A hybrid quantum algorithm to detect conical intersections

Emiel Koridon[1,2], Joana Fraxanet[3], Alexandre Dauphin[3,4], Lucas Visscher[2], Thomas E. O'Brien[5,1], and Stefano Polla[5,1]

[1]Instituut-Lorentz, Universiteit Leiden, 2300RA Leiden, The Netherlands
[2]Theoretical Chemistry, Vrije Universiteit, 1081HV Amsterdam, The Netherlands
[3]ICFO - Institut de Ciències Fotòniques, 08860 Castelldefels (Barcelona), Spain
[4] PASQAL SAS, 2 av. Augustin Fresnel Palaiseau, 91120, France
[5]Google Research, Munich, 80636 Bavaria, Germany

arXiv:2304.06070v2 [quant-ph] 12 Feb 2024

Conical intersections are topologically protected crossings between the potential energy surfaces of a molecular Hamiltonian, known to play an important role in chemical processes such as photoisomerization and non-radiative relaxation. They are characterized by a non-zero Berry phase, which is a topological invariant defined on a closed path in atomic coordinate space, taking the value π when the path encircles the intersection manifold. In this work, we show that for real molecular Hamiltonians, the Berry phase can be obtained by tracing a local optimum of a variational ansatz along the chosen path and estimating the overlap between the initial and final state with a control-free Hadamard test. Moreover, by discretizing the path into N points, we can use N single Newton-Raphson steps to update our state non-variationally. Finally, since the Berry phase can only take two discrete values (0 or π), our procedure succeeds even for a cumulative error bounded by a constant; this allows us to bound the total sampling cost and to readily verify the success of the procedure. We demonstrate numerically the application of our algorithm on small toy models of the formaldimine molecule ($H_2C{=}NH$).

Contents

1 Introduction **2**

2 Background **3**
2.1 Conical intersections 3
2.2 Berry phases in real Hamiltonians . . 3
2.3 Measuring Berry phase with a variational wavefunction 4

3 Methods **4**
3.1 Fixing the gauge with a real ansatz . . 4
3.2 Avoiding full optimization via Newton-Raphson steps 5
3.3 Regularization and backtracking . . . 5
3.4 Measuring the final overlap 6
3.5 Overview of the algorithm 6

4 Error analysis and bounding **6**
4.1 Bounding the NR error 7
4.2 Bounding the sampling noise 8
4.3 Scaling of the total cost 8

5 Adapting to an orbital-optimized PQC ansatz **8**
5.1 An OO-PQC ansatz with geometric continuity 9
5.2 Measuring boundary terms with the OO-PQC ansatz 9
5.3 Newton-Raphson updates of the OO-PQC ansatz 10

6 Numerical results **10**
6.1 Numerical simulation details 10
6.2 Minimal model with an degeneracy-free ansatz 11
6.3 Sampling noise 12
6.4 Larger basis and active space 13

7 Conclusion and outlook **13**
7.1 Paths towards improving convergence 14
7.2 Potential applications 15
7.3 Outlook 15

Acknowledgements **15**

References **16**

A Bounding overlaps by change in ansatz parameters **21**

B Bounding the norm of energy derivatives **21**

C Analytical orbital gradient and Hessian **22**

D Bounding the cumulative error due to Newton-Raphson updates **24**

E Bounding the sampling cost **26**

Stefano Polla: polla@lorentz.leidenuniv.nl

1 Introduction

Conical intersections (CI) are degeneracy points in the Born-Oppenheimer molecular structure Hamiltonians, where two potential energy surfaces cross. Similar to Dirac cones in graphene [1], these intersections are protected by symmetries of the Hamiltonian which guarantee that any loop in parameter space around a conical intersection has a quantized Berry phase [2]. CIs play an important role in photochemistry [3, 4], as they mediate reactions such as photoisomerization and non-radiative relaxation, which are key steps in processes such as vision [5] and photosynthesis [6]. Therefore, detecting the presence and resolving the properties of CIs is important for computing reaction and branching rates in photochemical reactions [7, 8]. Nevertheless, the study of such processes requires electronic structure methods capable of accurately modelling both the shape and the relative energies of the two intersecting potential energy surfaces, a requirement that poses challenges for the current available methods [9]. Given the need to develop novel methods for identifying and characterizing CIs, quantum computers present themselves as a highly promising option for this task.

Quantum computing has long been driven by the desire to simulate interacting physical systems, such as molecules, as a novel means of investigating their properties [10, 11]. This is typically achieved by preparing eigenstates of molecular Hamiltonians in quantum devices which can natively store and process quantum states. This task would otherwise require an exponentially-scaling classical memory. Recently, with the first noisy and intermediate-scale quantum devices (NISQ) [12] being built, it became increasingly important to research tailored and robust algorithms that minimize the quantum device requirements [12]. Variational quantum algorithms (VQA), such as the variational quantum eigensolver [13, 14] (VQE) and its variations, caught the spotlight in this context, as they allow to prepare and measure quantum states with circuits of relatively low depth. The key feature of VQAs is the repeated execution of short parameterized quantum circuits on the quantum device, from which measurement results are sampled. These results are used to estimate a cost function, which is then minimized by varying the parameters defining the gates of the quantum circuit. Due to the noise introduced by sampling, a relatively large number of circuit runs and measurements are typically needed to estimate the cost function accurately. In chemistry, where VQEs are often proposed as a method to resolve ground state energies to high accuracy, the number of required samples to achieve such accuracy can become prohibitively large [15]. Furthermore, the convergence of the cost function to an optimum is typically only suggested heuristically, and it is proven to be problematic in some cases that lack such heuristic structure [16]. Therefore, it is compelling to suggest VQAs that can access quantities that are less reliant on the precision of both the optimization process and the measurement procedure.

A promising target for VQAs is the computation of the Berry phase $\Pi_{\mathcal{C}}$, which can be used to resolve the existence of CIs. More specifically, Π_C is defined as the geometric phase acquired by an eigenstate of a parameterized Hamiltonian over a closed adiabatic path $\mathcal{C}$ in parameter space [2]. Most importantly, it is known that in the presence of certain symmetries, the Berry phase will be quantized to values 0 or π. This quantization is exactly what makes the Berry phase an attractive target for a VQAs, as it implies the final result of the computation need only be accurate to error $< \frac{\pi}{2}$. Quantum algorithms to compute Berry phases have been already proposed, both variational [17, 18] and Hamiltonian-evolution based [19]. Moreover, the long-known effects of Berry phase on nuclear dynamics around a conical intersection [20–22] have been explored recently in analog quantum simulation experiments [23–25]. Nevertheless, previous proposals did not attempt to detect CIs in realistic quantum chemistry problems with an efficient algorithm that can be run on NISQ devices.

In this paper, we propose a hybrid quantum algorithm to compute the quantized Berry phase for ground states of a family of parameterized real Hamiltonians. We focus on the specific application to molecular Hamiltonians, where we can identify a conical intersection by measuring the Berry phase along a loop in atomic coordinates space. We first review the definition of CIs and Berry phases in Sec. 2. Then, we present all the ingredients of the proposed algorithm in Sec. 3, which is similar to a VQA in spirit, but it does not require full optimization; rather, the variational parameters are updated by a single Newton-Raphson step for each molecular geometry along a discretization of the loop. In Sec. 4, we prove a convergence guarantee for the algorithm under certain assumptions on the ansatz, by providing sufficient condition bounds on the total number of steps and the acceptable sampling noise. Finally, we adapt our algorithm to a specific ansatz in Sec. 5 and we benchmark it on a model of the formaldimine molecule $H_2C{=}NH$ in Sec. 6. Section 7 presents our conclusions, a discussion of potential application cases for our algorithm and an outlook on possible enhancements.

The core code developed for the numerical benchmarks, which provides a flexible implementation of an orbital-optimized variational quantum ansatz, is made available in a GitHub repository at `https://github.com/emieeel/auto_oo` [26].

2 Background

2.1 Conical intersections

Let us consider a molecular electronic structure Hamiltonian $H(\mathbf{R})$ parameterized by the nuclear geometry $\mathbf{R}$ in some configuration space $\mathcal{R}$. A conical intersection is a point $\mathbf{R}^\times \in \mathcal{R}$ where two potential energy surfaces become degenerate, leading to non-perturbatively large non-adiabatic couplings, and thus a breakdown of the Born-Oppenheimer approximation [27, 28]. Conical intersections extend to a manifold of dimension $\dim[\mathcal{R}]-2$, and lead to the two potential energy surfaces taking the form of cones in the remaining two directions $\hat{x}$, $\hat{z}$. These two potential energy surfaces can be described as eigenstates of the effective Hamiltonian

$$H^{\text{eff}}(\mathbf{R}) = h_x[\mathbf{R}-\mathbf{R}^\times]_x\sigma_x + h_z[\mathbf{R}-\mathbf{R}^\times]_z\sigma_z. \quad (1)$$

The Pauli terms σ_x and σ_z form a complete basis for two-dimensional real symmetric matrices. Thus, a single conical intersection cannot be lifted by any real-valued and continuous perturbation of $H^{\text{eff}}(\mathbf{R})$; such a perturbation would only shift the value of $\mathbf{R}^\times$. Moreover, the presence of such an effective Hamiltonian implies that in any direction other than R_x or R_z the energies must be degenerate.

Simulations involving conical intersections are challenging due to the degeneracy of the two states involved, as the character of both states needs to be considered. Active space methods are often used for organic molecules, which involve selecting the chemical bonds that are formed or broken in the reaction pathway as well as the most significant spectator or correlating orbitals [29]. The situation is more complicated when transition metals are involved because the close energetic spacing of d-orbitals typically requires including all five d-orbitals of such a metal into the active space. An additional complication arises if the two crossing states correspond to different atomic configurations: a situation that is not uncommon for the early or late d- (or f-) metals, for which configurations with a different d- (f-) population are energetically close. In such cases, one may need to either work with non-orthogonal orbitals [30] or add an additional d-shell to the active space [31] to qualitatively describe the nature of both states. For cases of practical interest in which one wants to characterize and simulate the internal conversion processes in a complex photo-excited system, the presence of transition metals may easily lead to large active space requirements. These can not be met by classical algorithms and would be highly challenging for quantum algorithms as well.

One way to reduce the complexity of the problem is to first focus on the presence or absence of conical intersections that connect the ground and excited states. The measurement of the Berry phase in chemical systems allows this: without explicitly computing the excited state surface and non-adiabatic couplings, it should be possible to detect whether a loop in the nuclear coordinate space encloses a conical intersection or not. In this manner, one may alleviate the requirements for the active space selection and orbital optimization and quickly establish the region in the potential energy surface that contains an intersection with another surface and needs to be scrutinized further [32]. Information about the location of conical intersections is of interest also for ground-state dynamics; the CIs and the Berry phase they induce influence the propagation of nuclear wave packets on the adiabatic ground state surface and thereby affect the branching rates and efficiency of reactions or isomerizations [33]. For both types of applications, precise study of dynamics on ground state surfaces as well as characterizing the efficiency of radiationless decay, it is of interest to explore the possibilities offered by quantum algorithms.

2.2 Berry phases in real Hamiltonians

The Berry phase $\Pi_\mathcal{C}$ is the geometric phase acquired by some eigenstate $|\Phi(\mathbf{R})\rangle$ of a system with parameterized Hamiltonian $H(\mathbf{R})$ as it is adiabatically transported around a closed loop in parameter space $\mathcal{C} \subset \mathcal{R}$. $\Pi_\mathcal{C}$ can be defined as the closed line integral of the Berry connection along the loop $\mathcal{C}$

$$\Pi_\mathcal{C} = -i \oint_\mathcal{C} d\mathbf{R} \cdot \langle\Phi(\mathbf{R})|\nabla_\mathbf{R}|\Phi(\mathbf{R})\rangle. \quad (2)$$

The integrand must be imaginary, as $\nabla_\mathbf{R}\langle\Phi(\mathbf{R})|\Phi(\mathbf{R})\rangle = 2\text{Re}[\langle\Phi(\mathbf{R})|\nabla_\mathbf{R}|\Phi(\mathbf{R})\rangle] = 0$, thus $\Pi_\mathcal{C}$ is real. In this work, we assume $|\Phi(\mathbf{R})\rangle$ is the ground state of $H(\mathbf{R})$.

We need to parameterize the loop $\mathcal{C} = \{\mathbf{R}(t), t \in [0,1]\}$ in order to evaluate the integral. Moreover, it is possible to multiply the ground state $|\Phi(\mathbf{R})\rangle$ by a t-dependent phase, resulting in a $U(1)$-gauge a transformation which leaves all physical quantities invariant. We take

$$|\Psi(t)\rangle = e^{i\Theta(t)}|\Phi(\mathbf{R}(t))\rangle, \quad (3)$$

which allows us to rewrite the Berry phase as

$$\Pi_\mathcal{C} = -i\int_0^1 dt\langle\Psi(t)|\partial_t|\Psi(t)\rangle + \int_0^1 dt\,\partial_t\Theta(t). \quad (4)$$

If there is a representation for which each Hamiltonian $H(\mathbf{R})$ is real, it is possible to choose eigenstates that have all real components. In this case, we can choose $\Theta(t)$ such that $|\Psi(t)\rangle$ has real expansion, which implies the first integrand is real; as $\partial_t\langle\Psi(t)|\Psi(t)\rangle = 0$ this must also to be imaginary, therefore $\langle\Psi(t)|\partial_t|\Psi(t)\rangle = 0$. Under this choice, we can evaluate the Berry phase as a boundary term

$$\Pi_\mathcal{C} = \int_0^1 dt\,\partial_t\Theta(t) = \Theta(1)-\Theta(0) = \arg\Big[\langle\Psi(0)|\Psi(1)\rangle\Big], \quad (5)$$

where the last equality is obtained using the definition in Eq. (3). Furthermore, $|\Psi(1)\rangle$ and $|\Psi(0)\rangle$ are real by construction, which implies $\Pi_{\mathcal{C}}$ can only take two values (modulo 2π): 0 or π.

The quantization of $\Pi_{\mathcal{C}}$ implies that it is invariant for topological deformations of $\mathcal{C}$. If $\mathcal{C}$ can be contracted to a point, then $\Pi_{\mathcal{C}} = 0$. A non-trivial $\Pi_{\mathcal{C}} = \pi$ can only occur when $\mathcal{C}$ encircles a degeneracy. One can check with the effective Hamiltonian Eq. (1) that any loop encircling $\mathbf{R}^{\times}$ has a Berry phase of π. This extends by continuity to any region of $\mathcal{R}$ around the CI, as long as $\mathcal{C}$ does not enclose a second degeneracy point. Thus, we have a one-to-one correspondence between CIs and the nontrivial Berry phase.

2.3 Measuring Berry phase with a variational wavefunction

There have been various proposals in the literature for computing Berry phases using a gate-based quantum device [19, 17]. In this work, we propose to use a variational algorithm to track $|\Psi(t)\rangle$ when parallel-transported around the loop $\mathcal{C}$, in the spirit of the variational adiabatic method described in Ref. [15, 34]. We approximate

$$|\Psi(t)\rangle \approx |\psi(\boldsymbol{\theta}_t^*)\rangle := U(\boldsymbol{\theta}_t^*)|\psi_0\rangle, \tag{6}$$

where $|\psi(\boldsymbol{\theta})\rangle$ is a variational ansatz state, and $\boldsymbol{\theta}_t^*$ continuously tracks a local minimum $[\nabla_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta}_t^*) = 0]$ of the variational energy

$$E(t, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H(\mathbf{R}(t))|\psi(\boldsymbol{\theta})\rangle. \tag{7}$$

The angle $\boldsymbol{\theta}_t^*$ is well-defined as long as the Hessian $\nabla^2_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta})$ remains positive definite in a neighbourhood of $\boldsymbol{\theta}_t^*$ for all t, ensuring the $\boldsymbol{\theta}_t^*$ is continuous in t and non-degenerate. Although our treatment naturally extends to any variational ansatz that continuously parametrizes normalized states $|\psi(\boldsymbol{\theta})\rangle$ (including classical ansätze like e.g. matrix-product states), we assume the operator $U(\boldsymbol{\theta})$ is implemented by a parameterized quantum circuit (PQC) acting on an initial state $|\psi_0\rangle$; this implies that information about the state needs to be extracted from a quantum device through sampling.

3 Methods

In this section, we detail all the ingredients needed to implement our hybrid algorithm to resolve quantized Berry phases with a variational quantum ansatz. Initially, in Sec. 3.1, we discuss how selecting an ansatz that preserves the Hamiltonian's symmetries establishes a natural gauge, leading to the reduction of the Berry phase integral to the boundary term Eq. (5). In Sec. 3.2, we introduce our parameter update approach, which employs single Newton-Raphson steps to trace the variational state along a discretization of the loop $\mathcal{C}$. In Sec. 3.3 we explain how to employ a basic regularisation technique to handle the potential non-convexity of the cost function and in 3.4 we explain how to measure the final overlap using an ancilla-free Hadamard test. Finally, in Sec. 3.5, we provide a full overview of the algorithm.

3.1 Fixing the gauge with a real ansatz

As discussed in Sec. 2, the quantization of the Berry phase is granted by the symmetries of the Hamiltonian family $H(\mathbf{R})$, which ensure the existence of a basis for which each $H(\mathbf{R})$ has a real representation. When it comes to electronic structure Hamiltonians, it is always possible to find a real representation for time-reversal symmetric Hamiltonians with integer total spin [35]. Moreover, real noninteger-spin Hamiltonians are also found throughout nonrelativistic quantum chemistry.

As our variational ansatz state [in Eq. (6)] is defined by a family of unitary operators, it inherits a natural gauge from $U(\boldsymbol{\theta})$. In particular, if $U(\boldsymbol{\theta})$ is written as a product of real rotations in the basis in which $H(\mathbf{R})$ is real, then we force $|\psi(\boldsymbol{\theta})\rangle$ to have real components as well, which fixes a global $U(1)$ phase. This can be obtained by constructing the PQC with a sequence of parameterized unitaries such as

$$U_j(\theta_j) = e^{A_j \theta_j} \tag{8}$$

generated by antisymmetric operators A_j that are real in the chosen representation. (We choose dimensional units such that $\|A_j\| = 1$ without loss of generality, see Appendix A). Examples from electronic structure include real fermionic (de-)excitations, such as unitary singles ($A_{pq} = \hat{a}_p^\dagger \hat{a}_q - \hat{a}_q^\dagger \hat{a}_p$) and doubles ($A_{pqrs} = \hat{a}_p^\dagger \hat{a}_q \hat{a}_r^\dagger \hat{a}_s - \hat{a}_q^\dagger \hat{a}_p \hat{a}_s^\dagger \hat{a}_r$). Many PQC ansätze commonly proposed for quantum chemistry, such as unitary coupled cluster (UCC) [36, 13, 37] and quantum-number preserving gate fabrics (NPF) [38], are composed from these elementary rotations. Formally, our ansatz state can then be defined as

$$|\psi(\boldsymbol{\theta})\rangle = \prod_{j=1}^{n_p} U_{n_p - j}(\theta_{n_p - j})|\psi_0\rangle, \tag{9}$$

where U_k are the aforementioned parameterized rotations applied in circuit-composition order.

In this case, the Berry phase can be estimated by $\Pi_{\mathcal{C}} = \arg\left[\langle\psi(\boldsymbol{\theta}_{t=0}^*)|\psi(\boldsymbol{\theta}_{t=1}^*)\rangle\right]$, which corresponds to the boundary term in Eq. (5). This implies that, in the case of a non-trivial Berry phase, the path traced by $\boldsymbol{\theta}_t^*$ will not close up on itself (i.e. $\boldsymbol{\theta}_1^* \neq \boldsymbol{\theta}_0^*$), highlighting an important difference between the ansatz parameters $\boldsymbol{\theta}$, which fix the gauge of $|\psi(\boldsymbol{\theta})\rangle$, and the Hamiltonian parameters $\mathbf{R}$, which define a ground state $|\Phi(\mathbf{R})\rangle$ up to a $U(1)$ gauge freedom. However, the change in optimal parameters is insufficient to

prove the existence of a nontrivial Berry phase; we must both successfully track the minimum $\boldsymbol{\theta}_t^*$ as t traces from 0 to 1, and estimate the final overlap $\langle\psi(\boldsymbol{\theta}_0^*)|\psi(\boldsymbol{\theta}_1^*)\rangle$. While the argument (sign) of the overlap will yield the Berry phase, its absolute value is a proxy of success as it certifies the initial and final states are physically equivalent.

3.2 Avoiding full optimization via Newton-Raphson steps

As mentioned above, the Berry phase $\Pi_{\mathcal{C}}$ is a discrete quantity, and therefore we only need to estimate it to accuracy $< \frac{\pi}{2}$. In Appendix A, we show that this implies that we can accept an error on the estimate $\tilde{\boldsymbol{\theta}}_1$ of the final optimum $\boldsymbol{\theta}_1^*$ bounded in one-norm by $\|\tilde{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1^*\|_1 < 1$. Thus, we are not required to exactly track $|\psi(\boldsymbol{\theta}_t^*)\rangle$ and as a result, the variational energy Eq. (7) does not need to be fully re-optimized at every time-step t. Instead, it suffices to keep the estimate $\tilde{\boldsymbol{\theta}}_t$ of the optimal parameters within the basis of convergence of the true minimum $\boldsymbol{\theta}_t^*$.

To achieve this, we still need an initial optimum as an input, which is obtained by running one full optimization. If possible, the initial point is selected such that optimization is simplest. Then, we propose to use a single step of the Newton-Raphson algorithm at points $t \in \{\Delta t, 2\,\Delta t, \ldots, 1 - \Delta t, 1\}$.

The Newton-Raphson (NR) algorithm determines the update of the estimate of the minimum $\tilde{\boldsymbol{\theta}}$ through the gradient $\mathcal{G}(t, \boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta})$ and Hessian $\mathcal{H}(t, \boldsymbol{\theta}) := \nabla^2_{\boldsymbol{\theta}} E(t, \boldsymbol{\theta})$ of the variational energy Eq. (7). The derivatives can be computed using either finite-difference methods or parameter-shift rules [39]; either method requires sampling the variational energy $E(t, \boldsymbol{\theta})$ at a number of different parameter points $\boldsymbol{\theta}$. Given the estimate $\tilde{\boldsymbol{\theta}}_t$ of the optimum at point t as an initial guess, the NR step with cost $E(t+\Delta t, \boldsymbol{\theta})$ prescribes the update $\tilde{\boldsymbol{\theta}}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t + d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t}$, with

$$d\boldsymbol{\theta}^{\mathrm{NR}}_{t,\Delta t} = -\mathcal{H}^{-1}(t + \Delta t, \tilde{\boldsymbol{\theta}}_t)\boldsymbol{\mathcal{G}}(t + \Delta t, \tilde{\boldsymbol{\theta}}_t). \tag{10}$$

The Newton-Raphson method is well-known to have a finite-sized basin of quadratic convergence, as long as the cost function is strongly convex at the optimum,

$$m(\boldsymbol{\theta}^*) := \min_{\mathbf{v}} \frac{\mathbf{v}^T \mathcal{H}(\boldsymbol{\theta}^*)\mathbf{v}}{\mathbf{v}^T\mathbf{v}} > 0. \tag{11}$$

In other words, the lowest eigenvalue m of the Hessian of the cost function (which we call convexity) at the optimum $\boldsymbol{\theta}^*$ needs to be positive. Details on the convergence properties of NR are given in App. D. In section 4, we show that this quadratic convergence is fast enough to keep track of the minimum with only a single step for each t-point.

3.3 Regularization and backtracking

To ensure that we successfully track the minimum of the cost function $\boldsymbol{\theta}_t^*$, the estimates $\tilde{\boldsymbol{\theta}}_t$ need to remain in the strongly convex region of optimization space. The existence of such strongly convex region is not always guaranteed since it depends on the ansatz. A common cause of failure of this requirement is exemplified for ansätze with (local) over-parametrization of the state manifold. In fact, if the ansatz has redundant parameters the Hessian of the cost function Eq. (7) will always be singular ($m = 0$). Then, arbitrarily small perturbations of the cost function can then cause $m < 0$. When this occurs, the inversion of the Hessian needed for the Newton-Raphson step is ill-defined.

The most direct approach to solve this issue is to select an ansatz with no degeneracies, facilitating a strongly convex cost function at its minima. Nevertheless, this is only possible for very simple problems, and even in this case, quasi-degeneracies can make the convergence region extremely small. Since this is a well-known problem, many alternative solutions have been proposed in the literature. In this subsection, we will explore and implement two of them – back-tracking and regularization.

Back-tracking — Small positive eigenvalues of the Hessian can cause the standard Newton-Raphson step to overshoot along the relative parameter eigenmodes. This effectively reduces the size of the neighborhood of the minimum θ^* in which quadratic convergence is granted (see Appendix D). Since for positive convexity the direction provided by the Newton-Raphson step is guaranteed to be a descent direction, we can mitigate this overshoot by implementing line-search of the minimum on the segment defined by the NR step. In the common variant of back-tracking, the Newton step is iteratively damped by a constant $\beta \in (0, 1)$. At each iteration, the cost function in the new point is measured, until the cost function is reduced enough (the detailed condition is given in Algorithm 1). While some additional evaluations of the cost function are needed, the (more expensive) gradient and Hessian are only calculated once. Due to the repeated evaluation, one needs to consider extending line-search methods to cost functions evaluated with sampling noise.

Regularization — For realistic ansätze and Hamiltonians, it is difficult to avoid (quasi-) redundancies in some regions of parameter space. In this case, the cost function might not be strongly convex around the minimum, or the convexity might be too small to ensure a sufficiently-large convergence region. Furthermore, even for an ideally-convex cost function, noisy evaluation on a quantum device might result in a distorted Hessian with non-positive eigenvalues. To mitigate this issue, we can use a regularization technique that penalizes the change in parameters along the quasi-redundant directions. We propose to use

augmentation of the Hessian to regularize the NR step, obtaining a so-called quasi-Newton optimizer. Hessian augmentation is a common practice in quantum chemistry methods that feature orbital optimization, such as self-consistent field methods [29]. If the smallest eigenvalue of the Hessian λ_0 is smaller than a positive threshold convexity m_{thr}, we construct the augmented Hessian as follows

$$\mathcal{B} = \mathcal{H} + \nu\mathbb{1}, \tag{12}$$

where we add a constant $\nu > |\lambda_0|$. The augmented Hessian is then positive, and we can realize the NR update as

$$d\boldsymbol{\theta}^{\text{NR}} = -\beta\mathcal{B}^{-1}\boldsymbol{\mathcal{G}}. \tag{13}$$

Here, β is the damping constant from back-tracking line search and $\mathcal{G}$ is the gradient as in eq. (10). Regularization and back-tracking are typically used in tandem, as quadratic convergence is harder to guarantee when using regularization. The choice of ν is nontrivial: we want it to be large enough to suppress parameter changes along quasi-redundant directions, but we need to avoid exaggerating the damping along relevant directions. Common solutions include choosing $\nu = \rho|\lambda_0| + \mu$ with fixed positive constants ρ and μ, or using a trust-region method [40, 41] where the Newton step is constrained to lie within a ball of some radius h (such that $||d\boldsymbol{\theta}^{\text{NR}}||_2 \leq h$).

The augmented Hessian method does not require further evaluations of the cost function, although the damping of the parameter updates might imply that more t-steps are needed to successfully resolve the Berry phase.

Algorithm 1: NR step subroutine with regularization and backtracking

Input: Initial estimated parameters $\tilde{\boldsymbol{\theta}}_\tau$
Estimated Hessian $\tilde{\mathcal{H}}$
Estimated gradient $\tilde{\boldsymbol{\mathcal{G}}}$
Cost function $E(\theta)$ given by eq. (7)
Convexity requirement $m_{\text{thr}} > 0$
Positive constants μ, ρ, α, β
Output: Updated estimated parameters $\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}$.

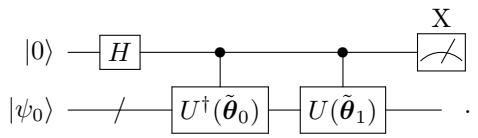
$\lambda_0 \leftarrow$ lowest eigenvalue of $\tilde{\mathcal{H}}$;
// Regularization
if $\lambda_0 < m$ **then**
  $\mathcal{B} \leftarrow \tilde{\mathcal{H}} + (\rho|\lambda_0| + \mu)\mathbb{1}$;
  $d\boldsymbol{\theta} \leftarrow \mathcal{B}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$
else
  $d\boldsymbol{\theta} \leftarrow \tilde{\mathcal{H}}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$
// Backtracking
while $E(\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}) > E(\tilde{\boldsymbol{\theta}}_\tau) + \alpha(\tilde{\boldsymbol{\mathcal{G}}} \cdot d\boldsymbol{\theta})$ **do**
  $d\boldsymbol{\theta} \leftarrow \beta d\boldsymbol{\theta}$
return $\tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}$

3.4 Measuring the final overlap

For a real ansatz, the overlap of the tracked states at $t = 0$ and $t = 1$ must be real and it can be rewritten as

$$\langle\psi(\tilde{\boldsymbol{\theta}}_0)|\psi(\tilde{\boldsymbol{\theta}}_1)\rangle = \text{Re}\left[\langle 0|U^\dagger(\tilde{\boldsymbol{\theta}}_0)U(\tilde{\boldsymbol{\theta}}_1)|0\rangle\right]. \tag{14}$$

This quantity can readily be measured by a Hadamard test, implemented as



The required number of samples is small, as we only need to resolve whether the sign of the overlap is $+1$ (trivial $\Pi_\mathcal{C} = 0$) or -1 (nontrivial $\Pi_\mathcal{C} = \pi$).

Implementing the circuit above requires to realize controlled-$U(\boldsymbol{\theta})$, which might increase significantly the depth of the compiled quantum circuit and make the implementation unfeasible on near-term hardware. However, this requirement can be bypassed by using the control-free echo verification technique [42, 43], in place of the standard Hadamard test, to sample Eq. (14). This technique requires access to a reference state $|\psi_{\text{ref}}\rangle$ orthogonal to $|\psi_0\rangle$, which should acquire a known eigenphase φ under the action of the PQC, $U(\boldsymbol{\theta})|\psi_{\text{ref}}\rangle = e^{i\varphi}|\psi_{\text{ref}}\rangle$. As most of the PQC ansätze used for electronic structure states preserve the total number of electrons (including UCC and NPF), the fully unoccupied state $|0...0\rangle$ can be used as reference. Control-free echo verification circuits only require implementing the non-controlled $U(\boldsymbol{\theta})$, and furthermore provide built-in error mitigation power.

3.5 Overview of the algorithm

We are now ready to formalize the proposed algorithm for resolving Berry phases, Algorithm 2. The formalization we present here will allow us to bound the number of steps and the sampling cost in the following Section 4. Given a path $\mathcal{C}$ and a number of steps $1/\Delta t$, the algorithm attempts to calculate $\Pi_\mathcal{C}$ yielding either $\Pi_\mathcal{C} = 0$, $\Pi_\mathcal{C} = \pi$, or a FAIL state. Again, in Section 4 we will bound the probability of the FAIL state occurring. Additional features that extend the practicality of the algorithm and mitigate the failure cases are presented in Sec. 5 and later implemented in Sec. 6.

If the algorithm fails, it can be re-run with a larger number of steps N and thus a smaller step size Δt. A smaller step size decreases additive NR error bound (the error per step scales as Δt^2, the total bound thus scales as Δt).

4 Error analysis and bounding

In this section, we find analytic upper bounds on the cost of estimating a quantized Berry phase $\Pi_\mathcal{C}$ on

Algorithm 2: Resolve quantized Berry phase

Input: Family of Hamiltonians $H(\mathbf{R})$, $\mathbf{R} \in \mathcal{R}$
Real ansatz PQC $|\psi(\boldsymbol{\theta})\rangle$
Set of initial optimal $\boldsymbol{\theta}_0^*$
Closed path $\mathcal{C} \in \mathcal{R}$, $\mathbf{R}(t) : [0,1] \mapsto \mathcal{C}$
Number of steps N to discretize $\mathcal{C}$
Precision requirements $\sigma_{\mathcal{G}} > 0$ and $\sigma_{\mathcal{H}} > 0$
Convexity requirement $m_{\text{thr}} > 0$
Regularization reg $\in \{\text{False}, \text{True}\}$
Final fidelity requirement $F \in (0,1)$

Output: $\Pi_{\mathcal{C}} = 0$ or $\Pi_{\mathcal{C}} = \pi$ or *FAIL*.

$\Delta t = 1/N$;
$\tilde{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_0^*$;
for $\tau \in \{0, \Delta t, 2\,\Delta t, ..., 1 - \Delta t\}$ **do**
  $E(t,\theta) \leftarrow$ define cost function as in Eq. (7);
  $\tilde{\boldsymbol{\mathcal{G}}}_j \leftarrow$ sample the gradient to precision $\sigma_{\mathcal{G}}$ $\left[\tilde{\boldsymbol{\mathcal{G}}}_j = \frac{\partial E}{\partial \theta_j}(\tau + \Delta t, \tilde{\boldsymbol{\theta}}_\tau)\right]$;
  $\tilde{\mathcal{H}}_{jk} \leftarrow$ sample the Hessian to precision $\sigma_{\mathcal{H}}$ $\left[\mathcal{H}_{jk} = \frac{\partial E}{\partial \theta_j \partial \theta_k}(\tau + \Delta t, \tilde{\boldsymbol{\theta}}_\tau)\right]$;
  if *reg* = *False* **then**
    $\lambda_0 \leftarrow$ lowest eigenvalue of $\tilde{\mathcal{H}}$;
    if $\lambda_0 < m_{thr}$ **then return** *FAIL* and exit;
    $d\boldsymbol{\theta}^{\text{NR}} \leftarrow -\tilde{\mathcal{H}}^{-1}\tilde{\boldsymbol{\mathcal{G}}}$ (see Eq. (10));
    $\tilde{\boldsymbol{\theta}}_{\tau+\Delta t} \leftarrow \tilde{\boldsymbol{\theta}}_\tau + d\boldsymbol{\theta}^{\text{NR}}$;
  if *reg* = *True* **then**
    $\tilde{\boldsymbol{\theta}}_{\tau+\Delta t} \leftarrow$ **Subroutine 1** ($\tilde{\boldsymbol{\theta}}_\tau$, $\tilde{\boldsymbol{\mathcal{G}}}$, $\tilde{\mathcal{H}}$);
$f \leftarrow$ final overlap as in Eq. (14) to precision F;
if $f^2 < F$ **then return** *FAIL* and exit;
return $\Pi_{\mathcal{C}} = \arg\{f\}$

a fixed curve $\mathcal{C}$ using Algorithm 2. To simplify the treatment regularization and back-tracking are not considered: instead, we require each estimate $\tilde{\boldsymbol{\theta}}_{j\,\Delta t}$ at the j-th step to be within the region of quadratic convergence of the cost function at the next t-step $E((j+1)\Delta t, \boldsymbol{\theta})$. We prove that this translates to a guarantee of convergence of the algorithm, under three conditions:

1. At the local minimum $\boldsymbol{\theta}_t^*$, where $\psi(\boldsymbol{\theta}_t^*)$ approximates the ground state, the cost function $E(t,\boldsymbol{\theta})$ is strongly convex [as described in Eq. (11)];
2. The number of discretization steps N is sufficiently large;
3. The sampling noise on each of the Hessian and gradient elements ($\sigma_{\mathcal{H}}$ and $\sigma_{\mathcal{G}}$ respectively) is sufficiently small.

The first point entails a requirement on the cost function, defined by the family of Hamiltonians $H(\mathbf{R})$ and the choice of ansatz $|\psi(\boldsymbol{\theta})\rangle$. This requirement is not satisfied if the ansatz state is defined with redundant parameters. We contend that, while strong convexity is a significant assumption, incorporating regularization (or one of the other techniques suggested in the outlook) can alleviate the necessity for such an assumption in practical applications. Our proof provides upper bounds on N and lower bounds on $\sigma_{\mathcal{H}}$ and $\sigma_{\mathcal{G}}$, which suffice to grant convergence. However, these are not to be considered practical prescriptions, as we do not believe them to be optimal; rather they show which are the relevant factors playing a role in the convergence of the algorithm. As the sampled gradient and Hessian are random variables, the guarantee of convergence for bounded error is to be understood in a probabilistic sense.

We first clarify natural assumptions and notation used in the calculation of the basin of convergence of Newton's method. We require the cost function $E(t,\boldsymbol{\theta})$ to be twice-differentiable by $\boldsymbol{\theta}$, for all t, in a region around the true minima $\boldsymbol{\theta}_t^*$. We require the Hessian to be Lipschitz continuous across this region,

$$\|\mathcal{H}(t,\boldsymbol{\theta}) - \mathcal{H}(t,\boldsymbol{\theta} + d\boldsymbol{\theta})\| < L\|d\boldsymbol{\theta}\|. \tag{15}$$

(Here, the Lipschitz constant L can be considered a bound $\|\mathcal{T}\| \leq L$ on the norm of the tensor of third derivatives $T(t,\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}E(t,\boldsymbol{\theta})$.) We also require that the gradient of the t-derivative $\dot{\boldsymbol{\mathcal{G}}} = \nabla_{\boldsymbol{\theta}}\frac{dE}{dt}$ is bounded by $\dot{\mathcal{G}}_{\max}$ in 2-norm. These regularity conditions are satisfied for the PQC ansätze we consider in Sec. 5, and in App. B we argue for bounds on L and $\dot{\mathcal{G}}_{\max}$. The strong convexity assumption described in the previous paragraph entails a constant lower bound m_{thr} on the smallest eigenvalue of the Hessian $\mathcal{H}(t,\boldsymbol{\theta}_t^*)$ at the minimizer $\boldsymbol{\theta}_t^*$ for all t.

4.1 Bounding the NR error

We will first calculate a lower bound on Δt that ensures the error $\delta\tilde{\boldsymbol{\theta}}_t = \tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t^*$ is bounded by a constant for all values of t. (as shown in Appendix A, a bound on $\|\delta\tilde{\boldsymbol{\theta}}_1\|_1$ is sufficient to ensure $\Pi_{\mathcal{C}}$ can be accurately resolved.) In this calculation, we will allow for an additive error $\sigma_{\boldsymbol{\theta}}$ on $\tilde{\boldsymbol{\theta}}_t$ due to sampling noise; we will simultaneously calculate an upper bound on $\sigma_{\boldsymbol{\theta}}$. We sketch the calculation here and defer details to App. D.

Firstly, it can be shown (Theorem 1 in Appendix D) that the Newton-Raphson step Eq. (10) with cost function $E(t+\Delta t, \boldsymbol{\theta})$ is guaranteed to converge quadratically [44] to the minimizer $\boldsymbol{\theta}_{t+\Delta t}^*$ as long as the initial guess $\tilde{\boldsymbol{\theta}}_t$ is within a ball centred in the minimizer of radius $\frac{m_{\text{thr}}}{4L}$. Quadratic convergence means that the distance of the updated guess from the minimizer will scale as the square of the distance of the initial guess from the minimizer,

$$\|\delta\tilde{\boldsymbol{\theta}}_{t+\Delta t}\| \leq \frac{L}{m_{\text{thr}}}\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_{t+\Delta t}^*\|^2. \tag{16}$$

The right-hand side of this equation can be bounded through the triangle inequality as

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| \leq \|\delta\tilde{\boldsymbol{\theta}}_t\| + \|\boldsymbol{\theta}^*_t - \boldsymbol{\theta}^*_{t+\Delta t}\|. \tag{17}$$

We can bound the second term in this equation by taking the total t-derivative of the optimality condition $\boldsymbol{\mathcal{G}}(t, \boldsymbol{\theta}^*_t) = 0$, yielding

$$\|\boldsymbol{\theta}^*_{t+\Delta t} - \boldsymbol{\theta}^*_t\| \leq m_{\text{thr}}^{-1}\dot{\mathcal{G}}_{\max}\Delta t. \tag{18}$$

If at step t we are within the radius given by Theorem 1, the NR step will quadratically converge, suppressing also the error from the previous step, and yielding $\|\delta\tilde{\boldsymbol{\theta}}_t\| \leq \frac{m_{\text{thr}}^2}{16L^2}$.

To account for sampling noise effect, we then consider a small additive error $\sigma_{\boldsymbol{\theta}}$ to $\tilde{\boldsymbol{\theta}}_t$. Maximising this allowed sampling noise at each step (as we will see, this becomes the bottleneck in our method) then yields the two bounds

$$\sigma_{\boldsymbol{\theta}} \leq \frac{\sqrt{2}-1}{4}\frac{m_{\text{thr}}}{L}, \quad \Delta t \leq \frac{m_{\text{thr}}^2}{8L\dot{\mathcal{G}}_{\max}}. \tag{19}$$

When both bounds are satisfied, we are guaranteed single steps of Newton's method will maintain convergence around the path $\mathcal{C}$.

4.2 Bounding the sampling noise

We now translate the bound on $\sigma_{\boldsymbol{\theta}}$ to bounds on the variance of estimates of each element of the gradient and Hessian ($\sigma^2_{\mathcal{G}}$ and $\sigma^2_{\mathcal{H}}$ respectively). This proceeds by simple propagation of variance through Eq. (10). We find

$$\begin{aligned}\sigma^2_{\boldsymbol{\theta}} := \text{Var}[d\boldsymbol{\theta}^{\text{NR}}_{t,\Delta t}] \leq \quad & \|\mathcal{H}^{-1}\|^2\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] + \\ & + \|\mathcal{H}^{-1}\|^2\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right]\|d\boldsymbol{\theta}^{\text{NR}}_{t,\Delta t}\|^2,\end{aligned} \tag{20}$$

where $\delta\boldsymbol{\mathcal{G}}$ and $\delta\mathcal{H}$ are the random variables representing the errors on gradient and Hessian. (a more detailed calculation is given in App. E.) Assuming $\boldsymbol{\theta}$ has n_p elements, each element of the gradient is i.i.d. with variance $\sigma^2_{\mathcal{G}}$, we get

$$\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] = n_p\,\sigma^2_{\mathcal{G}}. \tag{21}$$

As $\delta\mathcal{H}$ is a $n_p \times n_p$ real symmetric matrix, assuming its elements are i.i.d. with variance $\sigma^2_{\mathcal{H}}$, we can invoke Wigner's semicircle law [45] to approximate its norm by $\sqrt{n_p}\,\sigma_{\mathcal{H}}$, thus

$$\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \approx n_p\,\sigma^2_{\mathcal{H}}. \tag{22}$$

Combining these with Eq. (20), and requiring the resulting variance to be small compared to the square allowed additive error $\sigma^2_{\boldsymbol{\theta}}$ we obtain the bound

$$\sigma^2_{\mathcal{G}} + \sigma^2_{\mathcal{H}}\|d\boldsymbol{\theta}^{\text{NR}}_{t,\Delta t}\|^2 \ll \frac{3-2\sqrt{2}}{16}\frac{m^4_{\text{thr}}}{n_pL^2}. \tag{23}$$

We can then bound the norm of the NR update as $\|d\boldsymbol{\theta}^{\text{NR}}_{t,\Delta t}\| \leq m_{\text{thr}}^{-1}\|\boldsymbol{\mathcal{G}}\|$. Splitting the error budget in half we obtain.

$$\sigma^2_{\mathcal{G}} \ll \frac{3-2\sqrt{2}}{32}\frac{m^4_{\text{thr}}}{n_pL^2} \tag{24}$$

$$\sigma^2_{\mathcal{H}} \ll \frac{3-2\sqrt{2}}{32}\frac{m^6_{\text{thr}}}{n_pL^2\|\boldsymbol{\mathcal{G}}\|^2}. \tag{25}$$

Note that these bounds are not tight. For instance, by applying Cauchy-Schwartz inequality to bound $\|\mathcal{H}^{-1}\cdot\mathcal{G}\| \leq \|\mathcal{H}^{-1}\|\|\mathcal{G}\|$, we overlook the fact that the gradient will change more slowly along a lower-eigenvalue eigenmode of the Hessian. We believe further work might allow to define tighter bounds.

4.3 Scaling of the total cost

To give an estimate on how many measurements we need to sample gradient and Hessian to sufficient precision, we need to recast the quantities in Eq. (25) (the dominant term of the sampling variance) in terms of parameters of the problem. If we use an ansatz without redundancies (or if we can get rid of redundancies through e.g. regularization), and assuming we approximate the ground state well enough, the convexity m_{thr} will be larger than the ground state gap Δ, as every parameterized rotation in the PQC ansatz will introduce a state orthogonal to the ground state. The norm of the gradient and the Lipschitz constant can be bound proportionally to their max norm, as shown in Appendix B, thus $\|\boldsymbol{\mathcal{G}}\| \leq \sqrt{n_p}\|H\|$, $\|\dot{\boldsymbol{\mathcal{G}}}\| \leq \sqrt{n_p}\|\dot{H}\|$ and $L \leq n_p^{3/2}\|H\|$ (where $\|H\|$ is the spectral norm of the Hamiltonian). The number of measurements to sample the Hessian to precision Eq. (25) are proportional to the inverse of the bound, with proportionality constant $M_{\mathcal{H}}$ indicating the number of shots required to sample a single element of the Hessian to unit variance (this depends on details such as the decomposition taken to measure the Hamiltonian, and the specifics of the derivative estimation method). Multiplying this by the number of steps $\frac{1}{dt}$ [Eq. (19)] gives us the total number of shots required for convergence

$$M_{\text{tot}} = 10^3\frac{n_pL^3\|\boldsymbol{\mathcal{G}}\|^2\dot{\mathcal{G}}_{\max}}{\Delta^8}M_{\mathcal{H}} \tag{26}$$

$$< 10^3\frac{n_p^4\|H\|^7\|\dot{H}\|}{\Delta^8}M_{\mathcal{H}}. \tag{27}$$

5 Adapting to an orbital-optimized PQC ansatz

To achieve a good representation of the ground state character while minimizing depth and number of evaluations of quantum circuits, we employ a hybrid ansatz composed of classical orbital rotations and a parameterized quantum circuit (PQC) to represent

correlations within an active space. The concept of an orbital-optimized variational quantum eigensolver (OO-VQE) is explored in [46, 47], as an extension of VQE conceptually similar to a complete active space self-consistent field (CASSCF) calculation. In this section, we introduce the construction of the OO-PQC ansatz and discuss its specific use in our algorithm, where the orbitals need to continuously track a changing active space depending on the nuclear geometry.

5.1 An OO-PQC ansatz with geometric continuity

To represent the electronic structure state, we start by choosing an atomic orbital basis, i.e. a discretization of space defined by a set of N non-orthogonal atomic orbitals $\chi_\mu(\mathbf{R}, \boldsymbol{x})$ (functions of the electronic coordinate $\boldsymbol{x} \in \mathbb{R}^3$, where we make explicit the parametric dependence on the nuclear coordinates $\mathbf{R}$); these orbitals define the overlap matrix

$$S_{\mu\nu}(\mathbf{R}) = (\chi_\mu(\mathbf{R})|\chi_\nu(\mathbf{R})) \tag{28}$$

$$:= \int_{\mathbb{R}^3} \chi^*_\mu(\mathbf{R}, \boldsymbol{x})\chi_\nu(\mathbf{R}, \boldsymbol{x})\, d^3\boldsymbol{x}. \tag{29}$$

The atomic orbitals (AOs), along with the overlap matrix, depend on the geometry of the molecule specified by the nuclear coordinates $\mathbf{R}$. (For the sake of simplicity, we limit to considering real AOs.) From these, we could define a set of parameterized orthonormal molecular orbitals (MO)

$$\phi_p(\mathbf{R}, C^{\mathrm{AO}}) = \sum_\mu \chi_\mu(\mathbf{R}) C^{\mathrm{AO}}_{\mu p}, \tag{30}$$

which would allow for the definition of a parameterized active space. The downside of this parametrization is that, to ensure MO orthonormality, we need C^{AO} to satisfy the constraint

$$C^{\mathrm{AO}\dagger} S(\mathbf{R}) C^{\mathrm{AO}} = \mathbb{1}, \tag{31}$$

which depends nontrivially from $\mathbf{R}$. This implies that we cannot trivially use the same C^{AO} for different geometries $\mathbf{R}$.

In order to address this problem, we have opted to use orthonormalized atomic orbitals (OAO) that are derived from the AOs through symmetric Löwdin orthogonalization [48] as reference in the definition of parameterized MOs. The OAOs are defined as $\phi^{\mathrm{OAO}}_p(\mathbf{R}) = \sum_\mu \chi_\mu(\mathbf{R}) S^{-1/2}_{\mu p}(\mathbf{R})$. Building on these, we can define the MOs as

$$\phi_q(\mathbf{R}, C) = \sum_{\mu,p} \chi_\mu(\mathbf{R}) S^{-1/2}_{\mu p}(\mathbf{R}) C_{pq}, \tag{32}$$

where $C = S^{1/2}(\mathbf{R}) \cdot C^{\mathrm{AO}}$. The orthonormality constraint Eq. (31) then reduces to requiring C to be orthogonal, and it is independent on $\mathbf{R}$. In summary, Eq. (32) defines a set of orthonormal molecular orbitals parameterized by C, well-defined and continuous for changing $\mathbf{R}$.

To start up our algorithm, the matrix C^{AO} can be initialized by a Hartree-Fock (or any other molecular coefficient matrix, e.g. coming from a small CASSCF calculation) at some initial geometry $\mathbf{R}(0)$. From this we recover $C = S^{1/2}(\mathbf{R}(0)) \cdot C^{\mathrm{AO}}$, which is then treated as a variational parameter of the ansatz. Using the parameterized MO Eq. (32) we construct the electronic structure Hamiltonian $H(R, C)$ in the (parameterized) molecular basis.

Based on the initial Hartree-Fock orbital energies, we split the N orbitals into a core set with N_{O} doubly-occupied orbitals, an active set with N_{A} orbitals, and a virtual set with N_{V} empty orbitals. Although the split of the orbital indices remains constant throughout the algorithm, the orbitals themselves continuously change through their dependence on $\mathbf{R}$ and C. The correlations are treated only within an active space of η_{A} electrons in N_{A} orbitals. Tracing out the core and virtual orbitals yields the active-space Hamiltonian $H_{\mathrm{A}}(\mathbf{R}, C)$.

The correlated active-space state $|\psi(\boldsymbol{\theta})\rangle$ is represented on a quantum device, using a PQC ansatz of the form Eq. (9). The cost function then becomes

$$E(\mathbf{R}, C, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H_{\mathrm{A}}(\mathbf{R}, C)|\psi(\boldsymbol{\theta})\rangle, \tag{33}$$

and it can be evaluated by sampling the 1- and 2-electron reduced density matrix (RDM) of the state [49]. (Other efficient sampling schemes, e.g. based on double factorization [50, 51], can be used.)

5.2 Measuring boundary terms with the OO-PQC ansatz

When evaluating the final overlap [Eq. (14)] with an orbital-optimized ansatz, we have to consider that the states $|\psi(\boldsymbol{\theta}_0)\rangle$ and $|\psi(\boldsymbol{\theta}_1)\rangle$ are defined on different active space orbitals, determined by the MO matrices C_0 and C_1 respectively. The transformation between the two sets of orbitals, $\boldsymbol{\phi}(\mathbf{R}, C_1) = \boldsymbol{\phi}(\mathbf{R}, C_0) \cdot C_{0\to1}$, is represented by the orthogonal matrix

$$C_{0\to1} = C_0^\dagger C_1. \tag{34}$$

If the algorithms successfully tracked the lowest-energy active space state of the system, the Hilbert spaces spanned by the active orbitals defined by C_0 and C_1 should match. (The same is true for the core space and the virtual space.) This implies the matrix $C_{0\to1}$ will have block structure, with $[C_{0\to1}]_{pq} \neq 0$ only if p, q are both in the same set of orbitals (core, active or virtual). The orbital rotations within the core (virtual) subspace will not generate any phase on the state of the system, as all orbitals are doubly-occupied (doubly-unoccupied). The orbital rotation within the active space can then be translated to a

unitary transformation on the state by a Bogoliubov transformation

$$G_{0\to1} = \exp\Big\{ \sum_{\{p,q\}\in \mathrm{AS}} [\log(C_{0\to1})]_{pq} c_p^\dagger c_q \Big\}, \tag{35}$$

with $c_p^\dagger, c_p$ fermionic creation and annihilation operators on the p orbital. The final overlap

$$\mathrm{Re}\left[\langle 0|U^\dagger(\tilde{\boldsymbol{\theta}}_0) G_{0\to1} U(\tilde{\boldsymbol{\theta}}_1)|0\rangle\right] =: \omega_{\mathcal{C}} \tag{36}$$

can then be sampled with a Hadamard test, given an quantum circuit implementing the (eventually controlled) operation $G_{0\to1}$. Under Jordan-Wigner encoding, a quantum circuit for $G_{0\to1}$ can be implemented as a fabric of parameterized fermionic swap gates of depth N_A following a QR decomposition of the orbital rotation generator $[\log(C_{0\to1})]_{pq}$, also known as a givens rotation fabric [52]. These gates preserve the zero-electrons reference state, allowing to employ the ancilla-free echo verification technique mentioned in section 3.4 to measure the final overlap.

5.3 Newton-Raphson updates of the OO-PQC ansatz

The proposed OO ansatz has two sets of parameters, C and $\boldsymbol{\theta}$. As the MO matrix C is subject to the constraint Eq. (31), its elements cannot be freely updated with NR. Instead, for each NR update with initial MO matrix C, we reparametrize the MOs with a unitary transformation: $C \leftarrow Ce^{-\kappa}$, where κ is any antisymmetric matrix. The derivatives of the energy with respect to any element κ_{pq} can be evaluated analytically (see Appendix C). Furthermore, under this parametrization it can be shown that κ_{pq} where p, q are both core indices or both virtual indices are *redundant* [29] in the definition of the active space orbitals; these $N_\mathrm{O}^2 + N_\mathrm{V}^2$ parameters are set to zero without reducing the expressivity of the ansatz. We call the unraveled set of remaining parameters $\boldsymbol{\kappa}$. To implement the Newton-Raphson step, the gradient and Hessian with respect to the combined set of parameters $(\boldsymbol{\theta}, \boldsymbol{\kappa})$ is computed. In this manner the gradient splits into two components, and the Hessian into three components

$$\nabla_{(\boldsymbol{\theta},\boldsymbol{\kappa})} E = (\nabla_{\boldsymbol{\theta}} E, \nabla_{\boldsymbol{\kappa}} E) \tag{37}$$

$$\nabla^2_{(\boldsymbol{\theta},\boldsymbol{\kappa})} E = \begin{bmatrix} \nabla^2_{\boldsymbol{\theta}} E & \nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E \\ (\nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E)^\intercal & \nabla^2_{\boldsymbol{\kappa}} E \end{bmatrix}. \tag{38}$$

The PQC parameter derivatives $\nabla_{\boldsymbol{\theta}} E$ and $\nabla^2_{\boldsymbol{\theta}} E$ can be evaluated through parameter-shift rule [39, 53] or finite difference, by sampling on the quantum device. The derivatives with respect to the OO parameters $\nabla_{\boldsymbol{\kappa}} E$ and $\nabla^2_{\boldsymbol{\kappa}} E$ are linear functions of the 2-electron RDM, whose coefficients can be computed analytically [29]. The remaining component, the mixed Hessian $\nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E$ can be similarly evaluated as a linear function of $\boldsymbol{\theta}$-derivatives of the RDM; for this, we can use the same data sampled from the quantum device to evaluate $\nabla_{\boldsymbol{\theta}} E$. We detail the procedure of estimating these Hessian components in App. C. Thus, evaluating the derivatives with respect to the OO parameters does not require extra sampling on the quantum device.

6 Numerical results

In this section, we demonstrate the application of our method to a small model system: the formaldimine molecule $H_2C{=}NH$, an established model in the context of quantum algorithms for excited states in [46, 47, 54]. This molecule is known to have a conical intersection between the singlet ground state and first excited state potential energy surfaces [55]. This CI plays an important role in the photoisomerization process of formaldimine, which in turn can be considered a minimal model for the photoisomerization of the rhodopsin protonated Schiff-base (a key step in the visual cycle process [56, 57]). We consider geometries obtained from the equilibrium configuration by varying the direction of the N−H bond, defined by the bending angle α and the dihedral angle ϕ (see Fig. 1d). Varying these angles defines the considered plane in nuclear configuration space $\mathcal{R}$. First, we consider a minimal model of formaldimine (within the minimal basis and a small active space), on which we can test the properties of the algorithm 2. Then, we investigate the effects of sampling noise on these results. Finally, we study a more complex model of the same molecule (with a larger basis set and active space), and show that we can achieve similar results by employing regularization and backtracking to deal with the degeneracies of the ansatz manifold.

6.1 Numerical simulation details

For our simulations, we use the PennyLane [58] package with the PyTorch backend to construct the hybrid quantum-classical cost function of Eq. (33) supporting automatic differentiation (AD) with respect to the PQC parameters. To achieve this, we implement the transformation of the one- and two-body AO basis integrals to the parameterized MO basis [Eq. (32)] with AD support. The transformed integrals are projected onto the active space and contracted with the active space one- and two-electron RDM. The RDM elements and their derivatives with respect to the PQC parameters are obtained using PennyLane and its built-in AD scheme based on the parameter shift rule. In summary, gradients and Hessians of the cost function with respect to the PQC parameters are obtained using AD, the orbital gradient and Hessian components are estimated analytically (see Appendix C), and the off-diagonal block of the composite Hessian [Eq. (38)] is retrieved by automatic

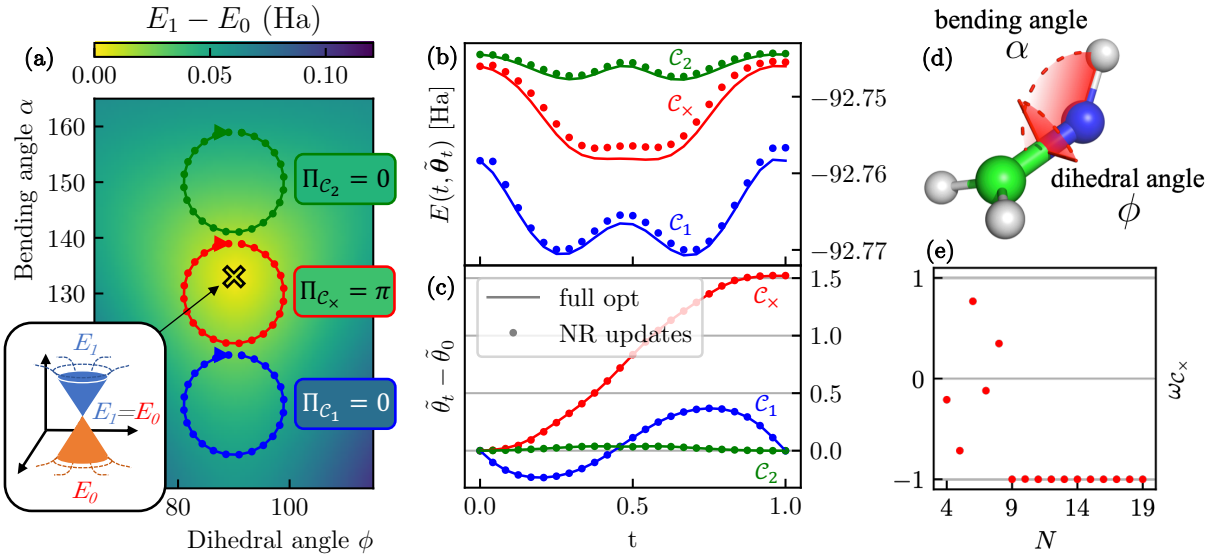


Figure 1: **(a)** Three loops in the nuclear configuration space of formaldimine; $\mathcal{C}_1$ (green), $\mathcal{C}_\times$ (red) and $\mathcal{C}_2$ (blue). $\mathcal{C}_\times$ encircles a CI, resulting in a non-trivial Berry phase. In this representation, the loops are discretized by $N = 25$ points. The color plot indicates the energy gap at the full configuration interaction (FCI) level. **(b)** Energy and **(c)** change in PQC parameter around the three loops, with the same color coding as in (a) and the same $N = 25$. These results refer to a (2,2) active space, and a minimal (STO-3G) basis set description of Formaldimine, with a OO-UCCD ansatz that has a single θ parametrizing the only double excitation. The continuous lines show the true optimum $\boldsymbol{\theta}_t^*$ and the relative energy (obtained by full optimization), while the markers show the progress of the estimate $\tilde{\boldsymbol{\theta}}_t$ from Algorithm 2, in absence of sampling noise. The Hessian stays positive throughout the path, no regularization is needed. **(e)** Final overlap computed by Algorithm 2 for the red loop containing a CI, for a varying number of total discretization points N. For $N < 9$, the Hessian is not always positive and regularization is needed to invert the Hessian, but no backtracking is used. **(d)** Schematic representation of Formaldimine, indicating the parameters used to define the nuclear geometries in this work.

differentiation of the analytical orbital gradient. We use PySCF [59] to generate the molecular integrals (i.e. the full space one- and two-electron integrals and overlap matrices) in the atomic orbital basis.

The core code developed for this project is made available as a python package in the GitHub repository [26]. This code provides a flexible implementation of the orbital-optimized PQC ansatz, which can find many applications in VQAs for chemistry. A tutorial Jupyter notebook showcasing a calculation of Berry phase in the minimal model of Formaldimine is provided in the `examples` folder in the repository.

6.2 Minimal model with an degeneracy-free ansatz

We first demonstrate the application of our algorithm to a minimal model of formaldimine, for which we can approximate the ground state with a simple ansatz with no degeneracies. The molecule is described in a minimal STO-3G basis-set, and we select an active space of $\eta_\mathrm{A} = 2$ electrons in $N_\mathrm{A} = 2$ spatial orbitals [i.e. CAS(2,2)]. As the orbital optimization already allows (spin-adapted) single excitations within the active space, the only parameterized gate we can include in our PQC ansatz is the double-excitation $U(\theta) = e^{\theta(c_0^\dagger c_1^\dagger c_2 c_3 - c_2^\dagger c_3^\dagger c_0 c_1)}$; this corresponds to the unitary coupled-cluster doubles [UCC(S)D] ansatz, where the singles (S) are not explicitly included because they would be redundant with the orbital optimization. This is enough to describe exactly any active space state compatible with the symmetries of the model, without over-parametrizing the ansatz state.

In Fig. 1 we demonstrate the application of our algorithm to this model. The minimal basis set is small enough that we can run a full configuration interaction (FCI) calculation to exactly resolve the ground and first excited state energies $E_0(\mathbf{R})$ and $E_1(\mathbf{R})$. Observing the gap $E_1(\mathbf{R}) - E_0(\mathbf{R})$ (portrayed in Fig. 1a), we can determine the location of the conical intersection $\phi^\times = 90°$, $\alpha^\times \approx 132°$. We then define three loops in the configuration space $\mathcal{R}$, one loop $\mathcal{C}_\times$ containing the CI and two "trivial" loops $\mathcal{C}_1$, $\mathcal{C}_2$. These loops are centered around $\phi = 90°$ and $\alpha = 110°$ ($\mathcal{C}_1$), $\alpha = 130°$ ($\mathcal{C}_\times$) and $\alpha = 150°$ ($\mathcal{C}_2$), and all have a radius of 10°. Fig. 1c shows the progress of the estimate $\tilde{\theta}_t$ (shifted by $\tilde{\theta}_0$) of the optimal PQC parameter θ_t^* throughout the $N = 25$ single-NR-update t-steps. Note that, while we only plot the single PQC parameter relative to the parameterized double excitation, the algorithm updates the MO coefficients $\tilde{C}_t$ as well. We observe that $\tilde{\boldsymbol{\theta}}_1 = \tilde{\boldsymbol{\theta}}_0$ for the trivial loops $\mathcal{C}_1, \mathcal{C}_2$, while $\tilde{\boldsymbol{\theta}}_1 \neq \tilde{\boldsymbol{\theta}}_0$ for the loop $\mathcal{C}_\times$ containing the CI. This is an indication of the effect of the Berry phase, but not the result of the algorithm yet; measuring the overlap Eq. (36) yields the correct Berry

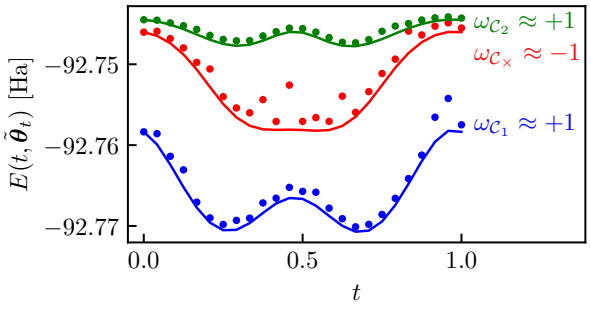


Figure 2: Energies throughout the loops in Fig. 1 (discretized by $N = 25$ steps), in the presence of sampling noise. Each element of the composite gradient Eq. (37) and Hessian Eq. (38) are perturbed by random gaussian noise with variance $\sigma^2 = 5 \times 10^{-6}$. In this instance the Hessian stays positive around the path, so no regularization is needed. The full lines, like in Fig. 1b, indicate the true optimum $E(t, \boldsymbol{\theta}_t^*)$ obtained by full optimization.

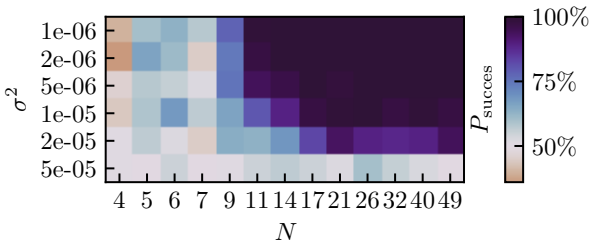


Figure 3: Success probability P_{succes} of the algorithm as a function of the number of discretization points N and the sampling noise variance σ^2 for the loop $\mathcal{C}_\times$ containing a CI. The details of the model and geometry of the loop $\mathcal{C}_\times$ match Fig. 1 (red loop). Success is defined by resolving a final overlap $\langle\psi(\tilde{\boldsymbol{\theta}}_0)|\psi(\tilde{\boldsymbol{\theta}}_1)\rangle < 0$, which returns $\Pi_{\mathcal{C}_\times} = \pi$. The success probability is computed over 100 simulated runs. A $P_{\text{success}} \approx 50\%$ indicates the algorithm returns random outcomes. Regularization by Hessian augmentation is enabled in these calculations, while no back-tracking is used.

phase $\Pi_{\mathcal{C}} = \arg[\omega_{\mathcal{C}}]$. The estimated energy $E(t, \tilde{\boldsymbol{\theta}}_t)$ and the optimal $E(t, \boldsymbol{\theta}_t^*)$ (obtained by full local optimization) are shown in Fig. 1b. We can observe a small deviation from the optimal energy in the region where the character of the state changes faster (along the line $\phi = \phi^\times$, $\alpha < \alpha^\times$), but this does not disrupt the tracking of the minimum. Finally, Fig. 1e shows that a number of discretization points $N \geq 9$ is needed to correctly resolve $\Pi_{\mathcal{C}_\times} = \pi$, through the evaluation of the overlap [Eq. (36)] $\omega_{\mathcal{C}_\times} = -1$.

6.3 Sampling noise

In this section, we explore the robustness of our algorithm with respect to the sampling noise characteristic of VQAs. We directly add a proxy of sampling noise η to each element of the gradient and Hessian. Each η is an independent Gaussian random variable with variance σ^2; a different η is added to each element of the gradient of (37) and Hessian of Eq. (38) to get the noisy estimates $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{H}}$. In this way, one avoids defining a specific sampling strategy, keeping our results general.

In Fig. 2 we show the energy profile of the three loops whose geometry is represented in Fig. 1a, for one such random realisation of the sampling noise. (The plotted energy expectation is evaluated exactly, noise is only added to the gradient and Hessian used in the NR updates.) These three loops yield the same Berry phase results as the noiseless case.

The probability P_{success} of Algorithm 2 correctly resolving the Berry phase $\Pi_{\mathcal{C}_\times}$ on the nontrivial loop $\mathcal{C}_\times$ is reported in Fig. 3, as a function of the number of discretization steps N and of the variance of the added noise on each sampled quantity σ^2. The expected final overlap Eq. (36) is -1 for this case, as the loop contains a CI. For each value of N and σ^2, we simulate 100 noisy runs of the algorithm and we declare as successful the ones that yield a negative final overlap (implying $\Pi_{C_\times} = \pi$). Finally we average these outcomes to retrieve the succes probabilities P_{succes}.

From these simulations we conclude that sampling noise reduces the accuracy of tracking the ground state and thus increases the probability of obtaining inaccurate energies. Nevertheless, for a moderate amount of sampling noise our algorithm still resolves the Berry phase correctly. We observe that an error on each gradient and Hessian element with variance of $\sigma^2 = 10^{-5}$ (or smaller) does not compromise the resolution of the Berry phase, as long as the number of discretization points is sufficiently large ($N > 10$, very close to the noiseless case portrayed in Fig. 1e). On the other hand, a large enough sampling error ($\sigma^2 \geq 5 \times 10^{-5}$) produces essentially random results ($P_{\text{success}} \approx 50\%$).

The number of shots required to achieve a given accuracy on gradient and Hessian depends on the chosen measurement optimization scheme and the technique used to estimate derivatives. We note that the energy can be expressed as a linear function of one- and two-electron RDMs,

$$E(t, \boldsymbol{\theta}) = \sum_{\xi} D(\boldsymbol{\theta})_\xi h(t)_\xi, \tag{39}$$

where $\xi = pq, pqrs \in [N_{\text{A}}]$ are the one- and two-body active-space orbital indices, $D(\boldsymbol{\theta})_\xi$ are the RDM elements and $h(t)_\xi$ are the one- and two-electron integrals. The RDM elements can be grouped and measured together with optimized shot allocation [49, 60–63], with the potential for further optimization based on basis selection [64] or on a more advanced factorization of the cost function [51, 65]. Derivatives are also linear in the RDMs evaluated at different parameter values (see App. C), and they can be reconstructed either through finite-difference methods or using pa-

rameter shift rules [39, 66].

To provide an upper bound on the number of shots necessary to achieve the required accuracy in our numerical benchmark, we consider an un-optimized measurement scheme which samples each element of the RDMs separately. In this scheme, optimal resource allocation requires each $D(\boldsymbol{\theta})_\xi$ to be sampled with a number of shots proportional to $|h(t)_\xi|$ [67]. This allows to estimate the cost function to accuracy σ with a number of shots upper bounded by $N_{\text{shots}} = \sigma^{-2}||\vec{h}(t)||_1^2$, with $||\vec{h}(t)||_1 = \sum_\xi |h(t)_\xi|$ the total one-norm of the electronic integrals. The Gradient and Hessian with respect to a single parameter are estimated through the parameter shift rule only requiring 3 parameter points. The same RDM samples can be used to estimate all elements of the mixed gradient and Hessian, and we assume that the one-norm of the derivatives of h with respect to the orbital rotations are of the same magnitude as $||\vec{h(t)}||_1$. In our simulation benchmark (on the path $\mathcal{C}_\times$), the one-norm of the integral takes values up to $\max_t ||\vec{h}(t)||_1 = 1.5376$. Thus, we estimate an upper bound on the total number of shots of the order of $\approx 3 \times 2.4\sigma^{-2}$.

6.4 Larger basis and active space

To test convergence for a more realistic case where the cost function is not always strongly convex at its minima, we simulate the algorithm on a more challenging model of formaldimine. The model is constructed employing the cc-pVDZ basis set (43 atomic orbitals), and an active space of four electrons in four spatial orbitals [CAS(4,4)]. As a PQC ansatz for the active space state, we use the number-preserving fabric (NPF) ansatz introduced in [38], consisting of a fabric of spin-adapted orbital rotations and double excitations on sets of two spatial orbitals (four spin-orbitals). Four layers of this ansatz are enough to recover the exact CASCI ground state energy inside the active space of 4 orbitals, resulting in 20 PQC parameters. This is an overparameterization of the ground state, implying a global redundancy in the ansatz and resulting in a singular hessian at every point. The goal of this numerical demonstration is to show that Algorithm 2 can still recover the Berry phase, in this case using regularization and backtracking.

In Fig. 4 the energies throughout two loops are shown. The location of the conical intersection $(\alpha^\times, \phi^\times)$ in the larger basis set moves compared the case shown in Fig. 1 (this is to be expected, as the cc-pVDZ and STO-3G models are effectively different); the basis is now too large to attempt an FCI calculation that would resolve the gap exactly. One could instead resort to a State-Averaged CASSCF calculation to resolve the location of the CI, however, the state-average approach might bias the location of the CI. For this demonstration, we manually select two

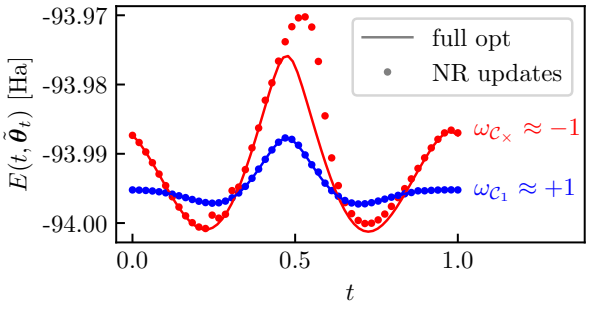


Figure 4: Energies throughout two loops, for a more challenging model of Formaldimine, described in the cc-pVDZ basis set and with a (4,4) active space. The (red) blue loop indicates a (non-)trivial Berry phase, centered around ($\alpha = 113°$) $\alpha = 145°$ and $\phi = 90°$ (both), with a radius of 15°, and discretized by $N = 50$ points (note these loops are different from Fig. 1). The basis set is too large to run a FCI calculation to locate the CI, and using another approximate method (e.g. state-average CASSCF) might bias the CI location. Instead, we manually choose larger loop geometries and use Algorithm 2 to find the value of $\Pi_{\mathcal{C}}$.

loops with a slightly larger radius of 15°, centered around $\alpha = 113°$ ($\mathcal{C}_\times$, red line) and around $\alpha = 145°$ ($\mathcal{C}_1$, blue line). These values are chosen based on the location of the CI at the level of theory of large state-average CAS(14,14)SCF calculation, which returns $\alpha^\times \approx 113°$. We choose $\phi = 90°$, as the CI is forced to lie on the $\phi^\times = 90$ hyperplane due to the Cs reflection point-group symmetry. Due to the single-step Newton-Raphson optimization, we can observe a difference between the energies evaluated at each iteration point of our algorithm (circle markers) and the variational optima (solid lines). The discrepancy is more pronounced after a large change in the state character, such as happens just after $t = 0.5$ in the topologically-nontrivial loop $\mathcal{C}_\times$. There, the variational parameters need to change more rapidly to track the state. As long as the discretization step is small enough, the single Newton-Raphson steps ensure the variational state keeps tracking the local minimum. Indeed, we can resolve the correct Berry phase with only $N = 50$ discretiation points, recovering a final overlap [Eq. (36)] of $\omega_{\mathcal{C}_\times} = -0.9994$ (loop containing the CI), and $\omega_{\mathcal{C}_1} = 0.99998$ (trivial loop).

7 Conclusion and outlook

In this work, we introduced a hybrid algorithm to resolve conical intersections through the Berry phase they induce. This is achieved by tracking the ground state with a variational quantum ansatz, along a closed path $\mathcal{C}$ in nuclear configuration space. This algorithm only requires approximating the ground state (in contrast to e.g. the state-average VQE [46]) for one nuclear geometry $\mathbf{R}$ at a time, reducing the expressivity requirements of the ansatz. The key re-

quirement of the algorithm is that the ansatz parameters are changed smoothly, tracking a local minimum of the cost function (7) and ensuring the $U(1)$-gauge (global phase) of the ansatz state remains well-defined.

As the output is quantized ($\Pi_{\mathcal{C}} \in \{0, \pi\}$), the result only needs to be estimated to a constant precision, and the algorithm is robust to some amount of error; we considered optimization error (the error on estimates of a variational parameter, due to the approximate minimizer) and sampling noise on the quantities measured on the quantum device. We showed that one can update the variational parameters with a single newton step for each geometry in a discretization of the loop $\mathcal{C}$. We proved analytically that the algorithm is granted to converge for a large enough number of discretization steps N, and small enough additive error, under the assumption that the cost function is strongly convex at its minimum. (We consider sampling noise explicitly, but the robustness results extends to any additive noise, including hardware noise.) We argue this result practically extends to cases where the strong convexity assumption is not satisfied, as long as some regularization technique is employed.

This reasoning is corroborated by numerical demonstrations of CI resolution on a small example system – the formaldimine molecule. Using a minimal description of formaldimine [STO-3G basis, a CAS(2,2), UCCD ansatz], for which we have a strongly convex cost function, we show convergence of Algorithm 2 without using regularization for a sufficiently large $N > 11$, as we expect from our analytical results. We also demonstrate the effect of sampling noise in this setting, showing that our algorithm is robust to a sizable amount of noise, achieving convergence for N comparable to the noiseless case. Finally, we demonstrate the application of Algorithm 2 with regularization on a more complicated and realistic model of formaldimine [cc-pVDZ basis, CAS(4,4), NPF ansatz]. This case shows that, even with a cost function that is never convex, we can employ regularization to resolve the Berry phase correctly.

7.1 Paths towards improving convergence

The key step in our algorithm, where most of the cost in terms of quantum resources is concentrated, is the evaluation of the (NR) parameter update. Ensuring the parameter estimates remain within the basin of convergence of the cost function is crucial, and it is the bottleneck in terms of the cost of our algorithm. As shown in Section 4, the size of the basin of convergence depends on the convexity of the cost function at the minimum (11). Overparametrizations (local or global) of the cost function are especially disrupting, as they produce singular Hessians ($m = 0$) even at optimal points. In this work, we proposed to use regularization by Hessian augmentation and back-tracking to solve this problem; this technique is practical and, as shown numerically in Section 6.4, it can produce convergent results for systems with overparametrized cost functions. However, the success of these techniques may depend on the choice of their hyperparameters (α, β, μ, ρ in Algorithm 1), and their application makes the analytical study of the algorithm convergence harder. In this section, we suggest alternative approaches to regularization and resource allocation for the algorithm, which would require further study.

Quantum natural gradients — Quantum natural gradient (QNG) descent is a recently-proposed parameter update technique [68, 69], which takes into account the geometry of the ground state manifold. Natural gradient techniques, already long in use in classical machine learning [70], are invariant with respect to reparametrizations of the cost function; more importantly for our case, they nullify the effect of overparametrizations [71]. The idea of QNG is to transform gradients with respect to the ansatz parameters into gradients with respect to Quantum Information Geometry. This reparametrization is achieved through the Fubini-Study metric tensor $g(\tilde{\boldsymbol{\theta}}_t)$ (to be evaluated at each update). A gradient descent step would then become:

$$\tilde{\boldsymbol{\theta}}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t - \eta g^{+}(\tilde{\boldsymbol{\theta}}_t)\boldsymbol{\mathcal{G}}(t + \Delta t, \tilde{\boldsymbol{\theta}}_t), \tag{40}$$

where η is the learning rate and $g^{+}(\tilde{\boldsymbol{\theta}}_t)$ is the pseudo-inverse of the metric tensor. Here $\boldsymbol{\mathcal{G}}(t+\Delta t, \tilde{\boldsymbol{\theta}}_t)$ is just the usual gradient of the energy as in Algorithm 2. The resulting QNG step updates the ansatz by a fixed amount in the norm induced by the distance between quantum states, instead of a parameter-space norm, solving the issues connected to overparametrization. Another option would be to use classical natural gradients (NG) [70] defined on the 1- and 2-RDM manifold, which rely on the classical Fisher information matrix.

Adaptive step selection — Choosing a sufficient number of steps N to discretize $\mathcal{C}$ is key to the success of our algorithm, as proven in Section 4 and shown in Fig. 1 (bottom right). This is because the minimum $\boldsymbol{\theta}^*_t$, along with its basin of convergence, changes between subsequent steps by an amount proportional to the step size $\Delta t = 1/N$. In Algorithm 2, we propose to linearly discretize a given parametrization of $\mathcal{C}$ for the sake of simplicity. An adaptive choice of Δt could greatly reduce the cost of the algorithm, letting the steps be larger in the regions where the ground state changes the least, while concentrating more points in the regions where the ground state character sharply shifts. An adaptive step selection technique that preserves the provable convergence could be easily implemented if the gradient and Hessian of the cost function are measured through the 1- and 2-electron RDM, as per Eq. (39). The derivatives with respect to $\boldsymbol{\theta}$ are then calculated by chain rule from the deriva-

tives of the RDMs, at the current parameter value of $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_t$. This allows to compute energy $E(t', \tilde{\boldsymbol{\theta}}_t)$, gradient $\mathcal{G}(t', \tilde{\boldsymbol{\theta}}_t)$ and Hessian $\mathcal{H}(t', \tilde{\boldsymbol{\theta}}_t)$ for any value t', without further evaluations of the PQC. The step size can be then chosen as the maximum Δt such that the Hessian convexity $m(t + \Delta t, \tilde{\boldsymbol{\theta}}_t)$ remains above some positive threshold value. Further research could quantify the improvement that adaptive step selection would bring to our algorithm, and identify a method to implement this for optimized energy derivatives sampling techniques, such as those using double factorization [65].

7.2 Potential applications

The algorithm we propose resolves the Berry phase along a given path $\mathcal{C}$; the description of the loop is an input of the algorithm. This loop construction will depend on the details of the considered application, and might involve chemical intuition and the consideration symmetries of the molecule where present. In realistic applications, we conceive our algorithm as a tool that can help to (1) certify CIs proposed by other methods, (2) determine whether a CI plays a role in a certain reaction, and/or (3) locate a point of the CI manifold in parameter space. In either case, an initial proposal of a path $\mathcal{C}$ that might contain the CI is necessary.

The case 1 is the most direct application of our algorithm. The loop $\mathcal{C}$ is chosen to surround a quasi-degeneracy previously identified by an approximate classical method. The result of our algorithm could then confirm or disprove the presence of the CI. In case 2, given a photochemical reaction whose geometry is approximately known, we can use our algorithm on a set of loops to understand wether a CI plays a role in the reaction. These loops can be constructed by variations of the reaction path along perpendicular coordinates, focusing on the modes that influence the orbitals involved in the reaction, thus greatly reducing the search space for the CI. Finally, to locate the CI (case 3) various search approaches can be considered. For example, starting from a large loop $\mathcal{C}$ that is known to contain the CI, binary-search can be used to iteratively shrink the loop and locate the CI to the desired precision. The considered loops could be defined on a plane, if there is heuristic information about the direction along which the potential energy surfaces split. In alternative, a mesh of orthogonal loops in a subspace of the nuclear configuration space $\mathcal{R}$ can be tested. This, in combination with the data about the ground state energy collected by running the optimization in our algorithm, could also be used to determine the approximate location of the minimal energy crossing point [i.e. the point $\mathbf{R}_{\mathrm{MECP}}$ on the CI manifold with minimum $E_0(\mathbf{R}_{\mathrm{MECP}}) = E_1(\mathbf{R}_{\mathrm{MECP}})$]. Further work is needed to explore these problems, develop procedures to solve them and define and test practical application cases in the three categories.

7.3 Outlook

The bounds presented in Section 4 have been calculated to provide a guarantee of convergence for our method, which is an atypical feature for variational quantum algorithms. These are not supposed to be tight bounds or resource estimates for a realistic application of our algorithm. Further research is needed to define better bounds. This, along with the choice of a specific method to extract the energy and its derivatives (e.g. RDM sampling and parameter shift rule) could allow estimating the cost of a practical application of this algorithm. Furthermore, a study of the errors due to the ansatz not perfectly reproducing the ground state, and those induced by circuit noise, could help to understand the practical limitations of the algorithm.

The computation of Berry phases is also central when characterizing topological phases of matter [72]. For the specific case of non-interacting Hamiltonians with chiral or inversion symmetry, the winding number is analytically computed through the Zak phase [73], which is related to the Berry phase that is accumulated after a closed loop through the Brillouin Zone. For interacting systems, in which one cannot access momentum space, there is a mechanism in which one introduces an external periodic perturbation to the Hamiltonian [74, 75]. As long as the perturbation does not close the gap and respects the symmetries of the system, the Berry phase can be computed by considering a closed loop in parameter space, similar to what is proposed in this work. As an outlook, one could consider extending the VQE approach to detect topological phases of matter through the computation of the Zak phase. Furthermore, VQA approaches to other topological invariants, such as the Chern number [76] could be considered (possibly inspired by methods related to their experimental detection, such as Thouless pumping [77]).

Finally, classical algorithms to resolve conical intersections from Berry phases inspired by this approach could be designed. If the approximate ground state is represented by a variational classical ansatz that fixes the $U(1)$ gauge, a simple extension of our method could be achievable. For example, this could be achieved with an extension of CASSCF (essentially a CASCI solver on top of an orbital optimization) that implements continuous local optimization of the SCF matrix, to allow enforcing the smoothness constrains that are crucial to keep the gauge of the state fixed.

Acknowledgements

We thank Joonhoo Lee, Bill Huggins, Nick Rubin, Ryan Babbush, Franco Buda and Maciej Lewenstein for stimulating discussions. We thank Dyon van

Vreumingen and Patrick Emonts for their useful feedback.

EK and SP acknowledge support from Shell Global Solutions BV. AD acknowledges funding from the European Union under Grant Agreement 101080142 and the project EQUALITY. ICFO group acknowledges support from: ERC AdG NOQIA; Ministerio de Ciencia y Innovation Agencia Estatal de Investigaciones (PGC2018-097027-B-100/10.13039/501100011033, CEX2019-000910-S/10.13039/501100011033, Plan National FIDEUA PID2019-106901GB-I00, FPI, QUANTERA MAQS PCI2019-111828-2, QUANTERA DYNAMITE PCI2022-132919, Proyectos de I+D+I "Retos Colaboración" QUSPIN RTC2019-007196-7); MICIIN with funding from European Union NextGenerationEU(PRTR-C17.I1) and by Generalitat de Catalunya; Fundació Cellex; Fundació Mir-Puig; Generalitat de Catalunya (European social fund FEDER and CERCA program, AGAUR Grant No. 2021 SGR 01452, QuantumCAT / U16-011424, co-funded by ERDF Operational Program of Catalonia 2014-2020); Barcelona Supercomputing Center MareNostrum (FI-2022-1-0042); EU Horizon 2020 FET-OPEN OPTOlogic (Grant No 899794); EU Horizon Europe Program (Grant Agreement 101080086 — NeQST), National Science Centre, Poland (Symfonia Grant No. 2016/20/W/ST4/00314); ICFO Internal "QuantumGaudi" project; European Union's Horizon 2020 research and innovation program under the Marie-Skłodowska-Curie grant agreement No 101029393 (STREDCH) and No 847648 ("La Caixa" Junior Leaders fellowships ID100010434: LCF/BQ/PI19/11690013, LCF/BQ/PI20/11760031, LCF/BQ/PR20/11770012, LCF/BQ/PR21/11840013). Views and opinions expressed in this work are, however, those of the author(s) only and do not necessarily reflect those of the European Union, European Climate, Infrastructure and Environment Executive Agency (CINEA), nor any other granting authority.

References

[1] A. K. Geim and K. S. Novoselov. The rise of graphene. *Nature Materials*, 6(3):183–191, March 2007. ISSN 1476-4660. DOI: 10.1038/nmat1849.

[2] Michael Victor Berry. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 392(1802):45–57, March 1984. DOI: 10.1098/rspa.1984.0023.

[3] Wolfgang Domcke, David Yarkony, and Horst Köppel, editors. *Conical Intersections: Theory, Computation and Experiment*. Number v. 17 in Advanced Series in Physical Chemistry. World Scientific, Singapore ; Hackensack, NJ, 2011. ISBN 978-981-4313-44-5.

[4] David R. Yarkony. Nonadiabatic Quantum Chemistry—Past, Present, and Future. *Chemical Reviews*, 112(1):481–498, January 2012. ISSN 0009-2665. DOI: 10.1021/cr2001299.

[5] Dario Polli, Piero Altoè, Oliver Weingart, Katelyn M. Spillane, Cristian Manzoni, Daniele Brida, Gaia Tomasello, Giorgio Orlandi, Philipp Kukura, Richard A. Mathies, Marco Garavelli, and Giulio Cerullo. Conical intersection dynamics of the primary photoisomerization event in vision. *Nature*, 467(7314):440–443, September 2010. ISSN 1476-4687. DOI: 10.1038/nature09346.

[6] Gloria Olaso-González, Manuela Merchán, and Luis Serrano-Andrés. Ultrafast Electron Transfer in Photosynthesis: Reduced Pheophytin and Quinone Interaction Mediated by Conical Intersections. *The Journal of Physical Chemistry B*, 110(48):24734–24739, December 2006. ISSN 1520-6106, 1520-5207. DOI: 10.1021/jp063915u.

[7] Howard E Zimmerman. Molecular Orbital Correlation Diagrams, Mobius Systems, and Factors Controlling Ground- and Excited-State Reactions. II. *Journal of the American Chemical Society*, 88(7):1566–1567, 1966. ISSN 0002-7863. DOI: 10.1021/ja00959a053.

[8] Fernando Bernardi, Massimo Olivucci, and Michael A. Robb. Potential energy surface crossings in organic photochemistry. *Chemical Society Reviews*, 25(5):321–328, 1996. ISSN 0306-0012. DOI: 10.1039/cs9962500321.

[9] Leticia González, Daniel Escudero, and Luis Serrano-Andrés. Progress and Challenges in the Calculation of Electronic Excited States. *ChemPhysChem*, 13(1):28–51, 2012. ISSN 1439-4235. DOI: 10.1002/cphc.201100200.

[10] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6-7):467–488, June 1982. ISSN 0020-7748, 1572-9575. DOI: 10.1007/BF02650179.

[11] Alán Aspuru-Guzik, Anthony D. Dutoi, Peter J. Love, and Martin Head-Gordon. Simulated Quantum Computation of Molecular Energies. *Science*, 309(5741):1704–1707, September 2005. DOI: 10.1126/science.1113479.

[12] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. ISSN 2521-327X. DOI: 10.22331/q-2018-08-06-79.

[13] Alberto Peruzzo, Jarrod R. McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, September 2014. ISSN 2041-1723. DOI: 10.1038/ncomms5213.

[14] Jarrod R. McClean, Jonathan Romero, Ryan Babbush, and Alán Aspuru-Guzik. The theory of

variational hybrid quantum-classical algorithms. *New Journal of Physics*, 18(2):023023, February 2016. ISSN 1367-2630. DOI: 10.1088/1367-2630/18/2/023023.

[15] Dave Wecker, Matthew B Hastings, and Matthias Troyer. Progress towards practical quantum variational algorithms. *Physical Review A*, 92(4):042303, October 2015. ISSN 1050-2947. DOI: 10.1103/PhysRevA.92.042303.

[16] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, November 2018. ISSN 2041-1723. DOI: 10.1038/s41467-018-07090-4.

[17] Shiro Tamiya, Sho Koh, and Yuya O. Nakagawa. Calculating nonadiabatic couplings and berry's phase by variational quantum eigensolvers. *Phys. Rev. Research*, 3:023244, Jun 2021. DOI: 10.1103/PhysRevResearch.3.023244.

[18] Xiao Xiao, J. K. Freericks, and A. F. Kemper. Robust measurement of wave function topology on NISQ quantum computers, October 2022. URL https://doi.10.48550/arXiv.2101.07283.

[19] Bruno Murta, G. Catarina, and J. Fernández-Rossier. Berry phase estimation in gate-based adiabatic quantum simulation. *Phys. Rev. A*, 101:020302, Feb 2020. DOI: 10.1103/PhysRevA.101.020302. URL https://link.aps.org/doi/10.1103/PhysRevA.101.020302.

[20] Hugh Christopher Longuet-Higgins, U. Öpik, Maurice Henry Lecorney Pryce, and R. A. Sack. Studies of the Jahn-Teller effect .II. The dynamical problem. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 244(1236):1–16, February 1958. DOI: 10.1098/rspa.1958.0022.

[21] C. Alden Mead and Donald G. Truhlar. On the determination of Born–Oppenheimer nuclear motion wave functions including complications due to conical intersections and identical nuclei. *The Journal of Chemical Physics*, 70(5):2284–2296, March 1979. ISSN 0021-9606. DOI: 10.1063/1.437734.

[22] Ilya G. Ryabinkin, Loïc Joubert-Doriol, and Artur F. Izmaylov. Geometric Phase Effects in Nonadiabatic Dynamics near Conical Intersections. *Accounts of Chemical Research*, 50(7):1785–1793, July 2017. ISSN 0001-4842. DOI: 10.1021/acs.accounts.7b00220.

[23] Jacob Whitlow, Zhubing Jia, Ye Wang, Chao Fang, Jungsang Kim, and Kenneth R. Brown. Simulating conical intersections with trapped ions, February 2023. URL https://doi.org/10.48550/arXiv.2211.07319.

[24] Christophe H. Valahu, Vanessa C. Olaya-Agudelo, Ryan J. MacDonell, Tomas Navickas, Arjun D. Rao, Maverick J. Millican, Juan B. Pérez-Sánchez, Joel Yuen-Zhou, Michael J. Biercuk, Cornelius Hempel, Ting Rei Tan, and Ivan Kassal. Direct observation of geometric phase in dynamics around a conical intersection. *Nature Chemistry*, 15(11):1503–1508, November 2023. ISSN 1755-4330, 1755-4349. DOI: 10.1038/s41557-023-01300-3.

[25] Christopher S. Wang, Nicholas E. Frattini, Benjamin J. Chapman, Shruti Puri, Steven M. Girvin, Michel H. Devoret, and Robert J. Schoelkopf. Observation of wave-packet branching through an engineered conical intersection. *Physical Review X*, 13(1):011008, January 2023. ISSN 2160-3308. DOI: 10.1103/PhysRevX.13.011008.

[26] Emiel Koridon and Stefano Polla. auto_oo: an autodifferentiable framework for molecular orbital-optimized variational quantum algorithms. Zenodo, February 2024. URL https://doi.org/10.5281/zenodo.10639817.

[27] E. Teller. The Crossing of Potential Surfaces. *The Journal of Physical Chemistry*, 41(1):109–116, January 1937. ISSN 0092-7325. DOI: 10.1021/j150379a010.

[28] G. Herzberg and H. C. Longuet-Higgins. Intersection of potential energy surfaces in polyatomic molecules. *Discussions of the Faraday Society*, 35 (0):77–82, January 1963. ISSN 0366-9033. DOI: 10.1039/DF9633500077.

[29] Trygve Helgaker, Poul Jørgensen, and Jeppe Olsen. *Molecular Electronic-Structure Theory*. Wiley, first edition, August 2000. ISBN 978-0-471-96755-2 978-1-119-01957-2. DOI: 10.1002/9781119019572.

[30] R. Broer, L. Hozoi, and W. C. Nieuwpoort. Non-orthogonal approaches to the study of magnetic interactions. *Molecular Physics*, 101(1-2):233–240, January 2003. ISSN 0026-8976. DOI: 10.1080/0026897021000035205.

[31] Valera Veryazov, Per Åke Malmqvist, and Björn O. Roos. How to select active space for multiconfigurational quantum chemistry? *International Journal of Quantum Chemistry*, 111 (13):3329–3338, 2011. ISSN 1097-461X. DOI: 10.1002/qua.23068.

[32] David R. Yarkony. Diabolical conical intersections. *Reviews of Modern Physics*, 68(4):985–1013, October 1996. DOI: 10.1103/RevModPhys.68.985.

[33] C. Alden Mead. The molecular Aharonov—Bohm effect in bound states. *Chemical Physics*, 49(1):23–32, June 1980. ISSN 0301-0104. DOI: 10.1016/0301-0104(80)85035-X.

[34] Stuart M. Harwood, Dimitar Trenev, Spencer T. Stober, Panagiotis Barkoutsos, Tanvi P. Gujarati, Sarah Mostame, and Donny Greenberg. Improving the Variational Quantum Eigensolver

Using Variational Adiabatic Quantum Computing. *ACM Transactions on Quantum Computing*, 3(1):1:1–1:20, January 2022. ISSN 2643-6809. DOI: 10.1145/3479197.

[35] C. Alden Mead. The "noncrossing" rule for electronic potential energy surfaces: The role of time-reversal invariance. *The Journal of Chemical Physics*, 70(5):2276–2283, March 1979. ISSN 0021-9606. DOI: 10.1063/1.437733.

[36] Rodney J. Bartlett, Stanislaw A. Kucharski, and Jozef Noga. Alternative coupled-cluster ansätze II. The unitary coupled-cluster method. *Chemical Physics Letters*, 155(1):133–140, February 1989. ISSN 0009-2614. DOI: 10.1016/S0009-2614(89)87372-5.

[37] Jonathan Romero, Ryan Babbush, Jarrod R. McClean, Cornelius Hempel, Peter J. Love, and Alán Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology*, 4(1):014008, October 2018. ISSN 2058-9565. DOI: 10.1088/2058-9565/aad3e4.

[38] Gian-Luca R. Anselmetti, David Wierichs, Christian Gogolin, and Robert M. Parrish. Local, expressive, quantum-number-preserving vqe ansatze for fermionic systems. *New Journal of Physics*, 23, 4 2021. DOI: 10.1088/1367-2630/ac2cb3.

[39] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, March 2019. ISSN 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.99.032331.

[40] Hans Jorgen Aa. Jensen and Poul Jorgensen. A direct approach to second-order MCSCF calculations using a norm extended optimization scheme. *The Journal of Chemical Physics*, 80 (3):1204–1214, February 1984. ISSN 0021-9606. DOI: 10.1063/1.446797.

[41] Benjamin Helmich-Paris. A trust-region augmented Hessian implementation for restricted and unrestricted Hartree–Fock and Kohn–Sham methods. *The Journal of Chemical Physics*, 154 (16):164104, April 2021. ISSN 0021-9606. DOI: 10.1063/5.0040798.

[42] Thomas E. O'Brien, Stefano Polla, Nicholas C. Rubin, William J. Huggins, Sam McArdle, Sergio Boixo, Jarrod R. McClean, and Ryan Babbush. Error Mitigation via Verified Phase Estimation. *PRX Quantum*, 2(2), oct 2021. DOI: 10.1103/prxquantum.2.020317.

[43] Stefano Polla, Gian-Luca R. Anselmetti, and Thomas E. O'Brien. Optimizing the information extracted by a single qubit measurement. *Physical Review A*, 108(1):012403, July 2023. DOI: 10.1103/PhysRevA.108.012403.

[44] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer, New York, 2nd ed edition, 2006. ISBN 978-0-387-30303-1.

[45] Eugene P. Wigner. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. ISSN 0003-486X. DOI: 10.2307/1970079.

[46] Saad Yalouz, Bruno Senjean, Jakob Günther, Francesco Buda, Thomas E O'Brien, and Lucas Visscher. A state-averaged orbital-optimized hybrid quantum–classical algorithm for a democratic description of ground and excited states. *Quantum Science and Technology*, 6(2):024004, jan 2021. ISSN 2058-9565. DOI: 10.1088/2058-9565/abd334.

[47] Saad Yalouz, Emiel Koridon, Bruno Senjean, Benjamin Lasorne, Francesco Buda, and Lucas Visscher. Analytical nonadiabatic couplings and gradients within the state-averaged orbital-optimized variational quantum eigensolver. *Journal of Chemical Theory and Computation*, 18(2): 776–794, 2022. DOI: 10.1021/acs.jctc.1c00995. PMID: 35029988.

[48] Per-Olov Löwdin. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *The Journal of Chemical Physics*, 18(3):365–375, 1950. DOI: 10.1063/1.1747632.

[49] Xavier Bonet-Monroig, Ryan Babbush, and Thomas E. O'Brien. Nearly Optimal Measurement Scheduling for Partial Tomography of Quantum States. *Physical Review X*, 10(3): 031064, September 2020. DOI: 10.1103/PhysRevX.10.031064.

[50] Vera von Burg, Guang Hao Low, Thomas Häner, Damian S. Steiger, Markus Reiher, Martin Roetteler, and Matthias Troyer. Quantum computing enhanced computational catalysis. *Physical Review Research*, 3(3):033055, July 2021. ISSN 2643-1564. DOI: 10.1103/PhysRevResearch.3.033055.

[51] Jeffrey Cohn, Mario Motta, and Robert M. Parrish. Quantum Filter Diagonalization with Compressed Double-Factorized Hamiltonians. *PRX Quantum*, 2(4):040352, December 2021. DOI: 10.1103/PRXQuantum.2.040352.

[52] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Benjamin Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Edward Farhi, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Matthew P. Harrigan, Alan Ho, Sabrina Hong, Trent Huang, William J Huggins, Lev Ioffe, Sergei V. Isakov, Evan Jeffrey, Zhang Jiang,

Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Erik Lucero, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Matthew Neeley, Charles Neill, Hartmut Neven, Murphy Yuezhen Niu, Thomas E. O'Brien, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Kevin J. Satzinger, Vadim Smelyanskiy, Doug Strain, Kevin J. Sung, Marco Szalay, Tyler Y. Takeshita, Amit Vainsencher, Theodore White, Nathan Wiebe, Z. Jamie Yao, Ping Yeh, and Adam Zalcman. Hartree-Fock on a superconducting qubit quantum computer. *Science*, 369 (6507):1084–1089, August 2020. ISSN 0036-8075. DOI: 10.1126/science.abb9811.

[53] Patrick Huembeli and Alexandre Dauphin. Characterizing the loss landscape of variational quantum circuits. *Quantum Science and Technology*, 6(2):025011, February 2021. ISSN 2058-9565. DOI: 10.1088/2058-9565/abdbc9.

[54] Hirotoshi Hirai. Excited-state molecular dynamics simulation based on variational quantum algorithms, November 2022. URL https://doi.org/10.48550/arXiv.2211.02302.

[55] Vlasta Bonačić-Koutecký and Josef Michl. Photochemicalsyn-anti isomerization of a Schiff base: A two-dimensional description of a conical intersection in formaldimine. *Theoretica chimica acta*, 68(1):45–55, July 1985. ISSN 1432-2234. DOI: 10.1007/BF00698750.

[56] Robert R. Birge. Nature of the primary photochemical events in rhodopsin and bacteriorhodopsin. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1016(3):293–327, April 1990. ISSN 0005-2728. DOI: 10.1016/0005-2728(90)90163-X.

[57] M Chahre. Trigger and Amplification Mechanisms in Visual Phototransduction. *Annual Review of Biophysics and Biophysical Chemistry*, 14(1):331–360, 1985. DOI: 10.1146/annurev.bb.14.060185.001555.

[58] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M. Sohaib Alam, Guillermo Alonso-Linaje, B. AkashNarayanan, Ali Asadi, Juan Miguel Arrazola, Utkarsh Azad, Sam Banning, Carsten Blank, Thomas R. Bromley, Benjamin A. Cordier, Jack Ceroni, Alain Delgado, Olivia Di Matteo, Amintor Dusko, Tanya Garg, Diego Guala, Anthony Hayes, Ryan Hill, Aroosa Ijaz, Theodor Isacsson, David Ittah, Soran Jahangiri, Prateek Jain, Edward Jiang, Ankit Khandelwal, Korbinian Kottmann, Robert A. Lang, Christina Lee, Thomas Loke, Angus Lowe, Keri McKiernan, Johannes Jakob Meyer, J. A. Montañez-Barrera, Romain Moyard, Zeyue Niu, Lee James O'Riordan, Steven Oud, Ashish Panigrahi, Chae-Yeun Park, Daniel Polatajko, Nicolás Quesada, Chase Roberts, Nahum Sá, Isidor Schoch, Borun Shi, Shuli Shu, Sukin Sim, Arshpreet Singh, Ingrid Strandberg, Jay Soni, Antal Száva, Slimane Thabet, Rodrigo A. Vargas-Hernández, Trevor Vincent, Nicola Vitucci, Maurice Weber, David Wierichs, Roeland Wiersema, Moritz Willmann, Vincent Wong, Shaoming Zhang, and Nathan Killoran. PennyLane: Automatic differentiation of hybrid quantum-classical computations, July 2022. URL https://doi.org/10.48550/arXiv.1811.04968.

[59] Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, Nick S. Blunt, Nikolay A. Bogdanov, George H. Booth, Jia Chen, Zhi-Hao Cui, Janus J. Eriksen, Yang Gao, Sheng Guo, Jan Hermann, Matthew R. Hermes, Kevin Koh, Peter Koval, Susi Lehtola, Zhendong Li, Junzi Liu, Narbe Mardirossian, James D. McClain, Mario Motta, Bastien Mussard, Hung Q. Pham, Artem Pulkin, Wirawan Purwanto, Paul J. Robinson, Enrico Ronca, Elvira R. Sayfutyarova, Maximilian Scheurer, Henry F. Schurkus, James E. T. Smith, Chong Sun, Shi-Ning Sun, Shiv Upadhyay, Lucas K. Wagner, Xiao Wang, Alec White, James Daniel Whitfield, Mark J. Williamson, Sebastian Wouters, Jun Yang, Jason M. Yu, Tianyu Zhu, Timothy C. Berkelbach, Sandeep Sharma, Alexander Yu. Sokolov, and Garnet Kin-Lic Chan. Recent developments in the PySCF program package. *The Journal of Chemical Physics*, 153 (2):024109, July 2020. ISSN 0021-9606. DOI: 10.1063/5.0006074.

[60] William J. Huggins, Jarrod R. McClean, Nicholas C. Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Information*, 7(1):1–9, February 2021. ISSN 2056-6387. DOI: 10.1038/s41534-020-00341-7.

[61] Andrew Zhao, Nicholas C. Rubin, and Akimasa Miyake. Fermionic partial tomography via classical shadows. *Physical Review Letters*, 127(11): 110504, September 2021. ISSN 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.127.110504.

[62] Seonghoon Choi, Tzu-Ching Yen, and Artur F. Izmaylov. Improving quantum measurements by introducing "ghost" Pauli products. *Journal of Chemical Theory and Computation*, 18(12):7394–7402, December 2022. ISSN 1549-9618, 1549-9626. DOI: 10.1021/acs.jctc.2c00837.

[63] Alexander Gresch and Martin Kliesch. Guar-

anteed efficient energy estimation of quantum many-body Hamiltonians using ShadowGrouping, September 2023. URL https://doi.org/10.48550/arXiv.2301.03385.

[64] Emiel Koridon, Saad Yalouz, Bruno Senjean, Francesco Buda, Thomas E. O'Brien, and Lucas Visscher. Orbital transformations to reduce the 1-norm of the electronic structure hamiltonian for quantum computing applications. *Phys. Rev. Res.*, 3:033127, Aug 2021. DOI: 10.1103/PhysRevResearch.3.033127.

[65] Edward G. Hohenstein, Oumarou Oumarou, Rachael Al-Saadon, Gian-Luca R. Anselmetti, Maximilian Scheurer, Christian Gogolin, and Robert M. Parrish. Efficient Quantum Analytic Nuclear Gradients with Double Factorization, July 2022. URL https://doi.org/10.48550/arXiv.2207.13144.

[66] David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, March 2022. ISSN 2521-327X. DOI: 10.22331/q-2022-03-30-677. URL https://doi.org/10.22331/q-2022-03-30-677.

[67] Nicholas C Rubin, Ryan Babbush, and Jarrod McClean. Application of fermionic marginal constraints to hybrid quantum algorithms. *New Journal of Physics*, 20(5):053020, may 2018. DOI: 10.1088/1367-2630/aab919. URL https://dx.doi.org/10.1088/1367-2630/aab919.

[68] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum Natural Gradient. *Quantum*, 4:269, May 2020. ISSN 2521-327X. DOI: 10.22331/q-2020-05-25-269. URL https://doi.org/10.22331/q-2020-05-25-269.

[69] Johannes Jakob Meyer. Fisher Information in Noisy Intermediate-Scale Quantum Applications. *Quantum*, 5:539, September 2021. ISSN 2521-327X. DOI: 10.22331/q-2021-09-09-539.

[70] Shun-ichi Amari. Natural Gradient Works Efficiently in Learning. *Neural Computation*, 10 (2):251–276, 02 1998. ISSN 0899-7667. DOI: 10.1162/089976698300017746.

[71] Tengyuan Liang, Tomaso Poggio, Alexander Rakhlin, and James Stokes. Fisher-Rao Metric, Geometry, and Complexity of Neural Networks, February 2019. URL https://doi.org/10.48550/arXiv.1711.01530.

[72] János K. Asóth, László Oroszlány, and András Pályi. *A short course on topological insulators: band structure and edge states in one and two dimensions*. Springer, 2016. ISBN 9783319256078 9783319256054.

[73] J. Zak. Berry's phase for energy bands in solids. *Phys. Rev. Lett.*, 62:2747–2750, Jun 1989. DOI: 10.1103/PhysRevLett.62.2747.

[74] Yasuhiro Hatsugai. Quantized berry phases as a local order parameter of a quantum liquid. *Journal of the Physical Society of Japan*, 75(12): 123601, 2006. DOI: 10.1143/JPSJ.75.123601.

[75] Takahiro Fukui, Yasuhiro Hatsugai, and Hiroshi Suzuki. Chern numbers in discretized brillouin zone: Efficient method of computing (spin) hall conductances. *Journal of the Physical Society of Japan*, 74(6):1674–1677, 2005. DOI: 10.1143/JPSJ.74.1674.

[76] Shiing-shen Chern. Characteristic Classes of Hermitian Manifolds. *Annals of Mathematics*, 47(1):85–121, 1946. ISSN 0003-486X. DOI: 10.2307/1969037.

[77] Roberta Citro and Monika Aidelsburger. Thouless pumping and topology. *Nature Reviews Physics*, 5(2):87–101, January 2023. ISSN 2522-5820. DOI: 10.1038/s42254-022-00545-0.

[78] D. J. Thouless. Stability conditions and nuclear rotations in the Hartree-Fock theory. *Nuclear Physics*, 21:225–232, November 1960. ISSN 0029-5582. DOI: 10.1016/0029-5582(60)90048-1.

A Bounding overlaps by change in ansatz parameters

The variational Berry phase for a real ansatz state (as introduced in section 3.1) is resolved as the argument of the boundary term $\arg[\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle]$. To obtain a nontrivial Berry phase, the initial and final parameters $\tilde{\boldsymbol{\theta}}(0)$ and $\tilde{\boldsymbol{\theta}}(1)$ need to be far enough to allow $\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle = -1$. This implies that the optimal parameters need to change enough along the parametrization of the path (for t going from 0 to 1). In this appendix, we translate this into a lower bound on the one-norm-distance between initial and final parameters $\|\tilde{\boldsymbol{\theta}}(1) - \tilde{\boldsymbol{\theta}}(0)\|_1$. As a consequence, we also find a lower bound on how much error can be allowed without on the final parameter $\tilde{\boldsymbol{\theta}}(1)$ compromising the Berry phase measurement.

We first state a lemma which will be useful in the proof:

Lemma 1. *Given two unitary operators U, U', each decomposed as a product of $N \geq 2$ unitary operators $U = U_{N-1}...U_1U_0$, there holds the bound $\|U - U'\| \leq \sum_{j\in[N]}\|U_j - U'_j\|$.*

We prove this by induction. For $N = 2$, the proof is by the triangle inequality

$$\begin{aligned}\|U'_1U'_0 - U_1U_0\| &= \|U'_1(U'_0 - U_0) + (U'_1 - U_1)U_0\| &(41)\\ &\leq \|U'_1\|\|(U'_0 - U_0)\| + \|(U'_1 - U_1)\|\|U_0\| &(42)\\ &= \|(U'_0 - U_0)\| + \|(U'_1 - U_1)\|. &(43)\end{aligned}$$

The proof for $N + 1$ can be similarly be reduced to the proof for N:

$$\|U'_NU'_{N-1}...U'_0 - U_NU_{N-1}...U_0\| \leq \|U'_N - U_N\| + \|U'_{N-1}...U'_0 - U_{N-1}...U_0\|. \tag{44}$$

Let us consider the case of real Ansatz Eq. (9). We remind $U_i = e^{A_i\theta_i}$ for antisymmetric real A_i. We can then bound the overlap between ansatz states using Lemma 1,

$$\begin{aligned}1 - \langle\psi(\boldsymbol{\theta})|\psi(\boldsymbol{\theta'})\rangle &= 1 - \langle\psi_0|U^\dagger(\boldsymbol{\theta})U(\boldsymbol{\theta'})|\psi_0\rangle &(45)\\ &= \langle\psi_0|U^\dagger(\boldsymbol{\theta})[U(\boldsymbol{\theta}) - U(\boldsymbol{\theta'})]|\psi_0\rangle &(46)\\ &\leq \|U(\boldsymbol{\theta}) - U(\boldsymbol{\theta'})\| &(47)\\ &\leq \sum_{j\in[N]}\|U_j(\theta_j) - U_j(\theta'_j)\| &(48)\\ &= \sum_{j\in[N]}\|e^{A_j\theta_j} - e^{A_j\theta'_j}\| &(49)\\ &= 2\sum_{j\in[N]}\left\|\sin\left[\frac{A_j}{2}(\theta_j - \theta'_j)\right]\right\| \leq \sum_{j\in[N]}\|A_j\||\theta_j - \theta'_j| &(50)\end{aligned}$$

The difference of parametrers $|\theta_j - \theta'_j|$ is always rescaled by the respective $\|A_j\|$; we can interpret this by considering θ_j and A_j as dimensionful quantities, with inverse dimension to each other. We can always redefine units rescaling θ_j and A_j – without loss of generality we choose units for which $\|A_j\| = 1$ (the only assumption being the boundedness of A_j). Under this choice,

$$\langle\psi(\tilde{\boldsymbol{\theta}}(0))|\psi(\tilde{\boldsymbol{\theta}}(1))\rangle = -1 \implies \|\tilde{\boldsymbol{\theta}}(0) - \tilde{\boldsymbol{\theta}}(1)\|_1 \geq 2 \tag{51}$$

Thus the parameters need to change (in one-norm) by at least 2 along the path to achieve the same state and nontivial Berry phase. By the same reasoning, an error on the final parameters $\delta\tilde{\boldsymbol{\theta}}(1)$ bounded by $\|\delta\tilde{\boldsymbol{\theta}}(1)\|_1 < 1$ will not change the argument (i.e. the sign) of the overlap, thus allowing to resolve the correct Berry phase.

B Bounding the norm of energy derivatives

Suppose we have a variational state parameterized by $\boldsymbol{\theta}$

$$|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|0\rangle = \prod_{k=0}^{n_p-1} U_k(\theta_k)V_k|0\rangle \tag{52}$$

on a n_p-dimensional manifold (within a larger Hilbert space), where the product is assumed to be taken in unitary composition order (right to left). Assume each U_k is generated by anti-hermitian operators iA_k with unit norm

$$U_k(\theta) = e^{iA_k\theta}, \quad A_k = A_k^\dagger, \quad \|A_k\| = 1. \tag{53}$$

Given H is an observable of known norm $\|H\|$, define

$$E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle. \tag{54}$$

Define the tensors of derivatives,

$$\mathcal{G}_j(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_j}E(\boldsymbol{\theta}) \tag{55}$$

$$\mathcal{H}_{jk}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial\theta_j\partial\theta_k}E(\boldsymbol{\theta}) \tag{56}$$

$$\mathcal{T}_{jkl}(\boldsymbol{\theta}) = \frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}E(\boldsymbol{\theta}) \tag{57}$$

$$\tag{58}$$

our goal is to bound their (vector-induced) 2-norms $\|\mathcal{G}\|$, $\|\mathcal{H}\|$, $\|\mathcal{T}\|$.

We first notice that

$$\frac{\partial}{\partial\theta_j}U(\theta) = \left(\prod_{k=j}^{n_p-1}U_k(\theta_k)V_k\right)iA_j\left(\prod_{k=0}^{j-1}U_k(\theta_k)V_k\right) = i\tilde{A}_jU(\theta) \tag{59}$$

where $\|\tilde{A}_j\| = \|A_j\| = 1$, as conjugation by a unitary preserves norm. Using this, we can get expressions for the tensors of derivatives in terms of commutators of Hermitian operators of known norm

$$\mathcal{G}_j(\boldsymbol{\theta}) = i\langle\psi(\boldsymbol{\theta})|[H,\tilde{A}_j]|\psi(\boldsymbol{\theta})\rangle \tag{60}$$

$$\mathcal{H}_{jk}(\boldsymbol{\theta}) = -\langle\psi(\boldsymbol{\theta})|[[H,\tilde{A}_j],\tilde{A}_k]|\psi(\boldsymbol{\theta})\rangle \tag{61}$$

$$\mathcal{T}_{jkl}(\boldsymbol{\theta}) = -i\langle\psi(\boldsymbol{\theta})|[[[H,\tilde{A}_j],\tilde{A}_k],\tilde{A}_l]|\psi(\boldsymbol{\theta})\rangle \tag{62}$$

A trivial bound on this involves bounding each commutator by its norm, e.g.

$$\|\mathcal{G}(\boldsymbol{\theta})\|^2 = \sum_{k=0}^{n_p+1}\|\langle\psi(\boldsymbol{\theta})|[H,\tilde{A}_j]|\psi(\boldsymbol{\theta})\rangle\|^2 \leq 4n_p\|H\|^2. \tag{63}$$

It is an open question wether we can improve on this bound.

We can get a similar result for the derivatives with respect to the orbital rotations, considering the reparametrization described in Sec. 5.3. We call $|\psi\rangle$ the PQC ansatz state in the full (active + core + virtual) space, padded with virtual (core) registers of qubits in the state $|0\rangle$ ($|1\rangle$). We drop explicit dependence on C and $\boldsymbol{\theta}$, and we make explicit the differential rotation parameters κ. The cost function is then

$$E(\kappa) = \langle\psi|e^{\sum_{pq}\kappa_{pq}E_{pq}}He^{-\sum_{rs}\kappa_{rs}E_{rs}}|\psi\rangle, \quad E_{pq} = c^\dagger_{p,\uparrow}c_{q,\uparrow} + c^\dagger_{p,\downarrow}c_{q,\downarrow}, \tag{64}$$

where E_{pq} is the generator of a spin-adapted orbital rotation. Its derivatives at $\kappa = 0$ [note that the index pairs (pq) are collected in one index for the purpose of rotating higher order derivatives] are easily calculated to be

$$\mathcal{G}_{(pq)}(\kappa = 0) = \langle\psi|[H,E_{pq}]|\psi\rangle, \tag{65}$$

$$\mathcal{H}_{(pq),(rs)}(\kappa = 0) = \langle\psi|[[H,E_{pq}],E_{rs}]|\psi\rangle, \tag{66}$$

$$\mathcal{T}_{(pq),(rs),(tu)}(\kappa = 0) = \langle\psi|[[[H,E_{pq}],E_{rs}],E_{tu}]|\psi\rangle. \tag{67}$$

Observing that $\|E_{pq}\| = 2$ we obtain the same result as above (up to constant factors 2, 4, 8 respectively, coming from this norm).

C Analytical orbital gradient and Hessian

In this section, we expand on the estimation of analytic orbital gradient [right block of the vector in Eq. (37)] and orbital-orbital and orbital-circuit Hessian [bottom right and top right blocks of the matrix in Eq. (38), respectively]. We show how after the reparametrization $C \leftarrow C\cdot e^{-\boldsymbol{\kappa}}$, the $\boldsymbol{\kappa}$-derivatives of the cost function $E(C\cdot e^{-\boldsymbol{\kappa}},\boldsymbol{\theta})$ can be expressed as a linear function of the 1- and 2-electron RDM, and the mixed Hessian $\nabla_{\boldsymbol{\kappa}}\nabla_{\boldsymbol{\theta}}E(C\cdot e^{-\boldsymbol{\kappa}},\boldsymbol{\theta})$ can be expressed in terms of $\boldsymbol{\theta}$-derivatives of the same RDM.

We first define the 1- and 2-electron reduced density matrix (RDM) in the spin-restricted formalism

$$\gamma_{pq}(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|E_{pq}|\psi(\boldsymbol{\theta})\rangle \tag{68}$$

$$\Gamma_{pqrs}(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|e_{pqrs}|\psi(\boldsymbol{\theta})\rangle, \tag{69}$$

where $E_{pq} = \sum_\sigma c^\dagger_{p\sigma}c_{q\sigma}$ and $e_{pqrs} = \sum_{\sigma\tau} c^\dagger_{p\sigma}c^\dagger_{r\tau}c_{s\tau}c_{q\sigma} = E_{pq}E_{rs} - \delta_{qr}E_{ps}$. Here, p, q, r, s are meant to be general indices (either occupied, active or virtual), where the state $|\psi(\boldsymbol{\theta})\rangle$ is to be intended as padded by two registers of qubits in the $|0\rangle^{\otimes 2N_\mathrm{V}}$ ($|1\rangle^{\otimes 2N_\mathrm{O}}$) state for the virtual (occupied) orbitals. In the molecular orbital basis defined by C [orbitals in Eq. (32)], we can write the Hamiltonian as

$$H = \sum_{pq} h_{pq}E_{pq} + \frac{1}{2}\sum_{pqrs} g_{pqrs}e_{pqrs} \tag{70}$$

where h_{pq} and g_{pqrs} are the one- and two-electron integrals (with spatial orbital indices p, q, r, s ordered according to the chemists' convention), and they implicitly depend on C through the MOs. The expectation value of the Hamiltonian can then be written as a contraction of the integrals with the RDM,

$$E(C, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H(C)|\psi(\boldsymbol{\theta})\rangle = \sum_{pq}[h(C)]_{pq}\gamma_{pq} + \frac{1}{2}\sum_{pqrs}[g(C)]_{pqrs}\Gamma_{pqrs} \tag{71}$$

where we made explicit the dependence on C.

To derive analytical orbital rotation derivatives, we closely follow Ref. [29]. We start by separating the dependence on κ of the reparametrized cost function $E(C\cdot e^{-\kappa}, \boldsymbol{\theta})$, by using the equivalent state transformation formalism provided by Thouless theorem [78]

$$E(C\cdot e^{-\kappa}, \boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H(C\cdot e^{-\kappa})|\psi(\boldsymbol{\theta})\rangle = \langle\psi(\boldsymbol{\theta})|e^{\hat{\kappa}}H(C)e^{-\hat{\kappa}}|\psi(\boldsymbol{\theta})\rangle \tag{72}$$

where $\hat{\kappa} = \sum_{pq}\kappa_{pq}E_{pq}$ is the operator that generates a unitary on the Hilbert state space equivalent to the orbital rotation. We know that the rotations where p, q are both virtual indices form a redundant subgroup, so we can freeze the corresponding $\kappa_{pq} = 0$; the same is true for p, q both core space indices. In other terms, $\kappa_{pq} = 0$ if $p, q \in V$ or $p, q \in O$, with V and O the sets of virtual and core indices. The remaining elements of κ_{pq} satisfy $\kappa_{pq} = -\kappa_{qp}$. We define a vector of unique non-redundant orbital rotation parameters

$$\boldsymbol{\kappa} = \{\kappa_{pq}, \forall p \in O\cup A, \forall q \in A\cup V : q > p\}, \tag{73}$$

and we redefine the cost function with respect to this vector,

$$E(C, \boldsymbol{\kappa}, \boldsymbol{\theta}) \equiv E(C\cdot e^{-\kappa}, \boldsymbol{\theta}). \tag{74}$$

We are interested in the derivative with respect to this vector; we can always switch from the matrix κ to the unraveled vector of unique non-redundant parameters $\boldsymbol{\kappa}$, and vice versa. By comparing the Taylor series in κ with the Baker-Campbell-Hausdorff expansion:

$$E(\boldsymbol{\theta}, \boldsymbol{\kappa}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle + \langle\psi(\boldsymbol{\theta})|[\hat{\kappa}, H]|\psi(\boldsymbol{\theta})\rangle + \frac{1}{2}\langle\psi(\boldsymbol{\theta})|[\hat{\kappa}, [\hat{\kappa}, H]]|\psi(\boldsymbol{\theta})\rangle + \ldots \tag{75}$$

One can readily verify that the analytical orbital derivatives at $\kappa_{pq} = 0$ are given by:

$$[\nabla_{\boldsymbol{\kappa}}E]_{pq} := \left.\frac{\partial E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial\kappa_{pq}}\right|_{\kappa=0} = \langle\psi(\boldsymbol{\theta})|[E^-_{pq}, H]|\psi(\boldsymbol{\theta})\rangle \tag{76}$$

$$[\nabla^2_{\boldsymbol{\kappa}}E]_{pqrs} := \left.\frac{\partial^2 E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial\kappa_{pq}\partial\kappa_{rs}}\right|_{\kappa=0} = \frac{1}{2}(1 + P_{pq,rs})\langle\psi(\boldsymbol{\theta})|[E^-_{pq}, [E^-_{rs}, H]]|\psi(\boldsymbol{\theta})\rangle \tag{77}$$

where $P_{pq,rs}$ permutes the pair of indices pq with rs. The calculation of the commutators in Eq. (76) and (77) can be found in common quantum chemistry textbooks [29], and they all one- or two-body operators; thus their expectation value can be written as a linear form in the RDM (γ, Γ). The gradient evaluates to

$$[\nabla_{\boldsymbol{\kappa}}E]_{pq} = 2(F_{pq}(\boldsymbol{\theta}) - F_{qp}(\boldsymbol{\theta})) \tag{78}$$

where F is the generalized Fock matrix,

$$F_{pq}(\boldsymbol{\theta}) = \sum_m \gamma_{pm}(\boldsymbol{\theta})h_{qm} + \sum_{mnk}\Gamma_{pmnk}(\boldsymbol{\theta})g_{qmnk} \tag{79}$$

The Hessian evaluates to

$$[\nabla^2_{\boldsymbol{\kappa}} E]_{pqrs} = (1 - P_{pq})(1 - P_{rs}) \left[2\gamma_{pr} h_{qs} - (F_{pr} + F_{rp})\delta_{qs} + 2Y_{pqrs}\right], \tag{80}$$

where we introduced

$$Y_{pqrs} = \sum_{mn} \Gamma_{pmrn} g_{qmns} + \Gamma_{pmnr} g_{qmns} + \Gamma_{prmn} g_{qsmn} \tag{81}$$

and dropped the explicit dependence on $\boldsymbol{\theta}$. For the composite Hessian, we simply take the gradient of Eq. (78) with respect to $\boldsymbol{\theta}$, by using the chain rule

$$[\nabla_{\boldsymbol{\kappa}} \nabla_{\boldsymbol{\theta}} E]_{pq} := \left.\frac{\partial^2 E(\boldsymbol{\theta}, \boldsymbol{\kappa})}{\partial \kappa_{pq} \partial \boldsymbol{\theta}}\right|_{\kappa=0} = 2\left(\frac{\partial F_{pq}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial F_{qp}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \tag{82}$$

where

$$\frac{\partial F_{pq}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_m \frac{\partial \gamma_{pm}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} h_{qm} + \sum_{mnk} \frac{\partial \Gamma_{pmnk}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} g_{qmnk}. \tag{83}$$

Thus, once we have the derivatives of the 1- and 2-RDM to sufficient precision, we can evaluate the orbital gradient, Hessian and composite Hessian analytically, recovering all terms in Eq. (37) and Eq. (38) without any additional quantum cost.

D Bounding the cumulative error due to Newton-Raphson updates

In this section, we prove that using a single Newton-Raphson (NR) parameter update per Δt-step is sufficient to achieve an error on the estimate of the minimizer scaling as $O(\Delta t^2)$ after any number of Δt-steps, as long as the cost function is strongly-convex at the minimum and Δt is small enough. First, we recall sufficient conditions for quadratic convergence of the Newton-Raphson step. We then use these to bound the error of a single NR-step when the cost function is changed from $E(t, \boldsymbol{\theta}) \to E(t + \Delta t, \boldsymbol{\theta})$. We translate this into an upper bound on Δt which guarantees the error stays bounded throughout the optimization path. Finally, we show that we can allow a sufficiently small additive error on the Newton-Raphson update and retain the bounded error throughout the path.

Quadratic convergence of NR — Consider a cost function $E(\boldsymbol{\theta})$ with gradient $\mathcal{G}_j = \frac{\partial E}{\partial \theta_j}$ and Hessian $\mathcal{H}_{jk} = \frac{\partial E}{\partial \theta_j \partial \theta_k}$, and an initial guess of a minimizer $\boldsymbol{\theta}^{(0)}$. The Newton-Raphson step prescribes the update $\boldsymbol{\theta}^{(0)} \mapsto \boldsymbol{\theta}^{\mathrm{NR}} = \boldsymbol{\theta}^{(0)} + d\boldsymbol{\theta}^{\mathrm{NR}}$ with $d\boldsymbol{\theta}^{\mathrm{NR}} = \mathcal{H}^{-1}(\boldsymbol{\theta}^{(0)})\mathcal{G}(\boldsymbol{\theta}^{(0)})$. Theorem 3.5 from Nocedal and Wright [44] gives sufficient conditions under which quadratic convergence of the NR update is guaranteed. We simplify these conditions, and obtain the following

Theorem 1. *Consider a cost function $E(\boldsymbol{\theta})$ with Lipschitz-continuous Hessian $\|\mathcal{H}(\boldsymbol{\theta}) - \mathcal{H}(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\| \leq L\|\delta\boldsymbol{\theta}\|$ and a local minimizer $\boldsymbol{\theta}^*$ with positive convexity $m := \|\mathcal{H}^{-1}(\boldsymbol{\theta}^*)\|^{-1} > 0$. Given an initial guess $\boldsymbol{\theta}^{(0)}$ which is close enough to the minimum, i.e.*

$$\|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^*\| \leq \frac{m}{4L}, \tag{84}$$

the NR update will converge quadratically towards the minimum with

$$\|\boldsymbol{\theta}^{NR} - \boldsymbol{\theta}^*\| \leq \frac{L}{m}\|\boldsymbol{\theta}^{(0)} - \boldsymbol{\theta}^*\|^2. \tag{85}$$

To prove this, we only need to show that a strong convexity condition is satisfied within a r-ball centered in $\boldsymbol{\theta}^*$ including all close-enough possible initial guesses ($r = \frac{m}{4L}$), i.e.

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})\| \leq 2m^{-1}, \quad \forall\|\delta\boldsymbol{\theta}\| \leq \frac{m}{4L}. \tag{86}$$

To prove this we expand $\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})$ using Taylor's theorem,

$$\exists 0 < s < 1: \quad \mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta}) = \mathcal{H}^{-1}(\boldsymbol{\theta}^*) + \delta\boldsymbol{\theta} \cdot \frac{\partial \mathcal{H}^{-1}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta}) \tag{87}$$

$$\frac{\partial \mathcal{H}^{-1}}{\partial \boldsymbol{\theta}} = -\mathcal{H}^{-1}\frac{\partial \mathcal{H}}{\partial \boldsymbol{\theta}}\mathcal{H}^{-1} \tag{88}$$

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + \delta\boldsymbol{\theta})\| \leq \|\mathcal{H}^{-1}(\boldsymbol{\theta}^*)\| + \|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\|^2 L\|\delta\boldsymbol{\theta}\| \tag{89}$$

$$\leq m^{-1} + \|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\|^2 \frac{m}{4}, \tag{90}$$

where we used the Lipschitz constant L as a bound on the derivative of the Hessian. This last condition holds if

$$\|\mathcal{H}^{-1}(\boldsymbol{\theta}^* + s\delta\boldsymbol{\theta})\| \leq 2m^{-1}. \tag{91}$$

As we can choose $s < 1$ and the result is clearly true at $\delta\boldsymbol{\theta} = 0$, the result holds recursively.

Single NR update with a changing cost function — We now consider a family of cost functions $E(t, \boldsymbol{\theta})$ continuously parameterized by t. Suppose we have an approximation $\tilde{\boldsymbol{\theta}}_t$ of the minimizer $\boldsymbol{\theta}^*_t$ of $E(t, \boldsymbol{\theta})$, with error $\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\|$. In each step of our method (Algorithm 2), we shift the cost function $E(t, \boldsymbol{\theta}) \to E(t + \Delta t, \boldsymbol{\theta})$ by Δt and we use the current minimizer estimate $\tilde{\boldsymbol{\theta}}_t$ as initial guess for the next step; $\boldsymbol{\theta}^{(0)}_{t+\Delta t} = \tilde{\boldsymbol{\theta}}_t$. We can bound the error of this initial guess by the triangle inequality,

$$\|\boldsymbol{\theta}^{(0)}_{t+\Delta t} - \boldsymbol{\theta}^*_{t+\Delta t}\| \leq \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| + \|\boldsymbol{\theta}^*_{t+\Delta t} - \boldsymbol{\theta}^*_t\|. \tag{92}$$

While the first term is the (given) error on the estimate, the second can be obtained by taking the total t-derivative of the minimum condition $\mathcal{G}(t, \boldsymbol{\theta}^*(t)) = 0$, and applying Taylor's theorem

$$\exists \tau \in [t, t + \Delta t] : \quad \|\boldsymbol{\theta}^*_{t+\Delta t} - \boldsymbol{\theta}^*_t\| = \left\| \frac{d\boldsymbol{\theta}^*_t}{dt}\Big|_{t=\tau} \right\| \Delta t = \|\mathcal{H}^{-1}(\tau, \boldsymbol{\theta}^*_\tau)\dot{\boldsymbol{\mathcal{G}}}(\tau, \boldsymbol{\theta}^*_\tau)\|\Delta t \tag{93}$$

with $\dot{\boldsymbol{\mathcal{G}}} = \nabla \frac{\partial}{\partial t} E(\boldsymbol{\theta}, t)$.

We now assume that the convexity at the minimum is bounded from below throughout the whole t-path by a constant $m \geq 0$,

$$m(t, \boldsymbol{\theta}^*_t) := \|\mathcal{H}^{-1}(t, \boldsymbol{\theta}^*_t)\|^{-1} > m \quad \forall t \in [0, 1], \tag{94}$$

and that the gradient of the change is never larger than $\dot{\mathcal{G}}_{\max}$. (while the first assumption imposes the nontrivial condition of strong convexity at the minimum, the second is always granted for cost functions from a continuous family of bounded Hamiltonians). We can then write

$$\|\boldsymbol{\theta}^{(0)}_{t+\Delta t} - \boldsymbol{\theta}^*_{t+\Delta t}\| \leq \|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| + m^{-1}\dot{\mathcal{G}}_{\max}\Delta t. \tag{95}$$

To ensure this initial guess is within the quadratic convergence region of the NR step, we require

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| + m^{-1}\dot{\mathcal{G}}_{\max}\Delta t \leq \frac{m}{4L}, \tag{96}$$

choosing an $\alpha, \beta \in (0, 1]$ this condition can be written as

$$\Delta t = \frac{m^2}{4L\dot{\mathcal{G}}_{\max}}\alpha\beta, \tag{97}$$

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| = (1 - \alpha)\beta\frac{m}{4L}. \tag{98}$$

This allows to apply Theorem 1, and bound the error after a single NR step,

$$\|\tilde{\boldsymbol{\theta}}_{t+\Delta t} - \boldsymbol{\theta}^*_{t+\Delta t}\| \leq \frac{L}{m}\left[\frac{\beta m}{4L}\right]^2 = \beta^2\frac{m}{16L}. \tag{99}$$

Multiple steps — We want to ensure the error on the minimizer estimate remains bounded for t taking subsequent values is $[0, \Delta t, 2\,\Delta t, ..., 1]$, while taking a single NR step at a time. We can do this by imposing the error after each step [Eq. (99)] is not larger than the error on the previous step estimate,

$$\beta^2\frac{m}{16L} \leq (1 - \alpha)\beta\frac{m}{4L}. \tag{100}$$

This is granted for any $\beta \in (0, 1]$ by choosing $\alpha = 1 - \frac{\beta}{4}$. The maximum $\Delta t = \frac{3}{4}\frac{m^2}{4L\dot{\mathcal{G}}_{\max}}$ is achieved by picking $\beta = 1$, and yields an error bounded by the constant $\frac{m}{16L}$.

Allowing an additive error — To account for sampling noise, it is useful to consider an additive error of magnitude $\sigma_{\boldsymbol{\theta}}$ on the estimate $\tilde{\boldsymbol{\theta}}_t$ of the minimizer $\boldsymbol{\theta}^*_t$ at each t-point, modifying Eq. (98) into

$$\|\tilde{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*_t\| = (1 - \alpha)\beta\frac{m}{4L} + \sigma_{\boldsymbol{\theta}}. \tag{101}$$

This yields the condition

$$\frac{L}{m}\left[\frac{\beta m}{4L} + \sigma_{\boldsymbol{\theta}}\right]^2 \leq (1 - \alpha)\beta\frac{m}{4L}. \tag{102}$$

If we define γ by $\sigma_{\boldsymbol{\theta}} = \frac{\gamma\beta m}{4L}$, we can write

$$(1+\gamma)^2 \frac{m}{16L}\beta^2 \leq (1-\alpha)\beta\frac{m}{4L}, \tag{103}$$

which is saturated by $\beta = 4\frac{1-\alpha}{(1+\gamma)^2}$. We then get

$$\Delta t = \frac{m^2}{L\dot{\mathcal{G}}_{\max}} \frac{\alpha(1-\alpha)}{(1+\gamma)^2}, \tag{104}$$

which is maximised (while keeping $\beta \leq 1$) by $\alpha = \max[\frac{1}{2}, 1 - \frac{(1+\gamma)^2}{4}]$. The allowed additive noise is then

$$\sigma_{\boldsymbol{\theta}} = \frac{\gamma}{(1+\gamma)^2}\frac{m}{L}(1-\alpha) = \frac{\gamma m}{2L}\min[\frac{1}{2}, \frac{1}{(1+\gamma)^2}], \tag{105}$$

maximised for the choice $\gamma = \sqrt{2}-1$ yielding

$$\sigma_{\boldsymbol{\theta}} = \frac{\sqrt{2}-1}{4}\frac{m}{L}, \quad \Delta t = \frac{m^2}{8L\dot{\mathcal{G}}_{\max}}. \tag{106}$$

E Bounding the sampling cost

We call $\sigma^2_{\mathcal{G}}$ and $\sigma^2_{\mathcal{H}}$ the variances of each element of the gradient and Hessian respectively, due to sampling noise. To compute the error on the parameter updates, we propagate these variances through the definition of the NR update Eq. (10). The first-order differential change (here denoted with δ) of the NR update $d\boldsymbol{\theta}^{\mathrm{NR}}$ with respect to changes in the gradient and Hessian is

$$\delta[d\boldsymbol{\theta}^{\mathrm{NR}}] = \mathcal{H}^{-1} \cdot [-\delta\boldsymbol{\mathcal{G}} + \delta\mathcal{H} \cdot d\boldsymbol{\theta}^{\mathrm{NR}}], \tag{107}$$

where we use the ordered matrix-product notation, with vectors in boldface. When $\delta\boldsymbol{\mathcal{G}}$ and $\delta\mathcal{H}$ are the random variables representing the errors on the gradient and Hessian, the expected mean square error on the NR update defining the norm of the covariance matrix

$$\left\|\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}]\right\| := \mathbb{E}\left[\|\delta[d\boldsymbol{\theta}^{\mathrm{NR}}]\|^2\right] = \mathbb{E}\left[\|\mathcal{H}^{-1}\cdot\delta\boldsymbol{\mathcal{G}}\|^2\right] + \mathbb{E}\left[\|\mathcal{H}^{-1}\cdot\delta\mathcal{H}\cdot d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right], \tag{108}$$

where we used the zero-average property of $\delta\boldsymbol{\mathcal{G}}$ and $\delta\mathcal{H}$ to drop the expectation values of mixed terms. This can further be bounded as

$$\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}] \leq \|\mathcal{H}^{-1}\|^2\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] + \|\mathcal{H}^{-1}\|^2\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right]\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2. \tag{109}$$

Assuming the same variance $\sigma^2_{\mathcal{G}}$ on each of the n_p elements $\mathcal{G}_j$ of the gradient, we get

$$\mathbb{E}\left[\|\delta\boldsymbol{\mathcal{G}}\|^2\right] = n_p\,\sigma^2_{\mathcal{G}}. \tag{110}$$

As the Hessian is a random real symmetric matrix with i.i.d. elements, each with a variance $\sigma_{\mathcal{H}}$, we can invoke Wigner's semicircle law [45] to bound the spectral norm as

$$\mathbb{E}\left[\|\delta\mathcal{H}\|^2\right] \leq (\sqrt{n_p}\,\sigma_{\mathcal{H}})^2. \tag{111}$$

Combining these with the strong convexity bound $\|\mathcal{H}^{-1}\| \leq m^{-1}$, we get

$$\mathrm{Var}[d\boldsymbol{\theta}^{\mathrm{NR}}] \leq m^{-2}\left[n_p\,\sigma^2_{\mathcal{G}} + n_p\,\sigma^2_{\mathcal{H}}\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right] \tag{112}$$

The calculations in Appendix D conclude that, for each Δt-step, we can afford an additive error on the NR update of at most $\sigma_{\boldsymbol{\theta}} \leq \frac{\gamma}{4}\frac{m}{L}$ with $\gamma = \sqrt{2}-1$. Comparing this result to the variance just calculated, we can formulate the requirement on the variance

$$m^{-2}n_p\left[\sigma^2_{\mathcal{G}} + \sigma^2_{\mathcal{H}}\|d\boldsymbol{\theta}^{\mathrm{NR}}\|^2\right] \ll \frac{\gamma^2}{16}\frac{m^2}{L^2}. \tag{113}$$

This gives conditions on the elementary sampling variances

$$\sigma_{\mathcal{G}}^2 \ll \frac{\gamma^2}{16}\frac{m^4}{L^2 n_p} \tag{114}$$

$$\sigma_{\mathcal{H}}^2 \ll \frac{\gamma^2}{16}\frac{m^4}{L^2 n_p \left\| d\boldsymbol{\theta}^{\mathrm{NR}} \right\|^2} < \frac{\gamma^2}{16}\frac{m^6}{L^2 n_p \|\dot{\mathcal{G}}_{\max}\|^2 \Delta t^2} \tag{115}$$

To recast this bound in terms of variables of the problem, we use the following relations derived in Appendix B: $L = \max\|\mathcal{T}\| < n_p^{3/2}\|H\|$ (the norm of the third derivative tensor $\mathcal{T}$ is bounded by $n_p^{3/2}$ times by its infinity norm), $\|\dot{\mathcal{G}}_{\max}\|\Delta t < \sqrt{n_p}\|\frac{dH}{dt}\|\Delta t \approx \sqrt{n_p}\|H\|$ (same infinity norm bound). Furthermore, we assume the convexity is larger than the ground state gap, $m > \Delta$; this holds if the ansatz approximates the ground state well enough, and changes in any ansatz parameter θ_k introduce a different excited state. Substituting these relations we obtain

$$\sigma_{\mathcal{G}}^2 \ll 0.01\frac{\Delta^4}{\|H\|^2 n_p^4}, \tag{116}$$

$$\sigma_{\mathcal{H}}^2 \ll 0.01\frac{\Delta^6}{\|H\|^4 n_p^5}. \tag{117}$$

The number of total required shots to sample the Hessian (gradient) for all the N steps will thus scale as $\sigma_{\mathcal{H}}^{-2}\Delta t^{-1}$ ($\sigma_{\mathcal{G}}^{-2}\Delta t^{-1}$). Picking the maximal $\Delta t = \frac{m^2}{8L\dot{\mathcal{G}}_{\max}}$, and considering only the dominant term (relative to sampling the Hessian) we can write

$$\#\text{shots} \propto n_p^7 \frac{\|H\|^7 \|\frac{dH}{dt}\|}{\Delta^8} \tag{118}$$