

PAPER

Ensemble variational Monte Carlo for optimization of correlated excited state wave functions

To cite this article: William A Wheeler *et al* 2024 *Electron. Struct.* **6** 025001

View the [article online](#) for updates and enhancements.

You may also like

- [Research on Extreme Points of RSI Expert System Based on Nonlinear Regression Analysis of Data](#)
Pin Wang and Feng Su
- [Sequential annealing-gradient Gamma-Knife radiosurgery optimization](#)
Roger Ove and Richard Popple
- [Regularized image denoising based on spectral gradient optimization](#)
Tibor Luki, Joakim Lindblad and Nataša Sladoje



PAPER

Ensemble variational Monte Carlo for optimization of correlated excited state wave functions

William A Wheeler¹ , Kevin G Kleiner² and Lucas K Wagner^{2,*} ¹ Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States of America² The Anthony J Leggett Institute for Condensed Matter Theory, Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States of America

* Author to whom any correspondence should be addressed.

E-mail: lkwagner@illinois.edu**Keywords:** optimization, excited states, variational Monte Carlo, correlated wave functions**Abstract**

Variational Monte Carlo methods have recently been applied to the calculation of excited states; however, it is still an open question what objective function is most effective. A promising approach is to optimize excited states using a penalty to minimize overlap with lower eigenstates, which has the drawback that states must be computed one at a time. We derive a general framework for constructing objective functions with minima at the the lowest N eigenstates of a many-body Hamiltonian. The objective function uses a weighted average of the energies and an overlap penalty, which must satisfy several conditions. We show this objective function has a minimum at the exact eigenstates for a finite penalty, and provide a few strategies to minimize the objective function. The method is demonstrated using *ab initio* variational Monte Carlo to calculate the degenerate first excited state of a CO molecule.

1. Introduction

Quantum Monte Carlo (QMC) has been established as a highly accurate method for computing ground-states of *ab initio* systems [1–5]. Progress has also been made towards efficiently computing accurate excited states in QMC [6–10]; however, there are limitations to the methods that have been applied to *ab initio* systems so far. A state-averaging approach has been used to simultaneously optimize a set of ground and excited states, where a common set of orbitals and Jastrow are used to construct all the states [11]. In many systems, optimal orbitals for describing an excited state are quite different from the optimal ground-state orbitals, so large determinant expansions are required to accurately represent these states using a shared set of orbitals. An optimization approach using variance of the local energy has been used to target excited states without state averaging [12–15]; however, relying on the variance leads to difficulty optimizing to the correct minimum in some situations [16]. An alternative method avoids state averaging and variance optimization by penalizing the wave function's overlap with a set of known eigenstates [17, 18]. This method only optimizes one excited state at a time, requiring several expensive calculations in sequence.

We note that there has been some work that optimizes all states simultaneously [19, 20] using a modification of the method in [17, 18]. Additionally, a recently proposed method simultaneously optimizes excited states by representing the ensemble as a determinant of many-body wave functions, avoiding the need to choose an overlap penalty [21]. In the approaches of [19, 21], there is a degenerate minimum in which all wave functions are linear combinations of the eigenstates [22]. While in principle one could rediagonalize the Hamiltonian in the subspace, this is inconvenient in variational Monte Carlo (VMC) since the wave function parameterizations are often nonlinear and the off-diagonal matrix elements of the Hamiltonian typically have high variance. Finally, [20] includes an overlap penalty between states i and j only if $i < j$, effectively

implementing the sequential optimization of [17, 18] in a single calculation by using separate, dynamic cost functionals for each state.

In this paper, we define a single, penalty-based cost functional to optimize an ensemble of states simultaneously. For a finite, given value of penalty (no free parameters), the functional has the eigenstates as its global minimum, with Hessian eigenvalues all greater than 0. Using a weighted average of energies, as proposed in [23], this cost functional yields the low energy eigenstates, and the Hamiltonian does not need to be rediagonalized in the final low-energy subspace. We derive conditions on the optimization parameters that ensure the ensemble optimizes to the lowest-energy eigenstates. As a practical demonstration, we optimize this cost functional to find excited states of a CO molecule. We believe that strategies based on this functional are likely to be efficient and practical.

2. Penalty-based cost functional

Summary and statement. We aim to find the lowest N eigenstates of a time-independent Hamiltonian whose eigenstates $|\Phi_i\rangle$ satisfy $\hat{H}|\Phi_i\rangle = E_i|\Phi_i\rangle$. We define an N -state functional of an arbitrary ensemble of states $\{\Psi_i\}$,

$$O[\{\Psi_i\}] = \sum_i w_i E[\Psi_i] + \lambda \sum_{i < j} |S_{ij}|^2, \quad (1)$$

where $E[\Psi_i] = \langle \Psi_i | \hat{H} | \Psi_i \rangle$ is the energy expectation, $S_{ij} = \langle \Psi_i | \Psi_j \rangle$ is the overlap, $w_i > 0$ are weights, and $\lambda > 0$ penalizes wave function overlap. O is minimized by the set of the lowest N eigenstates $\{\Phi_i\}$ of $\hat{H}$ under the following conditions: (1) $w_i > w_j$ for all i, j where $E_i < E_j$, and (2)

$$\lambda > \max_{i < j} \left[(E_j - E_i) \frac{w_i w_j}{w_i - w_j} \right]. \quad (2)$$

We prove this condition in sections 2.1 and 2.2. There are many ways of choosing the weights w_i ; in section 2.3.1, we discuss one strategy for choosing weights, as well as deriving the algorithm of [20] by choosing a particular limit of the weights approaching zero.

2.1. Preliminaries

If the states $\{\Psi_i\}$ are orthogonal, a few properties are known. An orthogonal ensemble satisfies the variational upper bound property [23]

$$\sum_i w_i E[\Psi_i] \geq \sum_i w_i E[\Phi_i]. \quad (3)$$

If the weights are all equal, every set of orthogonal states in the low-energy space $\mathcal{LE}_N = \text{span}[\Phi_0, \dots, \Phi_{N-1}]$ has the same cost [22], resulting in a continuum of critical points. More generally, if the weights in any subset $\{w_j\}$ are equal, then any rotation of the corresponding states $\{\Phi_j\}$ leaves the cost functional unchanged. We therefore require the weights to be strictly decreasing between nondegenerate states, $w_i > w_{i+1}$ if $E_i \neq E_{i+1}$, to ensure that an orthogonal ensemble optimizes to the target ensemble $\{\Phi_i\}$ [23]. Of course, within a degenerate subspace, any rotation of states leaves the energy unchanged and is an acceptable solution.

Now we turn to the case of $\{\Psi_i\}$ that are allowed to be nonorthogonal, and the minimization of O . Ψ_i are orthogonal in the limit $\lambda \rightarrow \infty$. At the opposite extreme, if $\lambda = 0$, O is minimized when all states are the ground state. We will show that there is a critical value λ_c , above which O is minimized by the ensemble of eigenstates $\{\Phi_i\}$.

We consider an ensemble $\{\Psi_i\}$ within the low-energy space $\mathcal{LE}_N$. For any choice of w_i and λ , the ensemble that minimizes O is guaranteed to be in $\mathcal{LE}_N$, [23] since any components outside $\mathcal{LE}_N$ can be rotated into $\mathcal{LE}_N$ without increasing the overlap term. We thus can parameterize the ensemble states as linear combinations of eigenstates in $\mathcal{LE}_N$,

$$\Psi_i = \sum_{j=0}^{N-1} c_{ij} \Phi_j, \quad (4)$$

where $c_{ii} = \sqrt{1 - \sum_j |c_{ij}|^2}$ is determined by the normalization. Our target ensemble $\{\Phi_i\}$ is represented by the parameters $c_{ij} = \delta_{ij}$. In general, the cost functional with this parameterization is represented as

$$O = \sum_{ik} w_i |c_{ik}|^2 E_k + \lambda \sum_{i < j} \left| \sum_k c_{ik} c_{jk} \right|^2. \quad (5)$$

O is fully parameterized by c_{ij} with $i \neq j$, which are free to vary within $|c_{ij}| \leq 1$.

2.2. Determining critical λ_c

2.2.1. λ_c for two-state system

We will start by determining the critical λ_c for a two-state system such that the eigenstates $\{\Phi_0, \Phi_1\}$ minimize O when $\lambda > \lambda_c$. We do this by showing that the Hessian of the cost function is positive definite. The cost functional for two states is

$$O = \sum_{i=0}^{N-1} w_i E_i + w_0 |c_{01}|^2 \Delta E - w_1 |c_{10}|^2 \Delta E \quad (6)$$

$$+ \lambda |c_{00} c_{10} + c_{01} c_{11}|^2, \quad (7)$$

where $\Delta E = E_1 - E_0$. The gradient and Hessian of O give conditions for the critical penalty λ_c . The gradient is

$$\frac{\partial O}{\partial c_{01}} = 2w_0 c_{01} \Delta E + \lambda 2S_{01} \left(c_{11} - \frac{c_{01} c_{10}}{c_{00}} \right), \quad (8)$$

using the fact that $\frac{dc_{00}}{dc_{01}} = -c_{01}/c_{00}$. In this form, it is clear that our target ensemble $c_{ij} = \delta_{ij}$ is indeed a critical point. The Hessian matrix at the target ensemble $c_{ij} = \delta_{ij}$ is

$$H_O = \begin{pmatrix} 2w_0 \Delta E + 2\lambda & 2\lambda \\ 2\lambda & -2w_1 \Delta E + 2\lambda \end{pmatrix}, \quad (9)$$

where the rows and columns correspond to the free parameters c_{01} and c_{10} . It has eigenvalues

$$\alpha = 2\lambda + (w_0 - w_1) \Delta E \pm \sqrt{(w_0 + w_1)^2 \Delta E^2 + 4\lambda^2}. \quad (10)$$

This critical point is a minimum when the eigenvalues are both positive, which constrains λ to

$$\lambda > \lambda_c = \Delta E \frac{w_0 w_1}{w_0 - w_1}. \quad (11)$$

From the denominator, we see that if the weights are equal, $w_0 = w_1$, the critical $\lambda_c \rightarrow \infty$; in this case, no choice of λ will minimize O at the target ensemble. In the other limit $w_1 \rightarrow 0$, we have $\lambda_c \rightarrow 0$. We expect the optimal choice of weights and λ to be intermediate for efficient optimization.

We verify the condition of equation (11) by optimizing the cost functional in a two-state Hilbert space using different values of λ and weights (w_0, w_1) . The results are shown in figure 1. The penalty $\lambda/\Delta E$ is plotted against the first weight w_0 , and the weights are normalized, $w_0 + w_1 = 1$. The blue line shows the critical condition $\lambda_c/\Delta E = w_0 w_1/(w_1 - w_0)$. As expected, the overlap is zero when $\lambda > \lambda_c$, and is nonzero when $\lambda < \lambda_c$. As $\lambda \rightarrow 0$, the overlap goes to one, as both states optimize to the ground state.

2.2.2. λ_c for N -state system

We now generalize to an N -state system, again by showing that the Hessian of the cost function is positive definite. For an ensemble of N states, the cost functional has the form given in equation (5). We set $c_{ii}^2 = 1 - \sum_j |c_{ij}|^2$. The gradients take the form

$$\frac{\partial O}{\partial c_{ij}} = 2w_i c_{ij} (E_j - E_i) + \sum_{k \neq i} 2\lambda S_{ik} \left(c_{kj} - \frac{c_{ij} c_{ki}}{c_{ii}} \right). \quad (12)$$

The derivative is zero when $\Psi_i = \Phi_i$, since $c_{ij} = \delta_{ij}$, and $S_{ij} = \delta_{ij}$.

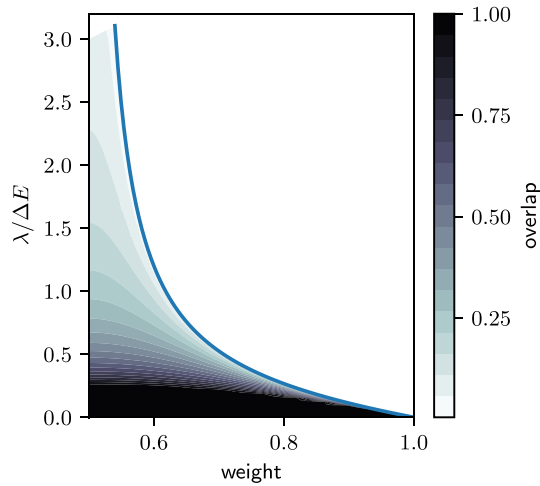


Figure 1. Overlap of the optimized ensembles in the two-state model system at different values of penalty λ and weight w_0 . The weights are normalized so that $w_0 + w_1 = 1$. The blue line is the calculated critical penalty λ_c , above which the overlap is strictly zero.

We now check that the Hessian matrix evaluated at the target ensemble state has positive eigenvalues. The Hessian matrix is made up of the second derivatives $\partial^2 O / \partial c_{ij} \partial c_{kl}$, where $i \neq j$ and $k \neq l$. We start by identifying the nonzero off-diagonal terms, where $(i, j) \neq (k, l)$:

$$H_O = 2\lambda \sum_{m \neq i} \left[\frac{\partial S_{im}}{\partial c_{kl}} \left(c_{mj} - \frac{c_{ij} c_{mi}}{c_{ii}} \right) \right. \quad (13)$$

$$\left. + S_{im} \frac{\partial}{\partial c_{kl}} \left(c_{mj} - \frac{c_{ij} c_{mi}}{c_{ii}} \right) \right] \quad (14)$$

$$= 2\lambda \sum_{m \neq i} \frac{\partial S_{im}}{\partial c_{kl}} \delta_{mj} \quad (15)$$

$$= 2\lambda \frac{\partial}{\partial c_{kl}} \sum_n c_{in} c_{jn} \quad (16)$$

$$= 2\lambda (\delta_{ik} c_{jl} + \delta_{jk} c_{il}), \quad (17)$$

where we are at the target ensemble, $c_{ij} = \delta_{ij}$. We see that the only nonzero off-diagonal term is $\partial^2 O / \partial c_{ij} \partial c_{ji}$, where $(k, l) = (j, i)$. These off-diagonal terms are the result of two-state mixing, the situation described in the previous section. Thus, we can apply the result of equation (11) and the critical λ_c satisfies the condition

$$\lambda_c = \max_{i < j} \left[(E_j - E_i) \frac{w_i w_j}{w_i - w_j} \right], \quad (18)$$

and when $\lambda > \lambda_c$, the target ensemble is a minimum of O .

2.3. Choosing weights

2.3.1. Equal critical λ

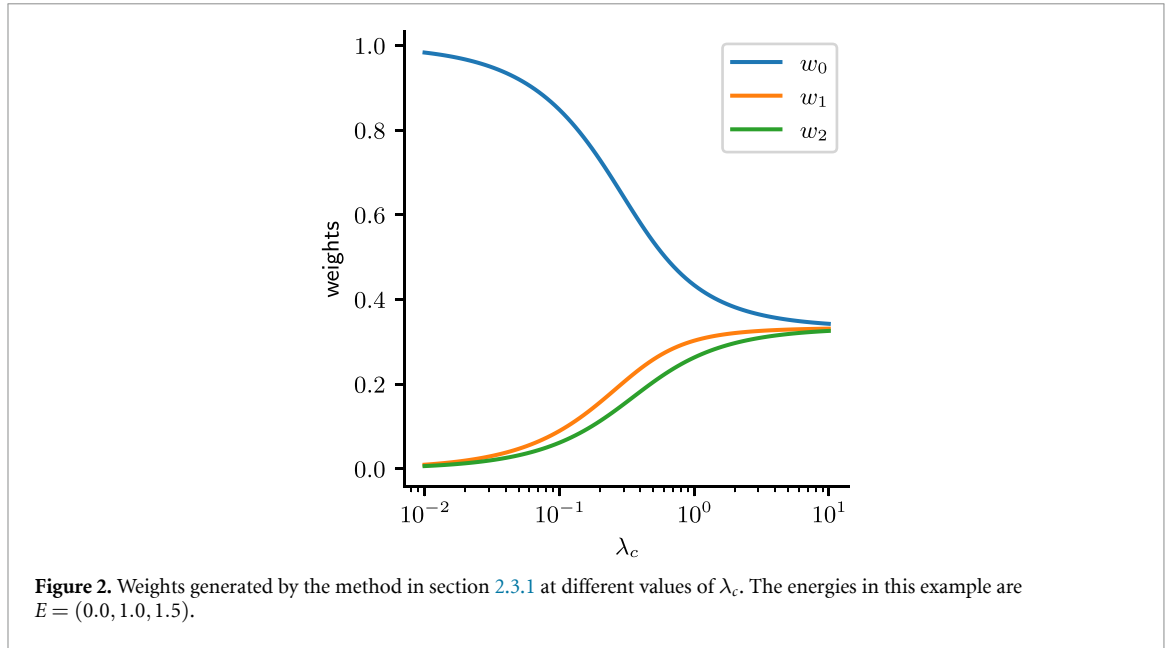
In the sections above, we derived the critical value λ_c that ensures the cost functional is minimized by eigenstates given a set of weights $\{w_i\}$. As defined in equation (18), λ_c is the maximum of the critical values for each pair of states. One possible choice of weights is to fix all the pairwise critical values to be equal. This strategy has one free parameter, the critical value itself.

Consider the ratios $\tilde{w}_i = w_i / w_0$ and the energy differences $\tilde{E}_i = E_i - E_0$. The critical λ_c^{i0} for states i and 0 are expressed as

$$\frac{\lambda_c^{i0}}{w_0} = \tilde{E}_i \frac{\tilde{w}_i}{1 - \tilde{w}_i}. \quad (19)$$

If we set all the λ_c^{i0} equal to a chosen constant λ_c , the weight ratios are given by

$$\tilde{w}_i = \frac{1}{1 + \tilde{E}_i w_0 / \lambda_c}. \quad (20)$$



With the weights determined this way, the critical λ_c^{ij} for any pair (i, j) also have the same value λ_c ,

$$\frac{\lambda_c^{ij}}{w_0} = (\tilde{E}_i - \tilde{E}_j) \frac{\tilde{w}_i \tilde{w}_j}{\tilde{w}_j - \tilde{w}_i} \quad (21)$$

$$= (\tilde{E}_i - \tilde{E}_j) \frac{1}{1/\tilde{w}_i - 1/\tilde{w}_j} \quad (22)$$

$$= (\tilde{E}_i - \tilde{E}_j) \frac{1}{1 + \tilde{E}_i w_0 / \lambda_c - (1 + \tilde{E}_j w_0 / \lambda_c)} \quad (23)$$

$$= \frac{\lambda_c}{w_0}. \quad (24)$$

The value of w_0 is determined by normalizing the sum of the weights to one,

$$w_0 = \left(\sum_i \tilde{w}_i \right)^{-1}, \quad (25)$$

and can be solved for given a desired λ_c and an estimate of the energies E_i . In practice, the energies E_i must be estimated by an approximate method, since the optimized energies are not yet known. The weights affect the efficiency of the optimization, but not the optimal solution, so exact energies are not necessary. Note that it is not necessary for the weights to be normalized; however, only the ratio of the weights to λ impacts the cost function, so we normalize the weights for interpretability.

As an example, we show the weights generated by this method for a range of values of λ_c in figure 2. At the extreme of small λ_c , w_0 is close to one, and all other weights are near zero. This has the obvious problem that the excited states will not be well distinguished by the cost functional and will be dominated by stochastic noise from the ground state energy. At the other extreme of large λ_c , the weights tend towards being equal. As discussed above, equal weights cause the cost functional to have a degenerate minimum, where any rotation of the first N eigenstates has the same energy as the eigenstates themselves. This effect is expected once noise in the overlap term (magnified by the large λ) dominates the differences in the weighted energies, even when the weights are not exactly equal. In practice, we find that even deterministic optimization on simple models has difficulty reaching the minimum at these extremes of λ_c .

2.3.2. Simultaneous ‘stack of states’

Another strategy for optimization is to simultaneously optimize all states using the cost functions of [17, 18], where each state Ψ_j is orthogonalized only to lower states Ψ_i , $i < j$. This strategy, used in [20], is equivalent to the limit of a certain choice of weights in equation (1), which we derive in this section.

Consider setting

$$\lambda_{ij} = \alpha_{ij} w_j, \quad (26)$$

keeping in mind that the overlap sum in equation (1) is only over $i < j$, and where α_{ij} is a weighting variable chosen so that

$$\alpha_{ij} > \frac{E_j - E_i}{1 - w_j/w_i}. \quad (27)$$

Then $\lambda_{ij} > \lambda_c$. The derivative of O with respect to the parameter set p_j of wave function Ψ_j is then

$$\nabla_j O = w_j \left(\nabla_j E[\Psi_j] + \sum_{i < j} 2\alpha_{ij} S_{ij} \nabla_j S_{ij} \right) \quad (28)$$

$$+ \sum_{k > j} 2w_k \alpha_{jk} S_{jk} \nabla_j S_{jk}. \quad (29)$$

Now take the limit as $w_j/w_i \rightarrow 0$ while keeping the sum of weights equal to 1. This satisfies the ensemble criterion $w_i > w_j$ for $E_i < E_j$ and $w_i > 0$. In this limit, α_{ij} only needs to be greater than $E_j - E_i$. The gradients then become

$$\nabla_j O = w_j \left(\nabla_j E[\Psi_j] + \sum_{i < j} 2\alpha_{ij} S_{ij} \nabla_j S_{ij} \right), \quad (30)$$

which is equivalent to the algorithm recently proposed in [20]. In gradient-based optimization, the parameter updates Δp at each iteration are determined by the gradient $\nabla_j O$. A common form is $\Delta p_j = \tau F^{-1} \nabla_j O$, where F is a Fisher information matrix [17, 18] and τ is the descent step length that can be optimized using correlated sampling [24] or reduced on a schedule. In this version of the algorithm, where the parameter gradients are given by equation (30), we can consider the time step to be scaled by the weight, $\tilde{\tau} = \tau/w_j$. With the scaled time step $\tilde{\tau}$, the parameter updates are completely independent of the weights.

We have shown that a practical algorithm proposed in a heuristic manner actually rigorously converges to the eigenstates of the system. One should note that this choice of λ_{ij} hides the upper bound property of O (equation (3)), making comparisons between different wave function ansatzes challenging. However, one can choose to evaluate any objective function on the optimized wave functions with finite weights and λ_{ij} satisfying our λ_c criterion to recover the upper bound property.

3. Application to *ab initio* system

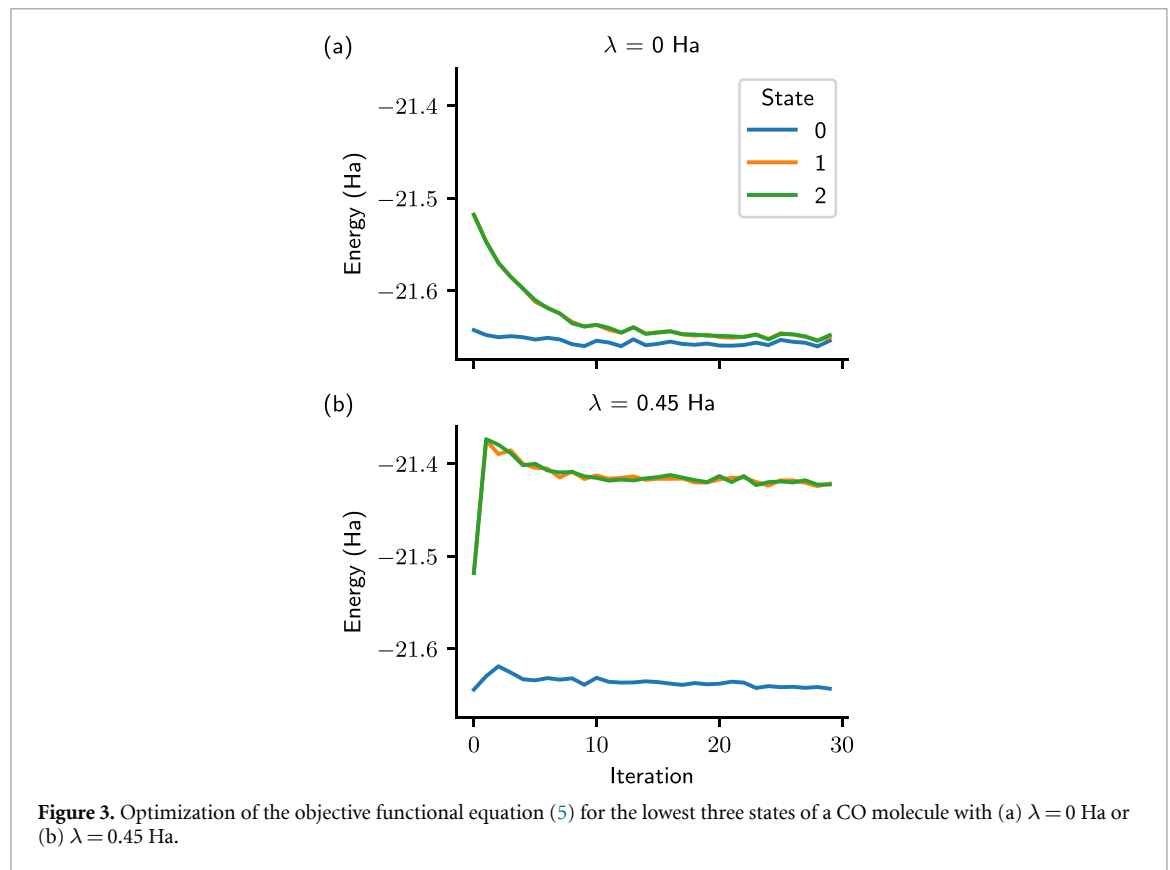
To demonstrate our penalty-based ensemble optimization on a simple, well-characterized system, we apply it to a CO molecule at the experimentally-determined equilibrium bond length of $r = 2.13$ Bohr [25]. We compute the ground state and degenerate first two excited states using VMC with different values of the overlap penalty λ , verifying the vertical excitation energy of absorption against experiment [25].

We represent our *ab initio* wave functions using a multi-Slater-Jastrow ansatz,

$$\Psi(\mathbf{R}) = e^{J(\mathbf{R};\alpha)} \sum_k c_k D_k(\mathbf{R};\beta), \quad (31)$$

where $\mathbf{R}$ represents all electron coordinates, J is a two-body Jastrow factor with parameters α , D_k are Slater determinants of one-body orbitals parameterized by β , and c_k are determinant expansion coefficients. Determinants and initial expansion coefficients for the lowest three states were generated from a complete active space configuration interaction (CASI) calculation with six electrons and six orbitals in the active space, giving a basis of 400 determinants. The orbitals were computed using Hartree Fock (HF) with correlation-consistent effective core potentials and corresponding triple- ζ basis functions [26]. The HF and CASI calculations were carried out in PySCF. We note that by using a Jastrow factor, we achieve accurate wave functions with orders of magnitude fewer determinants than typically needed in CASI calculations [24], allowing us to use unusually small CASI calculations in this step.

We initialize our three states for ensemble optimization (equation (1)) from the CASI determinant coefficients and a pre-optimized Jastrow factor. The Jastrow factor is initialized by a standard ground state optimization using the CASI ground state determinants; only the Jastrow parameters are optimized in this step. State 0 is initialized as the CASI ground state multiplied by the pre-optimized Jastrow. State 1 is initialized as an equal-weight linear combination of CASI states 0 and 1 multiplied by the same



pre-optimized Jastrow. State 2 is initialized similarly to state 1, using CASCI states 0 and 2. We start states 1 and 2 at superpositions of CASCI states to avoid the saddle points near eigenstates and ensure the method converges to the correct excited states, even starting far from them. We emphasize that these starting functions are constructed only from the orbitals and determinant coefficients of the CASCI calculations, and do not depend at all on the CASCI energies. The weights $\{w_i\}$ are set to (0.41, 0.29, 0.29) using equation (20) with $\lambda_c = 0.27$ Ha, the energy gap from the CASCI calculations. All Jastrow parameters α , orbital parameters β , and determinant coefficients c_k are optimized with 500 walkers. Vertical excitation energies are evaluated from the optimized wave functions using VMC with 10 000 walkers. The Jastrow optimization and all ensemble optimization calculations were carried out in PyQMC [24].

Ensemble optimization using our cost functional generates the expected solutions in the limiting values of λ . When $\lambda = 0$ Ha, all three states optimize to the ground state, shown in the upper plot of figure 3. All three energies decrease over the optimization and approach the same value. Our estimate for the critical penalty based on CASCI energies is $\lambda_c = 0.27$. When $\lambda = 0.45$ Ha, substantially larger than λ_c , state 0 optimizes to the ground state and states 1 and 2 optimize to the lowest two excited states, shown in the right plot of figure 3. The energies of states 1 and 2, known to be degenerate, approach the same value of 0.220(5) Ha above the state 0 energy, even though they were initialized to a state much lower in energy. The results confirm that the optimization works as expected on this system. The VMC-calculated vertical excitation energy is within error bars of the experimentally determined $X^1\Sigma^+ \rightarrow a^3\Pi_r$ value of 0.22 Ha [25].

We also confirm that our ensemble optimization yields the expected results for small and large values of λ . As λ increases from zero, the optimized states start out perfectly overlapping, and end up mutually orthogonal when λ/λ_c is large, shown in figure 4. We find that orthogonality is only consistently achieved when $\lambda/\lambda_c > 1.5$, although our derived condition only requires $\lambda/\lambda_c > 1$. We attribute the difference to difficulty optimizing in the region $1 < \lambda/\lambda_c < 1.5$. In this region, the overlaps are zero for some points and not others, a sign of inconsistent optimization. When $\lambda/\lambda_c = 1$, we showed that the Hessian of the cost functional has a zero eigenvalue, meaning the cost functional is flat along at least one direction at the minimum. For λ/λ_c just above 1, this direction has very low curvature, making the minimum difficult to distinguish within stochastic noise. For large enough λ/λ_c , the curvature of the cost functional along this overlap direction is higher, and the minimum is reliably found.

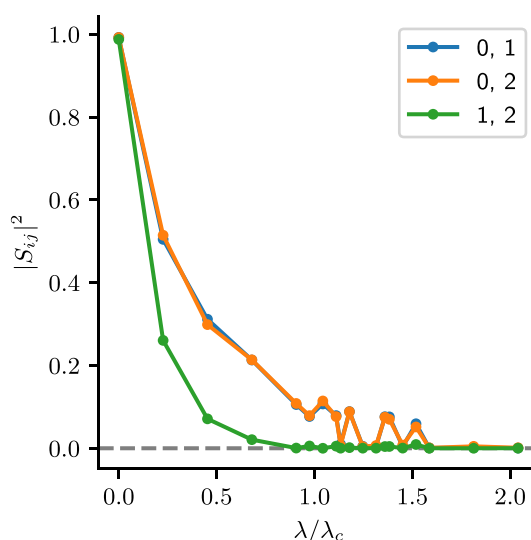


Figure 4. Overlap between three wave functions of the CO molecule as a function of the overlap penalty λ . As the penalty approaches the critical value λ_c , the overlaps go to zero. An overlap larger than the critical value is necessary to stabilize the stochastic optimization. A lighted dashed line is shown at zero overlap for reference.

4. Conclusion

We have presented a cost functional for simultaneously optimizing an ensemble of states to the low-energy eigenstates of a Hamiltonian in VMC. Our approach is state-specific, allowing all parameters of each ensemble state to vary independently without any restrictions to enforce orthogonality. The cost functional uses a weighted sum of energy expectations, distinguishing it from other recent approaches to ensemble optimization. Whether the cost functional has the correct minimum (at the low energy eigenstates) depends on the choice of λ and the weights. In particular, we showed that there is a critical value of the penalty λ_c , which λ must exceed; otherwise, the states that minimize the cost functional will not be orthogonal, and therefore not the correct eigenstates. Finally, we demonstrated the proposed cost functional on an *ab initio* VMC calculation of a carbon monoxide molecule. We verified the expected behavior when the penalty λ is below and above the estimated λ_c , and that the method reaches the correct energy eigenstates even when initialized far from them.

We showed that there is a family of numerical methods which can satisfy the constraints noted in equations (18) and (1), at least one of which has been proposed in the literature without proof in the context of neural network wave functions [20]. It is not clear which in this family are most optimal; an interesting future direction will be to determine the most convenient formulation of excited state optimization. For all these techniques, the total computational cost of the excited state optimization is approximately N times the cost of a single ground state calculation, where N is the number of excited states sought. We anticipate the cost functional presented here to be useful in future excited state QMC calculations with a variety of wave functions, including traditional as well as neural network and other machine-learning ansatzes.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This work was supported by U.S. National Science Foundation via Award No. 1931258 (W A W, L K W) and by the National Science Foundation Graduate Research Fellowship Program under Award No. DGE-1746047 (K G K). *Ab initio* calculations used computing resources from the Illinois Campus Cluster as well as the Flatiron Institute's Scientific Computing Core. We would like to thank Kieron Burke for bringing ensemble DFT to our attention, Robert Webber and Michael Lindsey for early conversations, and Claudia Filippi for many comments and discussions.

ORCID iDs

William A Wheeler  <https://orcid.org/0000-0002-8391-0530>

Kevin G Kleiner  <https://orcid.org/0000-0003-0155-1982>

Lucas K Wagner  <https://orcid.org/0000-0002-3755-044X>

References

- [1] Foulkes W M C, Mitas L, Needs R J and Rajagopal G 2001 *Rev. Mod. Phys.* **73** 33
- [2] Lühchow A 2011 *WIREs Comput. Mol. Sci.* **1** 388
- [3] Shulenburger L 2013 *Phys. Rev. B* **88** 245117
- [4] Wagner L K and Ceperley D M 2016 *Rep. Prog. Phys.* **79** 094501
- [5] Williams K T *et al* (Simons Collaboration on the Many-Electron Problem) 2020 *Phys. Rev. X* **10** 011041
- [6] Blunt N S, Alavi A and Booth G H 2015 *Phys. Rev. Lett.* **115** 050603
- [7] Blunt N S, Booth G H and Alavi A 2017 *J. Chem. Phys.* **146** 244105
- [8] Choo K, Carleo G, Regnault N and Neupert T 2018 *Phys. Rev. Lett.* **121** 167204
- [9] Benavides-Riveros C L, Chen L, Schilling C, Mantilla S and Pittalis S 2022 *Phys. Rev. Lett.* **129** 066401
- [10] Otis L and Neuscamman E 2023 *WIREs Comput. Mol. Sci.* **13** e1659
- [11] Filippi C, Zaccheddu M and Buda F 2009 *J. Chem. Theory Comput.* **5** 2074
- [12] Zhao L and Neuscamman E 2016 *J. Chem. Theory Comput.* **12** 3436
- [13] Shea J A R and Neuscamman E 2017 *J. Chem. Theory Comput.* **13** 6078
- [14] Pineda Flores S D and Neuscamman E 2019 *J. Phys. Chem. A* **123** 1487
- [15] Otis L and Neuscamman E 2023 *J. Chem. Theory Comput.* **19** 767
- [16] Cuzzocrea A, Scemama A, Briels W J, Moroni S and Filippi C 2020 *J. Chem. Theory Comput.* **16** 4203
- [17] Higgott O, Wang D and Brierley S 2019 *Quantum* **3** 156
- [18] Pathak S, Busemeyer B, Rodrigues J N B and Wagner L K 2021 *J. Chem. Phys.* **154** 034101
- [19] Shepard S, Panadés-Barrueta R L, Moroni S, Scemama A and Filippi C 2022 *J. Chem. Theory Comput.* **18** 6722
- [20] Entwistle M T, Schätzle Z, Erdman P A, Hermann J and Noé F 2023 *Nat. Commun.* **14** 274
- [21] Pfau D, Axelrod S, Sutterud H, von Glehn I and Spencer J S 2023 arXiv:2308.16848
- [22] Theophilou A K 1979 *J. Phys. C: Solid State Phys.* **12** 5419
- [23] Gross E K U, Oliveira L N and Kohn W 1988 *Phys. Rev. A* **37** 2805
- [24] Wheeler W A *et al* 2023 *J. Chem. Phys.* **158** 114801
- [25] Tobias I, Fallon R J and Vanderslice J T 1960 *J. Chem. Phys.* **33** 1638
- [26] Bennett M C, Melton C A, Annaberdiev A, Wang G, Shulenburger L and Mitas L 2017 *J. Chem. Phys.* **147** 224106