

Response properties in phaseless auxiliary field quantum Monte Carlo

Ankit Mahajan,^{1,2,*} Jo S. Kurian,² Joonho Lee,^{1,3} David R. Reichman,¹ and Sandeep Sharma^{2,†}

¹*Department of Chemistry, Columbia University, New York, NY 10027, USA*

²*Department of Chemistry, University of Colorado, Boulder, CO 80302, USA*

³*Department of Chemistry and Chemical Biology,
Harvard University, Cambridge, MA 02138, USA*

We present a method for calculating first-order response properties in phaseless auxiliary field quantum Monte Carlo (AFQMC) through the application of automatic differentiation (AD). Biases and statistical efficiency of the resulting estimators are discussed. Our approach demonstrates that AD enables the calculation of reduced density matrices (RDMs) with the same computational cost scaling as energy calculations, accompanied by a cost prefactor of less than four in our numerical calculations. We investigate the role of self-consistency and trial orbital choice in property calculations.

1. INTRODUCTION

The calculation of energy gradients is of great importance in electronic structure theory. Derivatives of the energy with respect to external perturbations can be used to evaluate many properties of interest. Nuclear gradients are used for geometry optimization and molecular dynamics simulations. Parameter derivatives are useful in wave function and basis set optimizations. Throughout the history of quantum chemistry, the development of efficient methods for evaluating energy gradients has thus been a focal point due to their central importance.^{1–5} This development often involved meticulous derivation of analytic expressions and manual coding of implementations. The development of efficient algorithms has led to cost-effective gradient evaluations, comparable in cost to the evaluation of the energy. In recent years, automatic or algorithmic differentiation (AD) has gained recognition as a powerful tool for calculating derivatives of functions defined by computer programs, notably in machine learning. By applying the chain rule to elementary steps in a computation, AD can calculate numerically exact derivatives, considerably reducing the implementation effort.⁶ The reverse-mode AD is particularly remarkable, enabling the evaluation of first-order derivatives with respect to an arbitrary number of parameters at the same cost scaling as the underlying function. As a result, it becomes an appealing choice for calculating all nuclear forces or reduced density matrices in quantum chemistry. Various studies have explored the utility of AD in conventional electronic structure methods.^{7–11} Reference 12 showcased the effectiveness of AD in differentiating tensor network-based functions.

Quantum Monte Carlo (QMC) approaches comprise a scalable and accurate suite of alternatives to traditional quantum chemistry methods.^{13–18} Among these, the auxiliary field quantum Monte Carlo (AFQMC) approach stands out as a powerful tool for investigating correlated

electronic systems.^{19–21} It has been used to study a wide variety of systems, including solids and molecules with a range of correlation strengths.^{22–27} However, the evaluation of energy derivatives has posed challenges for projection QMC methods. Mixed estimators are usually used in energy calculations with these methods, which are not accurate for general properties or nuclear gradients. An incorrect application of the Hellman-Feynman theorem leads to this mixed estimator that is easy to evaluate but leads to an uncontrolled bias for operators that do not commute with the Hamiltonian. Although various approximations can address this shortcoming,^{28–34} many challenges still remain due to uncontrolled biases or large variances. It is worth mentioning that within the context of variational Monte Carlo (VMC), AD was first employed by Sorella and Capriotti to calculate nuclear gradients.³⁵ Additionally, there have been studies of AD in diffusion Monte Carlo (DMC)³⁶ and lattice VMC³⁷, following an approach similar to that presented here.

The problem of calculating derivatives of complicated stochastic functions plays a crucial role in various domains ranging from sensitivity analysis in finance³⁸ and queueing theory in management³⁹ to optimizations in reinforcement learning.⁴⁰ An effective approach for formulating low-variance estimators involves utilizing correlated sampling, by employing common random numbers for example, to achieve small cost differences with high accuracy. When Monte Carlo simulations involve discrete random variables, like branching in QMC calculations, biases can arise in derivative estimators, and addressing these errors is an active area of research.⁴¹ In this paper, we leverage some of this technology to compute first-order response properties in AFQMC. Specifically, we demonstrate that reverse-mode AD enables the calculation of Reduced Density Matrices (RDMs) with the same cost scaling as energy calculations.

This paper is organized as follows: we first review relevant parts of the AFQMC algorithm (Section 2.1) and existing approaches for calculating properties (Section 2.2). Then we discuss use of AD in AFQMC, including various contributions to the derivatives as well as biases in our method (Section 2.3). We then present illustrative numerical results to demonstrate the accuracy and sta-

* ankitmahajan76@gmail.com

† sanshar@gmail.com

tistical efficiency of our method (Section 3.1). Finally, we compare AD-AFQMC dipole moments of some molecules with other quantum chemistry methods and experiments (Section 3).

2. THEORY

2.1. Review of AFQMC

In this section, we present an outline of the AFQMC algorithm, focusing on the aspects relevant to gradient calculations. We refer the reader to recent review articles^{42–44} for a more comprehensive description. In AFQMC, the ground state is represented as a weighted sum of non-orthogonal Slater determinants

$$|\Psi_0\rangle = \sum_i w_i \frac{|\phi_i\rangle}{\langle\psi_T|\phi_i\rangle}, \quad (1)$$

where w_i are weights, $|\phi_i\rangle$ are walker Slater determinants with complex orbitals, and $|\psi_T\rangle$ is the trial state. This state is obtained by applying an exponential form of projector onto the trial state as

$$e^{-\tau\hat{H}}|\psi_T\rangle \xrightarrow{\tau \rightarrow \infty} |\Psi_0\rangle, \quad (2)$$

where τ is the imaginary time, and we assume $\langle\psi_T|\psi_0\rangle \neq 0$. We use the quantum chemistry Hamiltonian given by

$$\hat{H} = \sum_{ij} h_{ij} \hat{a}_i^\dagger \hat{a}_j + \frac{1}{2} \sum_{\gamma} \left(\sum_{ij} L_{ij}^{\gamma} \hat{a}_i^\dagger \hat{a}_j \right)^2, \quad (3)$$

where h_{ij} are one-electron integrals and L_{ij}^{γ} are Cholesky decomposed two-electron integrals in an orthonormal orbital basis. To sample the action of the propagator, the Hubbard-Stratonovic transform is used to write the short-time propagator as

$$e^{-\Delta\tau\hat{H}} = \int d\mathbf{x} p(\mathbf{x}) \hat{B}(\mathbf{x}), \quad (4)$$

where $\mathbf{x}$ is the vector of auxiliary fields, $p(\mathbf{x})$ is the standard normal distribution of the auxiliary fields, and $\hat{B}(\mathbf{x})$ is a complex propagator given by the exponential of a one-body operator. Due to Thouless' theorem, $\hat{B}(\mathbf{x})$ acts on a Slater determinant $|\phi\rangle$ as

$$\hat{B}(\mathbf{x})|\phi\rangle = |\phi(\mathbf{x})\rangle, \quad (5)$$

where $|\phi(\mathbf{x})\rangle$ is another Slater determinant obtained by a linear transformation of the orbitals in $|\phi\rangle$. In the hybrid approximation, importance sampling is achieved by shifting the auxiliary fields with a force bias given as

$$\bar{x}_{\gamma} = -\sqrt{\Delta\tau} \frac{\langle\psi_T|\sum_{ij} L_{ij}^{\gamma} \hat{a}_i^\dagger \hat{a}_j|\phi\rangle}{\langle\psi_T|\phi\rangle}, \quad (6)$$

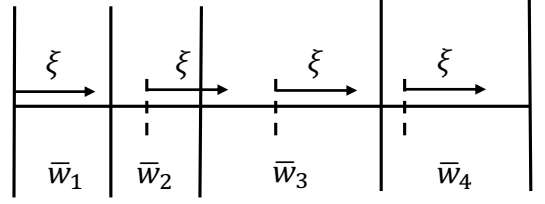


FIG. 1. A schematic showing the stochastic reconfiguration (SR) procedure with four walkers. ξ is drawn from a uniform distribution on $[0, \frac{1}{4}]$ and $\bar{w}_i = \frac{w_i}{\sum_i w_i}$ denote normalized weights. In this instance, SR leads to the reconfiguration $\{\phi_1, \phi_2, \phi_3, \phi_4\} \rightarrow \{\phi_1, \phi_3, \phi_3, \phi_4\}$.

and with the weights propagated as

$$w_i(\mathbf{x}) = \left| \frac{\langle\psi_T|\phi(\mathbf{x})\rangle}{\langle\psi_T|\phi\rangle} e^{\mathbf{x} \cdot \bar{\mathbf{x}} - \frac{\bar{\mathbf{x}}^2}{2}} \right| w_i. \quad (7)$$

Force bias is not enough, by itself, to control the large fluctuations stemming from the sign or phase problem. The phaseless constraint²¹ can be used to overcome the phase problem at the expense of a systematic bias in the sampled wave function by performing a cosine projection as

$$w_i(\mathbf{x}) = \max(0, \cos(\Delta\theta)), w_i(\mathbf{x}) \quad (8)$$

where

$$\Delta\theta = \text{Arg} \left(\frac{\langle\psi_T|\phi(\mathbf{x})\rangle}{\langle\psi_T|\phi\rangle} \right). \quad (9)$$

The size of this phaseless bias is dictated by the accuracy of the trial state.

For statistical efficiency, we periodically perform reconfigurations of the walker population. We use the stochastic reconfiguration (SR) method⁴⁵ where walkers with small weights are removed, and those with large weights are duplicated. Figure 1 illustrates this procedure schematically. It has the desirable property that when the walker weights are uniform, the walker population remains unchanged. It is important to note that SR is a discontinuous function, as small changes in weights lead to different reconfiguration instances. These discontinuities are a source of bias in our AD gradient calculations, as described in section 2.3.

2.2. Calculating properties in AFQMC

Given an observable $\hat{O}$, the simplest way to estimate its expectation value in AFQMC is to use the mixed estimator given by

$$\langle\hat{O}\rangle_{\text{mixed}} = \frac{\sum_{\phi} w_{\phi} \frac{\langle\psi_T|\hat{O}|\phi\rangle}{\langle\psi_T|\phi\rangle}}{\sum_{\phi} w_{\phi}}. \quad (10)$$

If $\hat{O}$ does not commute with the Hamiltonian, this estimator has a bias due to its non-variational nature. One

way to mitigate this error is to use the extrapolated estimator^{46–48}

$$\langle \hat{O} \rangle_{\text{extrapolated}} = 2\langle \hat{O} \rangle_{\text{mixed}} - \frac{\langle \psi_T | \hat{O} | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle}. \quad (11)$$

The simplicity of evaluating this estimator comes with a crucial caveat; its accuracy heavily relies on the trial state employed. For instance, achieving convergence of dipole moments in reference 49 required the use of a large multi-Slater trial state.

A more common approach to calculating properties in AFQMC is using backpropagation,¹⁹ which overcomes the non-variationality issue by propagating the trial state backward as

$$\langle \hat{O} \rangle_{\text{bp}} = \frac{\langle \psi_T | e^{-\tau_{\text{bp}} \hat{H}} \hat{O} e^{-\tau \hat{H}} | \psi_T \rangle}{\langle \psi_T | e^{-(\tau_{\text{bp}} + \tau) \hat{H}} | \psi_T \rangle}. \quad (12)$$

A direct simulation of two independent random walks results in a highly noisy and computationally intensive estimator. To control this noise, importance sampling is employed to correlate the two walks. However, in certain systems, this procedure has been observed to introduce a significant phaseless bias. This bias can be reduced by undoing the phaseless approximation in the backpropagation, albeit at the expense of noisier estimates.³³ Backpropagation has also been used to evaluate nuclear gradients based on the Hellman-Feynman theorem.⁵⁰

With the goal of bringing accuracy and computational cost of property calculations on par with that of calculating the energy, we pursue a response approach in this paper. Consider a perturbed Hamiltonian given by

$$\hat{H}(\lambda) = \hat{H} + \lambda \hat{O}, \quad (13)$$

where λ is a real parameter. The response estimator can be written as

$$\langle \hat{O} \rangle_{\text{response}} = \left. \frac{dE(\lambda)}{d\lambda} \right|_{\lambda=0}, \quad (14)$$

where $E(\lambda)$ is the AFQMC energy calculated with the perturbed Hamiltonian $\hat{H}(\lambda)$. For a variationally optimized wave function, this estimator is equivalent to the expectation value by virtue of the Hellman-Feynman theorem. For non-variational methods like coupled cluster theory, response properties are found to be very accurate and are routinely used. One way to evaluate energy derivatives is using the method of finite differences. However, as AFQMC is a stochastic method, it is necessary to correlate the finite difference energy calculations to minimize the noise in the energy difference. Such correlated sampling methods have been used in AFQMC⁵¹ as well as other QMC methods^{30,52}. Using a common stream of random numbers is an intuitively appealing strategy to maintain two Monte Carlo runs close to each other, but technical challenges arise in ensuring that the runs remain coherent. A promising approach to address this

challenge was recently reported in reference 53. A drawback of the correlated sampling finite difference method is the requirement for separate energy calculations for each perturbation, making the calculation of reduced density matrices and all nuclear forces computationally expensive, with the cost scaling linearly with the number of perturbations. In the next section, we will explore how reverse-mode automatic differentiation utilizes the common random number strategy, while also allowing calculations of responses to multiple perturbations at a cost comparable to that of a single energy calculation.

2.3. Derivatives of AFQMC energy

We can write the AFQMC energy expression making the dependence on the perturbation explicit as

$$\begin{aligned} E(\hat{H}_\lambda, \psi_T(\lambda), \{w_i(\lambda)\}, \{\phi_i(\lambda)\}) \\ = \frac{\sum_i w_i(\lambda) \frac{\langle \psi_T(\lambda) | \hat{H}(\lambda) | \phi_i(\lambda) \rangle}{\langle \psi_T(\lambda) | \phi_i(\lambda) \rangle}}{\sum_i w_i(\lambda)}. \end{aligned} \quad (15)$$

Differentiating this expression with respect to λ gives

$$\begin{aligned} \frac{dE}{d\lambda} = \frac{\sum_i w_i(\lambda) \frac{\langle \psi_T(\lambda) | \hat{O} | \phi_i(\lambda) \rangle}{\langle \psi_T(\lambda) | \phi_i(\lambda) \rangle}}{\sum_i w_i(\lambda)} + \frac{\partial E}{\partial \psi_T(\lambda)} \frac{d\psi_T(\lambda)}{d\lambda} \\ + \sum_i \left(\frac{\partial E}{\partial w_i(\lambda)} \frac{dw_i(\lambda)}{d\lambda} + \frac{\partial E}{\partial \phi_i(\lambda)} \frac{d\phi_i(\lambda)}{d\lambda} \right). \end{aligned} \quad (16)$$

The first term is the mixed estimator 10. The second term contains the explicit effect of the relaxation of the trial state in response to the perturbation on the energy. It also appears in the response calculation of many conventional quantum chemistry methods where it is evaluated (implicitly) using the coupled perturbed Hartree Fock (CPHF) equations.⁵⁴ The final term contains contributions from the derivatives of weights and walkers, which, in turn, are influenced by the response of the trial state. The dependence of weights and walkers on the perturbation is accumulated through many steps of AFQMC propagation and local energy measurements. One can evaluate their derivatives by repeated application of the chain rule. AD offers an efficient and convenient way to accomplish this task.

All these contributions can be evaluated using AD on an energy function that incorporates the effects of the perturbation on the final energy. If stochastic reconfiguration is not performed, the AFQMC energy can be written as the integral

$$E(\lambda) = \int d\mathbf{x} \mathcal{N}(\mathbf{x}) f(\mathbf{x}, \lambda), \quad (17)$$

where $\mathbf{x}$ is the collection of normally distributed auxiliary fields for all time steps and walkers, and f denotes the AFQMC procedure, including propagation and measurement. For the sake of notational convenience, we have

suppressed the dependence of f on other quantities that are not pertinent to the current discussion. The derivative of this integral can be obtained as

$$\frac{dE}{d\lambda} = \int d\mathbf{x} \mathcal{N}(\mathbf{x}) \frac{\partial f}{\partial \lambda}(\mathbf{x}, \lambda), \quad (18)$$

Note that the phaseless constraint (8) is not differentiable at the origin, but it is continuous and differentiable almost everywhere. The switching of derivative and integral is allowed in this case because all the operations performed in the evaluation of f are continuous and differentiable almost everywhere (within the region of interest).³⁸ As a result, (18) provides a means to sample the derivative through Monte Carlo sampling, akin to the energy, and is known as a pathwise estimator. This estimator exhibits desirable properties of low variance, thanks to correlated sampling, and avoids numerical issues stemming from finite difference step sizes. This pathwise estimator is precisely the one that is evaluated by performing automatic differentiation on the energy function. Upon introducing SR, the energy can be written as

$$E(\lambda) = \int d\mathbf{x} d\boldsymbol{\xi} \mathcal{N}(\mathbf{x}) \mathcal{U}(\boldsymbol{\xi}) \tilde{f}(\mathbf{x}, \boldsymbol{\xi}, \lambda), \quad (19)$$

where $\boldsymbol{\xi}$ are uniformly distributed variables. Due to the discrete nature of SR, the function $\tilde{f}$ is not continuous and the pathwise estimator is biased. Several methods have been developed to address biases arising from discontinuities, score function or likelihood ratio estimators being a popular choice.^{38,39,55} However, in preliminary numerical experiments, we observed this method to be quite noisy. Hence, we defer a detailed study of these bias mitigation techniques to future work. In this paper, we use the biased pathwise estimator to sample the derivative of the energy function. We present a numerical analysis of this bias in section 3.

2.4. Implementation details

In AFQMC energy calculations, a sequence of phaseless imaginary time propagation steps is executed before reconfiguration and energy measurement. At the end of this block of steps, the walker weights and local energies are averaged to yield estimates of the energy along with its associated stochastic error. To obtain energy derivative samples, AD is performed on the averaged energy from a sufficiently large number of such blocks. The number of blocks to be included in the AD calculation is dictated by the time required for the convergence of walker and weight derivatives (16). We present a numerical investigation of this convergence in section 3.1.

AD can be used in two modes: forward and reverse. In forward-mode AD, the derivative of the energy function is evaluated by propagating the perturbation through the energy function. This approach calculates the response

to one perturbation at a time, resulting in a computational cost scaling comparable to correlated finite difference methods. However, it offers the advantage of having no finite difference errors and requires low memory, making it suitable for evaluating a few properties in large systems. For example, one can evaluate the dipole moment of a system by using the perturbation given by

$$\hat{H}(\boldsymbol{\varepsilon}) = \hat{H} + \boldsymbol{\varepsilon} \cdot \hat{\mathbf{d}}, \quad (20)$$

where $\hat{\mathbf{d}}$ is the dipole operator, and calculating the derivative with respect to the electric field $\boldsymbol{\varepsilon}$, requiring at most three forward-mode AD calculations. In reverse-mode AD, gradients are computed by propagating changes in energy backward through to the perturbations. Remarkably, it shares the same computational cost scaling as the energy function itself when calculating gradients with respect to an arbitrary number of perturbations. Depending on the specific implementation, the cost is usually between 3 to 5 times that of the energy evaluation. This property makes reverse-mode AD well-suited for efficiently calculating many properties simultaneously. For example, one can evaluate one and two-particle reduced density matrices (RDM) by using the following perturbed Hamiltonian

$$\hat{H}(\lambda) = \hat{H} + \sum_{ij} \lambda_{ij} \hat{a}_i^\dagger \hat{a}_j + \sum_{iajb} \gamma_{ipjq} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_q \hat{a}_p, \quad (21)$$

and evaluating derivatives with respect to λ_{ij} and γ_{ipjq} , respectively.

Reverse-mode AD achieves its impressive computational cost scaling by storing intermediates during the evaluation of the energy function resulting in a linear increase in memory usage with the number of AFQMC time steps to be differentiated. Since the derivatives of weights and walkers need to be converged with respect to the propagation time, this leads to a substantial memory cost for reverse-mode AD with increasing system size. We use checkpointing,⁶ which recalculates some intermediates on the backward pass instead of storing them, to reduce this cost. This reduction in memory usage comes at the expense of an increased computational cost prefactor.

For capturing the orbital relaxation of a HF trial state, the AFQMC energy function can be written to include a call to a HF solver. We use a fixed and conservatively large number of SCF iterations to ensure that the derivatives are fully converged. Additionally, issues related to degeneracies in reverse-mode AD of eigendecompositions can be dealt with by ignoring rotations in degenerate spaces.^{10,12}

3. RESULTS

In this section, we present the results of our response ph-AFQMC calculations, referred to as AD-AFQMC, and analyze both systematic and stochastic errors. Our

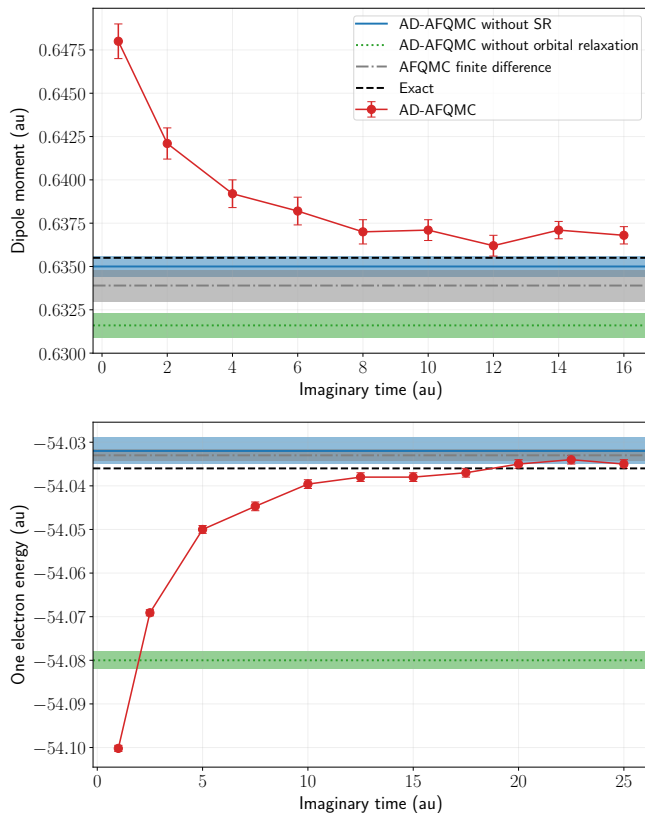


FIG. 2. Dipole moment of NH_3 in the cc-pVDZ basis (top) and one electron energy of a H_{20} chain in the STO-6G basis (bottom) as a function of propagation time.

analysis focuses on the accuracy of 1-RDM and dipole moment calculations using AFQMC with a HF trial state. We used PySCF⁵⁶ to obtain molecular integrals and to perform quantum chemistry wave function calculations. The near-exact Density Matrix Renormalization Group (DMRG) calculations on hydrogen chains were done using the block2 code.⁵⁷ We used QMCPACK⁵⁸ to do backpropagation AFQMC calculation. The code used to perform AD-AFQMC calculations is available in a public repository⁵⁹, with the Jax⁶⁰ library used for AD. Reverse-mode AD was employed for 1-RDM calculations, while forward-mode AD was used for dipole moments (except for self-consistent calculations). Input and output files for all calculations can be accessed from a public repository.⁶¹ We used a time step of 0.01 a.u. in all ph-AFQMC calculations. Cholesky decompositions were calculated up to a threshold error of 10^{-5} . Errors due to these approximations can be controlled systematically and we estimate them to be smaller than the statistical errors in the results presented here.

3.1. Illustrative results

First, we analyze various contributions to observables and their convergence with imaginary time. In the top panel of figure 2, we show the convergence of the AD-AFQMC dipole moment of the ammonia molecule in the cc-pVDZ basis at equilibrium geometry. The x-axis represents imaginary time, over which the AD is performed, starting with equilibrated walkers. After around 8 a.u., the dipole moment converges to a value outside the stochastic error bar of the finite difference AFQMC estimator calculated using common random numbers. This discrepancy arises from the bias due to SR in sampling AD derivatives, leading to a bias of 0.003(1) a.u. in this case. Since this is a relatively small system, we can perform AD-AFQMC calculations without SR and this value matches with the finite difference calculation within stochastic error. However, without SR, more sampling effort is required to achieve the same level of stochastic error. One can perform SR less frequently to control this bias, but we do not explore this approach further since the bias is relatively small. We also show the converged value of the AD-AFQMC dipole moment (with SR) without including trial orbital relaxation, illustrating that including orbital relaxation brings the dipole moment closer to the exact value.

In the bottom panel of figure 2, we conduct a similar analysis for the one-electron energy of a twenty-atom linear hydrogen chain with an interatomic separation of 2.4 a.u. in the minimal STO-6G basis. We use an unrestricted HF (UHF) trial state for the AFQMC calculations. In this case, the SR bias of AD-AFQMC is 0.001(1) a.u. The orbital relaxation contribution is substantial and increases the accuracy of the estimator significantly.

3.2. Hydrogen chains

In figure 3, we present a comparison of the accuracy of AD-AFQMC one-electron energies and 1-RDMs with coupled cluster methods for hydrogen chains. Near-exact reference benchmark results for this one-dimensional system were obtained using DMRG. Unrestricted coupled cluster calculations were performed on top of UHF references. We did not include orbital relaxation in CC calculations, as we found it not to significantly impact the results. All AFQMC calculations were done with UHF trial states and with SR. 1-RDM errors were calculated using the Frobenius norm. We consider two bond lengths (d) of 1.6 a.u. and 2.4 a.u.

For $d = 1.6$ a.u., out of the CC methods CCSD(T) performs very well, as one would expect for a system close to equilibrium. We note that for the same bond length, the chains become more correlated with increasing length, explaining the increasing error in both the one-electron energy and 1-RDM for the CC methods. AD-AFQMC one-electron energy systematic errors are smaller than CCSD(T) for all chain lengths, except for H_{10} . This dis-

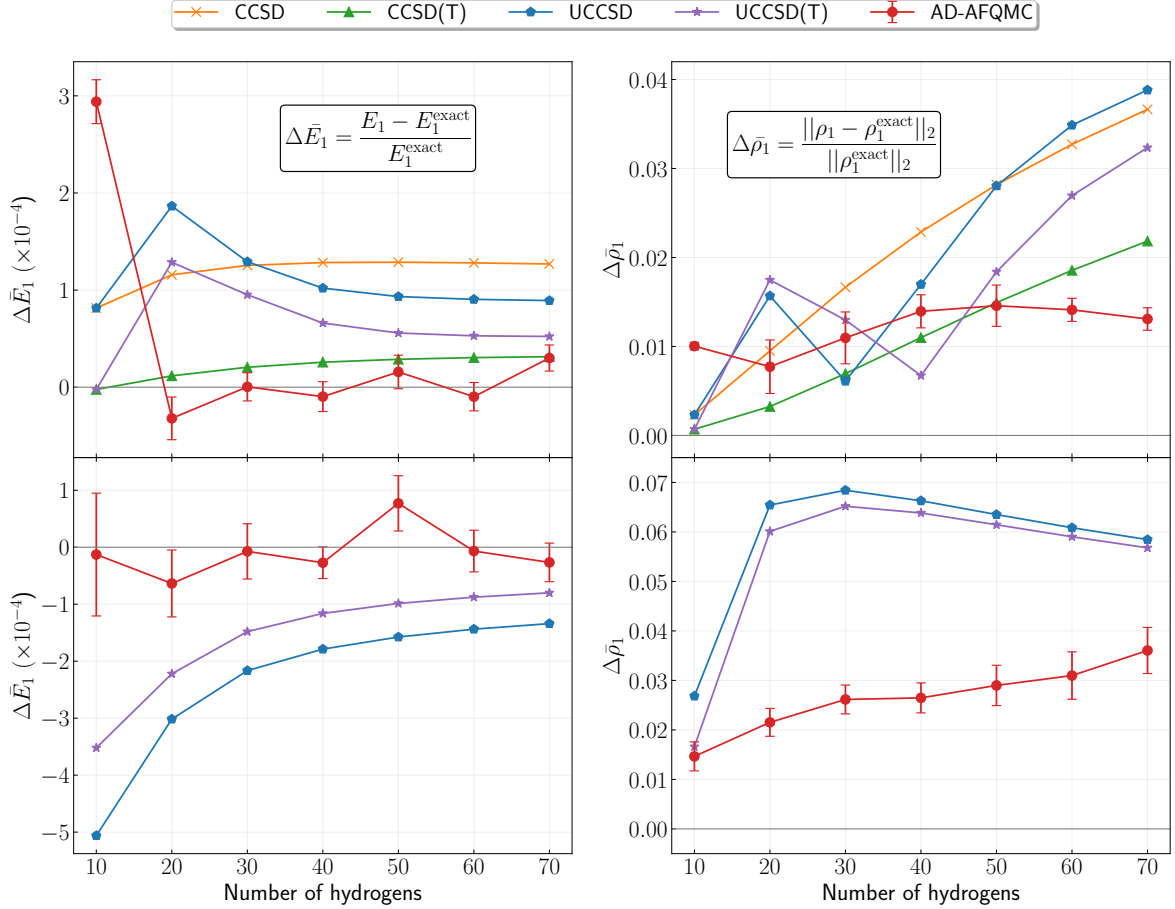


FIG. 3. Relative error in the one-electron energy (left) and 1-RDM (right) for a hydrogen chain (STO-6G basis) with an inter-hydrogen separation of $d = 1.6$ Bohr (top) and $d = 2.4$ Bohr (bottom).

crepancy is due to the proximity of $d = 1.6$ a.u. to the Coulson-Fischer point of this chain, beyond which the UHF solution becomes more stable. The AD-AFQMC 1-RDM error is also smaller than CCSD for all chains except H_{10} and smaller than CCSD(T) beyond H_{50} .

For $d = 2.4$ a.u., the CC methods are less accurate due to the increased correlation in the system. In this case, restricted methods have convergence issues, so we only present the results for unrestricted observables. For both one-electron energies and 1-RDMs, AD-AFQMC errors are significantly smaller compared to UCCSD and UCCSD(T), demonstrating the efficacy of AD-AFQMC in providing accurate estimates of properties across a range of correlation strengths.

Analyzing the system-size scaling of stochastic errors in AD-AFQMC estimates is crucial to assess their applicability in large systems. For a local observable, it is desirable that the stochastic error not grow with system size, since the observable itself does not scale. In figure 4, we show the stochastic error in the occupation of a hydrogen $1s$ STO-6G atomic orbital in the middle of a hydrogen chain for $d = 1.6$ a.u. and $d = 2.4$ a.u. for the same amount of sampling effort. Remarkably, it is

observed that the stochastic error does not increase with system size asymptotically. While the stochastic error in energy increases as $\sqrt{N_{\text{Hydrogens}}}$, the noise in the energy derivative does not increase. This behavior has previously been reported for correlated sampling estimators in 52.

In figure 5, we show the cost (wall time, including compilation time) of forward and reverse mode AD-AFQMC 1-RDM calculations relative to that of energy. As expected, the reverse mode AD costs are 3 to 4 times as expensive as the energy calculation for all chain lengths, while the forward mode AD costs are 2 to 3 times as expensive. We also show the size scaling of energy calculation cost, which is seen to be between quadratic and cubic. We note that computational cost scaling of the backpropagation algorithm for the calculation of properties and forces is also similarly favorable.^{33,50}

3.3. Dipole moments

In table 1, we show a comparison of dipole moments calculated using AD-AFQMC with those calculated using

TABLE 1. Dipole moments (in a.u.) of molecules at equilibrium geometries using the aug-cc-pVTZ basis set. BP-AFQMC refers to estimates obtained using backpropagation in AFQMC with partial path restoration.

Molecule	MP2	CCSD	BP-AFQMC	AD-AFQMC	CCSD(T)	Experiment
H ₂ O	0.7298	0.7335	0.707(4)	0.720(2)	0.7247	0.730(2) ⁶²
NH ₃	0.5996	0.6015	0.578(6)	0.592(2)	0.5938	0.581(1) ⁶³
HCl	0.4389	0.4318	0.400(5)	0.429(1)	0.4273	0.430 ⁶⁴
HBr	0.3379	0.3289	0.33(1)	0.329(2)	0.3245	0.325 ⁶⁴
CO	0.1050	0.0199	0.12(1)	0.019(4)	0.0429	0.048(1) ⁶⁵
CH ₂ O	0.9375	0.9666	0.82(1)	0.965(5)	0.9389	0.918(1) ⁶⁶
C ₄ H ₅ N	0.7354	0.7241	-	0.730(9)	0.7255	0.72(2) ⁶⁷

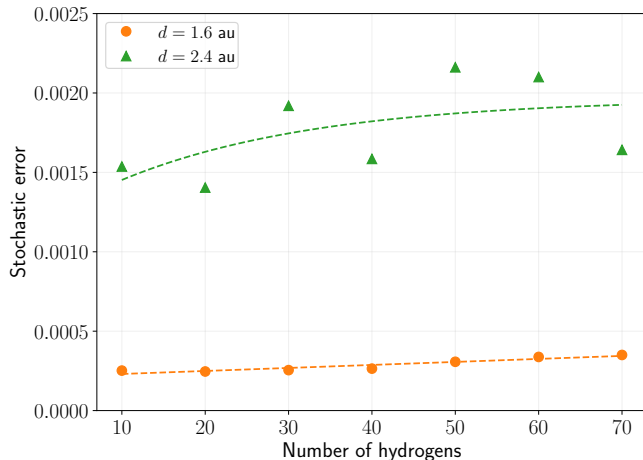


FIG. 4. Stochastic error in the occupation number in an atomic orbital in the middle of a hydrogen chain (STO-6G basis).

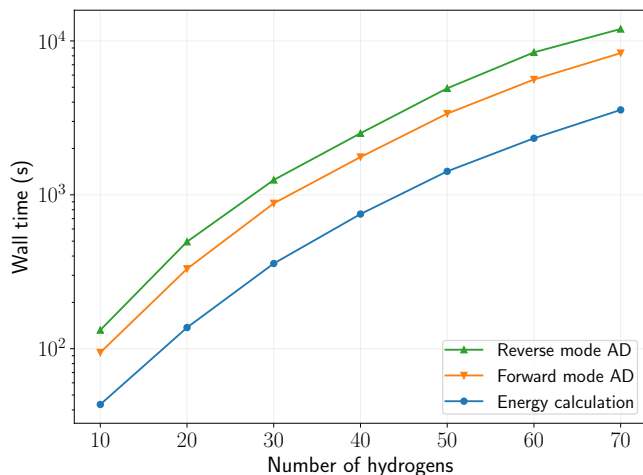


FIG. 5. Computational time scaling of AD-AFQMC calculations for hydrogen chains of different lengths with the same sampling effort. Reverse mode AD times correspond to the 1-RDM calculations shown in figure 3, while forward mode AD times correspond to the occupation number calculations in figure 4.

backpropagation (BP-AFQMC) and with conventional quantum chemistry methods. We use the aug-cc-pVTZ basis which is large enough to allow meaningful comparison with experimental values. The equilibrium geometries used here are provided in reference 61. For pyrrole, we used the geometry reported in reference 68. We include orbital relaxation contributions and use the frozen-core approximation in all methods. Orbital relaxation contributions are not relevant for BP-AFQMC.

Since these molecules are largely single reference, CCSD(T) values are expected to be very accurate. We see that AD-AFQMC dipole moments are in better agreement with CCSD(T) values than CCSD in most cases. They are also close to experimental dipole moments with some residual basis set error. This is an encouraging sign considering that AFQMC has a less steep cost scaling compared to the CC methods. Results obtained with BP-AFQMC using partial path restoration are significantly less accurate than AD-AFQMC for CO and CH₂O. While the sampling effort in BP and AD-AFQMC is not directly comparable, we used a similar allocation of computational resources for both sets of calculations. Note that one can use better trial states to obtain improved BP estimates as demonstrated in reference 33. Details about the convergence of BP-AFQMC are provided in the supporting information.

We see greater errors in both AD-AFQMC and CCSD dipole moments of CO and CH₂O molecules. To mitigate this error, we use the self-consistent AFQMC approach.^{69,70} We calculate the 1-RDM using reverse-mode AD-AFQMC and use the corresponding natural orbitals in the trial state of the subsequent AFQMC calculation. This procedure is repeated until the desired convergence threshold. Note that we include orbital relaxation contributions in the zeroth iteration to obtain an accurate 1-RDM initially, but not in the subsequent iterations. The results of this procedure for CO and CH₂O are shown in figure 6. In both cases, the self-consistent iterations can be seen to improve the dipole accuracy with convergence achieved in less than five steps. Availability of 1-RDM from AD-AFQMC can thus enable systematic improvement in AFQMC properties. We also show the results of a single AD-AFQMC calculation using DFT-B3LYP as the trial state without orbital relaxation. These dipole moments show better agreement with the CCSD(T) value suggesting that these orbitals are optimal

for AFQMC calculations. Thus it may be possible to devise orbital optimization schemes other than the one used in self-consistent AFQMC to obtain better trial states for property calculations.

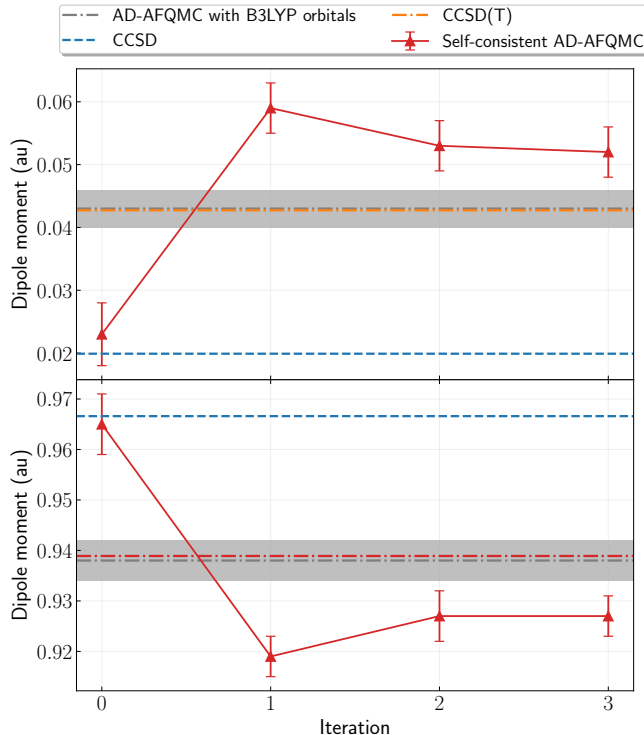


FIG. 6. Convergence of the dipole moment of CO (top panel) and CH₂O (bottom panel) with the number of self-consistent AFQMC iterations using the aug-ccpVTZ basis. Note that self-consistency performs worse than using B3LYP orbitals.

4. CONCLUSION

In this work, we investigated the utility of automatic differentiation (AD) for calculating response properties in auxiliary field quantum Monte Carlo (AFQMC) with the aim of achieving accuracy and computational cost comparable with energy calculations. We studied various contributions to energy derivatives and their convergence behavior. Our analysis demonstrated the accuracy of AD-AFQMC for computing one-particle reduced density matrices (1-RDM) in hydrogen chains, achieved at a computational cost no more than four times that of the energy calculation. We found that the stochastic error of the AD-AFQMC estimator for a local orbital density does not increase with system size. Finally, dipole moments of several molecules calculated using AD-AFQMC were found to be in good agreement with CCSD(T) and experiments. We were able to further improve the accuracy of these dipole moments by employing self-consistent AFQMC calculations, and alternatively by using trial states from DFT.

A way to improve AD-AFQMC involves addressing the bias due to stochastic reconfiguration. While for the systems studied in this paper, the SR bias was found to be small, it is possible that in systems where there are large fluctuations in the weights, this bias can be significant. We plan to explore bias mitigation techniques in future work. Another avenue for improvement lies in using more accurate trial states. In this work, we used a single determinant trial state, but it is possible to use multi-determinant trial states. Furthermore, to enable calculations on much larger systems, advancements in AD implementations are necessary. Developing more sophisticated checkpointing schemes may help manage the memory usage of reverse-mode AD, allowing for more extensive calculations without excessive memory constraints. Overall, these potential enhancements will contribute to the broader applicability of AD-AFQMC.

Despite these challenges, we believe that AD-AFQMC presents a viable and promising approach for calculating properties in AFQMC. A natural extension of this work is the calculation of nuclear gradients using AD-AFQMC, which can be accomplished by combining our code with differentiable molecular integral evaluators. The low cost-scaling and high-accuracy of AD-AFQMC open up possibilities for extremely accurate molecular dynamics simulations.

DATA AVAILABILITY

Code used for AD-AFQMC calculations is available in a public GitHub repository at reference 59, and the input and output files for these calculations are also available at reference 61.

ACKNOWLEDGEMENTS

We would like to thank Matthew Foulkes, Shiwei Zhang, and Miguel Morales for helpful discussions. AM and JSK were supported by NSF Career award CHE-2145209. SS was supported by a grant from the Camille and Henry Dreyfus Foundation. DRR acknowledges funding support from NSF CHE-2245592. This work utilized resources from the University of Colorado Boulder Research Computing Group, which is supported by the National Science Foundation (Award Nos. ACI-1532235 and ACI-1532236), the University of Colorado Boulder, and Colorado State University. Some calculations were also performed at the Advanced Research Computing at Hopkins (ARCH) core facility through allocation CHE230028 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

REFERENCES

- ¹ Pulay, P. Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules: I. Theory. *Molecular Physics* **1969**, *17*, 197–204.
- ² Pople, J.; Krishnan, R.; Schlegel, H.; Binkley, J. S. Derivative studies in hartree-fock and møller-plesset theories. *International Journal of Quantum Chemistry* **1979**, *16*, 225–241.
- ³ Handy, N. C.; Schaefer III, H. F. On the evaluation of analytic energy derivatives for correlated wave functions. *The Journal of chemical physics* **1984**, *81*, 5031–5033.
- ⁴ Salter, E.; Trucks, G. W.; Bartlett, R. J. Analytic energy derivatives in many-body methods. I. First derivatives. *The Journal of Chemical Physics* **1989**, *90*, 1752–1766.
- ⁵ Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent advances in wave function-based methods of molecular-property calculations. *Chemical reviews* **2012**, *112*, 543–631.
- ⁶ Griewank, A.; Walther, A. *Evaluating derivatives: principles and techniques of algorithmic differentiation*; SIAM, 2008.
- ⁷ Ekström, U.; Visscher, L.; Bast, R.; Thorvaldsen, A. J.; Ruud, K. Arbitrary-order density functional response theory from automatic differentiation. *Journal of chemical theory and computation* **2010**, *6*, 1971–1980.
- ⁸ Bast, R.; Ekström, U.; Gao, B.; Helgaker, T.; Ruud, K.; Thorvaldsen, A. J. The ab initio calculation of molecular electric, magnetic and geometric properties. *Physical Chemistry Chemical Physics* **2011**, *13*, 2627–2651.
- ⁹ Abbott, A. S.; Abbott, B. Z.; Turney, J. M.; Schaefer III, H. F. Arbitrary-order derivatives of quantum chemical methods via automatic differentiation. *The journal of physical chemistry letters* **2021**, *12*, 3232–3239.
- ¹⁰ Zhang, X.; Chan, G. K.-L. Differentiable quantum chemistry with PySCF for molecules and materials at the mean-field level and beyond. *The Journal of Chemical Physics* **2022**, *157*, 204801.
- ¹¹ Kasim, M. F.; Lehtola, S.; Vinko, S. M. DQC: A Python program package for differentiable quantum chemistry. *The Journal of chemical physics* **2022**, *156*.
- ¹² Liao, H.-J.; Liu, J.-G.; Wang, L.; Xiang, T. Differentiable programming tensor networks. *Physical Review X* **2019**, *9*, 031041.
- ¹³ Kalos, M. H.; Levesque, D.; Verlet, L. Helium at zero temperature with hard-sphere and other forces. *Physical Review A* **1974**, *9*, 2178.
- ¹⁴ Ceperley, D.; Chester, G. V.; Kalos, M. H. Monte Carlo simulation of a many-fermion study. *Physical Review B* **1977**, *16*, 3081.
- ¹⁵ Ceperley, D. M.; Alder, B. J. Ground state of the electron gas by a stochastic method. *Physical review letters* **1980**, *45*, 566.
- ¹⁶ Nightingale, M. P.; Umrigar, C. J. *Quantum Monte Carlo methods in physics and chemistry*; Springer Science & Business Media, 1998.
- ¹⁷ Foulkes, W.; Mitas, L.; Needs, R.; Rajagopal, G. Quantum Monte Carlo simulations of solids. *Reviews of Modern Physics* **2001**, *73*, 33.
- ¹⁸ Becca, F.; Sorella, S. *Quantum Monte Carlo approaches for correlated systems*; Cambridge University Press, 2017.
- ¹⁹ Zhang, S.; Carlson, J.; Gubernatis, J. E. Constrained path quantum Monte Carlo method for fermion ground states. *Physical review letters* **1995**, *74*, 3652.
- ²⁰ Zhang, S.; Carlson, J.; Gubernatis, J. E. Constrained path Monte Carlo method for fermion ground states. *Physical Review B* **1997**, *55*, 7464.
- ²¹ Zhang, S.; Krakauer, H. Quantum Monte Carlo method using phase-free random walks with Slater determinants. *Physical review letters* **2003**, *90*, 136401.
- ²² Al-Saidi, W.; Zhang, S.; Krakauer, H. Auxiliary-field quantum Monte Carlo calculations of molecular systems with a Gaussian basis. *The Journal of chemical physics* **2006**, *124*, 224101.
- ²³ Suewattana, M.; Purwanto, W.; Zhang, S.; Krakauer, H.; Walter, E. J. Phaseless auxiliary-field quantum Monte Carlo calculations with plane waves and pseudopotentials: Applications to atoms and molecules. *Physical Review B* **2007**, *75*, 245123.
- ²⁴ Hao, H.; Shee, J.; Upadhyay, S.; Ataca, C.; Jordan, K. D.; Rubenstein, B. M. Accurate predictions of electron binding energies of dipole-bound anions via quantum Monte Carlo methods. *The journal of physical chemistry letters* **2018**, *9*, 6185–6190.
- ²⁵ Lee, J.; Malone, F. D.; Morales, M. A. Utilizing essential symmetry breaking in auxiliary-field quantum Monte Carlo: Application to the spin gaps of the C36 fullerene and an iron porphyrin model complex. *Journal of chemical theory and computation* **2020**, *16*, 3019–3027.
- ²⁶ Malone, F. D.; Mahajan, A.; Spencer, J. S.; Lee, J. ipie: A Python-Based Auxiliary-Field Quantum Monte Carlo Program with Flexibility and Efficiency on CPUs and GPUs. *Journal of Chemical Theory and Computation* **2022**, *19*, 109–121.
- ²⁷ Shee, J.; Arthur, E. J.; Zhang, S.; Reichman, D. R.; Friesner, R. A. Singlet-triplet energy gaps of organic biradicals and polyacenes with auxiliary-field quantum Monte Carlo. *Journal of chemical theory and computation* **2019**, *15*, 4924–4932.
- ²⁸ Badinski, A.; Haynes, P.; Trail, J.; Needs, R. Methods for calculating forces within quantum Monte Carlo simulations. *Journal of Physics: Condensed Matter* **2010**, *22*, 074202.
- ²⁹ Casalegno, M.; Mella, M.; Rappe, A. M. Computing accurate forces in quantum Monte Carlo using Pulay’s corrections and energy minimization. *The Journal of Chemical Physics* **2003**, *118*, 7193–7201.
- ³⁰ Filippi, C.; Umrigar, C. Correlated sampling in quantum Monte Carlo: A route to forces. *Physical Review B* **2000**, *61*, R16291.
- ³¹ Moroni, S.; Saccani, S.; Filippi, C. Practical schemes for accurate forces in quantum Monte Carlo. *Journal of chemical theory and computation* **2014**, *10*, 4823–4829.
- ³² Assaraf, R.; Caffarel, M. Zero-variance zero-bias principle for observables in quantum Monte Carlo: Application to forces. *The Journal of Chemical Physics* **2003**, *119*, 10536–10552.
- ³³ Motta, M.; Zhang, S. Computation of ground-state properties in molecular systems: Back-propagation with auxiliary-field quantum Monte Carlo. *Journal of chemical theory and computation* **2017**, *13*, 5367–5378.
- ³⁴ Thomas, R. E.; Opalka, D.; Overly, C.; Knowles, P. J.; Alavi, A.; Booth, G. H. Analytic nuclear forces and molecular properties from full configuration interaction quantum Monte Carlo. *The Journal of Chemical Physics* **2015**, *143*.
- ³⁵ Sorella, S.; Capriotti, L. Algorithmic differentiation and the calculation of forces by quantum Monte Carlo. *The*

- Journal of chemical physics* **2010**, *133*.
- 36 Poole, T. Calculating derivatives within quantum Monte Carlo. **2014**,
 - 37 Zhang, S.-X.; Wan, Z.-Q.; Yao, H. Automatic differentiable Monte Carlo: theory and application. *arXiv preprint arXiv:1911.09117* **2019**,
 - 38 Glasserman, P. *Monte Carlo methods in financial engineering*; Springer, 2004; Vol. 53.
 - 39 Glynn, P. W. Likelihood ratio gradient estimation for stochastic systems. *Communications of the ACM* **1990**, *33*, 75–84.
 - 40 Bengio, Y.; Léonard, N.; Courville, A. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432* **2013**,
 - 41 Mohamed, S.; Rosca, M.; Figurnov, M.; Mnih, A. Monte carlo gradient estimation in machine learning. *The Journal of Machine Learning Research* **2020**, *21*, 5183–5244.
 - 42 Motta, M.; Zhang, S. Ab initio computations of molecular systems by the auxiliary-field quantum Monte Carlo method. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2018**, *8*, e1364.
 - 43 Shi, H.; Zhang, S. Some recent developments in auxiliary-field quantum Monte Carlo for real materials. *The Journal of Chemical Physics* **2021**, *154*, 024107.
 - 44 Lee, J.; Pham, H. Q.; Reichman, D. R. Twenty years of auxiliary-field quantum Monte Carlo in quantum chemistry: An overview and assessment on main group chemistry and bond-breaking. *Journal of Chemical Theory and Computation* **2022**, *18*, 7024–7042.
 - 45 Buonaura, M. C.; Sorella, S. Numerical study of the two-dimensional Heisenberg model using a Green function Monte Carlo technique with a fixed number of walkers. *Physical Review B* **1998**, *57*, 11446.
 - 46 Whitlock, P.; Ceperley, D.; Chester, G.; Kalos, M. Properties of liquid and solid He 4. *Physical Review B* **1979**, *19*, 5598.
 - 47 Ceperley, D. M.; Kalos, M. *Monte Carlo methods in statistical physics*; Springer, 1986; pp 145–194.
 - 48 Rothstein, S. M. A survey on pure sampling in quantum Monte Carlo methods. *Canadian Journal of Chemistry* **2013**, *91*, 505–510.
 - 49 Mahajan, A.; Lee, J.; Sharma, S. Selected configuration interaction wave functions in phaseless auxiliary field quantum Monte Carlo. *The Journal of Chemical Physics* **2022**, *156*.
 - 50 Motta, M.; Zhang, S. Communication: Calculation of interatomic forces and optimization of molecular geometry with auxiliary-field quantum Monte Carlo. *The Journal of Chemical Physics* **2018**, *148*.
 - 51 Shee, J.; Zhang, S.; Reichman, D. R.; Friesner, R. A. Chemical transformations approaching chemical accuracy via correlated sampling in auxiliary-field quantum Monte Carlo. *Journal of Chemical Theory and Computation* **2017**, *13*, 2667–2680.
 - 52 Assaraf, R.; Caffarel, M.; Kollias, A. Chaotic versus non-chaotic stochastic dynamics in monte carlo simulations: A route for accurate energy differences in n-body systems. *Physical Review Letters* **2011**, *106*, 150601.
 - 53 Chen, S.; Yang, Y.; Morales, M.; Zhang, S. Algorithm for branching and population control in correlated sampling. *arXiv preprint arXiv:2307.15203* **2023**,
 - 54 Gerratt, J.; Mills, I. M. Force constants and dipole-moment derivatives of molecules from perturbed Hartree–Fock calculations. I. *The Journal of Chemical Physics* **1968**, *49*, 1719–1729.
 - 55 Schulman, J.; Heess, N.; Weber, T.; Abbeel, P. Gradient estimation using stochastic computation graphs. *Advances in neural information processing systems* **2015**, *28*.
 - 56 Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, K.-L. G. PySCF: the Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1340.
 - 57 Zhai, H.; Chan, G. K. Low communication high performance ab initio density matrix renormalization group algorithms. *The Journal of Chemical Physics* **2021**, *154*.
 - 58 Kim, J.; Baczewski, A. D.; Beaudet, T. D.; Benali, A.; Bennett, M. C.; Berrill, M. A.; Blunt, N. S.; Borda, E. J. L.; Casula, M.; Ceperley, D. M., et al. QMCPACK: an open source ab initio quantum Monte Carlo package for the electronic structure of atoms, molecules and solids. *Journal of Physics: Condensed Matter* **2018**, *30*, 195901.
 - 59 https://github.com/ankit76/ad_afqmc/.
 - 60 Bradbury, J.; Frostig, R.; Hawkins, P.; Johnson, M. J.; Leary, C.; Maclaurin, D.; Necula, G.; Paszke, A.; VanderPlas, J.; Wanderman-Milne, S.; Zhang, Q. JAX: composable transformations of Python+NumPy programs. <http://github.com/google/jax>, 2018.
 - 61 https://github.com/ankit76/ad_afqmc_data.
 - 62 Lide, D. R. *CRC handbook of chemistry and physics*; CRC press, 2004; Vol. 85.
 - 63 Shimizu, F. Stark Spectroscopy of NH₃ ν 2 Band by 10- μ CO₂ and N₂O Lasers. *The Journal of Chemical Physics* **1970**, *52*, 3572–3576.
 - 64 Lovas, F.; Tiemann, E.; Coursey, J.; Kotochigova, S.; Chang, J.; Olsen, K.; Dragoset, R. Diatomic spectral database, NIST standard reference database 114. 2005.
 - 65 Meerts, W.; De Leeuw, F.; Dymanus, A. Electric and magnetic properties of carbon monoxide by molecular-beam electric-resonance spectroscopy. *Chemical Physics* **1977**, *22*, 319–324.
 - 66 Theulé, P.; Callegari, A.; Rizzo, T. R.; Muentner, J. Fluorescence detected microwave Stark effect measurements in excited vibrational states of H 2 CO. *The Journal of chemical physics* **2003**, *119*, 8910–8915.
 - 67 Nelson, R. D.; Lide, D. R.; Maryott, A. A., et al. *Selected values of electric dipole moments for molecules in the gas phase*; US National Bureau of Standards Washington, DC, 1967; Vol. 10.
 - 68 Loos, P.-F.; Lipparini, F.; Boggio-Pasqua, M.; Scemama, A.; Jacquemin, D. A mountaineering strategy to excited states: Highly accurate energies and benchmarks for medium sized molecules. *Journal of Chemical Theory and Computation* **2020**, *16*, 1711–1741.
 - 69 Qin, M.; Shi, H.; Zhang, S. Coupling quantum Monte Carlo and independent-particle calculations: Self-consistent constraint for the sign problem based on the density or the density matrix. *Physical Review B* **2016**, *94*, 235119.
 - 70 Qin, M. Self-consistent optimization of the trial wave function within the constrained path auxiliary field quantum Monte Carlo method using mixed estimators. *Physical Review B* **2023**, *107*, 235124.