

An improved transition matrix for variational quantum Monte Carlo

Massimo Mella^{a,b}, Arne Lüchow^b, James B. Anderson^b

^a *Dipartimento di Chimica Fisica ed Elettrochimica, Università degli Studi di Milano, via Golgi 19, 20133 Milan, Italy*

^b *Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA*

Received 22 November 1996

Abstract

An improved transition matrix for variational Monte Carlo calculations is proposed. This matrix allows the use of larger time steps than the usual Langevin-based transition matrix and provides efficient sampling of electron positions in both the core and valence regions. Its efficiency and accuracy in predictions of energies for hydrogen-like systems and for the neon atom are demonstrated.

1. Introduction

The variational Monte Carlo (VMC) method is a convenient method for computing expectation values for a given trial wavefunction Ψ_T using Monte Carlo (MC) techniques [1]. Since analytical integration is not required, complete freedom is allowed in the choice of the form of the wavefunction for a given system. Recently, with the use of sophisticated forms for Ψ_T , excellent results have been obtained for first-row atoms [2–4], small molecules [5] and clusters composed of rare gas atoms [6–8]. Trial wavefunctions are usually optimized by minimization of the variance of the local energy $E_{\text{loc}} = \mathcal{H}\Psi_T/\Psi_T$ or a combination of the variance with the mean energy [7,9–11].

To obtain expectation values for various quantities for a given trial wavefunction Ψ_T , VMC methods sample the distribution $|\Psi_T|^2$ by means of the generalized Metropolis algorithm. In this procedure, a set of sample points (walkers) at positions $\{R_i\}$ in configuration space is used to generate a new set at positions $\{R'_i\}$ by displacing each according to an

analytical stochastic transition matrix $T(R \rightarrow R')$. The displacement to the new position for each walker is accepted with a probability

$$P_{\text{acc}}(R \rightarrow R') = \min \left[1, \frac{T(R' \rightarrow R) |\Psi_T(R')|^2}{T(R \rightarrow R') |\Psi_T(R)|^2} \right]. \quad (1)$$

If the proposed new position is not accepted the walker remains at its old position.

This procedure, given sufficient time to equilibrate and provided the process is ergodic, samples the correct distribution $|\Psi_T|^2$. Different algorithms use different transition matrices to choose the proposed displacement and their efficiency is strongly related to the choice of matrix. The standard Metropolis algorithm uses a constant transition matrix over a hypercube, while the Langevin-based algorithm chooses the displacements using the short-time approximation of the Green's function of the Fokker–Planck equation [1]. This is given by (in

dimensionless form for particles having the mass of the electron)

$$T(\mathbf{R} \rightarrow \mathbf{R}', t) = \frac{1}{(2\pi t)^{3N/2N}} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R} - t\mathbf{V})^2}{2t}\right] \quad (2)$$

where the quantum velocity is

$$\mathbf{V}(\mathbf{R}) = \left[\frac{\nabla \Psi_T}{\Psi_T} \right]_{\mathbf{R}} \quad (3)$$

N is the number of particles, \mathbf{R} and \mathbf{R}' are the old and new points in $3N$ configuration space and t is the time step for the simulation.

These transition matrices can lead to low efficiencies which result from different time or length scales present in different regions of configuration space [1]. For a first-row atom, to get the same acceptance probability in all the regions of the configuration space, the length of the attempted step must be at least an order of magnitude smaller in the core than in the valence region [12]. In the Langevin-based algorithm, this is equivalent to using a smaller time step for a move in the core region. The same difficulty arises in VMC calculations for helium clusters if the Langevin algorithm is used. The time step in the region near the coalescence point, i.e. for two atoms in close proximity, must be much smaller than in the rest of the configuration space [6]. This requirement makes the application of VMC techniques to molecules containing atoms with atomic number $Z > 10$ or to larger clusters computationally challenging.

This raises the question of how one can efficiently use different time steps or length scales in VMC simulations. Several different solutions have been explored in previous work. Umrigar [13] proposed a factorization of the transition matrix for each electron into a radial part T_r and an angular part T_Ω , i.e. $T = T_r T_\Omega$. Sun et al. [14] suggested the use of a simplified quadratic approximation for the wavefunction to sample the electron displacements, while Belohorec et al. [15] used different time steps for electrons in different regions of space. For helium clusters, Barnett and Whaley resorted to the use of the simple Metropolis procedure [6] and to different guiding functions [16] in order to avoid the problems

generated by the divergence of the modulus of the quantum velocity near the coalescence points. The efficiencies of these algorithms have been tested using the correlation time T_{corr} for a given trial wavefunction Ψ_T , the local and global acceptance ratios of Eq. (1), the accepted displacement and the quality of the sampled distributions in the different regions of the configuration space. The last three quantities give us ‘‘microscopic’’ [14] insight into the performance of the algorithms, while the correlation time T_{corr} is a ‘‘macroscopic’’ measure of the simulation efficiency. Similar tests have been made for an algorithm proposed by Batrouni and Reynolds [17] in which an acceleration matrix was introduced in order to obtain comparable speed for different coordinates. We also note a different approach, based on biased sampling of a distribution different from $|\Psi_T|^2$, recently proposed by Alexander and Coldwell [18].

In this Letter, we propose a simple modification of the Langevin-based transition matrix of Eq. (2) for use in VMC as well as diffusion Monte Carlo (DMC) simulations. In the section following we describe the proposed algorithm and in succeeding sections we describe several applications and discuss the performance of the algorithm.

2. Algorithm

The basic idea for the new transition matrix is quite simple: in the core or coalescence regions, where the standard short-time approximation of Eq. (2) deviates most from the exact transition matrix, and near the nodal surfaces where the quantum velocity diverges, the time step must be reduced. To achieve this goal we replace, for each particle, in the transition matrix of Eq. (2) the time t by two functions of t , $\tau_d(t)$ and $\tau_v(t)$, having the limiting form $\tau(t) = t + \mathcal{O}(t^2)$:

$$T(\mathbf{R} \rightarrow \mathbf{R}', \tau) = \prod_{i=1}^N \frac{1}{(2\pi\tau_{d,i})^{3/2}} \exp\left[-\frac{(\mathbf{r}'_i - \mathbf{r}_i - \tau_{v,i}\mathbf{v}_i)^2}{2\tau_{d,i}}\right], \quad (4)$$

where \mathbf{r}_i and \mathbf{r}'_i refer to the initial and final positions of the i th particle, respectively, and

$$\mathbf{v}_i(\mathbf{R}) = \left[\frac{\nabla_i \Psi_T}{\Psi_T} \right]_{\mathbf{R}}. \quad (5)$$

We define the effective time steps

$$\tau_{d,i} = (1 - c) \frac{e^{2a_i(\mathbf{R})t} - 1}{2a_i(\mathbf{R})} + \tau_{\min} \quad (6)$$

and

$$\tau_{v,i} = (1 - c) \frac{e^{a_i(\mathbf{R})t} - 1}{a_i(\mathbf{R})} + \tau_{\min} \quad (7)$$

where t is the usual time step, and $\tau_{\min} = ct$ is the minimum time step allowed for the simulation whatever the analytical form of $a_i(\mathbf{R})$. Here, indices d and v mean “diffusion” and “velocity”, respectively. In preliminary tests we found the efficiency of the transition matrix to be independent of the value of c for $c \leq 10^{-1}$. In this work we use $c = 10^{-2}$. The form of τ_d and τ_v is motivated by the exact Green’s function for the Ornstein–Uhlenbeck process with a linear drift term [19,20] and by the mapping proposed by Drozdov [21]. Using a Taylor expansion for $t \rightarrow 0$ it may be shown that $\tau_{d,v} = t + \mathcal{O}(t^2)$. In this limit the new matrix reduces to the standard Langevin short-time approximation, Eq. (2).

The function $a_i(\mathbf{R})$ is arbitrary except that it must be large and negative in those regions where a reduced time step is desired. We chose

$$a_i(\mathbf{R}) = k \nabla_i^2 \ln |\Psi_T(\mathbf{R})|, \quad (8)$$

where k is an adjustable parameter. For $k \rightarrow 0$ Eq. (4) reduces to the standard approximation, Eq. (2). The value of k may be optimized with respect to the efficiency of the simulation. All new quantities are defined using values available from the calculation of the local energy and the quantum velocity $\mathbf{V}(\mathbf{R})$ without additional computational effort.

It can be shown by means of a Taylor expansion of Ψ_T in Eq. (8) that both $\tau_{v,i}$ and $\tau_{d,i}$ approach τ_{\min} at nodal surfaces and at electron–nucleus cusps. As a result the new transition matrix is able to reduce the attempted step sizes in these regions. In a similar way one can show that for the general repulsive coalescence point described in the trial wavefunction by the analytical form e^{-h/r^n} with $n > 1$, the transi-

tion matrix reduces both $\tau_{v,i}$ and $\tau_{d,i}$ to τ_{\min} , for $r \rightarrow 0$.

The new transition matrix could also be used in diffusion Monte Carlo simulations because it reduces in the short-time limit to the standard short-time approximation as shown above. This is in contrast to other improved transition matrices [13,14].

3. Results and discussion

As an initial test of the accuracy of the new transition matrix we compared the accuracy of the walker distributions for a hydrogen-like atom with $Z = 6$ obtained in running VMC simulations (without an acceptance step) using the new matrix of Eq. (4) with $k = 1$ and using the standard Langevin matrix of Eq. (2). Since both transition matrices are correct only in the limit of $\tau \rightarrow 0$, all computed mean values $\langle \hat{O}_{\text{app}} \rangle$ are dependent on the time step chosen to carry out the simulation and differ from the exact values $\langle \hat{O}_{\text{exact}} \rangle$ by a quantity that is smaller for the matrix that gives the best approximation to the exact one. In Fig. 1 and Fig. 2 the relative errors $(\langle \hat{O}_{\text{app}} \rangle - \langle \hat{O}_{\text{exact}} \rangle) / \langle \hat{O}_{\text{exact}} \rangle$ for the mean local energy and the mean potential energy for the two transition matrices are plotted as a function of the time step. These were obtained employing the trial wavefunction $\Psi_T = e^{-5.9r}$, whose mean total energy $\langle E \rangle$ and potential energy $\langle V \rangle$ are -17.995 and -35.4 hartree, respectively. As illustrated in Figs. 1 and 2 the modified transition matrix is significantly better than the usual one in terms of the relative error. The

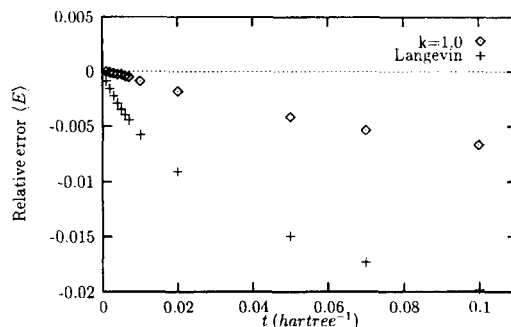


Fig. 1. Relative errors in mean local energy for a hydrogen-like atom with $Z = 6$ versus τ for the Langevin and improved matrices.

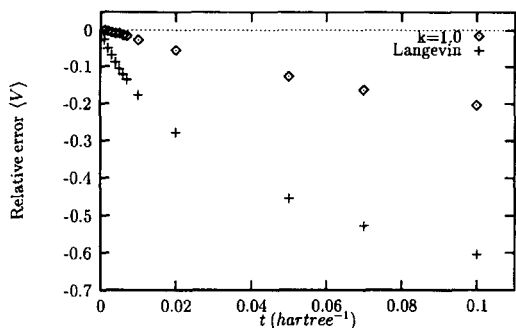


Fig. 2. Relative errors in mean potential energy for a hydrogen-like atom with $Z = 6$ versus τ for the Langevin and improved matrices.

error in each case is lower for the new matrix by a factor of six at the larger values of τ . Both the sign of the relative errors for $\langle V_{\text{app}} \rangle$ in Fig. 2 and plots of the electronic distribution generated using the new matrix show that the proposed matrix improves the sampling in the electron–nucleus cusp region increasing the walker density in the vicinity of the singularity of the electron–nucleus potential. This produces an increase in the acceptance ratio and thus decreases the correlation between successive samples, as will be shown later. A similar behavior was observed for all hydrogen-like systems with $1 \leq Z \leq 10$.

To compare the performances and efficiencies of the Langevin and proposed matrices in a realistic chemical problem, the neon atom was chosen as a second model system. The trial wavefunction Ψ_T was the double-zeta SCF wavefunction of Clementi and Roetti [22], multiplied by an electron–electron Jastrow factor in order to correctly satisfy the electronic cusp condition:

$$\Psi_T = \det \Phi^{(\alpha)} \det \Phi^{(\beta)} \prod_{i < j = 1}^{N_{el}} \exp\left(\frac{a_{ij} r_{ij}}{1 + b r_{ij}}\right), \quad (9)$$

where $a_{ij} = 1/4$ for like-spin electron pairs, $a_{ij} = 1/2$ for unlike-spin electron pairs and $b = 4.0$.

Like Sun et al. [14] we employed a microdiagnostic method to evaluate the efficiency of the algorithm. The space around the nucleus was divided into spherical shells and different quantities for each shell were accumulated during VMC runs. The statistical error, estimated from repeated runs, was found to be less than 1% for all the computed quantities. Unless

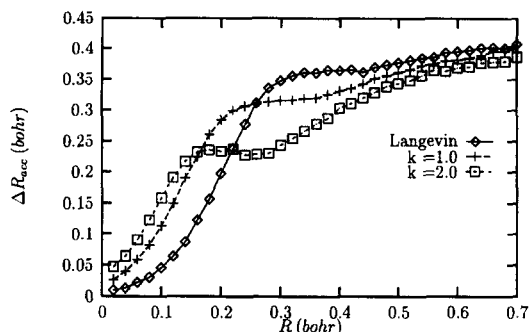


Fig. 3. Mean accepted displacement $\langle \Delta R_{\text{acc}} \rangle$ versus R at $\tau = 0.03$ hartree $^{-1}$ for the Langevin and improved matrices ($k = 1.0, 2.0$).

stated differently, all algorithms moved one electron at a time.

For the typical case illustrated in Fig. 3 the accepted displacement $\langle \Delta R_{\text{acc}} \rangle$ inside each shell is plotted for the time step $\tau = 0.03$ hartree $^{-1}$ for the standard and for the proposed matrix with different values of the parameter k . Increasing the value of k increases the mean accepted displacement and the acceptance ratio for $R \leq 0.2$ bohr, R being the distance from the nucleus to the upper limit of the radial shells. For $0.2 \text{ bohr} \leq R \leq 0.6 \text{ bohr}$ the new matrix gives slightly lower accepted displacements than the usual one. This behavior must be attributed to the reduction of the attempted displacement and to a slight reduction of the local acceptance probability, both due to the analytical form of the new matrix. For $R \geq 0.6$ bohr all the matrices are almost equivalent and have similar accepted displacements. In the same way as for the algorithm proposed by Sun et al.

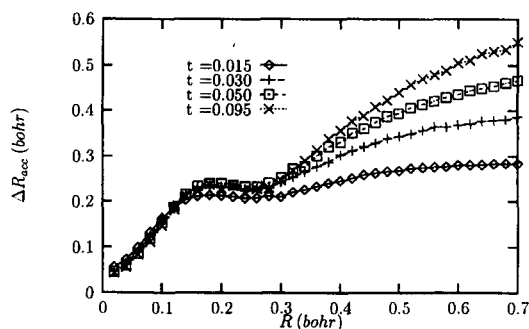


Fig. 4. Mean accepted displacement $\langle \Delta R_{\text{acc}} \rangle$ versus R for the improved matrix ($k = 2.0$), $0.015 \text{ hartree}^{-1} \leq \tau \leq 0.095 \text{ hartree}^{-1}$.

Table 1
Correlation time for neon from the Langevin matrix, $T_{\text{corr}}^{\text{Lang}}$, and proposed matrix, $T_{\text{corr}}^{\text{New}}$, with $k = 2$

τ (hartree ⁻¹)	$T_{\text{corr}}^{\text{New}}$	$T_{\text{corr}}^{\text{Lang}}$
0.007		22.65(6)
0.015	11.61(4)	12.66(6)
0.020	9.06(3)	
0.030	8.62(2)	17.62(8)
0.040	8.16(2)	
0.050	6.70(2)	
0.060	6.61(2)	32.57(9)
0.070	6.65(3)	
0.080	6.18(2)	
0.090	6.17(2)	
0.095	6.48(2)	

[14], an increase in the accepted displacement in the core region is obtained by means of a reduction of the local time step and attempted displacement. In Fig. 4 the mean accepted displacement is plotted for $k = 2$ and various time steps ranging from 0.015 to 0.095 hartree⁻¹. The quantity $\langle \Delta R_{\text{acc}} \rangle$ was found to be relatively insensitive to changes in time step used in the simulation in the region $R \leq 0.15$ bohr. This allows one to use a large time step to speed up the sampling of the valence region without reducing the efficiency of the sampling of the core region. To facilitate a quantitative comparison of the relative efficiency of the two algorithms, correlation times T_{corr} were computed by means of the expression [13]

$$T_{\text{corr}} = \lim_{N_s \rightarrow \infty} N_s \left(\frac{\sigma_b}{\sigma} \right)^2, \quad (10)$$

in which the standard deviation σ is that for individual data, σ_b is that for block averages and the block length is N_s . The correlation times for the Langevin algorithm are listed in Table 1 for different time steps along with those determined for the proposed matrix using $k = 2$. These data show that the proposed matrix is also improved relative to the standard Langevin algorithm in terms of the efficiency. The ratio of the best correlation times for the two algorithms is almost equal to two, and this results in a reduction of the computational effort needed to obtain a given statistical accuracy. Even more important is the insensitivity of the correlation time to the time step size τ and the possibility of using a time step four times larger. In this way a faster sampling

of the valence region can be obtained without the problems associated with electron trapping in the core region. This allows one to obtain a lower serial correlation in the sampling for observables different from the energy and dependent on the distribution in the valence region. We stress that one must be careful to check all possible violations of ergodicity during the simulation. In testing the Langevin algorithm we repeatedly found, using $\tau = 0.06$ hartree⁻¹, a null mean accepted displacement and oscillating electronic distributions for some inner shells. A similar problem was reported for helium clusters by Barnett and Whaley [6] who found oscillations in the sampled particle separation probability due to the magnitude of the quantum force near the coalescence point. Their solution of employing the standard Metropolis method has the drawback that this algorithm diffuses slowly in comparison to its guided alternatives.

4. Conclusions

We have proposed a modified Langevin transition matrix and presented a comparison with the standard matrix. Tests on hydrogen-like atoms show the new matrix to have a smaller time step error than the usual one. The new matrix, due to a reduction of the local time step used in the simulation, permits a more efficient sampling of the core region without slowing down the sampling of the valence properties and without extra computational effort. The easy implementation and the absence of any special conditions on the particle moves are the two major advantages of this algorithm. Finally, this method is applicable to quantum clusters and liquids as well as to atoms and molecules, and can be used in DMC as well as VMC simulations.

Acknowledgement

The authors thank Gabriele Morosi and Alessandro Ponti for careful reading of the manuscript and Dario Bressanini for many useful comments. MM acknowledges financial support by the Università degli Studi di Milano. AL is indebted to the Alexan-

der von Humboldt foundation for financial support by the Feodor-Lynen program. Computer resources for this work have been provided in part by the Centro CNR per lo Studio delle Relazioni tra Struttura e Reattività Chimica.

References

- [1] B.L. Hammond, W.A. Lester Jr. and P.J. Reynolds, Monte Carlo methods in ab initio quantum chemistry (World Scientific, Singapore, 1994).
- [2] K.E. Schmidt and J.W. Moskowitz, *J. Chem. Phys.* 93 (1990) 4172.
- [3] J.W. Moskowitz and K.E. Schmidt, *J. Chem. Phys.* 97 (1992) 3382.
- [4] C.J. Umrigar, K.G. Wilson and J.K. Wilkins, *Phys. Rev. Lett.* 60 (1988) 1719.
- [5] C. Filippi and C.J. Umrigar, *J. Chem. Phys.* 105 (1996) 213.
- [6] R.N. Barnett and K.B. Whaley, *Phys. Rev. A* 47 (1993) 4082.
- [7] A. Mushinski and M.P. Nightingale, *J. Chem. Phys.* 101 (1994) 8831.
- [8] S.W. Rick, D.L. Lynch and J.D. Doll, *J. Chem. Phys.* 95 (1991) 3506.
- [9] A.A. Frost, *J. Chem. Phys.* 10 (1942) 240.
- [10] H. Conroy, *J. Chem. Phys.* 41 (1964) 1327.
- [11] S.A. Alexander, R.L. Coldwell, H.J. Monkhorst and J.D. Morgan III, *J. Chem. Phys.* 95 (1991) 6622.
- [12] D. Bressanini and P.J. Reynolds, private communication.
- [13] C.J. Umrigar, *Phys. Rev. Lett.* 71 (1993) 408.
- [14] Z. Sun, M.M. Soto, R.N. Barnett and W.A. Lester Jr., *J. Chem. Phys.* 101 (1994) 394.
- [15] P. Belohorec, S.M. Rothstein and J. Vrbik, *J. Chem. Phys.* 98 (1993) 6401.
- [16] R.N. Barnett and K.B. Whaley, *J. Chem. Phys.* 96 (1992) 2953.
- [17] P.J. Reynolds, *Int. J. Quantum Chem. Symp.* 24 (1990) 679; G.G. Batrouni and P.J. Reynolds, unpublished; M.M. Soto, Ph. D. Thesis, University of California, Berkeley (1993).
- [18] S.A. Alexander and R.L. Coldwell, *J. Chem. Phys.* 103 (1995) 2572.
- [19] H. Haken, *Z. Phys. B* 24 (1976) 321.
- [20] S.Y. Leu and C.Y. Mou, *J. Chem. Phys.* 101 (1994) 5910.
- [21] A.N. Drozdov, *J. Chem. Phys.* 105 (1996) 515.
- [22] E. Clementi and C. Roetti, *At. Data Nuc. Data Tables* 14 (1974) 177.