No-compromise reptation quantum Monte Carlo

2007 J. Phys. A: Math. Theor. 40 F639

(http://iopscience.iop.org/1751-8121/40/27/F09)

IOP FTC
IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL
J. Phys. A: Math. Theor. **40** (2007) F639–F646 doi:10.1088/1751-8113/40/27/F09

FAST TRACK COMMUNICATION

No-compromise reptation quantum Monte Carlo

Wai Kong Yuen[1], **Thomas J Farrar**[1] **and Stuart M Rothstein**[2]

[1] Department of Mathematics, Brock University, St. Catharines, ON L2S 3A1, Canada
[2] Departments of Chemistry and Physics, Brock University, St. Catharines, ON L2S 3A1, Canada

E-mail: jyuen@brocku.ca and srothste@brocku.ca

Received 3 February 2007, in final form 11 May 2007
Published 20 June 2007
Online at stacks.iop.org/JPhysA/40/F639

Abstract
Since its publication, the reptation quantum Monte Carlo algorithm of Baroni and Moroni (1999 *Phys. Rev. Lett.* **82** 4745) has been applied to several important problems in physics, but its mathematical foundations are not well understood. We show that their algorithm is not of typical Metropolis–Hastings type, and we specify conditions required for the generated Markov chain to be stationary and to converge to the intended distribution. The time-step bias may add up, and in many applications it is only the middle of a reptile that is the most important. Therefore, we propose an alternative, 'no-compromise reptation quantum Monte Carlo' to stabilize the middle of the reptile.

PACS numbers: 02.70.Ss, 31.15.−p, 32.10.Dk

1. Reptation quantum Monte Carlo

In this paper, we show that Baroni and Moroni's [1] (BM), [2] so-called reptation quantum Monte Carlo algorithm is not of typical Metropolis–Hastings (MH) type, and we specify conditions required for the generated Markov chain to be stationary and to converge to the intended distribution:

$$\Pi(X) \propto \Phi_0^2(x_0)\mathcal{W}_\epsilon(x_N|x_{N-1})\cdots\mathcal{W}_\epsilon(x_1|x_0)\exp(-\mathcal{S}[X]). \tag{1}$$

Here, $\mathcal{W}_\epsilon(x_{i+1}|x_i)$ is the transition probability for $x_i \to x_{i+1}$ during the time interval $\tau \to \tau+\epsilon$, X is a time-discretized path generated by the Langevin diffusion: $x_0x_1\cdots x_N$, Φ_0 is an inputted trial function and

$$\mathcal{S}(X) = \epsilon\left(\tfrac{1}{2}\mathcal{E}(x_0) + \mathcal{E}(x_1) + \cdots + \mathcal{E}(x_{N-1}) + \tfrac{1}{2}\mathcal{E}(x_N)\right) \tag{2}$$

is the sum of local energies, $\mathcal{E}$, accumulated over the path X.

Nevertheless, we prove that the Markov chain X^t is convergent to Π by coupling (see, e.g., [3]) X^t with a genuine MH algorithm Z^t with a small modification. As a consequence of

its convergence to Π, the corresponding acceptance probability is

$$\begin{aligned} A[Y, X] &= \min\{1, \Pi(Y)W(X, Y)/\Pi(X)W(Y, X)\} \\ &= \min\{1, \exp(-\mathcal{S}(Y))/\exp(-\mathcal{S}(X))\}, \end{aligned} \tag{3}$$

BM's equation (5). Here the Φ_0^2 factors cancel, leaving just the ratio of exponentials. This assumes microscopic reversibility, formally allowing one to 'move' the point at which Φ_0^2 is evaluated to the same location for reptiles X and Y. More importantly, we show that by relaxing this assumption, we can stabilize the middle of the reptile, thereby somewhat ameliorating the time-step bias when estimating quantities that are sensitive to that region, such as moments of the sampled distribution.

This paper is organized as follows: in section 2 we show that the original reptation quantum Monte Carlo algorithm is irreversible but by considering a similar MH algorithm in section 3, it converges to the intended distribution, section 4. Since the time-step bias may add up and in many applications it is only the middle of a reptile that is the most important, we propose an algorithm, 'no-compromise reptation quantum Monte Carlo' to stabilize the middle of the reptile, section 5. We conclude our paper with numerical examples to demonstrate the power and limitations of this algorithm, section 6.

2. Irreversibility of the original reptation algorithm

The state space of the original algorithm is $\mathbb{R}^{d(N+1)}$ and a state of the algorithm is in the form of a *path* $x_0x_1\cdots x_N$, where $x_k \in \mathbb{R}^d$. The basic version of the algorithm with a fixed M proceeds as follows: let $X = x_0x_1\cdots x_N$ be the current state of the Markov chain and $\overline{X} = x_Nx_{N-1}\cdots x_0$. With probability $1/2$, a new reptile $Y = y_0y_1\cdots y_N$ is generated from X with proposal density given by $W^0(Y, X)$; otherwise, Y is generated from $\overline{X}$ with proposal density given by $W^0(Y, \overline{X})$, where

$$W^0(Y, X) = W_\epsilon(y_N, y_{N-1})W_\epsilon(y_{N-1}, y_{N-2}), \ldots, W_\epsilon(y_{N-M+1}, y_{N-M}), \tag{4}$$

whenever $y_0 = x_M$, $y_1 = x_{M+1}, \ldots, y_{N-M} = x_N$ and 0 otherwise. The new reptile is accepted with probability $A[Y, X]$. We denote the state of this Markov chain at time t by X^t.

This algorithm is irreversible, not a typical MH as claimed in the derivation of BM's equation (5). Figure 1 illustrates its steps (1) and (2), described below BM's equation (7). Given reptiles X, Y such that the proposal density $W(X, Y)$ 0, with 50% probability in case (1b), $W(Y, X) = 0$. In this case, a typical MH algorithm will result in automatic rejection of the proposed move instead of an acceptance with probability $A[Y, X]$ in (3).

As a result, X^t is not convergent in general, unless we make the following assumptions.

- *Micro-reversibility*: for any $x, y \in \mathbb{R}^d$,

$$W_\epsilon(y, x)\Phi_0(x)^2 = W_\epsilon(x, y)\Phi_0(y)^2, \tag{5}$$

  a common assumption that we will relax.
- *Symmetric initial distribution*: X^0 satisfies

$$P(X^0 \in B) = P(\overline{X^0} \in B) \tag{6}$$

  for any measurable set B, i.e., for any arbitrary state X, we have an equal probability density of starting the algorithm from X and $\overline{X}$.

Without reversibility, typical approaches to prove the convergence of X^t to Π break down. Instead, we compare it to a similar MH algorithm.

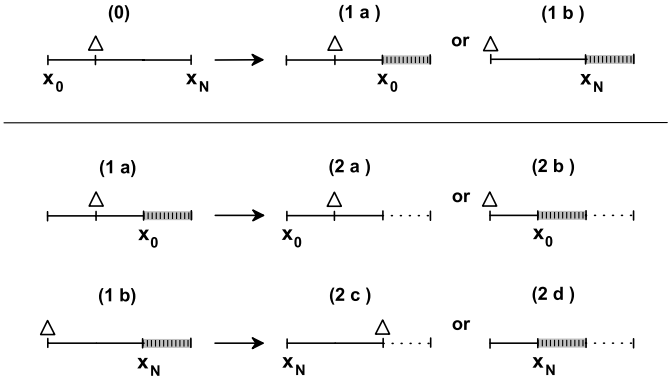

Figure 1. Irreversibility of BM's algorithm. Δ is at the $(M+1)$ st position on the initial path (0). Shaded and dotted segments are sequences of length M. These are added to the end of the path, after first time-reversing the path with 50% probability, and then deleting segments from its head. Top: X $\to$ Y. Bottom: Y $\to$ X and Y $\to \bar{\text{X}}$, where $\bar{\text{X}}$ is X with time-reversed ordering of configurations. Neither (1b) $\to$ (2c) nor (1b) $\to$ (2d) recovers X with positive probability.

3. Metropolis–Hastings algorithm

We consider a reversible algorithm Z^t that we denote as Metropolis–Hastings reptation quantum Monte Carlo, which is very similar to X^t.

This algorithm proceeds as follows: let $Z = z_0 z_1 \cdots z_N$ be the current state of the Markov chain. With probability $1/2$, a new reptile $Y = y_0 y_1 \cdots y_N$ is generated from Z with proposal density given by $W^0(Y, Z)$ from (4), whenever $y_0 = z_M, y_1 = z_{M+1}, \ldots, y_{N-M} = z_N$; otherwise, $Y = y_0 y_1 \cdots y_N$ is generated from Z with proposal density given by $W^1(Y, Z)$, where

$$W^1(Y, Z) = W_\epsilon(y_0, y_1) W_\epsilon(y_1, y_2), \ldots, W_\epsilon(y_{M-1}, y_M), \tag{7}$$

whenever $y_M = z_0, y_{M+1} = z_1, \ldots, y_N = z_{N-M}$. Then the new reptile is accepted with probability $A[Y, Z]$ from (3) again.

Note that the proposed moves of X^t and Z^t are identical if they both move forward with W^0. If the proposed moves are backward, X^t will 'flip' the reptile and add to the tail, i.e., it is adding to the head of the original reptile; whereas Z^t will simply add to the head with the same distribution without flipping the reptile. However, it is this subtle flipping that removes the reversibility of the algorithm X^t, figure 1. On the other hand, Z^t is a *genuine* MH algorithm reversible with respect to Π, i.e., the acceptance probability for a proposed move in the MH algorithm targeting Π can be simplified to $A[Y, Z]$. The key observation is that for all reptiles Y, Z,

$$W^0(Y, Z) > 0 \quad \Leftrightarrow \quad W^1(Z, Y) > 0,$$

eliminating the 'automatic rejections' in the original algorithm. With this property and micro-reversibility (5), it is routine to check that the acceptance probabilities for the proposed moves in both directions are in fact identical to $A[Y, Z]$.

Finally, Z^t is clearly Π-irreducible and aperiodic and hence converges to Π in a total variation distance from any initial value of the chain (see, e.g., [4]).

4. Convergence of the original reptation algorithm

Now to prove that X^t converges to Π as well, observe that we can couple X^t and Z^t together to form a joint Markov chain (X^t, Z^t) on $\mathbb{R}^{d(N+1)} \times \mathbb{R}^{d(N+1)}$ such that $X^0 = Z^0$ with the crucial property that

$$X^t = Z^t \quad \text{or} \quad X^t = \overline{Z^t}$$

for all t. Intuitively, such a coupling is possible because the only difference between the two chains is the subtle 'flipping'. This construction is routine but somewhat technical (see [5]). Besides, the symmetry of both algorithms and (6) imply

$$P(X^t \in B) = P(\overline{X^t} \in B) \qquad \text{and} \qquad P(Z^t \in B) = P(\overline{Z^t} \in B)$$

for any measurable B. So,

$$\begin{aligned}
P(X^t \in B) &= \tfrac{1}{2}[P(X^t \in B) + P(\overline{X^t} \in B)] \\
&= \tfrac{1}{2}[P(X^t \in B | X^t = Z^t)P(X^t = Z^t) + P(X^t \in B | X^t = \overline{Z^t})P(X^t = \overline{Z^t}) \\
&\quad + P(\overline{X^t} \in B | X^t = Z^t)P(X^t = Z^t) + P(X^t \in \overline{B} | X^t = \overline{Z^t})P(X^t = \overline{Z^t})] \\
&= \tfrac{1}{2}[P(Z^t \in B | X^t = Z^t)P(X^t = Z^t) + P(\overline{Z^t} \in B | X^t = \overline{Z^t})P(X^t = \overline{Z^t}) \\
&\quad + P(\overline{Z^t} \in B | X^t = Z^t)P(X^t = Z^t) + P(Z^t \in B | X^t = \overline{Z^t})P(X^t = \overline{Z^t})] \\
&= \tfrac{1}{2}[P(Z^t \in B) + P(\overline{Z^t} \in B)] \\
&= P(Z^t \in B).
\end{aligned}$$

Since Z^t converges to Π in a total variation distance, so does X^t.

Remarks

(1) In the original algorithm, the authors assumed (5) but did not state assumption (6) explicitly. However, it is pointed out that one must start the algorithm by first obtaining a sample path from the stationary Langevin diffusion. As a result, X^0 has a density proportional to $L_N(X)$, so that (6) is satisfied. Without this initialization step, however, X^t may not converge to Π. On the other hand, Z^t converges to Π from *any* starting point X^0. From a practical point of view, the algorithm Z^t will provide extra stability on the convergence as we only have $L_N(X) \approx L_N(\overline{X})$.

(2) Suppose that our goal is to estimate $E(f(X))$ where X is distributed as Π for some function f with the property that $f(X) = f(\overline{X})$ for all X. Then the two algorithms X^t and Z^t are essentially identical as the 'flipping' does not affect the value of f.

(3) Another approach to prove the convergence of a Markov chain is to find an invariant distribution but it can be shown that Π is only an invariant distribution approximately with respect to X^t; see [5] for details. If the initial sampled reptile X^0 is an outlier, there is a potential stability problem.

5. No-compromise algorithm

Since the time-step bias may add up and it is only the middle of a reptile that is the most important, we propose the *no-compromise* algorithm to stabilize the middle of the reptile. First, we assume that N is even and $M \leqslant N/2$ to simplify the algorithm. It is artificial to have even-length reptiles and experience suggests that for the algorithm to be efficient, M is usually much less than $N/2$. To improve our samples, we use the original MH algorithm to generate our samples *without* assuming micro-reversibility (5). However, if we use Π in the simulation as it is, we can only stabilize the head of the reptiles and the time-step bias may

still add up near the middle of the reptiles. For this reason, we use the original MH algorithm for the distribution:

$$\hat{\Pi}(X) \propto W_\epsilon(x_N, x_{N-1})W_\epsilon(x_{N-1}, x_{N-2}), \ldots, W_\epsilon\left(x_{\frac{N}{2}+1}, x_{\frac{N}{2}}\right)\Phi_0\left(x_{\frac{N}{2}}\right)^2 \times W_\epsilon\left(x_{\frac{N}{2}-1}, x_{\frac{N}{2}}\right)W_\epsilon\left(x_{\frac{N}{2}-2}, x_{\frac{N}{2}-1}\right), \ldots, W_\epsilon(x_0, x_1)\,\mathrm{e}^{-S[X]}.$$

Clearly, this distribution is identical to Π if we assume micro-reversibility (5). On the other hand, if the time-step bias is substantial, $\hat{\Pi}$ will guarantee that the marginal density of the *centre* is proportional to the desired density of the limiting Langevin diffusion multiplied by $\mathrm{e}^{-S[X]}$. The proposed moves are identical to those for Z^t but we must recalculate the acceptance probabilities.

This new algorithm proceeds as follows: let Z be the current state of the Markov chain. With probability $1/2$, a new reptile Y is generated *forward* from Z with proposal density given by $W^0(Y, Z)$ from (4) and accepted with probability

$$A^0[Y, Z] = \min\left\{1, \frac{W_\epsilon\left(z_{\frac{N}{2}}, z_{\frac{N}{2}+1}\right), \ldots, W_\epsilon\left(z_{\frac{N}{2}+M-1}, z_{\frac{N}{2}+M}\right)\Phi_0\left(z_{\frac{N}{2}+M}\right)^2\mathrm{e}^{-S[Y]}}{W_\epsilon\left(z_{\frac{N}{2}+M}, z_{\frac{N}{2}+M-1}\right), \ldots, W_\epsilon\left(z_{\frac{N}{2}+1}, z_{\frac{N}{2}}\right)\Phi_0\left(z_{\frac{N}{2}}\right)^2\mathrm{e}^{-S[Z]}}\right\}; \tag{8}$$

otherwise, Y is generated from Z *backward* with proposal density given by $W^1(Y, Z)$ from (7) and accepted with probability

$$A^1[Y, Z] = \min\left\{1, \frac{W_\epsilon\left(z_{\frac{N}{2}}, z_{\frac{N}{2}-1}\right), \ldots, W_\epsilon\left(z_{\frac{N}{2}-M+1}, z_{\frac{N}{2}-M}\right)\Phi_0\left(z_{\frac{N}{2}-M}\right)^2\mathrm{e}^{-S[Y]}}{W_\epsilon\left(z_{\frac{N}{2}-M}, z_{\frac{N}{2}-M+1}\right), \ldots, W_\epsilon\left(z_{\frac{N}{2}-1}, z_{\frac{N}{2}}\right)\Phi_0\left(z_{\frac{N}{2}}\right)^2\mathrm{e}^{-S[Z]}}\right\}. \tag{9}$$

These calculations are similar to that for A in section 3. In the case of A^0, the factors concerning the points between $z_{\frac{N}{2}}$ and $z_{\frac{N}{2}+M}$ can no longer be cancelled, due to the change of direction. Therefore, A^0 (and similarly A^1) cannot be reduced to A (3). We also note that if $M > N/2$, A^0 and A^1 will become more complicated as there is an additional change of direction.

6. Numerical examples

We modified the original reptation quantum Monte Carlo algorithm (RQMC) to be Metropolis–Hastings type (RQMC-MH), section 3, and then one to stabilize the middle of the reptile, the no-compromise version (RQMC-NC), section 5. In this section, we estimate moments of the electron-nuclear distribution for the ground-state hydrogen atom and compare the results obtained for this system using the three algorithms. We account for the *time-step* bias in the simulated data by fitting the biased estimates to a polynomial model. Atomic units are used throughout.

Independent runs were performed for each of the following trial functions, $\psi = \exp(-\alpha r)$, where $\alpha = 1.0$, 0.9 and 0.4 and r is the distance of the electron from the nucleus. The first trial function corresponds to the exact ground-state wavefunction, the second is approximate, but of good quality, and the third is of poor quality.

Algorithmic parameters were set as follows.

- The selected time steps are $\epsilon = 0.10 \ldots 0.30[0.05]$ au.
- The ensemble size is given by $100\left(\frac{\epsilon_0}{\epsilon}\right)$, where ϵ_0 is the largest time-step value (see, e.g., [6]).
- The reptile length is given by $N = \mathrm{int}\left[80\left(\frac{\epsilon_0}{\epsilon}\right)^{1.5}\right]$ (see, e.g., [6]).
- $M = 30$. In this simple model, since all three algorithms converge very quickly, there is no need to randomize the choice of M as in the original algorithm.

Table 1. Moments of the electron-nuclear distance for the hydrogen atom ground state estimated from the reptation quantum Monte Carlo algorithms: original [1] (RQMC), Metropolis–Hastings (RQMC-MH) and no-compromise (RQMC-NC) versions. Quoted are the intercepts and standard errors from polynomial fits of property versus time step: linear (L), quadratic (Q) or cubic (C). The trial functions are in the form $\psi = \exp(-\alpha r)$, where r is in bohr. Also quoted is their overlap with the exact wavefunction (S), their variational energy ($E_{\rm var}$) and the expected value of moment in their variational distribution: ψ^2 (variational). All entries are given in atomic units. Parentheses denote one standard error in the last digit recorded.

Algorithm	$\langle r \rangle$	$\langle r^2 \rangle$	$\langle r^3 \rangle$	$\langle 1/r \rangle$
$\alpha = 1.0$ (exact)	$S = 1.000$	$E_{\rm var} = -0.5$		
RQMC	1.5023(2)Q	3.0115(72)Q	7.573(62)Q	0.9877(26)Q
RQMC–MH	1.5028(43)Q	3.0126(196)Q	7.558(74)Q	0.9849(10)C
RQMC–NC	1.5008(3)L	3.0023(21)L	7.505(14)L	0.9994(15)Q
Variational	1.5	3.0	7.5	1.0
$\alpha = 0.9$ (good quality)	$S = 0.996$	$E_{\rm var} = -0.405$		
RQMC	1.5012(5)Q	3.0014(36)Q	7.4994(266)Q	0.9883(7)Q
RQMC–MH	1.5053(18)Q	3.0211(132)Q	7.5800(779)Q	0.9899(37)Q
RQMC–NC	1.5093(129)C	3.0284(262)Q	7.6020(74)L	0.9911(62)L
Variational	1.667	3.704	10.289	0.900
$\alpha = 0.4$ (crude)	$S = 0.738$	$E_{\rm var} = -0.080$		
RQMC	1.506(45)Q	3.014(166)Q	7.50(61)Q	1.054(142)Q
RQMC–MH	1.497(17)Q	2.989(80)Q	7.44(35)Q	1.069(44)Q
RQMC–NC	1.509(15)L	3.030(47)L	7.59(15)L	0.951(99)L
Variational	18.75	468.75	14 648.0	0.080

- To control the drift velocity, $\mathbf{F} \equiv \nabla\psi/\psi$, that can push the electron too far from the region of reasonable probability, we truncate the velocity components [7]:

$$F_i = \begin{cases} F_i & \text{if } |F_i| \leqslant 1/\epsilon, \\ \text{sign}[1/\epsilon, F_i] & \text{otherwise.} \end{cases} \tag{10}$$

  This truncation vanishes in the $\epsilon \to 0$ limit.

To generate an initial distribution for the no-compromise runs, reptiles were generated with the Metropolis–Hastings algorithm for several iterations, acceptance probability A (3), prior to implementing the no-compromise scheme, acceptance probability A^0 (8) or A^1 (9).

The following moments were estimated at each time step: $\langle r^n \rangle$, $n = \pm 1$, 2 and 3. As the hydrogen ground state is a nodeless system, in principle, each algorithm will recover the exact value of the property, independent of α, in the limit of zero-time-step.

Table 1 shows the results from fitting the simulated properties by weighted polynomial regression in ϵ. We account for *model-bias* by selecting the lowest-order polynomial model whose intercept $\pm$ two standard errors overlaps with that of the next higher-order polynomial.

The estimated moments from the no-compromise algorithm (RQMC-NC) are almost always (see below) within the statistical error of the exact values, even for the crude trial function, whose variational positive moments are incorrect by one to three orders of magnitude! Note that, with two exceptions, the no-compromise data are consistently fit by a lower-degree polynomial than those for the other algorithms, normally with a corresponding gain in precision. Naturally, the most precise results are generally obtained by employing the most accurate wavefunctions, as is the case for competing 'exact sampling' algorithms (see, e.g., [8]).

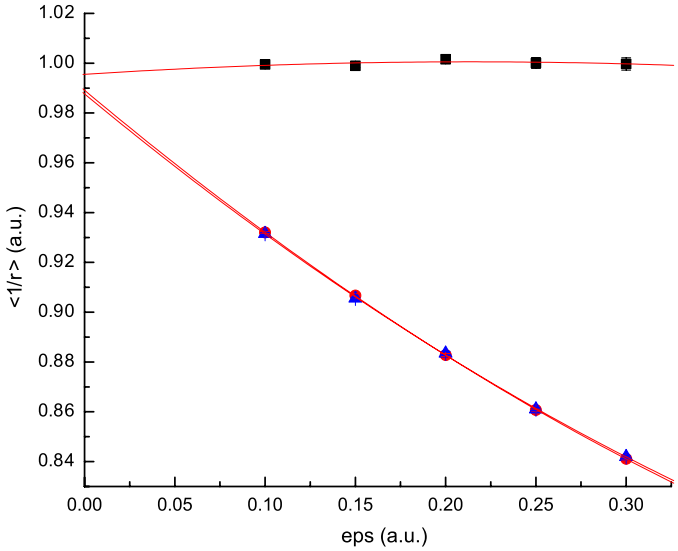

Figure 2. Hydrogen atom ground state: simulated inverse first moment of the electron-nuclear distance for the exact wavefunction. Upper curve: no-compromise reptation quantum Monte Carlo. Overlapping lower curves: reptation quantum Monte Carlo [1] and Metropolis–Hastings reptation quantum Monte Carlo. The error bars are smaller than the size of the symbols plotted.
(This figure is in colour only in the electronic version)

The no-compromise estimate of $\langle r^3 \rangle$ for the good-quality wavefunction ($\alpha = 0.9$) is not accurate. Here the data have oscillations that cannot be well-fit by a polynomial model. This same behaviour has occasionally been observed in previous attempts to reduce the time-step bias in quantum Monte Carlo simulations [7]. In practice, this is not a severe disadvantage, as often simulations are performed at sufficiently small values of the time step that the bias can be ignored relative to the statistical error, forgoing the need to extrapolate the data to zero-time-step.

As is expected, BM's original algorithm (RQMC) and our Metropolis–Hastings variant (RQMC-MH) give similar results. Notably, both fail to correctly estimate the inverse moment for the exact and good-quality wavefunctions. Here, in marked contrast to no-compromise, the time-step bias is so large that one must extrapolate well beyond the range of the simulated data, giving an unreliable estimate for the intercept, e.g., figure 2. In this case, one would need to do simulations at smaller values of ϵ to get agreement with the true value.

Acknowledgments

This work was supported, in part, by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] Baroni S and Moroni S 1999 *Phys. Rev. Lett.* **82** 4745
[2] Baroni S and Moroni S 1999 *Quantum Monte Carlo Methods in Physics and Chemistry* ed M P Nightingale and C J Umrigar (Dordrecht: Kluwer)

[3] Lindvall T 1992 *Lectures on the Coupling Method* (Mineola, NY: Dover)
[4] Tierney L 1994 *Ann. Stat.* **22** 1701
[5] Yuen W, Farrar T and Rothstein S M 2007 unpublished manuscript, available at http://spartan.ac.brocku.ca/~jyuen/preprints/rep_proof.pdf
[6] Langfelder P, Rothstein S M and Vrbik J 1997 *J. Chem. Phys.* **107** 8525
[7] DePasquale M F, Rothstein S M and Vrbik J 1988 *J. Chem. Phys.* **89** 3629
[8] Bosá I and Rothstein S M 2004 *J. Chem. Phys.* **121** 4486