Quantum natural gradient generalised to non-unitary circuits

Bálint Koczor\(^1\) and Simon C. Benjamin\(^1\)

\(^1\)Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, United Kingdom

Variational quantum circuits are promising tools whose efficacy depends on their optimisation method. For noise-free unitary circuits, the quantum generalisation of natural gradient descent was recently introduced. The method can be shown to be equivalent to imaginary time evolution, and is highly effective due to a metric tensor reconciling the classical parameter space to the device’s Hilbert space. Here we generalise quantum natural gradient to consider arbitrary quantum states (both mixed and pure) via completely positive maps; thus our circuits can incorporate both imperfect unitary gates and fundamentally non-unitary operations such as measurements. Whereas the unitary variant relates to classical Fisher information, here we find that quantum Fisher information defines the core metric in the space of density operators. Numerical simulations indicate that our approach can outperform other variational techniques when circuit noise is finite.

I. INTRODUCTION

Variational techniques are ubiquitous in physics and mathematics. More specifically, variational algorithms involving the incremental update of parameters that describe a many-body quantum state have been widely used for decades [1–5]. The technique involves using a tractable set of parameters to describe a quantum state in an exponentially larger Hilbert space, and therefore relies on an understanding that the states of importance (e.g. the low-energy states of some Hamiltonian) lie within a relatively small subspace.

With the rise of quantum computers as realistic technologies, naturally attention has been given to the question of how such a machine could perform as a variational tool [6–24]. The resulting model is hybrid with an iterative loop: a classical machine determines how to update the parameters describing a quantum state (the ‘ansatz state’) while the quantum coprocessor generates and characterises that state (using an ‘ansatz circuit’). This paradigm is of particular interest in the context of noisy, intermediate-scale quantum devices (NISQ devices) [24], because quite complex ansatz states can be prepared with shallow circuits [26–29], thus raising the possibility that resource-intensive quantum fault tolerance methods might not be needed.

The most well-studied application is the variational quantum eigensolver (VQE), where one seeks a final circuit configuration that minimises a cost function – normally the energy of some system of interest. For the optimisation of the classical parameters, one might employ any one of a range of methods: for example a direct search such as NelderMead (demonstrated experimentally in 2014 [7]), or a systematic scan if the number of parameters is small [9], or direct gradient descent (see e.g. Ref. [30]).

Recently an imaginary time principle was used to govern the parameter evolution [34], so that the ansatz state follows (as closely as possible) the trajectory $e^{-\mathrm{i}H\Delta t}\ket{\psi_0}$. The approach was found to outperform others in accuracy and convergence speed according to numerical simulations [31], and was subsequently demonstrated experimentally [32]. The derivation, which proceeds from McLachlan’s variational principle (and had been employed previously in the context of simulating real-time quantum evolution [33]) introduced a novel feature: a matrix object that characterises the sensitivity of the ansatz state to changes in each possible pair of parameters, but without reference to the cost function.

It was evident that this matrix had a crucial role in enabling the high performance of the technique, which in turn enabled studies in mixed state evolution [35] and general process evolution [34]. However an elucidation of the deeper meaning of this matrix was reached very recently, in relation to a concept called natural gradient. A paper by Stokes et al [35] considered the generalisation of this concept from the machine learning field [36–38] to the context of quantum computing. Natural gradient accounts for the non-trivial relationship between a translation in parameter space, and the corresponding translation in the problem space (for our case, the Hilbert space of the quantum processor). This is achieved through the use of a metric tensor which, as we presently discuss, proves to be exactly the matrix object responsible for the accurate performance of the real and imaginary time methods.

Specifically, Stokes et al [35] showed that if a state-vector is isomorphic to a classical probability distribution $p(n|\theta)$ (i.e., the state contains no phase information) then a variational quantum optimisation results in the update rule

$$\theta(t+1) = \theta(t) - \lambda |F_C|^{-1} g.$$  

(1)

Here the classical Fisher information matrix $F_C$ is a metric tensor that is related to the probability distribution

$$[F_C]_{kl} = \sum_n p(n|\theta) \frac{\partial^2 \ln [p(n|\theta)]}{\partial \theta_k \partial \theta_l}$$  

(2)

and corrects the gradient vector by accounting for the co-dependent and non-uniform effect of the parameters.

\* balint.koczor@materials.ox.ac.uk
simon.benjamin@materials.ox.ac.uk
on $p(n|\theta)$. The authors identified this form as identical to natural gradient as studied in the context of machine learning \[36\] [38]. A subsequent paper further elucidated the connection [39].

In the present paper we propose a novel quantum variational optimisation method which is directly analogous to Eq. (1) but can be applied to arbitrary mixed or pure states. Our approach thus generalises and unifies previously obtained state-vector evolutions to the non-trivial and most general case of density operators (both mixed and pure states). This is particularly relevant when considering the effect of imperfections on variational quantum circuits as illustrated on Fig. 1 but allows for additional degrees of freedom in the optimisation, such as non-unitary transformations including measurements.

This manuscript is organised as follows. We first recapitulate prior work on variational quantum optimisations using idealised quantum circuits in Sec. III. We then show in Sec. III how noise affects a variational quantum circuit and outline a prior approach that uses imaginary time evolution of mixed quantum states. We finally state our main results in Sec. IV using the quantum Fisher information matrix for the natural gradient optimisation of variational quantum circuits. We then compare our approach to previous ones in Sec. V and numerically simulate noisy variational circuits as a potential application.

II. OPTIMISING IDEALISED VARIATIONAL QUANTUM CIRCUITS

An idealised variational circuit is usually modelled as a unitary transformation $|\psi(\theta)\rangle = U_c(\theta)|00\ldots0\rangle$ applied to the reference state $|00\ldots0\rangle$ which is usually chosen as the computational zero state. The unitary operator $U_c(\theta)$ represents the entire quantum circuit and depends on a set of gate parameters $\theta \in \mathbb{R}^\nu$. As elucidated by the circuit model, it decomposes into a series of individual gates (typically acting on a small subset of the system, i.e., on one or two qubits)

$$U_c(\theta) = U_c(\theta_\nu) \ldots U_c(\theta_2)U_c(\theta_1)$$

each of which depends on a parameter $\theta_i$ with $i = \{1, 2, \ldots \nu\}$. It is typically the aim of a variational algorithm to find the minimum or maximum of an expectation value $E(\theta) := \langle \psi(\theta) | H | \psi(\theta) \rangle$ over the parameters where the observable $H$ is a Hermitian operator, typically the Hamiltonian of a simulated physical system. A hybrid approach assumes that a quantum processor can efficiently estimate the expectation value $E(\theta)$ for a set parameters and these parameters are optimised externally according to an update rule calculated by a classical computer.

As outlined in the introduction, numerous optimisation methods have been proposed for finding parameters that minimise this expectation value and indeed some have been demonstrated experimentally. The literature
is evolving rapidly; for example the stochastic gradient
descent approach was recently proven to converge when
assuming only a finite number of measurements for de-
termining the gradient vector \[10\].

We focus on two approaches: imaginary time evolution
as in \[31,33\] and natural gradient evolution as in \[35\].
These two methods are in fact equivalent in the unitary
context, and optimise the parameters \(\theta(t)\) iteratively in
steps \(\Delta t\) following the update rule

\[
\theta(t+1) = \theta(t) - \Delta t A^{-1} g. \tag{3}
\]

Here the inverse of the matrix \(A := A(\theta)\) is applied to the
gradient of the expectation value \(g_k = \partial_k E(\theta)\) and we will
consistently use the short-hand notation \(\partial_k := \frac{\partial}{\partial \theta_k}\). The
matrix object in the update rule corrects the gradient vector
to account for the non-uniform and co-dependent effect of the
parameters on the quantum states, refer to Sec. \[IV.A\] Its
matrix elements are given by the state-vector overlaps

\[
[A]_{kl} = \text{Re}[ \langle \partial_k \psi | \partial_l \psi \rangle - \langle \partial_k \psi | \psi \rangle \langle \psi | \partial_l \psi \rangle ], \tag{4}
\]

and we will consistently use the short-hand notation
throughout this paper \(\partial_k \psi := \frac{\partial \psi}{\partial \theta_k}\).

This update rule and the matrix \(A\) in Eq. \(3\) have been originally
derived in refs. \[31,33\] to simulate the imaginary time evolution of a
pure state vector

\[
|\psi(t)\rangle = \frac{e^{-iH t}|\psi(0)\rangle}{\langle \psi(0)|e^{-2iH t}\rangle \psi(0)} \tag{5}
\]

using a variational quantum circuit that can efficiently
estimate both \(A\) and \(g\). The exact evolution in Eq. \(3\)
converges to the ground state of the system for \(t \to \infty\) if
\(|\psi(0)\rangle\) has a non-zero overlap with the ground state.

As noted earlier, the matrix \(A\) was recently indepen-
dently derived \[35\] and shown to be equivalent to the
Fubini-Study metric tensor \[11,12,33\] which is a metric ten-
sor in complex vector spaces. Its analogy to the classi-
cal Fisher information was also shown, i.e., \(A = F_C/4\)
if the state vector is isomorphic to a classical proba-
bility distribution \(p(n|\theta)\). In this case the state vector
\(|\psi(\theta)\rangle = \sum_n \sqrt{p(n|\theta)} |k\rangle\) contains no phase information.
This was shown to be analogous to the natural gradient
optimisation from Eq. \(1\) which uses the classical Fisher
information matrix \(F_C\) as a metric tensor and is widely
used in the context of machine learning \[30,38\].

In contrast to the above discussed methods which were
derived for idealised perfect quantum circuits and pure
states, here we aim to take into account imperfections in
the optimisation. Moreover, our work is not restricted to
classical probability distributions or to state vectors as
in \[33\], but applicable to arbitrary quantum states as
density matrices \(\rho\).

III. VARIATIONAL ALGORITHMS WITH
IMPERFECT CIRCUITS

We first generalise the previously introduced idealised
unitary circuit model to the more realistic case taking
experimental imperfections of the variational circuit into
account. We describe this variational circuit as a com-
pletely positive mapping of density matrices as \(\rho(\theta) = \Phi(\theta) \rho_0\)
that depends on the parameters \(\theta \in \mathbb{R}^n\) where \(\rho_0\) is usual-
ly the computational zero state. Here \(\Phi(\theta)\) is the
superoperator that represents the realistic quantum
circuit. This quantum circuit only approximately de-
composes into

\[
\Phi(\theta) \approx \Phi_1(\theta_1) \ldots \Phi_k(\theta_k)
\]

a sequence of imperfect gate operations \(\Phi_k(\theta_k)\) due to
possible correlated noise.

Our approach is, however, not restricted to imperfect
quantum circuits. The only assumption we make about
the mapping \(\Phi(\theta)\) is that it is continuous in each of the
parameters \(\theta_k\) such that differentials \(\partial_k \rho(\theta)\) of the density
matrix as \(\partial_k \Phi(\theta) \rho_0\) exist for any state. This is natu-
really the case for quantum circuits undergoing Markovian
or time-dependent Markovian decoherence \[11\] but more
general mappings can satisfy this condition too. These
include, e.g., allowing measurements in the circuit inde-
dependently of the parametrisation or a Markovian or time-
dependent Markovian decoherence whose length depends
on a parameter.

A. Imaginary time evolution of mixed quantum
states

Let us consider now the imaginary time evolution of an
initial mixed quantum state \(\rho\) as

\[
\rho(t) = e^{-iH t} \rho e^{-iH t} / \text{Tr}[e^{-2iH t} \rho]
\]

This evolution, in general, goes to a direction that in-
creases mixedness of the density operator.

It has been shown in \[33\] that the closest unitary evolu-
tion can be simulated efficiently using variation quantum
circuits. We assume that these circuits produce quantum
states via a mapping \(\rho := \rho(\bar{\theta})\) as discussed in Sec \[III\]
The corresponding parameter-update rule is analogous to
Eq. \(3\) and results in \[33\]

\[
\bar{\theta}(t+1) = \bar{\theta}(t) - \Delta t M^{-1} Y , \tag{6}
\]

but here the vector \(Y\) appears instead of the energy gra-
dient which has the entries \(Y_k = -2 \text{Re}[\text{Tr}((\partial_k \rho) H \rho)]\). The
matrix \(M\) contains the Hilbert-Schmidt scalar prod-
ucts between differentials of the mixed state

\[
[M]_{kl} = \text{Tr}[(\partial_k \rho)(\partial_l \rho)]. \tag{7}
\]

This update rule results in an improved performance
when compared to simple gradient descent as illustrated
on Fig. 1 (pink). However, this circuit-based simulation of imaginary time evolution of a mixed state is vulnerable to becoming stuck in a point away from the optimum. This is because the exact gradient of imaginary time evolution is non-zero but points to a purely non-unitary direction. In the following we aim to develop an alternative approach that does not rely on the simulation of imaginary time evolution (yet reduces to that in case of pure states) and finds the true optimum as illustrated on Fig. 2 (green). Our approach is based on the quantum Fisher information.

**IV. QUANTUM FISHER INFORMATION AS NATURAL GRADIENT**

Before stating our main results, we briefly recall basic notions related to quantum Fisher information which is a concept extensively used in the field of quantum metrology for determining the metrological usefulness of a quantum state, refer to, e.g., 15,17 for more details. Here we introduce the quantum Fisher information for the first time in the context of variational quantum circuits as a measure that quantifies how much and in what way changing parameters in a quantum circuit affects the underlying quantum state.

**A. Sensitivity to parameters via the quantum Fisher information**

Assume now for simplicity a one-parameter quantum circuit as a one-parameter mapping \( \rho_0 = \Phi(\theta) \) acting on a reference state \( \rho \) and let us consider the resulting continuous family of quantum states \( \rho_\theta \). Here \( \Phi(\theta) \) can be, e.g., an imperfect Mølmer-Sørensen gate. The quantum Fisher information is a generalisation of the classical Fisher information and quantifies how different a state \( \rho_0 \) becomes under an infinitesimal variation \( \rho_{0 + \delta \theta} \) of this gate parameter \( \theta \).

This parametrised state produces a continuous family of probability distributions

\[
p(n|\theta) = \text{Tr}[\rho_\theta |b_n\rangle \langle b_n|] = \langle b_n|\rho_0|b_n\rangle
\]

when measured in a basis \( \{ |b_n\rangle \} \), where \( n = 1, 2, \ldots, d \), \( \sum_n |b_n\rangle \langle b_n| = \text{Id}_d \) and \( d \) is the dimensionality of the system. The classical Fisher information quantifies how different the probability distributions become under a variation of \( \theta \)

\[
F_c(\{ |b_n\rangle \}) = \sum_n p(n|\theta) \left( \frac{\partial \ln p(n|\theta)}{\partial \theta} \right)^2.
\]

and depends on the choice of measurement basis \( \{ |b_n\rangle \} \).

For example, if the parameter \( \theta \) corresponds to a z-rotation \( U_z(\theta) := \exp(-i\theta \sigma_z/2) \) of a qubit state \( \rho_0 = U_z(\theta) \rho_0 U_z(\theta)\dagger \), then measuring in the computational basis results in, e.g., \( \theta_0 \langle 0|\rho_0|0\rangle = 0 \) and therefore the Fisher information is \( F_c(\{ |0\rangle, |1\rangle \}) = 0 \). However, measuring in the \( \{ |\pm\rangle \} \) basis results in \( F_c(\{ |+\rangle, |-\rangle \}) \geq 0 \).

The quantum Fisher information is the maximum of \( F_c(\{ |b_n\rangle \}) \) when optimised over all possible measurement bases as \( \{ |b_n\rangle \} \) (including generalised POVM measurements). This quantum Fisher information \( F_Q := F_Q(\rho_0) \geq 0 \) can be computed for any member of the parametrised family of quantum states as the expectation value \( F_Q(\rho_0) = \text{Tr}[\rho_0 L_\theta^2] \). Here the Hermitian symmetric logarithmic derivative \( L_\theta \) contains the optimal measurement bases as eigenvectors and is defined via

\[
\frac{\partial \rho_\theta}{\partial \theta} = \frac{1}{2} (L_\theta \rho_\theta + \rho_\theta L_\theta).
\]

Decomposing a density matrix into \( \rho_0 = \sum_n p_n |\psi_n\rangle \langle \psi_n| \) projectors onto its eigenstates \( |\psi_n\rangle \) with \( p_n > 0 \) allows for explicitly computing matrix elements of the symmetric logarithmic derivative

\[
\langle \psi_i | L_\theta | \psi_j \rangle = \frac{2}{p_i + p_j} \langle \psi_i | \frac{\partial \rho_\theta}{\partial \theta} | \psi_j \rangle.
\]

**B. Quantum Fisher information matrix**

Let us now consider the matrix form of the quantum Fisher information which is the quantum generalisation of the classical Fisher information matrix from Eqs. (2). We will now consider noisy quantum circuits (or more general continuous mappings) from Sec. III that span a continuous family of density matrices as \( \rho(\theta) \). As in Eq. (8), the partial derivative of \( \rho(\theta) \) with respect to \( \theta_k \) defines the symmetric logarithmic derivative

\[
\frac{\partial \rho(\theta)}{\partial \theta_k} := \frac{1}{2} (L_k \rho(\theta) + \rho(\theta) L_k),
\]

and eigenvectors of \( L_k \) are the most sensitive measurement bases to detect variations in \( \theta_k \). Entries of the quantum Fisher information matrix \( [F_Q]_{kl} := [F_Q(\rho(\theta))]_{kl} \) are then the expectation values

\[
[F_Q]_{kl} = [F_Q]_{lk} = \frac{1}{2} \text{Tr}[\rho(\theta) (L_k L_l + L_l L_k)]
\]

of these symmetric logarithmic derivatives. Diagonal entries of the matrix \( F_Q \) correspond to the scalar quantum Fisher information \( \text{Tr}[\rho_0 L_\theta^2] \) and quantify the sensitivity of a quantum state with respect to individual parameters \( \theta_k \). And off-diagonal entries account for the co-dependence of parameters.

**C. Natural gradient descent for arbitrary quantum states**

We are now equipped to propose our generalisation of the natural gradient evolution from Eq. 1 using the quantum generalisation of \( F_C \), the quantum Fisher information matrix, which by definition quantifies parameter
sensitivity of the quantum states via the probability distributions they produce.

Our aim is to minimise the expectation value 

\[ E(\theta) = \text{Tr}[\rho(\theta) \mathcal{H}] \]

of a Hermitian observable \( \mathcal{H} \) over the parameters \( \theta \) using a variational quantum circuit that depends on these parameters. This circuit produces the quantum states via a mapping \( \rho(\theta) = \Phi(\theta) \rho_0 \) as discussed in Sec. III and might, for example, involve non-unitary transformations due to experimental imperfections or indeed intentional non-unitary transformations.

**Result 1.** The natural gradient update rule for parameters

\[ \theta(t+1) = \theta(t) - \frac{1}{\Delta t} [\mathbf{F}_Q]^{-1} q. \]  

The quantum Fisher information matrix \( \mathbf{F}_Q \) corrects the gradient vector \( q_k := \partial_{\theta_k} E(\theta) \) of the expectation value \( E(\theta) := \text{Tr}[\rho(\theta) \mathcal{H}] \) to account for the co-dependent and non-uniform effect of the parameters on an arbitrary quantum state \( \rho(\theta) \) (mixed or pure).

Computing the matrix \( \mathbf{F}_Q \) can be involved for arbitrary quantum states and there are numerous expressions available in the literature [45, 46]. Here we state one expression that is valid for arbitrary rank- \( r \)-density matrices \( \rho(\theta) = \sum_{n=1}^r p_n |\psi_n\rangle \langle \psi_n| \) with \( p_n > 0 \) where both \( p_n := p_n(\theta) \) and \( |\psi_n\rangle := |\psi_n(\theta)\rangle \) depend on the parameters. Matrix entries of \( \mathbf{F}_Q \) are given by

\[
[F_Q]_{kl} = \sum_{n=1}^r \left( \frac{\partial_k p_n}{p_n} \right) \left( \frac{\partial_l p_n}{p_n} \right) + \sum_{n=1}^r 4 p_n \text{Re}[\langle \partial_k \psi_n | \partial_l \psi_n \rangle] \\
- \sum_{n,m=1}^r \frac{8 p_n}{p_n + p_m} \text{Re}[\langle \partial_k \psi_n | \psi_m \rangle \langle \psi_n | \partial_l \psi_m \rangle] 
\]

(12)

As an important special case of this equation, let us restrict ourselves now to rank-one density matrices \( r = 1 \) as pure quantum states.

**Result 2.** For pure quantum states as \( \rho(\theta) = |\psi\rangle \langle \psi| \), the quantum fisher information matrix \( \mathbf{F}_Q \) simplifies to the matrix \( \mathbf{A} \) which appears in Eq. (9) of imaginary time evolution

\[
[F_Q]_{kl} = 4 \text{Re}[\langle \partial_k \psi | \partial_l \psi \rangle - \langle \partial_k \psi | \psi \rangle \langle \psi | \partial_l \psi \rangle] = 4 [\mathbf{A}]_{kl}. 
\]

**FIG. 2.** Example of a 4-qubit ansatz circuit used for finding the ground state energy of the Hamiltonian in Eq. (13). It consists of single qubit rotations and nearest neighbour coupling evolutions \( \exp[-i \theta_i \sigma^{(i)}_n \sigma^{(i+1)}_n] \) with \( \alpha = \{x, y, z\} \).

**FIG. 3.** Finding the ground-state energy (black) of the Hamiltonian in Eq. (13) using noisy variational circuits as in Fig. 2. Gradient descent (blue), quantum Fisher information (green) and imaginary time evolution (pink) optimisations were started from the same randomly chosen initial positions assuming different severity \( p_{\text{error}} \) of depolarising noise. Solid and dashed lines show the mean of the obtained energies while shading represents its standard deviation. The optimal energy \( E(\theta_{\text{opt}}) := \text{Tr}[\rho(\theta_{\text{opt}}) \mathcal{H}] \) underestimates the exact ground-state energy due imperfections of the quantum circuit.

The update rule in Result 1 therefore reduces to the imaginary time simulation of the state vector \(|\psi\rangle\) which was analysed in detail in the works [31, 33].

**V. NUMERICAL SIMULATIONS**

We now consider the explicit example of a spin-1/2 Hamiltonian and compare the different optimisation methods using a noisy ansatz circuit. Our aim is to find the ground-state energy of the Hamiltonian

\[
\mathcal{H} = \sum_{i=1}^{N-1} J [\sigma_x^{(i)} \sigma_x^{(i+1)} + \sigma_y^{(i)} \sigma_y^{(i+1)} + \sigma_z^{(i)} \sigma_z^{(i+1)}] + J [\sigma_x^{(N)} \sigma_x^{(1)} + \sigma_y^{(N)} \sigma_y^{(1)} + \sigma_z^{(N)} \sigma_z^{(1)}] + \sum_{i=1}^{N} \omega_i \sigma_z^{(i)}, \tag{13}
\]

where \( \sigma_x^{(i)}, \sigma_y^{(i)}, \sigma_z^{(i)} \) are the Pauli matrices.
which contains identical couplings $xx$, $yy$ and $zz$ between nearest neighbours with a constant which we set $J = 1$. We select on-site frequencies $\omega_i$ randomly according to a uniform distribution with values between $-1$ and $1$. The resulting Hamiltonian has a non-trivial, highly entangled ground state that we aim to approximate using the (not necessarily optimal) ansatz circuit shown on Fig. 2. This ansatz is composed of single qubit $x$ and $z$ rotations with arbitrary rotation angles and evolutions under nearest neighbour two-qubit Hamiltonians of the form $\exp[-i\theta_k \sigma^y_i \sigma^y_{i+1}]$ with $\alpha = \{x, y, z\}$. We assume a depolarising error after every gate in Fig. 2 with an error probability $p$ after single-qubit gates and $10p$ after two-qubit gates. This quantum circuit therefore produces a parametrised mixed state $\rho(\theta)$ as discussed in Sec. III.

We have simulated optimisations of the ansatz parameters and compare techniques which are illustrated on Fig. 4. We start the evolution from randomly chosen initial points $\theta(0)$ in a close proximity of the optimal parameters $\theta^{\text{opt}}$ that locally minimise the energy $E^{\text{opt}} = \text{Tr}[\rho(\theta^{\text{opt}})\mathcal{H}]$. Parameters of the ansatz circuit are then evolved for a fixed number of steps 30 using a step size 0.4 for the simple gradient descent and 0.1 otherwise. Fig. 3 (black) shows exact ground-state energies of this Hamiltonian as a function of the number of qubits. Fig. 3 blue, green and pink lines show energies obtained by evolving the same initial parameters using different methods. Solid and dashed lines show the obtained energies averaged over repeated optimisations (number of repetitions is 25) starting from randomly chosen initial points while shading represents its standard deviation.

Simple gradient descent Fig. 3 (blue) results in a poor performance (average energy is significantly above the exact) and only rarely gets close to the ground state (large standard deviation). Fig. 3 (green and pink) Imaginary time evolution and the quantum Fisher information methods could always get sufficiently close to the optimum. Note that the optimal energy $E^{\text{opt}} = \text{Tr}[\rho(\theta^{\text{opt}})\mathcal{H}]$ is slightly above the exact ground-state energy due to imperfections of the variational circuit which can only produce mixed states.

We are now interested in quantifying the accuracy and precision of our approach. Fig. 4 shows the distance $\Delta E$ from the optimal energy $E^{\text{opt}}$ after the previously discussed evolutions. Note that we have fixed the number of steps at 30. Solid lines on Fig. 4 show the average distance from the optimal energy over multiple runs while shading represents its standard deviation. Our approach using the quantum Fisher information matrix is both precise (small standard deviation) and accurate as its mean distance is dramatically smaller than in case of the other two methods. Note that imaginary time evolution (of mixed states as in Sec. III A) gives a biased estimate of the optimal energy as illustrated on Fig. 4.

VI. DISCUSSION AND CONCLUSION

In this work we have extended the work of Stokes et al. [35], which introduced quantum natural gradient evolution from the well-studied classical analog in machine learning [56, 58]. Whereas Stokes et al consider general unitary (and therefore noise-free) quantum circuits, we have considered the most general scenario of optimising arbitrary quantum states as density matrices. We have shown that the quantum Fisher information, a quantity much-studied in the context of quantum metrology, can be used to correct the gradient vector in a variational quantum algorithm to account for the non-uniform effect of the parameters on the underlying quantum states.

Moreover, since prior studies had used the same evolution rule as Stokes et al’s quantum natural gradient but in the context of imaginary time evolution [31], and those studies found the method to be superior to other optimisation methods, we speculate that the present generalisation is a robust solution for general noisy or otherwise non-unitary circuits. We numerically simulated noisy variational circuits and demonstrated that indeed...
our approach significantly outperforms simple gradient descent evolutions. Previously proposed imaginary time evolutions of mixed states result in a similar performance, but do not necessarily converge to the true optimum.

When compared to previous studies, our approach has the advantage that it explicitly takes into account imperfections of the variational quantum circuit. It is therefore appropriate for seeking the optimum when the quantum circuits to be employed are imperfect; however we emphasise that the applicability of the method is not restricted to noisy unitary circuits, but can be applied to the far-reaching scenario when a circuit contains intentional non-unitary transformations, such as measurements or variable-time decoherence.

**Note prior to arXiv submission:** We note that immediately prior to our submission another work has appeared on arXiv that discusses the effect of noise on variational quantum algorithms [51]. There are evidently commonalities since their work identifies the scalar quantum Fisher information as an important measure of performance, while our approach uses the quantum Fisher information matrix as a metric tensor. We look forward to exploring this relation further.

**ACKNOWLEDGMENTS**

S. C. B. acknowledges financial support from the NQIT UK National Hub, EPSRC grant EP/M013243/1. B. K. and S. C. B. acknowledge funding received from EU H2020-FETFLAG-03-2018 under the grant agreement No 820495 (AQTION). The numerical modelling involved in this study made use of the Quantum Exact Simulation Toolkit (QuEST), and the recent development QuESTLink [52] which permits the user to use Mathematica as the integrated front end. The authors are grateful to those who have contributed to both these valuable tools.

[1] R. Balian and M. Vénéroni, Static and dynamic variational principles for expectation values of observables, Annals of Physics 187, 29 (1988)
[2] F. Dalfovo, S. Giorgini, L. P. Pitaevskii, and S. Stringari, Theory of bose-einstein condensation in trapped gases, Rev. Mod. Phys. 71, 463 (1999)
[3] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pirzyn, H. Ver-schelde, and F. Verstraete, Time-dependent variational principle for quantum lattices, Phys. Rev. Lett. 107, 070601 (2011)
[4] T. Shi, E. Demler, and J. I. Cirac, Variational study of fermionic and bosonic systems with non-gaussian states: Theory and applications, Annals of Physics 390, 245 (2018)
[5] L. Vanderstraeten, J. Haegeman, and F. Verstraete, Tangent-space methods for uniform matrix product states, arXiv preprint arXiv:1810.07006 (2018)
[6] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv preprint arXiv:1411.4048 (2014).
[7] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O’Brien, A variational eigenvalue solver on a photonic quantum computer, Nature communications 5 (2014).
[8] Y. Wang, F. Dolde, J. Biamonte, R. Babbush, V. Bergholm, S. Yang, I. Jakobi, P. Neumann, A. Aspuru-Guzik, J. D. Whitfield, et al., Quantum simulation of helium hydride cation in a solid-state spin register, ACS nano 9, 7769 (2015).
[9] P. J. J. O’Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Covone, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, Scalable quantum simulation of molecular energies, Phys. Rev. X 6, 031007 (2016).
[10] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure, Phys. Rev. A 95, 020501 (2017).
[11] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New Journal of Physics 18, 023023 (2016).
[12] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O’Brien, and M. G. Thompson, Experimental bayesian quantum phase estimation on a silicon photonic chip, Phys. Rev. Lett. 118, 100503 (2017).
[13] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X 7, 021050 (2017).
[14] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Computation of molecular spectra on a quantum processor with an error-resilient algorithm, Phys. Rev. X 8, 011021 (2018).
[15] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, D. P. Tew, X. Zhou, J. L. O’Brien, and M. G. Thompson, Witnessing eigenstates for quantum simulation of hamiltonian spectra, Science Advances 4, 10.1126/sciadv.aap9646 (2018).
[16] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature 549, 242 (2017).
[17] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature 567, 491 (2019).
[18] C. Hempel, C. Maier, J. Romero, J. McClean, T. Monz,
H. Shen, P. Jurcevic, B. P. Lanyon, P. Love, R. Babbush, A. Aspuru-Guzik, R. Blatt, and C. F. Roos, Quantum chemistry calculations on a trapped-ion quantum simulator. Phys. Rev. X \textbf{8}, 031022 (2018).

[19] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. Love, and A. Aspuru-Guzik, Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz, arXiv preprint arXiv:1701.02691 (2017).

[20] O. Higgott, D. Wang, and S. Brierley, Variational quantum computation of excited states, arXiv preprint arXiv:1805.08138 (2018).

[21] R. Santagati, J. Wang, A. A. Gentile, S. Paesani, N. Wiebe, J. R. McClean, S. Morley-Short, P. J. Shadbolt, D. Bonneau, J. W. Silverstone, et al., Witnessing eigenstates for quantum simulation of hamiltonian spectra, Science advances \textbf{4}, eaap9646 (2018).

[22] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, W. A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, Physical Review A \textbf{95}, 042308 (2017).

[23] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, J. R. McClean, J. Carter, W. A. de Jong, and I. Siddiqi, Robust determination of molecular spectra on a quantum processor, arXiv preprint arXiv:1707.06408 (2017).

[24] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, et al., Self-verifying variational quantum simulation of the lattice schwinger model, arXiv preprint arXiv:1810.03421 (2018).

[25] J. Preskill, Quantum computing in the nisq era and beyond, arXiv preprint arXiv:1801.00862 (2018).

[26] I. Kassal, J. D. Whitfield, A. Perdomo-Ortiz, M.-H. Yung, and A. Aspuru-Guzik, Simulating chemistry using quantum computers, Annual review of physical chemistry \textbf{62}, 185 (2011).

[27] D. Lu, B. Xu, N. Xu, Z. Li, H. Chen, X. Peng, R. Xu, and J. Du, Quantum chemistry simulation on quantum computers: theories and experiments, Phys. Chem. Chem. Phys. \textbf{14}, 9411 (2012).

[28] K. B. Whaley, A. R. Dinner, and S. A. Rice, Quantum information and computation for chemistry (John Wiley & Sons, 2014).

[29] S. McArdle, S. Endo, A. Aspuru-Guzik, S. Benjamin, and Yuan, Quantum computational chemistry, arXiv preprint arXiv:1808.10402 (2018).

[30] P. Rebentrost, M. Schuld, L. Wossnig, F. Petruccione, and S. Lloyd, Quantum gradient descent and newton's method for constrained polynomial optimization, New Journal of Physics \textbf{21}, 073023 (2019).

[31] S. McArdle, T. Jones, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, Variational ansatz-based quantum simulation of imaginary time evolution, npj Quantum Information \textbf{5}, 75 (2019).

[32] M.-C. Chen, M. Gong, X.-S. Xu, X. Yuan, J.-W. Wang, C. Wang, C. Ying, J. Lin, Y. Xu, Y. Wu, S. Wang, H. Deng, F. Liang, C.-Z. Peng, S. C. Benjamin, X. Zhu, C.-Y. Lu, and J.-W. Pan, Demonstration of adiabatic variational quantum computing with a superconducting quantum coprocessor, (2019), arXiv:1905.03450 [quant-ph].

[33] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Theory of variational quantum simulation, Quantum \textbf{3}, 191 (2019).

[34] S. Endo, Y. Li, S. Benjamin, and X. Yuan, Variational quantum simulation of general processes (2018), arXiv:1812.08778 [quant-ph].

[35] J. Stokes, J. Izac, N. Killoran, and G. Carleo, Quantum natural gradient, arXiv preprint arXiv:1909.02108 (2019).

[36] S.-i. Amari, Neural learning in structured parameter spaces-natural riemannian gradient, in Advances in neural information processing systems (1997) pp. 127–133.

[37] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning (MIT press, 2016).

[38] S.-I. Amari, H. Park, and K. Fukumizu, Adaptive method of realizing natural gradient learning for multilayer perceptrons, Neural Computation \textbf{12}, 1399 (2000).

[39] N. Yamamoto, On the natural gradient for variational quantum eigensolver, (2019), arXiv:1909.05074 [quant-ph].

[40] R. Sweke, F. Wilde, J. Meyer, M. Schuld, P. K. Fährmann, B. Meynard-Pignoneau, and J. Eisert, Stochastic gradient descent for hybrid quantum-classical optimization, arXiv preprint arXiv:1910.01155 (2019).

[41] E. Study, Kürzeste wege im komplexen gebiet, Mathematische Annalen \textbf{60}, 321 (1905).

[42] F. Wilczek and A. Shapere, Geometric phases in physics, Vol. 5 (World Scientific, 1989).

[43] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, Quantum metrology with nonclassical states of atomic ensembles, Rev. Mod. Phys. \textbf{90}, 035005 (2018).

[44] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Phot. \textbf{5}, 222 (2011).

[45] B. Koczor, S. Endo, T. Jones, Y. Matsuizaki, and S. C. Benjamin, Variational-state quantum metrology, arXiv preprint arXiv:1908.08904 (2019).

[46] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, Quantum fisher information matrix and multiparameter estimation, Journal of Physics A: Mathematical and Theoretical (2019).

[47] D. Šafránek, Simple expression for the quantum fisher information matrix, Physical Review A \textbf{97}, 042322 (2018).

[48] This step size is slightly below the largest stable step size.

[49] L. Gentini, A. Cuccoli, S. Pirandola, P. Verrucchi, and L. Maccone, Witnessing eigenstates for quantum simulation of hamiltonian spectra, arXiv preprint arXiv:1909.02108 (2019).

[50] B. Cerezo, S. Shankar, A. Neufeld, M.ứt, P. Schauss, E. Micheli, A. Synak-Radtke, J. Mutus, A. Wossnig, and J. Kelly, Quantum algorithms for ab initio molecular energy calculations on a trapped-ion quantum simulator, Science advances \textbf{5} (2019).

[51] T. Jones and S. C. Benjamin, QuESTlink – Mathematica embiggend by a hardware-optimised quantum emulator, (2019), arXiv:1912.07904 [quant-ph].

[52] M.-C. Chen, M. Gong, X.-S. Xu, X. Yuan, J.-W. Wang, C. Wang, C. Ying, J. Lin, Y. Xu, Y. Wu, S. Wang, H. Deng, F. Liang, C.-Z. Peng, S. C. Benjamin, X. Zhu, C.-Y. Lu, and J.-W. Pan, Demonstration of adiabatic variational quantum computing with a superconducting quantum coprocessor, (2019), arXiv:1905.03150 [quant-ph].