JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

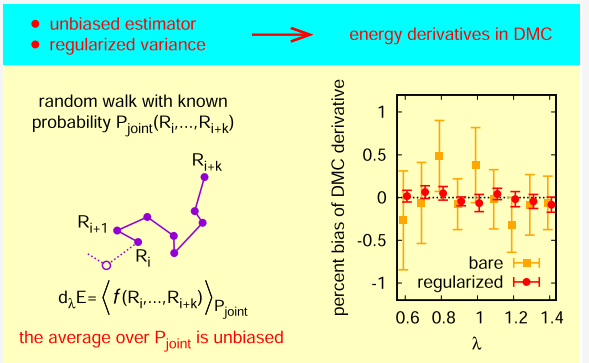
Energy Derivatives in Real-Space Diffusion Monte Carlo

Jesse van Rhijn,* Claudia Filippi,* Stefania De Palo,* and Saverio Moroni*

Cite This: J. Chem. Theory Comput. 2022, 18, 118−123 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We present unbiased, finite-variance estimators of energy derivatives for real-space diffusion Monte Carlo calculations within the fixed-node approximation. The derivative $d_\lambda E$ is fully consistent with the dependence $E(\lambda)$ of the energy computed with the same time step. We address the issue of the divergent variance of derivatives related to variations of the nodes of the wave function both by using a regularization for wave function parameter gradients recently proposed in variational Monte Carlo and by introducing a regularization based on a coordinate transformation. The essence of the divergent variance problem is distilled into a particle-in-a-box toy model, where we demonstrate the algorithm.



1. INTRODUCTION

Variational Monte Carlo (VMC) and diffusion Monte Carlo (DMC) are numerical stochastic approaches based on a real-space representation of a correlated trial wave function to study many-body quantum systems, including electronic structure problems. VMC calculates expectation values of quantum operators on the trial function, which in turn is optimized via minimization of a suitable cost function such as the variational energy. DMC further improves the VMC results through a stochastic implementation of the power method, which projects the lowest-energy component of the trial function. Its accuracy, in the fixed-node (FN) approximation almost invariably adopted to avoid the sign problem, is ultimately limited by the error in the nodal surface of the trial function.[1]

In the past decade, the efficient calculation of analytic energy derivatives,[2,3] leveraging modern optimization methods,[4−6] spawned impressive progress in both accuracy and scope of the VMC method.[7−10]

DMC largely benefits from advances in VMC because improved trial functions tend to have better nodes. However, it would be desirable to have efficient and unbiased estimators of derivatives in DMC as well to perform such tasks as the direct optimization of the nodal surface or DMC structural relaxation. This is still an open issue, with the latest developments featuring uncontrolled approximations and/or a very low efficiency.[11−13] Here, we present an algorithm to calculate unbiased energy derivatives in FN-DMC with finite variance and demonstrate it on a model where the effect of the nodes is dominant. In the Supporting Information, we also include an application to the lithium dimer.

2. ENERGY DERIVATIVES

In both VMC and DMC, the energy is calculated as

$$E = \int P(R)E_L(R)\mathrm{d}R / \int P(R)\mathrm{d}R \equiv \langle E_L \rangle_P \tag{1}$$

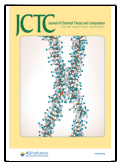
where R represents the coordinates of all the particles, $E_L(R) = H\Psi(R)/\Psi(R)$ is the local energy of the trial function $\Psi(R)$, and $P(R)$ is proportional to the underlying probability distribution: in VMC, $P(R) = \Psi^2(R)$, and in DMC, $P(R) = \Psi(R)\Phi(R)$ with $\Phi(R)$ being the FN solution. The derivative with respect to a parameter λ is

$$d_\lambda E = \langle d_\lambda E_L + (E_L - E) d_\lambda \ln P \rangle_P \tag{2}$$

The variance of this naïve estimator is 0 if both Ψ and its derivative $d_\lambda\Psi$ are exact; however, for an approximate trial function, $E_L(R)$ diverges at the nodes as $1/d(R)$, where $d = |\Psi|/\|\nabla\Psi\|$, and the variance diverges as well.[14,15]

2.1. Regularized Estimators. In VMC, this problem was fully solved by Attaccalite and Sorella[14] with a reweighting scheme, hereafter dubbed AS, whereby one samples the square of a modified trial function $\tilde{\Psi}$ which differs from Ψ only for d smaller than a cutoff parameter ϵ and stays finite on the nodal surface of Ψ. A similar sampling scheme was proposed by Trail.[16] The AS estimator has the same average of the bare estimator of eq 2 for any value of ϵ and finite variance.

Received: May 19, 2021
Published: December 20, 2021



An alternative regularized estimator, recently proposed by Pathak and Wagner,[15] simply consists in multiplying the term in brackets of eq 2 by the polynomial $f_\epsilon(x) = 7x^6 - 15x^4 + 9x^2$, with $x = \mathrm{d}/\epsilon$, whenever $x < 1$. This estimator, hereafter dubbed PW, has finite variance for any finite ϵ and a bias which vanishes as $\epsilon \to 0$. The polynomial is chosen in such a way to remove from the bias the linear term, shown[15] to be $\propto \int_0^1 (f_\epsilon - 1)\mathrm{d}x$, and to be continuously differentiable in (0, 1). The odd parity of fermionic wave functions near a node then implies a cubic leading term in the bias. Since various values of ϵ can be used in the same simulation, the bias can be eliminated at no cost by extrapolation. The PW estimator has been proposed for parameter gradients in the VMC optimization of Ψ, but it is equally applicable to VMC interatomic forces and, as will be shown, to generic derivatives in DMC.

We introduce a third regularized estimator, which we denote as "warp" by analogy with the space−warp transformation of ref 17 devised to reduce the statistical noise of the forces. There, as a nucleus is displaced, a transformation is applied to the coordinates of nearby electrons, in such a way to maintain the electron−nucleus distances approximately constant. Here, the goal is to maintain constant the value of $\mathrm{d}(R)$ when the nodal surface is displaced by a variation of the parameter λ. In this way, the diverging term in the local energy does not change, and the variance of the derivative is finite.

The warp transformation, illustrated in Figure 1, is defined as

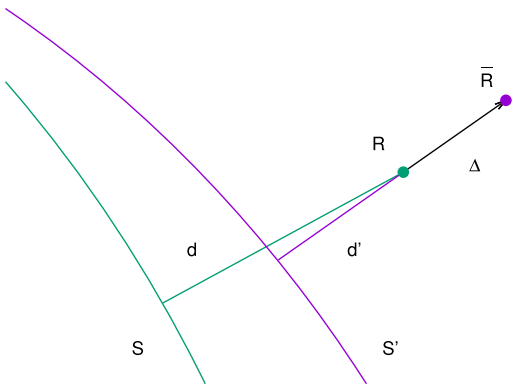


Figure 1. Schematic picture of the warp transformation. The curve S is the nodal surface for the value λ_0 of the parameter λ. The value of d for the current configuration R is pictorially represented as the distance from S (this is strictly true only if Ψ is linear, i.e., close enough to the node). When a variation of λ from λ_0 to λ' shifts S to S', the value of d changes to d′. Hence, we displace R by an amount $\Delta = \mathrm{d} - \mathrm{d}'$ to $\bar{R}$ in the direction of $\mathrm{sign}(\Psi')\nabla\Psi'(R)$ so that the value of $|\Psi'(\bar{R})|/\|\nabla\Psi'(\bar{R})\|$ is approximately equal to d.

$$\bar{R} = R + [\mathrm{d}(R) - \mathrm{d}'(R)]\mathrm{sign}(\Psi'(R))n'(R)u(\mathrm{d}(R)) \quad (3)$$

where primed quantities are calculated for the value λ' of the parameter, n' is the unit vector in the direction of $\nabla\Psi'(R)$, and $u(\mathrm{d})$ is a cutoff function with support $[0, \epsilon]$ which decreases smoothly from 1 to 0, restricting the warp transformation to a region close to the nodal surface. We use the quintic polynomial with zero first and second derivatives at the boundaries of the support.

For a finite increment of λ, the energy is

$$E' = \int E'_{\mathrm{L}}(\bar{R})P'(\bar{R})J\ \mathrm{d}R \Big/ \int P'(\bar{R})J\ \mathrm{d}R \quad (4)$$

where $J = \det J_{ij} = \det \partial\bar{R}_i/\partial R_j$ is the Jacobian of the transformation. For the practical implementation, we formulate a differential expression of the finite energy difference $E' - E$ so that no actual transformation is needed: the analytic derivative $\mathrm{d}_\lambda E$, calculated at the value of the parameter $\lambda = \lambda_0$, is

$$\mathrm{d}_\lambda E|_{\lambda_0} = \langle \partial_\lambda E_{\mathrm{L}} + \nabla E_{\mathrm{L}}\cdot\partial_\lambda\bar{R} + (E_{\mathrm{L}} - E)[\partial_\lambda \ln(PJ) + \nabla \ln P\cdot\partial_\lambda\bar{R}]\rangle_P \quad (5)$$

This warp regularized estimator has finite variance and no bias for any value of ϵ (see the Supporting Information).

Note that all the functions in eq 5 are evaluated at $\lambda = \lambda_0$, where $\bar{R} = R$ and $J = 1$. Therefore, the warp transformation only contributes to the estimator through the derivatives $\partial_\lambda\bar{R}_i|_{\lambda_0}$ and $\partial_\lambda \ln J|_{\lambda_0}$, while the sampling is done over one and the same distribution $P(R)$ for any parameter we may vary.

Furthermore, for $\lambda = \lambda_0$, the cofactors of the Jacobi matrix J_{ij} are $M_{ij} = \delta_{ij}$, and the seemingly awkward derivative of the Jacobian greatly simplifies, $\partial_\lambda \ln J|_{\lambda_0} = \sum_{ij} M_{ij}\partial_\lambda J_{ij}|_{\lambda_0} = \sum_i \partial_\lambda J_{ii}|_{\lambda_0}$, so that the implementation of eq 5 is not overly complicated. In particular, most of the derivatives needed are already present—or very similar to those already present—in VMC codes with analytic derivatives for structural and full variational optimization. The only exceptions are the off-diagonal components of the Hessian $\partial^2\Psi/\partial R_i\partial R_j$ and their derivatives with respect to λ, which contribute to $\partial_\lambda J$. These extra derivatives are only needed when the nodal distance is smaller than the cutoff ϵ, that is, for a very small fraction of the sampled configurations. Furthermore, we will show (heuristically) that the bias incurred by neglecting these terms can be extrapolated out at no cost.

2.2. Variational Monte Carlo. Before addressing the derivatives in DMC, we compare the three regularized estimators PW, AS, and warp in VMC. To this purpose, it is expedient to consider a system stripped of all complexities of external and interparticle potentials so that we can focus exclusively on the divergence of the local energy at the nodal surface. Our toy model is a free particle in an elliptic box with hard walls, meant to represent the configuration of a generic system within a nodal pocket. Atomic units are used throughout. We choose $\Psi = \Psi_0(x,y) = a^2 - x^2/C - y^2/(C - 1)$ with $C = [\cosh(1)]^2 = 2.3810978$, which is positive inside the ellipse, vanishes at the border, and is not the true ground state. Therefore, the ellipse is defined through the wave function, and we take the derivative with respect to the parameter a, which changes the size of the ellipse at constant eccentricity.

The average and variance of the various estimators, calculated by quadrature, are shown in Figure 2. None of the regularized estimators entails uncontrolled approximations. In particular, although there is no rigorous way to establish the range where the leading correction in ϵ is sufficient, or the number of powers in ϵ needed over an arbitrary range, PW can be accurately extrapolated to the unbiased result. Here, the bias of the PW estimator has a leading contribution of ϵ^2 because Ψ_0 does not have odd parity across the node. The second-order bias can be removed with a different choice of the polynomial, for example, $\tilde{f}_\epsilon(x) = 60x^2 - 200x^3 + 225x^4 - 84x^5$, at the expense of a larger variance for the given value of the cutoff ϵ. The right panel shows that the relative efficiency may

https://doi.org/10.1021/acs.jctc.1c00496
J. Chem. Theory Comput. 2022, 18, 118−123

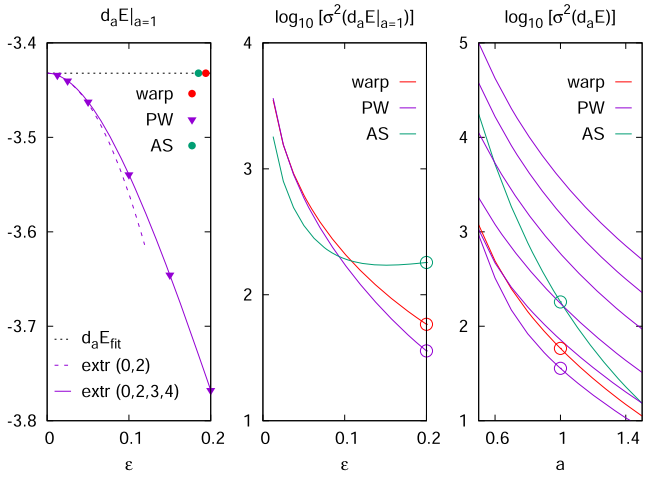


Figure 2. VMC calculations of d_aE, with integrals performed by quadrature. Left panel: derivative at $a = 1$ obtained with the warp, PW, and AS estimators as a function of the cutoff ϵ, compared to the "exact" result d_aE_{fit}, defined as the derivative of a fit to $E(a)$ calculated separately for several values of a. AS and warp, shown here only near $\epsilon = 0.2$, are unbiased; PW can be extrapolated to the unbiased result either including only the leading term in ϵ, here ϵ^2, on a sufficiently small range (dashed line) or using a sufficiently large number of terms on an extended range (solid line). Middle panel: variance (in a logarithmic scale) of the various regularized estimators as a function of the cutoff. As ϵ vanishes, all schemes regress to the infinite variance of the bare estimator. Right panel: variance (in a logarithmic scale) of the various regularized estimators as a function of the parameter a. For the warp and AS estimators, $\epsilon = 0.2$; for PW, data are reported for $\epsilon = 0.0125$, 0.025, 0.05, 0.15, and 0.2 in the order of decreasing variance. Common data in the middle and the right panels are circled.

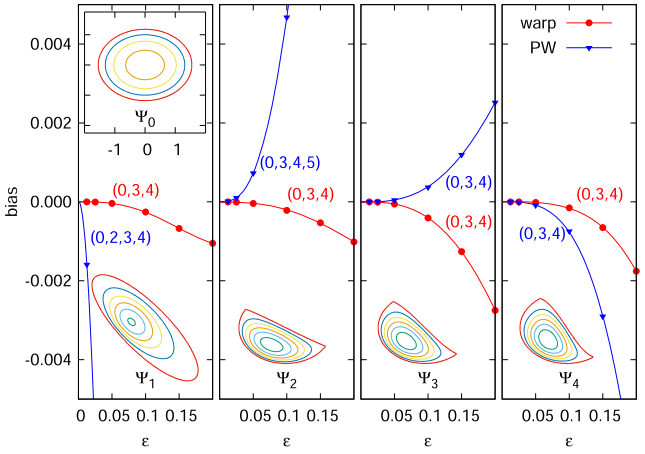


Figure 3. Bias of the VMC derivative for PW and warp with approximate Hessian, calculated by quadrature for various wave functions Ψ_1–Ψ_4. In each case, we show a contour plot of the normalized Ψ_i with level lines from 0 in steps of 0.2 (the contour plot of Ψ_0 in the top left inset defines the (x, y) scale). The colored labels near each curve indicate the powers in ϵ needed to extrapolate to the unbiased value with a five-digit accuracy. The bias of PW for Ψ_2–Ψ_4 has a leading term ϵ^3 because of the use of the modified polynomial $\tilde{f}_\epsilon$. Empirically, we see that the warp estimator has a cubic leading term in all cases.

depend on the system at hand; in this example, it varies over the range of a considered.

The bias of the warp estimator when the off-diagonal elements of the Hessian are neglected is compared to the bias of the PW estimator in Figure 3. Since we need a non-diagonal Hessian to start with, we consider (i) a rotated, more eccentric ellipse with a further non-symmetrical distortion of the wave function, and (ii–iv) the positive lobe of a wave function limited to half of the original ellipse by the nodal line $y + \sin(\alpha x) = 0$, with $\alpha = 0.5$, 1, and 1.5 (see the contour plots in Figure 3). The derivative is taken with respect to a for Ψ_1 and with respect to α for Ψ_2–Ψ_4. Within this (very limited) set of test cases, the bias is smaller and less system-dependent for the warp than for the PW estimator. The important result is that neither involves uncontrolled approximations as both can be extrapolated to the unbiased result in a single run. We have also verified that the warp estimator with the full Hessian is unbiased for finite ϵ.

2.3. Diffusion Monte Carlo. We now consider the derivative in DMC. The FN-DMC algorithm is a branching random walk of many weighted walkers, generated by a short-time approximation $G(R',R)$ to the importance-sampled Green's function, which asymptotically samples the distribution $P(R) = \Psi(R)\Phi(R)$.[1] The problem with the derivative estimator, eq 2, is the presence of the logarithmic derivative of $P(R)$, which is not a known function of R. However, P is the marginal distribution of the joint probability density P_{joint} of the whole random walk, which does have an explicit expression as a product of Green's functions,

$$P(R_n) = \int dR_0 \ldots dR_{n-1} P_{joint}(R_n, R_{n-1} \ldots, R_0) \equiv \int dR_0 \ldots dR_{n-1} \prod_{i=0}^{n-1} G(R_{i+1}, R_i) P_0(R_0) \quad (6)$$

where P_0 is the (largely arbitrary) probability distribution of the initial configuration R_0. Therefore, it is sufficient to consider the estimator of $d_\lambda E$ in eq 2 as an average over the whole trajectory of the random walk, rather than over the current configuration, to bring an explicitly known probability distribution to the fore.[13] This is similar to the calculation of forces in path integral Monte Carlo.[18]

In practice, the inclusion of the entire trajectory in the estimator is not necessary. As shown in ref 13, the logarithmic derivative of the DMC density distribution P in the estimator of eq 2 can be replaced with the summation

$$d_\lambda \ln P(R_n) = d_\lambda \sum_{i=n-k}^{n-1} \ln G(R_i', R_i) \quad (7)$$

over the last k steps of the random walk, with R_n being the current configuration R of eq 2. The omitted term,[13] $\langle [E_L(R_n) - E] d_\lambda \ln P(R_{n-k}) \rangle$, vanishes for sufficiently large k because $E_L(R_n)$ and $d_\lambda \ln P(R_{n-k})$ become statistically independent variables and $\langle E_L - E \rangle = 0$.

In eq 7, $G(R_i',R_i)$ is the transition rule from R_i to R_i' of the random walk. It includes a Metropolis test to reduce the time-step error;[1] therefore, R_i' is the configuration proposed when the walker is at R_i, and the next configuration R_{i+1} is R_i' or R_i if the move is accepted or rejected, respectively. Note that in the formal expression of P_{joint}, eq 6, the arguments of Green's functions are integrated over, whereas in the contribution to the estimator of the derivative, eq 7, they are the particular values of the particles' coordinates effectively sampled by the random walk. Correspondingly, the actual value taken by $G(R', R)$ is

https://doi.org/10.1021/acs.jctc.1c00496
J. Chem. Theory Comput. 2022, 18, 118–123

$$\begin{cases} T(R', R)p(R', R)W(R', R) & \text{for an accepted move} \\ T(R', R)[1 - p(R', R)]W(R, R) & \text{for a rejected move} \end{cases} \tag{8}$$

where T is the a priori transition probability, p is the probability of accepting the move, and W is the branching factor (see below).

The inclusion of rejected configurations in eq 7 and of the factor p or $1 - p$ in eq 8 in the derivative of the full Green's function is instrumental to obtain an estimate of $d_\lambda E$ completely consistent with the DMC energy $E(\lambda)$ calculated at the same time step. Their omission still can give an unbiased result in the limit $\tau \to 0$, but it may cause an unacceptably large time step error on the derivative.[13]

The functions T, p, and W are standard[19]

$$\begin{aligned} T(R', R) &= \exp\{-[R' - R - F(R)V(R)\tau]^2/2\tau\} \\ p(R', R) &= \begin{cases} 0 & \text{if the move crosses a node} \\ \min\left\{1, \dfrac{\Psi^2(R')T(R, R')}{\Psi^2(R)T(R', R)}\right\} & \text{otherwise} \end{cases} \\ W(R', R) &= \exp\{[S(R') + S(R)]\tau/2\} \end{aligned} \tag{9}$$

Here, τ is the time step, $V = \nabla \ln \Psi$ is the so-called velocity, and $F(R) = \sqrt{2V^2\tau - 1}/(V^2\tau)$ is the damping factor of its divergence near the nodes; the logarithm of the branching factor is also damped at the nodes, $S(R) = [E_{est} - E_L(R)]F(R) - \ln(N/N_0)$, where E_{est} is the best current estimate of the energy and N and N_0 are the current and the target number of walkers.

The presence of E_{est}, the Monte Carlo estimate of E, in the branching term S implies that the calculation of $d_\lambda E$ includes a contribution proportional to $d_\lambda E$ itself,

$$\tau \left\langle \sum_i [E_L(R_n) - E]F(R_{n-i}) \right\rangle d_\lambda E \equiv \bar{F} d_\lambda E \tag{10}$$

This does not require prior knowledge of the result: we can calculate the factor $\bar{F}$ and $(d_\lambda E)_0$, the derivative when the contribution of eq 10 is omitted, and combine them to get the unbiased result as $d_\lambda E = (d_\lambda E)_0 + \bar{F} d_\lambda E$, or $d_\lambda E = (d_\lambda E)_0/(1 - \bar{F})$.

The main technical complication in DMC is the need to store a few quantities for each derivative and for each value of ϵ over the last k steps of each walker, namely, $d_\lambda \ln G$ and $E_L d_\lambda \ln G$, to implement eq 2 with the probability distribution P of eq 7 and F and $E_L F$ to implement eq 10.

The AS regularized estimator has been applied to approximate DMC forces in refs 13 and 20. However, it pushes a finite density of walkers on the nodes, which is presumably not optimal in DMC. Furthermore, unlike in VMC, it requires[13] an extrapolation to $\epsilon \to 0$ which—at difference with the PW and warp estimators—cannot be done in a single run.

Therefore, for the DMC derivatives, we consider only the PW and the warp estimators. For the former, we insert eq 7 into eq 2 and multiply each term of the resulting summation by a polynomial f_ϵ calculated at the appropriate configuration. For the latter, we just insert eqs 7 in 5. For the warp estimator, the argument of the Jacobian needs some care: we evaluate $\partial_\lambda \ln J$ in the proposed configuration R' for both accepted and rejected moves; alternatively, we can include $\partial_\lambda \ln J$ only for the accepted moves, provided the warp transformation is not considered in the derivatives at R' when the move is rejected.

We present results of DMC simulations with the wave function Ψ_0, time step $\tau = 0.1$, and target number of walkers $N_0 = 100$. We have verified that in the limit $\tau \to 0$, we recover the analytic results[21] for the ground-state energy, $E = 2q/a^2$ with $q = 0.825352549$, within a statistical error of less than 1 part in 10,000. To this level of accuracy, the population control bias[1] is negligible.

Figure 4 exposes the drawback of the bare estimator: the probability distribution $p(\xi)$ for the block averages of the derivative features a right heavy tail, consistent with the expected[11,22] leading decay $\propto|\xi - \xi_0|^{-5/2}$. In the data trace, shown in the inset, heavy tails result in large spikes that would mar the smooth convergence of structural or variational optimization. Meaningful averages and statistical uncertainties of heavy-tailed distributions with known tail indices can be computed with a tail regression analysis.[23] This technique, however, requires a heavy post-processing not very practical for large-scale applications. The regularized estimators PW and warp, instead, have nearly Gaussian distributions amenable to standard statistical analysis with significantly smaller statistical errors and, most importantly, no large spikes in the data trace.

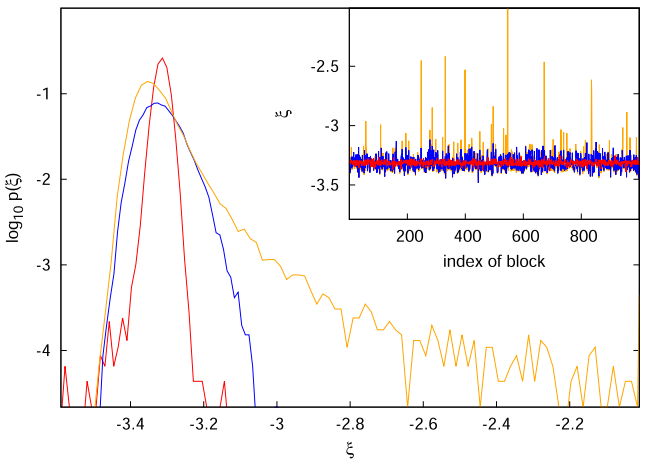


Figure 4. Histogram of 46,000 block averages, each of which taken over 10,000 steps of 100 walkers in a DMC calculation of $d_a E|_{a=1}$. The warp and PW estimators (with $\epsilon = 0.2$ and 0.0125, respectively) have nearly Gaussian distributions. The bare estimator has a heavy-tailed distribution; the largest value in the present sample exceeds $x = 10$. Inset: the data trace of the first 1000 block averages.

The central result of this work is shown in Figure 5. We calculate the energy E and its derivative $d_a E$ for a set of values of a and compare the DMC derivatives with the derivative of a fit to the DMC energies. All the estimators (bare, extrapolated PW, and warp) are unbiased, which demonstrates the correctness of the proposed algorithm. For comparison, the variational drift-diffusion (VD) approximation of ref 13 gives for $a = 1$ a bias of ~0.2, twice the full scale of the right panel, and it gets even worse for smaller time steps (although the VD approximation is devised to exploit good wave functions, while our Ψ_0 is poor on purpose to test the unbiased estimators).

For given a, the same run is used for all the estimators. Therefore, the statistical error is a direct measure of the square

https://doi.org/10.1021/acs.jctc.1c00496
J. Chem. Theory Comput. 2022, 18, 118−123

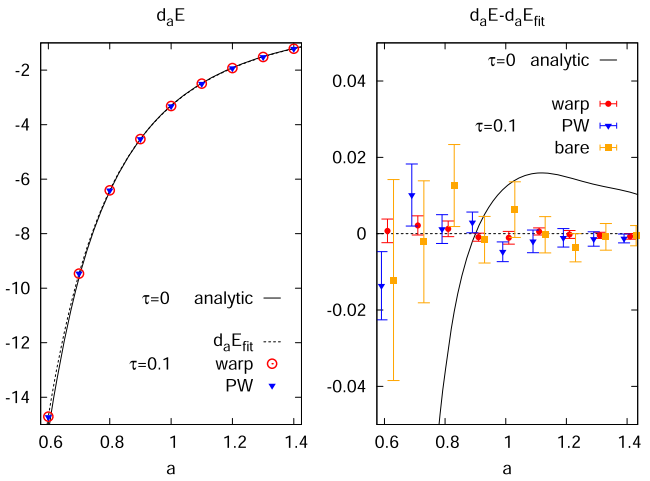


Figure 5. DMC calculations of d_aE with $\tau = 0.1$. Left panel: the warp and PW results compare favorably with the "exact" result, defined as the derivative d_aE_{fit} of a fit to DMC calculations of $E(a)$. The analytic result $d_aE = -4q/a^3$ differs from d_aE_{fit} because the latter has a finite time step error. Right panel: the difference of the calculated derivatives with d_aE_{fit} is shown on an expanded scale by subtracting the latter (hence, the solid black line is minus the time step error of d_aE_{fit}). The PW derivatives are extrapolated to $\epsilon \to 0$. We also include the bare-estimator derivatives, with averages and statistical uncertainties obtained with the tail regression estimator analysis toolkit made available in ref 23. Small horizontal shifts are applied to the same—a data for clarity.

root of their relative efficiency. The statistical errors of the bare, PW, and warp estimators, averaged over the values of a shown in Figure 5, are in the ratio of 4.9:2.4:1. These figures may belittle the PW estimator somewhat because in this particular example, a large quadratic bias needs to be eliminated by extrapolation, but they convey the relevant message that both PW and warp are significantly more efficient than the bare estimator.

3. CONCLUSIONS

In summary, we have presented an algorithm to calculate unbiased, finite-variance derivatives in DMC. The estimate of the derivative with respect to a given parameter is fully consistent with the dependence on that parameter of the FN energy, calculated with the same time step. The tail regression statistical analysis[23] can cope with the problem of the infinite variance of the bare estimator. Alternatively, and more efficiently, both the recently proposed PW regularization[15] and the warp regularization introduced in this work can be used to good effect to eliminate the divergence of the variance.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00496.

Analysis of the divergence of different local estimators of the energy derivatives and derivatives of DMC energy with respect to a variational parameter and internuclear distance for the lithium dimer (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Jesse van Rhijn − *MESA+ Institute for Nanotechnology, University of Twente, 7500 AE Enschede, The Netherlands*; Email: j.vanrhijn@utwente.nl

Claudia Filippi − *MESA+ Institute for Nanotechnology, University of Twente, 7500 AE Enschede, The Netherlands*; orcid.org/0000-0002-2425-6735; Email: c.filippi@utwente.nl

Stefania De Palo − *CNR-IOM DEMOCRITOS, Istituto Officina dei Materiali, and SISSA Scuola Internazionale Superiore di Studi Avanzati, I-34136 Trieste, Italy*; Email: depalo@iom.cnr.it

Saverio Moroni − *CNR-IOM DEMOCRITOS, Istituto Officina dei Materiali, and SISSA Scuola Internazionale Superiore di Studi Avanzati, I-34136 Trieste, Italy*; Email: moroni@iom.cnr.it

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00496

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank K. Nakano, S. Sorella, and L. Wagner for useful discussions. C.F. and S.M. acknowledge support from the European Centre of Excellence in Exascale Computing TREX, funded by the European Union's Horizon 2020—Research and Innovation program—under grant no. 952165.

■ REFERENCES

(1) Martin, R. M.; Reining, L.; Ceperley, D. M. *Interacting Electrons*; Cambridge University Press, 2016.

(2) Sorella, S.; Capriotti, L. Algorithmic differentiation and the calculation of forces by quantum Monte Carlo. *J. Chem. Phys.* **2010**, *133*, 234111.

(3) Filippi, C.; Assaraf, R.; Moroni, S. Simple formalism for efficient derivatives and multi-determinant expansions in quantum Monte Carlo. *J. Chem. Phys.* **2016**, *144*, 194105.

(4) Umrigar, C. J.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R. G. Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions. *Phys. Rev. Lett.* **2007**, *98*, 110201.

(5) Toulouse, J.; Umrigar, C. J. Optimization of quantum Monte Carlo wave functions by energy minimization. *J. Chem. Phys.* **2007**, *126*, 084102.

(6) Sorella, S.; Casula, M.; Rocca, D. Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods. *J. Chem. Phys.* **2007**, *127*, 014105.

(7) Kent, P. R. C.; Annaberdiyev, A.; Benali, A.; Bennett, M. C.; Landinez Borda, E. J.; Doak, P.; Hao, H.; Jordan, K. D.; Krogel, J. T.; Kylänpää, I.; Lee, J.; Luo, Y.; Malone, F. D.; Melton, C. A.; Mitas, L.; Morales, M. A.; Neuscamman, E.; Reboredo, F. A.; Rubenstein, B.; Saritas, K.; Upadhyay, S.; Wang, G.; Zhang, S.; Zhao, L. QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion quantum Monte Carlo. *J. Chem. Phys.* **2020**, *152*, 174105.

(8) Nakano, K.; Attaccalite, C.; Barborini, M.; Capriotti, L.; Casula, M.; Coccia, E.; Dagrada, M.; Genovese, C.; Luo, Y.; Mazzola, G.; Zen, A.; Sorella, S. TurboRVB: A many-body toolkit for ab initio electronic simulations by quantum Monte Carlo. *J. Chem. Phys.* **2020**, *152*, 204121.

(9) Needs, R. J.; Towler, M. D.; Drummond, N. D.; López Ríos, P.; Trail, J. R. Variational and diffusion quantum Monte Carlo calculations with the CASINO code. *J. Chem. Phys.* **2020**, *152*, 154106.

(10) Feldt, J.; Filippi, C. *Quantum Chemistry and Dynamics of Excited States*; John Wiley & Sons, Ltd, 2020; Chapter 8, pp 247−275.
(11) Badinski, A.; Haynes, P. D.; Trail, J. R.; Needs, R. J. Methods for calculating forces within quantum Monte Carlo simulations. *J. Phys.: Condens. Matter* **2010**, *22*, 074202.
(12) Assaraf, R.; Caffarel, M.; Kollias, A. C. Chaotic versus Nonchaotic Stochastic Dynamics in Monte Carlo Simulations: A Route for Accurate Energy Differences in N-Body Systems. *Phys. Rev. Lett.* **2011**, *106*, 150601.
(13) Moroni, S.; Saccani, S.; Filippi, C. Practical Schemes for Accurate Forces in Quantum Monte Carlo. *J. Chem. Theory Comput.* **2014**, *10*, 4823−4829.
(14) Attaccalite, C.; Sorella, S. Stable Liquid Hydrogen at High Pressure by a Novel Ab Initio Molecular-Dynamics Calculation. *Phys. Rev. Lett.* **2008**, *100*, 114501.
(15) Pathak, S.; Wagner, L. K. A light weight regularization for wave function parameter gradients in quantum Monte Carlo. *AIP Adv.* **2020**, *10*, 085213.
(16) Trail, J. R. Heavy−tailed random error in quantum Monte Carlo. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2008**, *77*, 016703.
(17) Filippi, C.; Umrigar, C. J. Correlated sampling in quantum Monte Carlo: A route to forces. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2000**, *61*, R16291−R16294.
(18) Zong, F.; Ceperley, D. M. Path integral Monte Carlo calculation of electronic forces. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1998**, *58*, 5123−5130.
(19) Umrigar, C. J.; Nightingale, M. P.; Runge, K. J. A diffusion Monte Carlo algorithm with very small time−step errors. *J. Chem. Phys.* **1993**, *99*, 2865−2890.
(20) Valsson, O.; Filippi, C. Photoisomerization of Model Retinal Chromophores: Insight from Quantum Monte Carlo and Multi-configurational Perturbation Theory. *J. Chem. Theory Comput.* **2010**, *6*, 1275−1292.
(21) Moon, P.; Spencer, D. E. *Field Theory Handbook*; Springer: Berlin, Heidelberg, 1988; p 17ff.
(22) Trail, J. R. Alternative sampling for variational quantum Monte Carlo. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2008**, *77*, 016704.
(23) López Ríos, P.; Conduit, G. J. Tail-regression estimator for heavy-tailed distributions of known tail indices and its application to continuum quantum Monte Carlo data. *Phys. Rev. E* **2019**, *99*, 063312.

https://doi.org/10.1021/acs.jctc.1c00496
J. Chem. Theory Comput. 2022, 18, 118−123