave function $\rho_0(r)$. The second term on the right is at least $\Delta t^{3/2}$ and can be neglected. Denoting the Gaussian expectation values in (45) as $\langle\ \rangle_{\Delta t}$, this gives

$$\begin{aligned}\rho(0)&=N_0 e^{-\zeta\Delta r}\langle(1+\Delta r/r)^2 e^{-\zeta r}\rangle_{\Delta t}\\&\approx N_0\left[1+2\Delta r\left\langle\frac{1}{r}\right\rangle_{\Delta t}-\zeta\langle r\rangle_{\Delta t}+\cdots\right]\\&\approx N_0\left[1-\left[\frac{2}{\pi}\right]^{1/2}\zeta\Delta t^{1/2}+\cdots\right].\end{aligned} \tag{46}$$

If one were to sample a more general function

$$\rho_0(r)=N_0\exp\left[-\frac{1}{n}(\zeta r)^n\right], \tag{47}$$

a similar estimate would give

$$\rho(0)\approx N_0\left[1-\frac{1}{2n}(2\zeta^2\Delta t)^{n/2}\frac{\Gamma\left[\frac{n+1}{2}\right]}{\Gamma(\frac{3}{2})}+\cdots\right]. \tag{48}$$

Thus the $\Delta t^{1/2}$ convergence of the central density is peculiar to the case of Slater orbitals. For LGV1, the estimated slope is $(2/\pi)^{1/2}=0.797\,89$. The fitted value from Fig. 12 is ≈ 1.05. For LGV2b, the corresponding estimate is $(1/\pi)^{1/2}=0.564\,19$, to be compared with the observed value of ≈ 0.49. Unfortunately, I was not able to derive analytic expressions for these exact coefficients.

As noted earlier, the wave function after each random walk is necessarily a Gaussian near the origin. The characteristic range of this Gaussian can be easily ob-

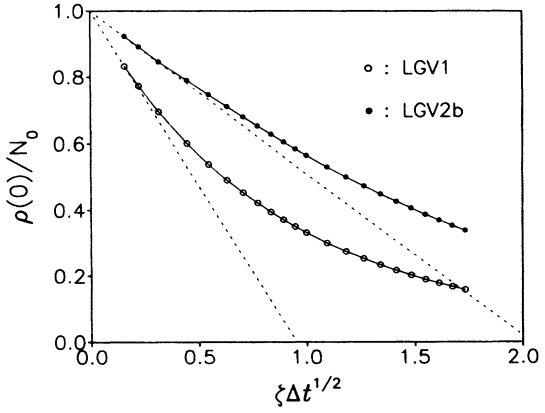

FIG. 12. The convergence of the center of the wave functior from directly iterating the algorithm's integral equation. Dotted lines are straight lines with slopes 1.05 and 0.49.

tained. Approximating the stationary wave function as

$$\rho(r)\approx\rho(0)\exp\left[-\frac{1}{2}\frac{r^2}{s^2}\right] \tag{49}$$

and substituting it into (45), yields, to leading order,

$$\rho(0)=\rho(0)\left[1+2\Delta r\left\langle\frac{1}{r}\right\rangle_{\Delta t}-\frac{1}{2s^2}\langle r^2\rangle_{\Delta t}+\cdots\right]. \tag{50}$$

In order for this stationary condition to hold, we must have

$$s^2\approx\frac{1}{4}\frac{1}{\Delta r}\frac{\langle r^2\rangle_{\Delta t}}{\langle 1/r\rangle_{\Delta t}}\,,$$
$$\zeta s\approx\left[\frac{3}{2}\right]^{1/2}\left[\frac{\pi}{2}\zeta^2\Delta t\right]^{1/4}=1.37(\zeta^2\Delta t)^{1/4}\,. \tag{51}$$

Thus, surprisingly, the range of the Gaussian only vanishes as $\Delta t^{1/4}$. For the more general case of (47), one obtains

$$\zeta^2 s^2\approx\frac{3}{2}\frac{\Gamma(\frac{3}{2})}{\Gamma\left[\frac{n+1}{2}\right]}(2\zeta^2\Delta t)^{1-n/2}+\cdots\,. \tag{52}$$

Again, the $\Delta t^{1/4}$ dependence is special to Slater orbitals.

In Fig. 13, we compare the approximate Gaussian wave function (49), with exact results from directly iterating (41). The range used is $\zeta s=1.30(\zeta^2\Delta t)^{1/4}$, which fits the data slightly better than the derived result (51). Since this is only to check (51), we use the exact central density for $\rho(0)$. The agreement at small r is excellent for all Δt considered. For the case of LGV2b, one obtains

$$\zeta s\approx\left[\frac{3}{4}\right]^{1/2}(\pi\zeta^2\Delta t)^{1/4}=1.15(\zeta^2\Delta t)^{1/4}\,. \tag{53}$$

The coefficient that improves the fit with the corresponding exact result is $\zeta s=1.00(\zeta^2\Delta t)^{1/4}$. The comparison is shown in Fig. 14. Again, the agreement is excellent at small r for all Δt considered.

Although Figs. 13 and 14 look rather similar, there is a subtle difference. Since the exact wave function $\rho_0(r)$ is linear in r near the origin, the Gaussian part of the approximate wave function is simply wrong. Hence the range of the Gaussian is an estimate of the error range r_e of the algorithm, implying that $r_e\approx s\approx\Delta t^{1/4}$. From Fig. 13, it is certainly the case that the range at which the evolving wave function departs from the exact wave function is comparable to the range of the Gaussian. In the case of LGV2b, this argument must also hold for large Δt, implying also that $r_e\approx\Delta t^{1/4}$. However, for small Δt, when the evolving central density is very close to the exact central density and the Gaussian is hemmed in by the sloping exact wave function, the error range is forced to be

$$\rho(0)=N_0(1-\zeta r_e+\cdots)\,,$$
$$N_0[1-0.49(\zeta^2\Delta t)^{1/2}+\cdots]=N_0(1-\zeta r_e+\cdots)\,, \tag{54}$$
$$r_e\approx 0.49\Delta^{1/2}\,.$$

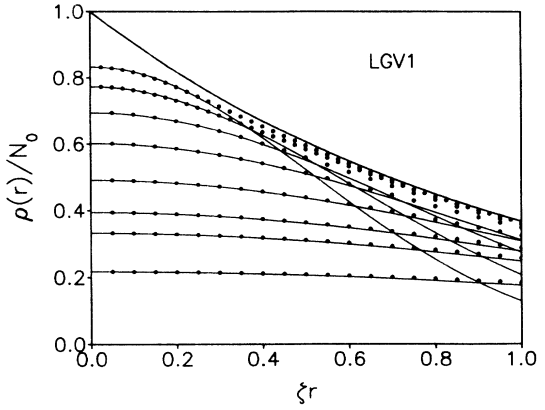

FIG. 13. Comparing the approximate Gaussian function (49) with algorithm LGV1's evolving wave function at finite Δt. Solid lines are Gaussian functions with $\zeta s=1.30(\zeta^2\Delta t)^{1/4}$. Dots are wave functions from directly iterating the algorithm's integral equation. Time-step sizes compared are, from top to bottom, $\zeta^2\Delta t=0.025$, 0.05, 0.1, 0.2, 0.4, 0.7, 1.0, and 2.0.

This is supported by Fig. 14; in the smallest Δt case, the error range is clearly much smaller than the range of the Gaussian. This argument can certainly account for the observed differences between LGV1 and LGV2b. However, we have not shown that at extremely small values of Δt, the same argument cannot be applied to LGV1. On this point, I can only appeal to numerical evidence. As we shall see, numerical results below strongly support the finding that while the error range changes from $r_e\propto\Delta t^{1/4}$ to $r_e\propto\Delta t^{1/2}$ for LGV2b, the error range is strictly $r_e\propto\Delta t^{1/4}$ for LGV1.

The above analysis suggests that, near the origin, the stationary wave function can be written as

$$\rho(r)=\rho_0(r)+\Delta\rho(r)\,, \tag{55}$$

where the roughly triangular defect function $\Delta\rho(r)$ can be approximated by

$$\Delta\rho(r)=\begin{cases}\approx[\rho(0)-\rho_0(0)](1-r/r_e), & 0\le r\le r_e\\ \approx 0, & r>r_e.\end{cases} \tag{56}$$

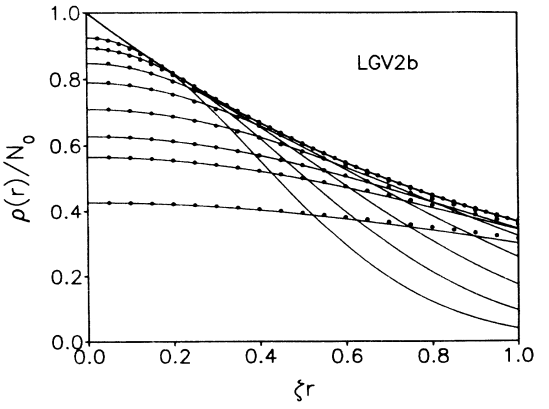

FIG. 14. Same as Fig. 13 but for algorithm LGV2b and $\zeta s=1.00(\zeta^2\Delta t)^{1/4}$.

For our purpose, it is adequate to recall that $[\rho(0)-\rho_0(0)] \propto -\Delta t^{1/2}$. Denoting expectation values with respect to ρ_0 as $\langle\ \rangle_0$, the convergence error in computing $\langle r^n \rangle$ can now be estimated as

$$\delta\langle r^n\rangle = \frac{\int d^3r\, r^n(\rho_0+\Delta\rho)}{\int d^3r(\rho_0+\Delta\rho)} - \langle r^n\rangle_0$$
$$\approx \int d^3r\, r^n\Delta\rho - \langle r^n\rangle_0 \int d^3r\Delta\rho$$
$$\propto -\Delta t^{1/2}\left[r_e^{n+3} - \langle r^n\rangle_0\left[1+\frac{n}{3}\right]\left[1+\frac{n}{4}\right]r_e^3\right], \tag{57}$$

For LGV1, this yields

$$\delta^{(1)}\langle r^n\rangle \propto \left[-\Delta t^{(n+5)/4} + \langle r^n\rangle_0\left[1+\frac{n}{3}\right]\left[1+\frac{n}{4}\right]\Delta t^{5/4}\right]. \tag{58}$$

For LGV2b, the convergence error must change from that of $\delta^{(1)}$ to

$$\delta^{(2)}\langle r^n\rangle \propto \left[-\Delta t^{(n+4)/2} + \langle r^n\rangle_0\left[1+\frac{n}{3}\right]\left[1+\frac{n}{4}\right]\Delta t^2\right]. \tag{59}$$

Note that for $n<0$, the first term is of lower order and is negative; for $n>0$, the second term is of lower order and is positive. Thus (58) and (59) immediately predict that for $n<0$, the convergence is from below, whereas for $n>0$ the convergence is from above. Also, these errors are conspicuous only when they are of lower order than that of the algorithm. In Figs. 15–18, we examine in detail cases for $n=-1$, -2, 1, and 2. In each case, the scaled, dimensionless expectation value $\langle(\zeta r)^n\rangle$ is plotted as a function of the dimensionless time-step parameter $\zeta^2\Delta t$.

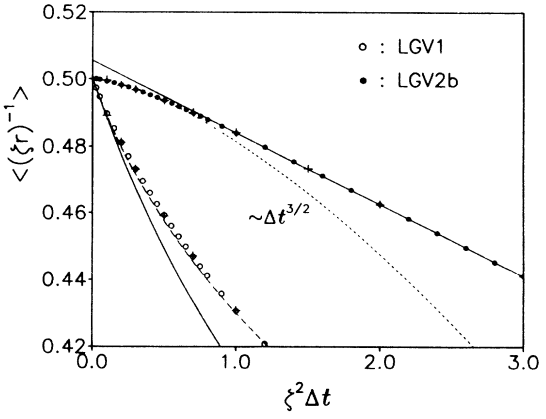

FIG. 15. The convergence of $\langle(\zeta r)^{-1}\rangle$. Dots and circles are expectation values obtained from directly iterating the algorithm's integral equation. Crosses are corresponding Monte Carlo results. The dashed and solid lines through LGV1 data are expectation values computed with the first- and second-order perturbative wave functions (43) and (44), respectively. The solid straight line and the dotted curve through LGV2b data are simple fits to demonstrate the Δt and the $\Delta t^{3/2}$ convergence behaviors, respectively.

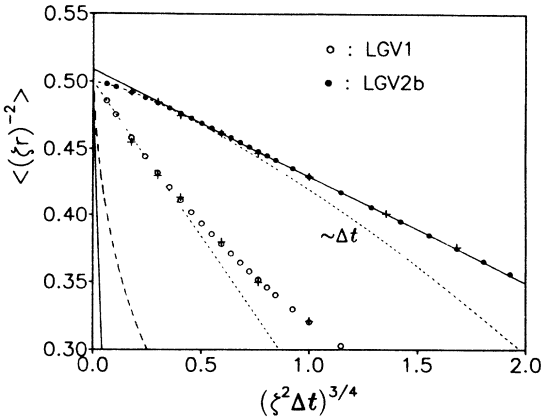

FIG. 16. Same as Fig. 15 but for $\langle(\zeta r)^{-2}\rangle$. The additional straight dotted line through LGV1 data is a fit to demonstrate the expected $\Delta t^{3/4}$ convergence.

Figure 15 shows the convergence of $\langle(\zeta r)^{-1}\rangle$, which is of special importance for atomic calculations. LGV1 converges linearly as expected, and is well described by the first-order perturbative wave function (43). The second-order wave function (44), which is a poorer wave function at small r, does not describe data well at larger values of Δt. Note that their common range of linear convergence is rather narrow. According to (58), the nonperturbative error is also linear in Δt. In this case, to the extent that the observed error is well described by the perturbative wave functions at small Δt, the nonperturbative contribution of (58) does not appear to be important. LGV2b converges linearly for $\zeta^2\Delta t \approx 1$–4, then converges as $\Delta t^{3/2}$ for $0<\Delta t \lesssim 1$. (I have checked that, in the latter case, the data definitely cannot be fitted with a quadratic dependence.) In the Appendix, I show that there is no second-order perturbative corrections to the wave function, hence this behavior is inexplicable in terms of perturbative analysis but is in precise agreement with (58)

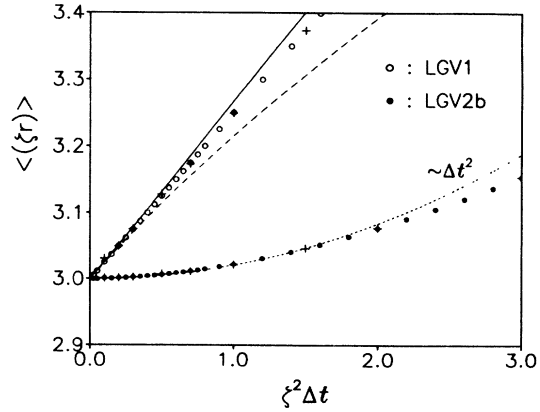

FIG. 17. Same as Fig. 15 but for $\langle\zeta r\rangle$. There is only one dotted curve through LGV2b data to indicate the expected quadratic convergence. See text for details.

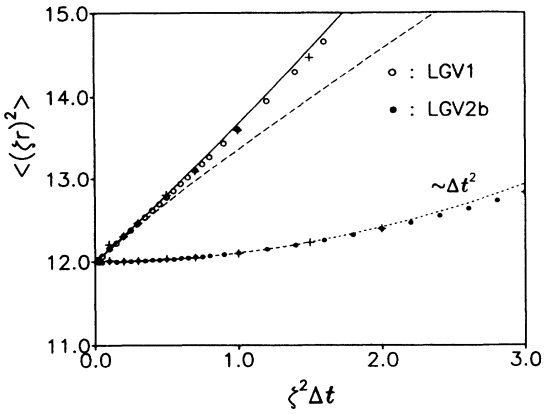

FIG. 18. Same as Fig. 17 but for $\langle(\zeta r)^2\rangle$.

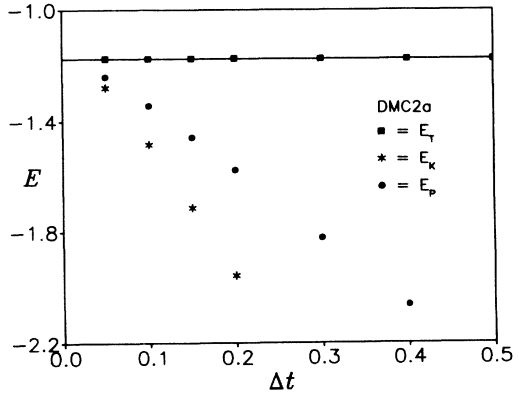

FIG. 19. Comparing the convergence of three ground-state energy estimators E_T, E_K, and E_P associated with trial function $\Phi_0^{\rm II}$ as obtained by algorithm DMC2a. See text for details.

and (59). The $\Delta t^{3/2}$ convergence is easily missed if one is not looking for it or if Δt is not small enough. More seriously, an incorrect result is obtained if one naively extrapolates just its linear dependence. (Even more deceiving is that fact that the wrong answer is only slightly incorrect.) This explains why second-order DMC algorithms which do not obey the cusp condition apparently only converge linearly. It is probable that they also converge as $\Delta t^{3/2}$ for small enough Δt.

Figure 16 shows the convergence of $\langle(\zeta r)^{-2}\rangle$. This is a crucial check on our nonperturbative analysis (58), which predicts that *not even a first-order algorithm can converge linearly.* For the ease of comparison, data are plotted as a function of $(\zeta\Delta t)^{3/4}$. For LGV1, perturbative corrections utterly fail to account for the observed step-size error. The convergence is $(\zeta\Delta t)^{3/4}$, in accordance with (58), rather than linear. For LGV2b, one again observes a clear crossover from that of $\Delta t^{3/4}$ to Δt, as predicted by (58) and (59).

To demonstrate that these nonperturbative errors are conspicuous only when they are of lower order than the order of the algorithm, Fig. 17 shows the convergence of $\langle\zeta r\rangle$. For LGV1, since (58) predicts that $\delta^{(1)}=\frac{5}{4}>1$, the convergence is dominated by perturbative effects and the expectation value should converge linearly. This is indeed observed, and data are well described by perturbative wave functions, especially by $\rho_0^{(2)}(r)$. For LGV2, since there is no second-order perturbative correction, one can only ascribe the observed quadratic convergence to (59). Unlike the previous two cases, over the same range of Δt, there is no abrupt change of convergence from that of $\Delta t^{5/4}$ to Δt^2. However, since this expectation is sensitive to the wave function at large r, the defect function (56) may simply to too crude to describe this crossover in detail. Similar remarks may also be applied to results obtained for $\langle(\zeta r)^2\rangle$, as shown in Fig. 18.

On the basis of (59), which is verified by Figs. 15–18, one concludes that the expectation value $\langle O\rangle$ computed via second-order Langevin algorithms can converge quadratically, provided that the operator O contains no negative powers of r. Thus without fundamentally curing the convergence failure of the algorithm, one can nevertheless circumvent it by avoiding operators that are sensitive to the wave function near the origin. In the case of atomic calculations, this means that one must choose trial functions such that the local energy is free from $1/r$ singularities; i.e., one must enforce all attractive cusp conditions. This explains why, in Sec. V, the trial energy converges quadratically in the case of $\Phi_0^{\rm II}$ and $\Phi_0^{\rm III}$ but not in the case of $\Phi_0^{\rm I}$.

As a further check, one notes that even when the trial function obeys all the cusp conditions and the local energy is nonsingular, the kinetic and the potential energies individually contain $1/r$ singularities. Thus the above explanation also predicts that, whereas the total energy converges quadratically, the kinetic and the potential energies individually do not. This is easy to check and is shown in Figs. 19–21 for the case of $\Phi_0^{\rm II}$. Instead of directly plotting for each second-order algorithm the mixed expectations $K_m=\langle\Psi_0|K|\Phi_0\rangle/\langle\Psi_0|\Phi_0\rangle$ and $V_m=\langle\Psi_0|V|\Phi_0\rangle/\langle\Psi_0|\Phi_0\rangle$, where $E_T=K_m+V_m$, I plotted

$$E_K=-(2K_m-K_v)\,,\qquad E_P=\tfrac{1}{2}(2V_m-V_v)\,, \tag{60}$$

where K_v and V_v are the corresponding variational kinetic and potential energies, respectively. Since E_K and E_P are just linear transformations of K_m and V_m, the order

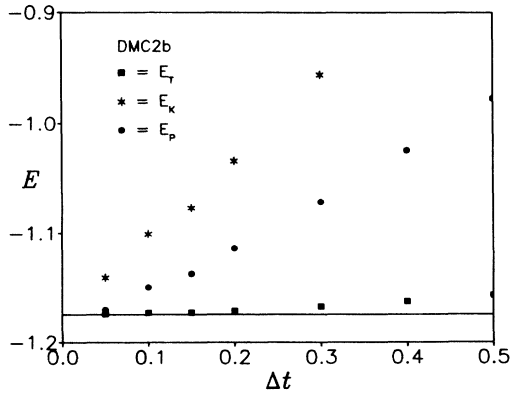

FIG. 20. Same as Fig. 19 but for algorithm DMC2b.

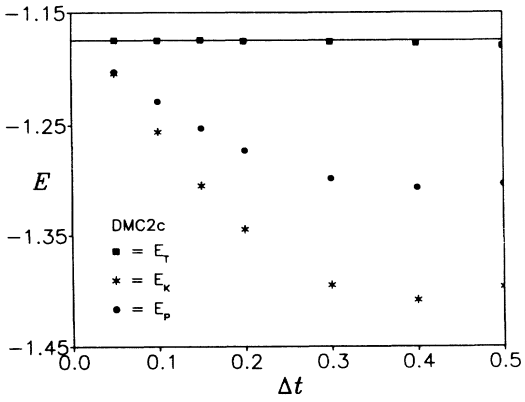

FIG. 21. Same as Fig. 20 but for algorithm DMC2c.

of convergence is obviously not affected. However, since $2O_m - O_v \approx \langle \Psi_0 | O | \Psi_0 \rangle / \langle \Psi_0 | \Psi_0 \rangle$ constitutes a perturbative estimate of the exact ground-state expectation value,[1] the virial theorem[39] implies that both E_K and E_P should also converge toward the exact ground-state energy. Their systematic derivations from E_0 would then give us another assessment of the quality of the trial function. Figure 19–21 clearly confirm the expectation that E_K and E_P are not converging quadratically. The convergence appears to be linear, but a $\Delta t^{3/2}$ dependence as $\Delta t \to 0$ is expected. Moreover, the steep slopes in the case of DMC2a imply that quadratic convergence is achieved through large but precise cancellations.

VII. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, I have used the operator approach to analyzing the diffusion Monte Carlo algorithm in terms of its underlying transfer matrix. The key observation is that by virtue of the Campell-Baker-Hausdorff formula, any approximate transfer matrix can be reconstituted to yield an approximate Hamiltonian. This approximate Hamiltonian then determines the perturbative convergence behavior of the corresponding algorithm and can be used to systematically derive second-order algorithms. I have also made explicit the essential connection between the importance-sampling process and the Langevin algorithm. It is the convergence of the Langevin algorithm that dictates the overall convergent behavior of the DMC algorithm. As a practical matter, one should first check the convergence of the embedded Langevin algorithm before running the full DMC algorithm.

The failure of the second-order DMC algorithm to converge quadratically is now understood quantitatively in terms of nonperturbative convergence errors due to the intrinsic failure of the Langevin algorithm in sampling Slater orbitals. Fortunately, when computing the energy, the effect of these errors can be neutralized by enforcing cusp conditions. When this is done, I show that the ground-state energy of He and H_2 do in fact converge quadratically. In this work, the existence of nonperturbative errors is partly based on numerical evidence, e.g., the $\Delta t^{1/2}$ convergence of $\rho(0)$, the $\Delta t^{1/4}$ and $\Delta t^{1/2}$ dependence of r_e, the form of the defect function $\Delta\rho(r)$, etc. This is not completely satisfactory; a more direct, analytical demonstration of these effects would be valuable. Moreover, it would be very important to find out whether alternative DMC algorithms can be devised such that these nonperturbative errors are fundamentally eliminated rather than just avoided.

The calculation of He and H_2 clearly shows that second-order DMC algorithms are superior to the first-order algorithm. This is amply demonstrated in Figs. 1 and 2 and in Figs. 4 and 5. With good trial functions, as shown in Figs. 1 and 4, and in Tables I and II, it appears that these second-order algorithms can directly compute the ground-state energy at a reasonably large value of Δt without extrapolation. For He and H_2, I have shown that Δt can be as large as 0.05. The efficiency of these new algorithms has been further verified by calculations done on model nuclei[27,28] and helium droplets.[29] In view of the simplicity of these algorithms as compared to the GFMC algorithm, it is of interest to test them further on problems of greater complexity, such as many-fermion problems and boson problems with spin correlations. Some works in these areas are in progress.

Of the three second-order algorithms, the fastest is DMC2c. However, its sometimes erratic convergent behavior is disconcerting. In the case of Fig. 1, one is sure that it actually overshoots the exact result by a small amount and converges from above. This peculiar behavior is traceable to its corresponding Langevin algorithm LGV2c and is known to occur also in lattice gauge calculations.[31] Thus when one is using DMC2c, one must be careful in deciding whether the algorithm has actually converged at a seemingly large value of Δt. A more-detailed analysis of the step-size error of LGV2c would be very valuable. (Since it is noncanonical, it cannot be analyzed by the transfer matrix method advocated here.) By contrast, in all cases considered, the canonical algorithms DMC2a and DMC2b converge monotonically. However, DMC2a is very slow, usually $\approx 50\%$ slower than DMC2c. Moreover, since LGV2b ends with a random walk, DMC2b's wave function is always nonsingular. There is no such guarantee for DMC2a. In fact, for cases considered in Sec. VI, the wave function generated by LGV2a contains explicit $1/r$ and $1/r^2$ singularities. We thus find that DMC2b is a good compromise between speed and reliability.

ACKNOWLEDGMENTS

I am grateful to my colleagues, D. Ernst, C. R. Hu, and C. M. Ko, E. Krotscheck, and R. Smith, for the use of their work stations in carrying out the calculations of this paper. This work was supported in part by National Science Foundation Grant Nos. PHY86-08149 and PHY89-07986.

APPENDIX: PERTURBATIVE CONVERGENCE ERRORS

As discussed in Sec. II, the convergence behavior of an algorithm can be studied via its reconstituted Hamiltoni-

an H'. For LGV1, where the transfer matrix is $T_L = e^{-\Delta t K} e^{-\Delta t D}$, the latter is given by

$$H' = H + \Delta t H_1 + \Delta t^2 H_2 \cdots , \tag{A1}$$

where $H = K + D$, $H_1 = -\frac{1}{2}[K,D]$, and $H_2 = \frac{1}{12}[(K-D),[K,D]]$. In solving for the ground state of H' perturbatively, one obtains a wave function expanded in integral powers of Δt. We shall refer to these Δt-dependent terms as perturbative convergence errors, distinct from those discussed in Sec. VI, which are trial-function specific and can involve nonintegral powers of Δt.

For the special case of sampling the Slater orbital $\rho_0(r) = \Phi_0^2(r) = e^{-\zeta r}$, with $\mathbf{G}(\mathbf{r}) = -\frac{1}{2}\zeta\hat{\mathbf{r}}$, the effect of H_1 on a spherical symmetric function $\rho(r)$ is

$$\begin{aligned} H_1\rho &= \tfrac{1}{2}[D,K]\rho \\ &= \tfrac{1}{4}\nabla_i[-G_i\nabla^2\rho + \nabla_i\nabla_j(G_j\rho)] \\ &= \tfrac{1}{4}\nabla_i\left[\tfrac{1}{2}\zeta\hat{r}_i\left[\rho'' + \frac{2}{r}\rho'\right] - \nabla_i\left[\frac{1}{r}\zeta\rho + \tfrac{1}{2}\zeta\rho'\right]\right] \\ &= \tfrac{1}{4}\nabla_i\left[\zeta\frac{\hat{r}_i}{r^2}\rho\right] . \end{aligned} \tag{A2}$$

Similarly, the effect of H_2 is

$$\begin{aligned} H_{2\rho} &= \tfrac{1}{12}[(K-D),[K,D]]\rho \\ &= \tfrac{1}{6}[H_1,(K-D)]\rho \\ &= \tfrac{1}{24}\nabla_i\left[\zeta\frac{\hat{r}_i}{r^3}(\zeta\rho - 2\rho')\right] . \end{aligned} \tag{A3}$$

In deriving these results, we have repeatedly used the identity

$$\nabla_i\hat{r}_j = \frac{1}{r}(\delta_{ij} - \hat{r}_i\hat{r}_j) , \tag{A4}$$

and have consistently ignored δ functions arising from $\nabla^2(1/r) = -4\pi\delta^3(r)$.

The first-order zero-energy ground state of H' is determined by

$$\begin{aligned} &(H + \Delta t H_1)\rho_0^{(1)}(r) = 0 , \\ &\nabla_i\left[-\tfrac{1}{2}\nabla_i - \tfrac{1}{2}\zeta\hat{r}_i + \tfrac{1}{4}\zeta\Delta t\frac{\hat{r}_i}{r^2}\right]\rho_0^{(1)}(r) = 0 , \end{aligned} \tag{A5}$$

which can be readily integrated to give

$$\rho_0^{(1)}(r) = N_1\exp\left[-\zeta r - \frac{\zeta\Delta t}{2r}\right] . \tag{A6}$$

The normalization constant N_1 is expressible in terms of modified Bessel functions $K_n[(2\zeta^2\Delta t)^{1/2}]$. However, in practice, we find it simpler to compute it by directly integrating $\rho_0^{(1)}(r)$.

The second-order ground state of H' is determined by

$$\begin{aligned} &(H + \Delta t H_1 + \Delta t^2 H_2)\rho_0^{(2)}(r) = 0 , \\ &\nabla_i\left[-\tfrac{1}{2}\nabla_i - \tfrac{1}{2}\zeta\hat{r}_i + \tfrac{1}{4}\zeta\Delta t\frac{\hat{r}_i}{r^2} + \tfrac{1}{24}\zeta\Delta t^2\frac{\hat{r}_i}{r^3}\left[\zeta - 2\frac{\partial}{\partial r}\right]\right]\rho_0^{(2)}(r) = 0 . \end{aligned} \tag{A7}$$

Writing $\rho_0^{(2)}(r) = e^{S(r)}$, the above implies that

$$S'(r) = -\zeta\left[1 - \tfrac{1}{2}\frac{\Delta t}{r^2} - \tfrac{1}{12}\zeta\frac{\Delta t^2}{r^3}\right]\left[1 + \tfrac{1}{6}\zeta\frac{\Delta t^2}{r^3}\right]^{-1} .$$

Keeping terms only up to second order in Δt and integrating gives

$$\rho_0^{(2)}(r) = N_2\exp\left[-\zeta r - \frac{\zeta\Delta t}{2r} - \frac{1}{2}\left[\frac{\zeta\Delta t}{2r}\right]^2\right] . \tag{A8}$$

Again, we compute the normalization constant N_2 by numerically integrating $\rho_0^{(2)}(r)$.

For algorithm LGV2b, whose transfer matrix is $T = e^{-\Delta t K/2} e^{-\Delta t D} e^{-\Delta t K/2}$, its reconstituted Hamiltonian is

$$H' = H + \tfrac{1}{24}\Delta t^2[(D+H),[D,K]] + \cdots . \tag{A9}$$

In this case, the effect of the second-order Hamiltonian is

$$\begin{aligned} H_2\rho &= \tfrac{1}{24}[(D+H),[D,K]]\rho \\ &= \tfrac{1}{12}[D+H,H_1]\rho \\ &= \tfrac{1}{24}\nabla_i\left[\zeta\frac{\hat{r}_i}{r^3}(\zeta\rho + \rho')\right] . \end{aligned} \tag{A10}$$

The corresponding second-order ground state is determined by

$$\begin{aligned} &(H + \Delta t^2 H_2)\rho_0^{(2)}(r) = 0 , \\ &\nabla_i\left[-\tfrac{1}{2}\nabla_i - \tfrac{1}{2}\zeta\hat{r}_i + \tfrac{1}{24}\zeta\Delta t^2\frac{\hat{r}_i}{r^3}\left[\zeta + \frac{\partial}{\partial r}\right]\right]\rho_0^{(2)}(r) = 0 . \end{aligned} \tag{A11}$$

Again, if $\rho_0^{(2)}(r) = e^{S(r)}$, then

$$S'(r) = -\zeta\left[1 - \tfrac{1}{12}\zeta\frac{\Delta t^2}{r^3}\right]\left[1 - \tfrac{1}{12}\zeta\frac{\Delta t^2}{r^3}\right]^{-1} = -\zeta . \tag{A12}$$

Thus surprisingly, there are no second-order corrections. Since LGV2b is even in Δt, this second-order cancellation implies that the perturbative error in sampling $\rho_0(r) = e^{-\zeta r}$ must be at least fourth order in Δt.

[1] D. Ceperly and M. Kalos, in *Monte Carlo Methods in Statistical Mechanics*, edited by K. Binder (Springer, New York, 1979).

[2] J. G. Zabolitzky, Prog. Part. Nucl. Phys. **16**, 103 (1986).

[3] J. Anderson, J. Chem. Phys. **63**, 1499 (1975); **65**, 4121 (1976); **73**, 3897 (1980); **74**, 6307 (1981); **76**, 5150 (1982).

[4] M. Kalos, D. Levesque, and L. Verlet, Phys. Rev. A **9**, 2178 (1974).

[5]P. A. Whitlock, M. H. Kalos, G. V. Chester, and D. M. Ceperley, Phys. Rev. B **19**, 5598 (1979).
[6]M. Kalos, M. Lee, P. Whitlock, and G. Chester, Phys. Rev. B **24**, 115 (1981).
[7]D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[8]P. J. Reynolds, D. M. Ceperley, B. J. Alder, and W. A. Lester, J. Chem. Phys. **77**, 5593 (1982).
[9]J. W. Moskowitz, K. E. Schmidt, M. E. Lee, and M. H. Kalos, J. Chem. Phys. **77**, 349 (1982).
[10]D. Ceperley, J. Comput. Phys. **51**, 404 (1983).
[11]D. Ceperley and B. Alder, J. Chem. Phys. **81**, 5833 (1984).
[12]J. G. Zabolitzky and M. H. Kalos, Nucl. Phys. A **356**, 114 (1981).
[13]J. Zabolitzky, K. Schmidt, and M. Kalos, Phys. Rev. C **25**, 1111 (1982).
[14]J. Carlson, Phys. Rev. C **36**, 2026 (1987).
[15]S. E. Koonin, in *Nuclear Theory 1981*, edited by G. F. Bertsch (World Scientific, Singapore, 1983).
[16]B. Serot, S. Koonin, and J. Negele, Phys. Rev. C **28**, 1679 (1983).
[17]J. W. Negele, J. Stat. Phys. **43**, 991 (1986).
[18]D. W. Heys and D. R. Stump, Phys. Rev. D **30**, 1315 (1984).
[19]S. A. Chin, J. W. Negele, and S. E. Koonin, Ann. Phys. (N.Y.) **157**, 140 (1984).
[20]T. DeGrand and J. Potvin, Phys. Rev. D **31**, 871 (1985).
[21]S. A. Chin, O. S. van Roosmalen, E. A. Umland, and S. E. Koonin, Phys. Rev. D **31**, 3201 (1985).
[22]S. A. Chin, C. Long, and D. Robson, Phys. Rev. D **37**, 3001 (1988).
[23]S. A. Chin, in *International Review of Nuclear Physics*, edited by T. T. S. Kuo and E. Osnes (World Scientific, Singapore, 1986), Vol. 4.
[24]J. Vrbit, J. Phys. A **18**, 1327 (1985).
[25]J. Vrbit and S. Rothstein, J. Comput. Phys. **63**, 130 (1986).
[26]S. Rothstein, N. Patil, and J. Vrbit, J. Comp. Chem. **8**, 412 (1987).
[27]S. A. Chin (unpublished).
[28]S. A. Chin and E. Kroscheck, Texas A&M University Report No. CTP07/89, 1989.
[29]S. A. Chin and E. Kroscheck, Phys. Rev. Lett. **65**, 2658 (1990).
[30]H. Rishen, *The Fokker-Planck Equation* (Springer-Verlag, Berlin, 1984).
[31]S. A. Chin, Nucl. Phys. (Proc. Suppl.) B **9**, 498 (1989).
[32]E. Helfand, Bell Syst. Tech. J. **58**, 2289 (1979).
[33]A. Ukawa and M. Fukugita, Phys. Rev. Lett. **55**, 1854 (1985).
[34]G. G. Batrouni *et al.* Phys. Rev. D **32**, 2736 (1985).
[35]I. Drummond, S. Duane, and R. Horgan, Nucl. Phys. B **220**, 119 (1983).
[36]C. Umrigar, K. Wilson, and J. Wilkins, Phys. Rev. Lett. **60**, 1719 (1988).
[37]D. Freund, B. Huxtable, and J. Morgan, Phys. Rev. A **29**, 980 (1984), and references therein.
[38]W. Kolos and L. Wolniewicz, J. Chem. Phys. **49**, 404 (1968), and references therein.
[39]J. C. Slater, *Quantum Theory of Matter*, 2nd ed. (McGraw-Hill, New York, 1968), Chap. 20.