- Variational quantum algorithms
  - VQE
  - Quantum natural gradient descent
  - Real-time VQA based on McLachlan principle

Variational quantum algorithms (VQAs)

What is a variational quantum algorithm?

Key idea: represent a quantum state of interest via a parametrized quantum circuit (= parametrized unitary operator):

reference state, e.g., $|0\rangle$.

$$|\psi(\vec{\theta})\rangle = U(\vec{\theta})\,|\psi_0\rangle$$

vector of real parameters $\vec{\theta} = (\theta_1, \ldots, \theta_{N_\theta}) \in \mathbb{R}^{N_\theta}$

$\vec{\theta}$ is a classical representation of the quantum state $|\psi(\vec{\theta})\rangle$.

Then, determine parameters $\vec{\theta}$ by classically optimizing an objective cost function $C(\vec{\theta})$, which can be computed by preparing $|\psi(\vec{\theta})\rangle$ on a QC and measuring expectation values. Typically choose $U(\vec{\theta})$ to consist of single & two-qubit gates.

VQAs are ideally tailored to NISQ conditions as we restrict $U(\vec{\theta})$ to circuits that can be efficiently implemented on hardware.

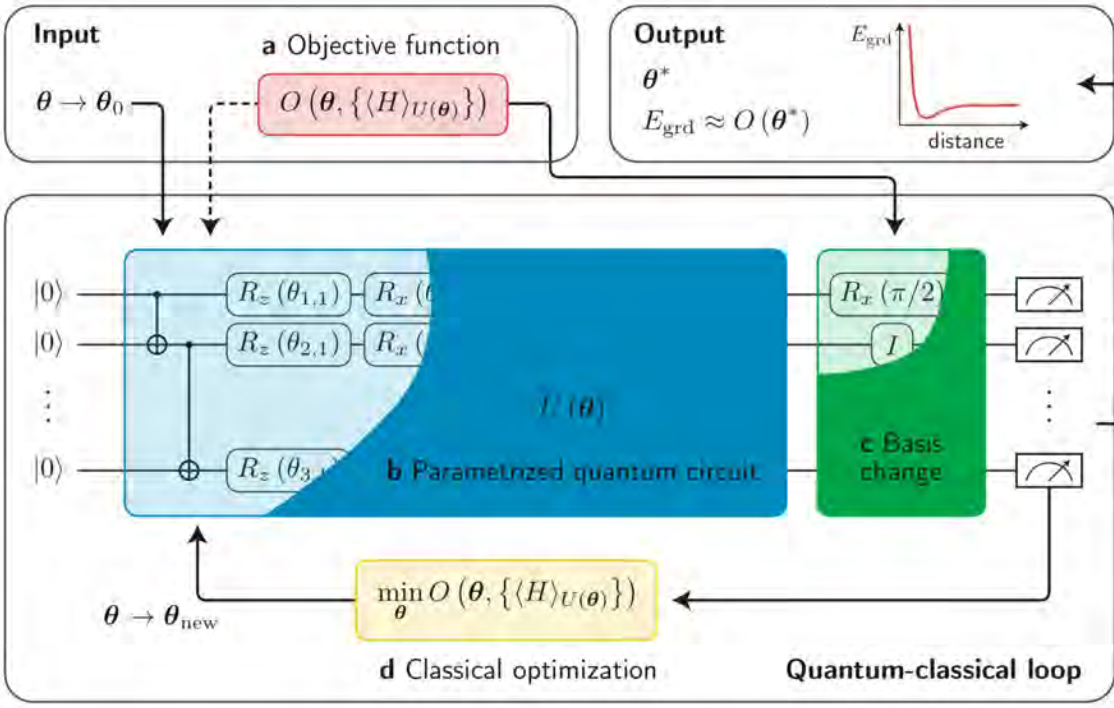


Figure 2 Diagrammatic representation of a Variational Quantum Algorithm (VQA). A VQA workflow can be divided into four main components: *a)* the objective function O that encodes the problem to be solved; *b)* the parameterized quantum circuit (PQC) U, which variables $\boldsymbol{\theta}$ are tuned to minimize the objective; *c)* the measurement scheme, which performs the basis changes and measurements needed to compute expectation values that are used to evaluate the objective; and *d)* the classical optimizer that minimizes the objective. The PQC can be defined heuristically, following hardware-inspired ansätze, or designed from the knowledge about the problem Hamiltonian H. Inputs of a VQA are the circuit ansatz $U(\boldsymbol{\theta})$ and the initial parameter values $\boldsymbol{\theta}_0$. Outputs include optimized parameter values $\boldsymbol{\theta}^*$and the minimum of the objective.

Caveat 1: only certain states can be reached with shallow circuits,

but circuit complexity ≠ entanglement (unlike in classical TN approaches)

Caveat 2: trade circuit depth for number of measurements (often this is a bottleneck in practice).

Examples:

Hamiltonian (or general hermitian operator = observable)

① $\cdot\ C(\vec{\theta}) = \langle \psi(\vec{\theta}) | H | \psi(\vec{\theta}) \rangle \equiv E(\theta)$

$\rightarrow$ goal is to find GS energy.

$$E_{GS} \leq \min_{\vec{\theta}} E(\vec{\theta}) \quad \text{[variational principle]}$$

classical minimization over noisy cost function!

This is VQE.

nonlocal even if H is local

② $\cdot\ C(\vec{\theta}) = \langle \psi_\theta | H^2 | \psi_\theta \rangle - \langle \psi_\theta | H | \psi_\theta \rangle^2$

minimize the energy variance $\Rightarrow$ prepare any eigenstate, also (highly) excited ones.

This is VQE-X.

③ $\cdot\ C(\vec{\theta}) = \langle \psi_\theta | \underbrace{(H - \lambda)^2}_{= H^2 - 2\lambda H + \lambda^2} | \psi_\theta \rangle =$

$$= \langle \psi_\theta | H^2 | \psi_\theta \rangle - \langle \psi_\theta | H | \psi_\theta \rangle^2 + \langle \psi_\theta | (H - \lambda) | \psi_\theta \rangle^2$$

Minimize sum of energy variance and energy difference to specified target λ (can be generalized

This is the folded spectrum method.

④ $\cdot\ C(\vec{\theta}) = 1 - |\langle \psi_{target} | \psi_\theta \rangle|^2$

Minimize infidelity with a target state.

$\Rightarrow C(\vec{\theta}) = 1 - |\langle \psi_0 | U^\dagger_{target} U_\theta | \psi_0 \rangle|^2 = 1 - p_{\psi_0}$

$\Rightarrow$ measure probability to return to $|\psi_0\rangle$

$\hat{=}$ measure global projector $P_G = \mathbb{1} - |\psi_0\rangle\langle\psi_0|$:

$$C(\vec{\theta}) = \langle \psi_0 | U^\dagger_\theta U_T \left(\mathbb{1} - |\psi_0\rangle\langle\psi_0|\right) U^\dagger_T U_\theta |\psi_0\rangle$$
$$= 1 - |\langle \psi_0 | U^\dagger_T U_\theta | \psi_0 \rangle|^2 = 1 - p_{\psi_0}$$

Global observables often difficult to optimize for large systems (one out of exponentially many states, barren plateaus)

Alternative: local cost function that exhibits the same minimum: $\underset{\vec{\theta}}{\operatorname{argmin}}\, C_G(\vec{\theta}) = \underset{\vec{\theta}}{\operatorname{argmin}}\, C_L(\vec{\theta})$

$$P_L = \mathbb{1} - \frac{1}{N}\sum_{j=1}^{N} |\psi_0\rangle_j\, {}_j\langle\psi_0| \otimes \mathbb{1}_{\bar{k}}$$

identity matrix for all qubits except k

<u>Example</u>: $|\psi_0\rangle = |0000\rangle$

$$U_T^{\dagger} U_\theta |\psi_0\rangle = \sqrt{p_0}\,|0000\rangle + \sqrt{p_2}\,|0010\rangle + \sqrt{p_5}\,|0101\rangle + \sqrt{p_8}\,|1000\rangle + \sqrt{p_{14}}\,|1110\rangle + \sqrt{p_{15}}\,|1111\rangle$$

$$\Rightarrow\quad C_G(\vec{\theta}) = 1 - p_0$$

$$C_L(\vec{\theta}) = 1 - \frac{1}{4}\left[4p_0 + 3p_2 + 2p_5 + 3p_8 + p_{14}\right]$$

$\Rightarrow$ C_L is more robust to small errors / fluctuations than C_G (more forgiving). $\square$

Common ansatz strategies:

① • hardware efficient ansatz (use native gates of QPU)

② • model specific ansatz

Hamiltonian variational ansatz (HVA)

$$H = \sum_i a_i h_i = \sum_{a=1}^{d} F_a$$

Choose F_a such that $e^{-iF_a t}$ can be easily implemented in circuit for any t, e.g., sum of pairwise commuting local terms

$$U(\vec{\theta}) = \prod_{\ell=1}^{\ell} \left[U_{\ell,m}(\theta_{\ell,m}) \cdots U_{\ell,1}(\theta_{\ell,1}) \right]$$

layers

with $U_{\ell,a} = e^{-i\theta_{\ell,a} F_a}$

③ • adaptive ansatz strategy: grow ansatz dynamically during optimization: ADAPT-VQE: append ansatz by an operator from a predefined pool based on maximizing energy gradient.

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-019-10988-2

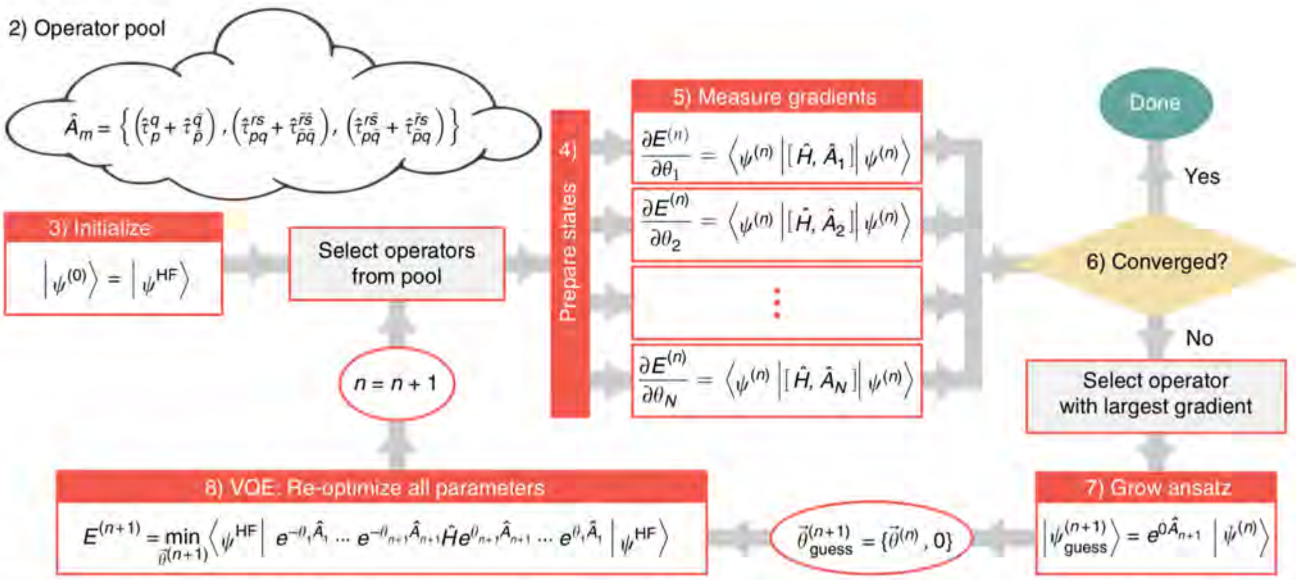
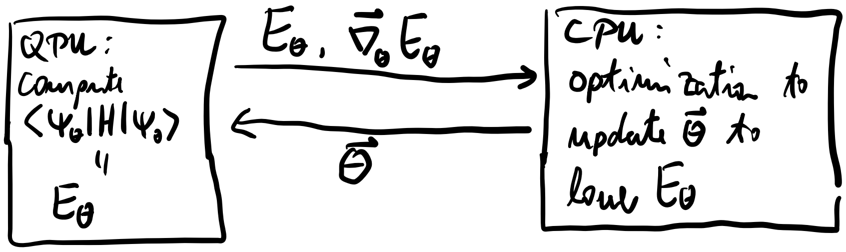


Fig. 1 Schematic depiction of the ADAPT-VQE algorithm described presented. Since step 1 occurs on classical hardware, it is not included in the illustration. $\vec{\theta}^{(n)}$ is the list of ansatz parameters at the nth iteration. The number of parameters, $len(\vec{\theta}^{(n)})$, is equal to the number of operators in the ansatz. "Operator Pool" refers to the collection of operators which are used to grow the ansatz one-at-a-time. Each τ_p^q represents a generalized single or double excitation, and these operators are then spin-complemented. The orbital indices refer to spatial orbitals, and the overbar indicates β spin. Orbital indices without overbars have α spin. Note that growing the ansatz does not drain the pool, and so operators can show up multiple times if selected by the algorithm

From Grimsley et al.,
Nat. Commun. (2019)

Classical optimization:

Classical-quantum feedback loop



Optimization strategies:

- gradient-based:
  - (stochastic) gradient descent
  - quantum natural gradient descent ($\hat{=}$ imaginary time evolution)
  - ADAM
  - BFGS (nonlinear optimization with constraints)
  - SPSS
- non-gradient based:
  - Nelder-Mead
  - BOBYQA
  - Evolutionary algorithms like CMA-ES

Parameter shift rule to compute gradients.

For ansätze $U(\vec{\theta}) = \prod_{i=1}^{N} e^{-i\theta_i P_i}$ with Pauli string P_i,

fulfilled by generalized HVA

(can be relaxed to any hermitian operator with just two eigenvalues)

one can use the parameter shift rule to analytically compute gradients (Mitarai et al. (2018), Schuld et al. (2018)):

$$e^{-i\theta_i P_i} = \cos(\theta_i) - i P_i \sin(\theta_i)$$

$$\Rightarrow \frac{\partial}{\partial \theta_i} U(\vec{\theta}) = U_{j>i}(\vec{\theta}_{j>i}) \left[-\sin(\theta_i) - i P_i \cos\theta_i \right] \cdot U_{j<i}(\vec{\theta}_{j<i})$$

The energy is given by

$$E_\theta = \langle \psi_\theta | H | \psi_\theta \rangle =$$

$$= \langle \psi_0 | U_1^\dagger \cdots U_N^\dagger H U_N \cdots U_1 | \psi_0 \rangle =$$

$$= \langle \psi_{i-1} | U_i^\dagger \underbrace{\tilde{H}_{i+1}}_{= U_{i+1}^\dagger \cdots U_N^\dagger H U_N \cdots U_{i+1}} U_i \underbrace{|\psi_{i-1}\rangle}_{= U_{i-1} \cdots U_1 |\psi_0\rangle}$$

Gradient of energy thus reads:

$$\frac{\partial}{\partial\theta_i} E_\theta = i \langle\psi_{i-1}|\, U_i^\dagger \left[P_i, \tilde{H}_{i+1}\right] U_i \,|\psi_{i-1}\rangle$$

$$\frac{\partial}{\partial\theta_i} \underbrace{e^{-i\theta_i P_i}}_{=U_i} = -iP_i\, e^{-i\theta_i P_i} = -i\, e^{-i\theta_i P_i} P_i$$

The commutator can be expressed as:

$$\left[P_i, \tilde{H}_{i+1}\right] = P_i \tilde{H}_{i+1} - \tilde{H}_{i+1} P_i =$$

$$= -i\left(\frac{1}{2}[1+iP_i]\,\tilde{H}_{i+1}\,[1-iP_i] - \frac{1}{2}[1-iP_i]\,\tilde{H}_{i+1}\,[1+iP_i]\right) =$$

$$e^{\mp i\frac{\pi}{4}P_i} = \frac{1}{\sqrt{2}}(1 \mp iP_i)$$

only cross-terms $1\cdot\tilde{H}_{i+1}P_i$ and $P_i\tilde{H}_{i+1}\cdot 1$ survive

$$= -i\,\frac{1}{2}\left[2iP_i\tilde{H}_{i+1} - 2i\tilde{H}_{i+1}P_i\right] =$$

$$= \left[P_i, \tilde{H}_{i+1}\right].$$

$$\Rightarrow \left[P_i, \tilde{H}_{i+1}\right] = \underbrace{-i\left[U_i^\dagger\left(\tfrac{\pi}{4}\right)\tilde{H}_{i+1}U_i\left(\tfrac{\pi}{4}\right) - U_i^\dagger\left(-\tfrac{\pi}{4}\right)\tilde{H}_{i+1}U_i\left(-\tfrac{\pi}{4}\right)\right]}$$

The gradient then becomes:

$$\frac{\partial}{\partial\theta_i} E_\theta = i \langle\psi_{i-1}|\, U_i^\dagger \overbrace{\left[P_i, \tilde{H}_{i+1}\right]}^{=-i[\dots]} U_i \,|\psi_{i-1}\rangle =$$

$$\Leftrightarrow \boxed{\frac{\partial}{\partial \theta_i} E_\theta = \langle \psi_{i-1} | U_i^\dagger(\theta_i + \tfrac{\pi}{4}) \tilde{H}_{i+1} U_i(\theta_i + \tfrac{\pi}{4}) - U_i^\dagger(\theta_i - \tfrac{\pi}{4}) \tilde{H}_{i+1} U_i(\theta_i - \tfrac{\pi}{4}) | \psi_{i-1} \rangle.}$$

Obtain energy gradient by shifting parameter $\theta_i \pm \frac{\pi}{4}$ and subtract. Note, these are all expectation values of PQCs.

Quantum natural gradient descent (QNG):

introduced by Stokes et al. (2019).

Use quantum geometry of variational pure states to accelerate convergence to local minima (compared to vanilla GD). Similar method very popular in ML (natural GD), where the Fisher information matrix plays the role of the quantum metric g_{ij}.

Vanilla gradient descent (to find local minima, use multiple starting points to reach global minimum):

Cost function

$$C(\theta) = \frac{1}{2} \langle \psi_\theta | H | \psi_\theta \rangle$$

evaluated using **PQC**

$$|\psi_\theta\rangle = \prod_{i=1}^{N} U_i(\theta_i) \, |\psi_0\rangle$$

$$\vec{\theta} = (\theta_1, \ldots, \theta_N)$$

$e^{-i\theta_i P_i}$, can also include gates that are θ independent.

Vanilla gradient descent update rule

$$\vec{\theta}_{t+1} := \vec{\theta}_t - \underbrace{\eta}_{\text{step size } \eta} \vec{\nabla} C(\theta_t)$$

$$= \underset{\vec{\theta}}{\operatorname{argmin}} \left[\langle \vec{\theta} - \vec{\theta}_t , \vec{\nabla} C(\vec{\theta}_t) \rangle + \frac{1}{2\eta} \| \vec{\theta} - \vec{\theta}_t \|_2^2 \right]$$

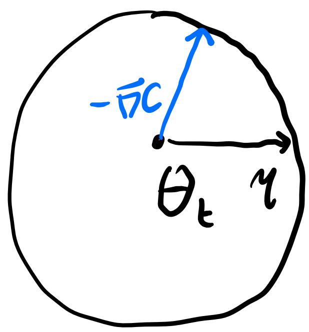
move in direction of steepest descent

fix step size to be η.

$\langle \vec{a}, b \rangle =$ Scalar product of vectors $\vec{a} \cdot \vec{b}$

Important: $\| \vec{\theta} - \vec{\theta}_t \|_2^2$ is w.r.t. Euclidean l_2-norm in parameter $\vec{\theta}$ space.

$$\| \vec{\theta} - \vec{\theta}_t \|_2^2 = \sum_{i=1}^{N} (\theta_i - \theta_{t,i})^2$$



Equivalence of updating rule & argmin:

$$\frac{\partial}{\partial \theta_i} \left[\langle \vec{\theta} - \vec{\theta}_t , \vec{\nabla} C \rangle + \frac{1}{2\eta} \| \vec{\theta} - \vec{\theta}_t \|_2^2 \right] =$$

$$= (\vec{\nabla} C)_i + \frac{1}{\cancel{2}\eta} \cancel{2} (\theta_i - \theta_{t,i}) = 0$$

$$\Rightarrow \theta_i = \theta_{t,i} - \eta (\vec{\nabla} C)_i \qquad \square$$

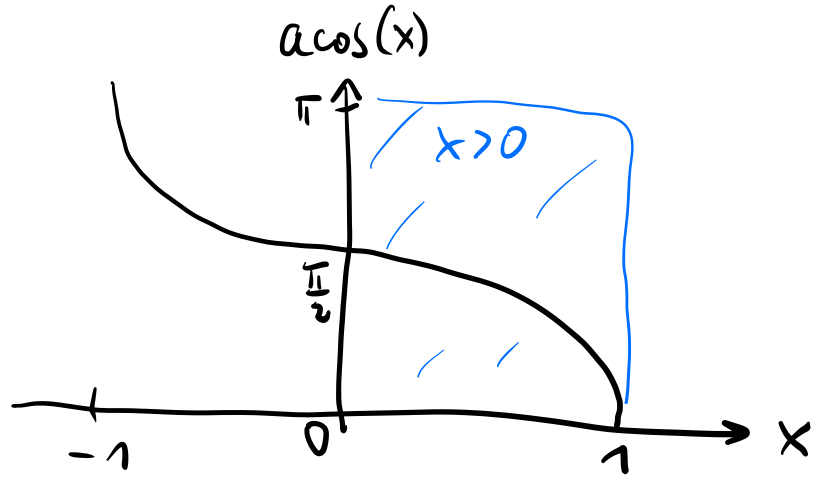
The Euclidean norm $\|\vec{a}\|_2$ in parameter space does not properly reflect the geometry of the states $|\psi_\theta\rangle$. This can slow down convergence.

The Fubini-Study metric is unique, unitarily invariant metric on state of pure states $|\psi\rangle\langle\psi| \in \mathbb{C}P^{d-1}$, $d = 2^n$ for qubits:

We saw earlier that one can define a metric on space of states (both pure & mixed) that is based on the fidelity

$$\boxed{d(P_\psi, P_\phi) = \operatorname{acos}(|\langle\psi|\phi\rangle|)} \in [0, \tfrac{\pi}{2}].$$

normalized vectors of pure states



The Fubini-Study metric tensor can then be obtained by computing distances b/w infinitesimally close states.

For our variational states ($\vec{\theta} = (\theta_1, \dots, \theta_N)$):

$$\boxed{d^2(P_\theta, P_{\theta+\partial\theta}) = \sum_{i,j=1}^{N} g_{ij}(\theta)\, \delta\theta_i\, \delta\theta_j}$$

Fubini-Study quantum metric on variational states

Derivation: (Einstein summation)

$$|\psi_{\theta+\partial\theta}\rangle = |\psi_\theta\rangle + \frac{\partial|\psi_\theta\rangle}{\partial\theta_i}\delta\theta_i + \frac{\partial^2|\psi_\theta\rangle}{\partial\theta_j\partial\theta_i}\delta\theta_j\,\delta\theta_i + \dots$$

Two useful identities:

1 $\langle\psi_\theta|\psi_\theta\rangle = 1$

$$\Rightarrow \underbrace{\frac{\partial\langle\psi_\theta|}{\partial\theta_i}|\psi_\theta\rangle}_{\text{can also be written as } \langle \frac{\partial\psi_\theta}{\partial\theta_i}, \psi_\theta\rangle} + \overbrace{\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle}^{= -i\langle\psi_\theta|i\frac{\partial}{\partial\theta_i}|\psi_\theta\rangle = -iA_i,\ A_i \in \mathbb{R}} = 0$$

$$\Leftrightarrow \frac{\partial\langle\psi_\theta|}{\partial\theta_i}|\psi_\theta\rangle + c.c. = 0$$

$\Rightarrow \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle = i\mathbb{R}$ is purely imaginary.

We write $\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle = -i\langle\psi_\theta|i\partial_{\theta_i}|\psi_\theta\rangle = -iA_i$, $A_i \in \mathbb{R}$

The real quantity $A_j = \langle \psi_\theta | i\partial_{\theta_j} | \psi_\theta \rangle$ is called Berry connection.

Taking one more derivative:

$$② \quad \frac{\partial^2 \langle \psi_\theta |}{\partial \theta_j \partial \theta_i} |\psi_\theta\rangle + \langle \psi_\theta | \frac{\partial^2 \psi_\theta}{\partial \theta_j \partial \theta_i} \rangle + \frac{\partial \langle \psi_\theta |}{\partial \theta_i} \frac{\partial |\psi_\theta\rangle}{\partial \theta_j} + \frac{\partial \langle \psi_\theta |}{\partial \theta_j} \frac{\partial |\psi_\theta\rangle}{\partial \theta_i} = 0$$

We want to calculate the distance (squared):

$$\Rightarrow d^2(P_\theta, P_{\theta+\delta\theta}) = \operatorname{acos}^2\left(|\langle \psi_\theta | \psi_{\theta+\delta\theta} \rangle| \right) =$$

We thus need the overlap

$$\langle \psi_\theta | \psi_{\theta+\delta\theta} \rangle = 1 + \langle \psi_\theta | \frac{\partial \psi_\theta}{\partial \theta_i} \rangle \delta\theta_i + \frac{1}{2} \langle \psi_\theta | \frac{\partial^2 \psi_\theta}{\partial \theta_j \partial \theta_i} \rangle \delta\theta_i \delta\theta_j$$

Thus,

$$|\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle|^2 = \langle\psi_{\theta+\delta\theta}|\psi_\theta\rangle\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle =$$

$$= \left[1 + \langle\frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle\delta\theta_i + \frac{1}{2}\langle\frac{\partial^2\psi_\theta}{\partial\theta_j\partial\theta_i}|\psi_\theta\rangle\delta\theta_i\delta\theta_j\right]$$

$$\cdot\left[1 + \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle\partial\theta_j + \frac{1}{2}\langle\psi_\theta|\frac{\partial^2\psi_\theta}{\partial\theta_k\partial\theta_l}\rangle\delta\theta_k\delta\theta_l\right]$$

$$= 1 + \underbrace{\left[\langle\frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle + \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle\right]}_{=0}\partial\theta_i$$

$$+\left(\underbrace{\langle\frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle}_{=iA_i}\underbrace{\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle}_{=-iA_j} + \frac{1}{2}\langle\frac{\partial^2\psi_\theta}{\partial\theta_j\partial\theta_i}|\psi_\theta\rangle\right.$$

$$\left.+\frac{1}{2}\langle\psi_\theta|\frac{\partial^2\psi_\theta}{\partial\theta_j\partial\theta_i}\rangle\right)\partial\theta_i\delta\theta_j =$$

$$= 1 + \Big[\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle \langle \psi_\theta | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle$$

$$- \frac{1}{2} \Big(\langle \frac{\partial \psi_\theta}{\partial \theta_j} | \frac{\partial \psi_\theta}{\partial \theta_i} \rangle + \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle \Big) \Big] \partial\theta_i \partial\theta_j + \dots$$

Taking the square root and using that $\sqrt{1+x} = 1 + \frac{x}{2} + O(x^2)$,

we find

$$\Rightarrow |\langle \psi_\theta | \psi_{\theta+\delta\theta} \rangle| = \sqrt{|\langle \psi_\theta | \psi_{\theta+\partial\theta} \rangle|^2} =$$

$$= 1 + \frac{1}{2} \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle \langle \psi_\theta | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle \partial\theta_i \partial\theta_j$$

$$- \frac{1}{4} \Big(\langle \frac{\partial \psi_\theta}{\partial \theta_j} | \frac{\partial \psi_\theta}{\partial \theta_i} \rangle + \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle \Big) \partial\theta_i \partial\theta_j$$

Using that $\mathrm{acos}^2(1-x) = 2x + O(x^2)$,

we finally obtain

$$d^2(P_\theta, P_{\theta+\delta\theta}) = \mathrm{acos}^2\left(|\langle \psi_\theta | \psi_{\theta+\delta\theta}\rangle|\right) =$$

$$= \overbrace{\left[\frac{1}{2}\left(\left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \frac{\partial \psi_\theta}{\partial \theta_i}\right\rangle + \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \frac{\partial \psi_\theta}{\partial \theta_j}\right\rangle\right)\right.}^{= \mathrm{Re}\left[\left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \frac{\partial \psi_\theta}{\partial \theta_i}\right\rangle\right]}$$

$$\left. - \underbrace{\left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \psi_\theta \right\rangle \left\langle \psi_\theta \middle| \frac{\partial \psi}{\partial \theta_j}\right\rangle}_{= A_i A_j \text{ also purely real as } \left\langle \psi_\theta \middle| \frac{\partial \psi}{\partial \theta_j}\right\rangle = -i A_j}\right] \delta\theta_i \, \delta\theta_j$$

$$\Rightarrow d^2(P_\theta, P_{\theta+\delta\theta}) = \mathrm{Re}\left[\left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \frac{\partial \psi_\theta}{\partial \theta_i}\right\rangle - \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \psi_\theta \right\rangle \left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_j}\right\rangle\right] \delta\theta_i \, \delta\theta_j =$$

$$= \mathrm{Re}\left[\underbrace{G_{ij}(\theta)}_{\text{Quantum geometric tensor}}\right] \delta\theta_i \, \delta\theta_j = \underbrace{g_{ij}(\theta)}_{\text{Fubini-Study metric}} \delta\theta_i \, \delta\theta_j .$$

$$\Rightarrow \boxed{g_{ij}(\theta) = \frac{1}{2}\left(\left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \frac{\partial \psi_\theta}{\partial \theta_i}\right\rangle + \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \frac{\partial \psi_\theta}{\partial \theta_j}\right\rangle\right) - A_i A_j}$$

Note that $\operatorname{Im}(G_{ij}) = \frac{1}{2i}\left(\left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \frac{\partial \psi_\theta}{\partial \theta_i}\right\rangle - \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \frac{\partial \psi_\theta}{\partial \theta_j}\right\rangle\right)$

$$= -\frac{i}{2}\left[\frac{\partial}{\partial \theta_j} \underbrace{\left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_i}\right\rangle}_{= -i A_i \text{ (Berry connection)}} - \frac{\partial}{\partial \theta_i}\left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_j}\right\rangle\right] =$$

$$= -\frac{1}{2}\underbrace{\left[\frac{\partial}{\partial \theta_j} A_i - \frac{\partial}{\partial \theta_i} A_j\right]}_{= \vec{\nabla}_\theta \times \vec{A}} = \text{curl of Berry connection}$$

$= \boxed{\text{Berry curvature.}}$

$$\boxed{\Rightarrow \underbrace{G_{ij}}_{\text{Quantum geometric tensor}} = \underbrace{g_{ij}}_{\text{Fubini-Study quantum metric}} - \frac{i}{2} \underbrace{\Omega_{ij}}_{\text{Berry curvature}}}$$

Recap:

The Fubini-Study metric describes distances of pure states. It is unitarily invariant, i.e., this distance measure is invariant under reparametrizations of $|\psi_\theta\rangle$.

Quantum natural gradient descent (QNG):

$$\theta_{t+1} := \underset{\vec{\theta}}{\operatorname{argmin}} \Big[\langle \theta - \theta_t , \vec{\nabla} C(\theta) \rangle + \frac{1}{2\eta} \| \theta - \theta_t \|^2_{g(\theta)} \Big]$$

step size determined using Fubini-Study (FS) metric

First-order optimality:

$$\frac{\partial}{\partial \theta_i} \Big[\langle \theta - \theta_t , \vec{\nabla} C(\theta) \rangle + \frac{1}{2\eta} \underbrace{\| \theta - \theta_t \|^2_{g(\theta)}}_{= \langle \theta - \theta_t , g(\theta - \theta_t) \rangle} \Big] = 0$$

$$\Leftrightarrow \big(\nabla C(\theta)\big)_i + \frac{1}{2\eta} \Big[g_{ij} (\theta_j - \theta_{t,j}) + \underbrace{(\theta_j - \theta_{t,j}) g_{ji}}_{\substack{g_{ij} = g_{ji} \\ \Rightarrow \text{ equal to} \\ g_{ij}(\theta_j - \theta_{t,j})}} \Big] = 0$$

$$\Rightarrow g_{ij}(\theta_t) \big(\theta_i - \theta_{t,i} \big) = -\eta \big[\vec{\nabla} C(\theta_t) \big]_i$$

$$\Leftrightarrow \boxed{\begin{aligned} & g(\theta_t)\left(\theta_{t+1} - \theta_t\right) = -\eta \vec{\nabla} C(\theta_t) \\ \Leftrightarrow\; & \theta_{t+1} = \theta_t - \eta\, \bar{g}^{-1}(\theta_t)\, \vec{\nabla} C(\theta_t) \end{aligned}}$$

Updating rules for QNG

generalized inverse

$$\left[g + \xi \mathbb{1}\right]^{-1} = \bar{g}^{-1}$$

with $\xi \ll 1$.

Relation to quantum imaginary time evolution:

It turns out that QNG update rule corresponds to quantum imaginary time evolution (QITE), projected onto the variational manifold and in the infinitesimal step size limit.

Derivation:

QITE is defined by $\left[|\psi_{\bar{\theta}}\rangle = e^{-H\delta\tau} |\psi_\theta\rangle \right]$:

$$\underset{\delta\theta \in \mathbb{R}^{N_\theta}}{\operatorname{argmin}} \underbrace{\left\| |\psi_{\bar{\theta}}\rangle - |\psi_{\theta+\delta\theta}\rangle\langle\psi_{\theta+\delta\theta}|\psi_{\bar{\theta}}\rangle \right\|_2^2}_{\substack{= \langle\psi_{\bar{\theta}}|\psi_{\bar{\theta}}\rangle + \langle\psi_{\theta+\delta\theta}|\psi_{\bar{\theta}}\rangle \\ -2|\langle\psi_{\theta+\delta\theta}|\psi_{\bar{\theta}}\rangle|^2}} =$$

minimize by maximizing the last term.

$$= \underset{\delta\theta \in \mathbb{R}^{N_\theta}}{\operatorname{argmax}} \left| \langle\psi_{\bar{\theta}}|\psi_{\theta+\delta\theta}\rangle \right|^2 .$$

Expand expression to quadratic order in $\delta\theta$, $\delta\tau$ and take 1st order optimality condition.

First we find

$$\langle \psi_{\bar\theta} | \psi_{\theta+\delta\theta}\rangle = \langle \psi_{\bar\theta}|\psi_\theta\rangle + \langle \psi_{\bar\theta}|\frac{\partial \psi_\theta}{\partial \delta\theta_i}\rangle \delta\theta_i + \frac{1}{2} \langle \psi_{\bar\theta}|\frac{\partial^2\psi_\theta}{\partial\delta\theta_j\, \partial\delta\theta_i}\rangle \delta\theta_j\, \delta\theta_i + \dots$$

$$\Rightarrow |\langle \psi_{\bar\theta} | \psi_{\theta+\delta\theta}\rangle|^2 =$$

$$= \left[\langle \psi_{\bar\theta}|\psi_\theta\rangle + \langle \psi_{\bar\theta}|\frac{\partial \psi_\theta}{\partial \delta\theta_i}\rangle \delta\theta_i + \frac{1}{2} \langle \psi_{\bar\theta}|\frac{\partial^2\psi_\theta}{\partial\delta\theta_j\, \partial\delta\theta_i}\rangle \delta\theta_j\, \delta\theta_i\right] \cdot \left[\langle \psi_\theta|\psi_{\bar\theta}\rangle + \langle \frac{\partial \psi_\theta}{\partial \delta\theta_i}|\psi_{\bar\theta}\rangle \delta\theta_i + \frac{1}{2} \langle \frac{\partial^2\psi_\theta}{\partial\delta\theta_j\, \partial\delta\theta_i}|\psi_{\bar\theta}\rangle \delta\theta_j\, \delta\theta_i\right]$$

$$= \underbrace{|\langle \psi_{\bar\theta}|\psi_\theta\rangle|^2}_{=\langle \psi_{\bar\theta}|\psi_\theta\rangle\langle \psi_\theta|\psi_{\bar\theta}\rangle} + \left[\langle \psi_{\bar\theta}|\psi_\theta\rangle \langle \frac{\partial \psi_\theta}{\partial \delta\theta_i}|\psi_{\bar\theta}\rangle + \langle \psi_\theta|\psi_{\bar\theta}\rangle\langle \psi_{\bar\theta}|\frac{\partial \psi_\theta}{\partial \delta\theta_i}\rangle\right] \delta\theta_i$$
$$+ \left[\langle \psi_{\bar\theta}|\frac{\partial \psi_\theta}{\partial \delta\theta_i}\rangle \langle \frac{\partial \psi_\theta}{\partial \delta\theta_j}|\psi_{\bar\theta}\rangle + \langle \psi_{\bar\theta}|\psi_\theta\rangle \frac{1}{2} \langle \frac{\partial^2\psi_\theta}{\partial\delta\theta_j\, \partial\delta\theta_i}|\psi_{\bar\theta}\rangle + \langle \psi_\theta|\psi_{\bar\theta}\rangle \frac{1}{2} \langle \psi_{\bar\theta}|\frac{\partial^2\psi_\theta}{\partial\delta\theta_j\, \partial\delta\theta_i}\rangle\right\}$$

$\cdot\, \delta\theta_i\, \delta\theta_j\,.$

Now expand $|\psi_{\bar\theta}\rangle = e^{-H\delta\tau}|\psi_\theta\rangle = (1-H\delta\tau)|\psi_\theta\rangle$

and keep terms up to second order in $\delta\tau, \delta\theta_i$:

does not depend on $\delta\theta_i$, so drops out when computing 1st order optimality $\frac{\partial}{\partial\delta\theta_i}(\ldots)=0$

$$|\langle\psi_{\bar\theta}|\psi_{\theta+\delta\theta}\rangle|^2 = |\langle\psi_{\bar\theta}|\psi_\theta\rangle|^2$$

$$+\left(\langle\psi_\theta|\psi_\theta\rangle - 2\delta\tau\langle\psi_\theta|H|\psi_\theta\rangle\right)\Big[\underbrace{\langle\frac{\partial\psi_\theta}{\partial\delta\theta_i}|\psi_\theta\rangle + \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\delta\theta_i}\rangle}_{=0}\Big]\delta\theta_i$$

$$-\underbrace{\Big[\langle\frac{\partial\psi_\theta}{\partial\delta\theta_i}|H|\psi_\theta\rangle + \langle\psi_\theta|H|\frac{\partial\psi}{\partial\delta\theta_i}\rangle\Big]}_{=\frac{\partial}{\partial\delta\theta_i}\langle\psi_\theta|H|\psi_\theta\rangle = \frac{\partial}{\delta\theta_i}E_\theta}\delta\tau\,\delta\theta_i$$

$$\underbrace{+\Big[\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\delta\theta_i}\rangle\langle\frac{\partial\psi_\theta}{\partial\delta\theta_j}|\psi_\theta\rangle - \frac{1}{2}\Big(\langle\frac{\partial\psi_\theta}{\partial\delta\theta_i}|\frac{\partial\psi_\theta}{\partial\delta\theta_j}\rangle + \langle\frac{\partial\psi_\theta}{\partial\delta\theta_j}|\frac{\partial\psi_\theta}{\partial\delta\theta_i}\rangle\Big)\Big]}_{=-\mathrm{Re}[G_{ij}]}\delta\theta_i\,\delta\theta_j$$

$$\Leftrightarrow |\langle \psi_{\bar{\theta}} | \psi_{\theta+\delta\theta} \rangle|^2 = |\langle \psi_{\bar{\theta}} | \psi_\theta \rangle|^2$$

$$- \frac{\partial}{\partial \delta\theta_i} \langle \psi_\theta | H | \psi_\theta \rangle \delta\theta_i \, \delta\tau$$

$$- g_{ij}(\theta) \, \delta\theta_i \, \delta\theta_j$$

To find $\underset{\delta\theta}{\operatorname{argmax}} |\langle \psi_{\bar{\theta}} | \psi_{\theta+\delta\theta} \rangle|^2$, we look at first order optimality

$$\frac{\partial}{\partial \delta\theta_i} |\langle \psi_{\bar{\theta}} | \psi_{\theta+\delta\theta} \rangle|^2 = 0$$

$$\Leftrightarrow - \frac{\partial}{\partial \delta\theta_i} \langle \psi_\theta | H | \psi_\theta \rangle \delta\tau - 2 g_{ij} \, \delta\theta_j = 0$$

$$\Leftrightarrow \quad g_{ij}(\theta)\,\delta\theta_j = -\frac{\partial}{\partial \delta\theta_i} \underbrace{\frac{1}{2}\langle\psi_\theta|H|\psi_\theta\rangle}_{=C(\theta)}\delta\tau$$

QITE

$$\Rightarrow \quad \boxed{g_{ij}(\theta)\,\delta\theta_j = -\frac{\partial}{\partial \delta\theta_i} C(\theta)\,\delta\tau}$$

This is exactly the QNG update rule

$\Rightarrow$ QITE $\hat{=}$ QNG

Using vector notation, we find

$$\Rightarrow \quad g\,\delta\theta = -\vec{\nabla} C(\theta)\,\delta\tau$$

In the limit $\delta\tau \to 0$, we find

$$g[\theta(\tau)]\,\dot{\theta}(\tau) = -\vec{\nabla} C[\theta(\tau)] \,.$$

This corresponds to the VQITE EOM derived from McLachlan's principle as we show next.

McLachlan's variational principle (Li, Benjamin et al, (2018))

Density matrix under imaginary time evolution:

$$\rho(\tau) = \frac{e^{-H\tau}\,\rho(0)\,e^{-H\tau}}{\mathrm{Tr}\left[e^{-2H\tau}\rho(0)\right]}$$

$$\Rightarrow \frac{\partial \rho}{\partial \tau} = -H\rho(\tau) - \rho(\tau)H - \overbrace{\frac{e^{-H\tau}\rho(0)\,e^{-H\tau}}{\mathrm{Tr}\left[e^{-2H\tau}\rho(0)\right]}}^{=\rho(\tau)}\;\overbrace{\frac{\mathrm{Tr}\left[-2He^{-2H\tau}\rho(0)\right]}{\mathrm{Tr}\left[e^{-2H\tau}\rho(0)\right]}}^{=-2\,\mathrm{Tr}[H\rho(\tau)]}$$

$$\Leftrightarrow \boxed{\frac{\partial \rho}{\partial \tau} = -\left\{H, \rho(\tau)\right\} + 2\langle H\rangle_{\rho(\tau)}\,\rho(\tau)}$$

Imag. time evolution of density matrix

Parametrize DM $\rho[\theta(\tau)]$ and then derive equation of motion of variational parameters using the variational principle.

What we want to minimize is the (McLachlan) distance between the time evolution of the variational parameters and the exact imag. time evolution:

$\|\rho\|^2 = \mathrm{Tr}[\rho^\dagger \rho]$ (Frobenius norm of matrices)

$$L^2 = \left\| \frac{\partial \rho[\theta(\tau)]}{\partial \theta_i} \dot{\theta}_i - \left(\frac{\partial \rho}{\partial \tau}\right)_{exact} \right\|^2 =$$

$$= \left\| \frac{\partial \rho[\theta(\tau)]}{\partial \theta_i} \dot{\theta}_i + \{H, \rho(\tau)\} - 2\langle H\rangle \rho(\tau) \right\|^2 =$$

$$= \mathrm{Tr}\left\{\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^\dagger \dot{\theta}_i + \{H,\rho\} - 2\langle H\rangle\rho\right]\left[\left(\frac{\partial \rho}{\partial \theta_j}\right)\dot{\theta}_j + \{H,\rho\} - 2\langle H\rangle \rho\right]\right\} =$$

$\rho^\dagger = \rho$, $H^\dagger = H$

$\theta \in \mathbb{R}$

$$= \mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^\dagger \left(\frac{\partial \rho}{\partial \theta_j}\right) \dot{\theta}_i \dot{\theta}_j\right]$$

$$+ \mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^\dagger \left(H\rho + \rho H - 2\langle H\rangle \rho\right) + \left(H\rho + \rho H - 2\langle H\rangle\rho\right)\left(\frac{\partial \rho}{\partial \theta_i}\right)\right]\dot{\theta}_i$$

$$+ \mathrm{Tr}\Big[\big(H\rho + \rho H - 2\langle H \rangle \rho \big) \big(H\rho + \rho H - 2 \langle H \rangle \rho \big) \Big].$$

We consider pure states in the following:

$$\rho(\tau) = |\psi(\tau)\rangle\langle\psi(\tau)| .$$

Then,

$$\frac{\partial}{\partial\theta_j}|\psi(\tau)\rangle\langle\psi(\tau)| = \Big|\frac{\partial\psi(\tau)}{\partial\theta_j}\Big\rangle\langle\psi(\tau)| + |\psi(\tau)\rangle\Big\langle\frac{\partial\psi(\tau)}{\partial\theta_j}\Big|$$

$$① \ \mathrm{Tr}\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^{\dagger}\underbrace{\Big(\frac{\partial\rho}{\partial\theta_j}\Big)}_{=\frac{\partial}{\partial\theta_j}|\psi(\tau)\rangle\langle\psi(\tau)|}\dot{\theta}_i\dot{\theta}_j\Big] =$$

$$= \mathrm{Tr}\Big[\Big(|\psi\rangle\Big\langle\frac{\partial\psi}{\partial\theta_i}\Big| + \Big|\frac{\partial\psi}{\partial\theta_i}\Big\rangle\langle\psi|\Big)\Big(\Big|\frac{\partial\psi}{\partial\theta_j}\Big\rangle\langle\psi| + |\psi\rangle\Big\langle\frac{\partial\psi}{\partial\theta_j}\Big|\Big)$$

$$\dot{\theta}_i\dot{\theta}_j =$$

$$= \Big[\Big\langle\frac{\partial\psi}{\partial\theta_i}\Big|\frac{\partial\psi}{\partial\theta_j}\Big\rangle + \underbrace{\Big\langle\frac{\partial\psi}{\partial\theta_i}\Big|\psi\Big\rangle}_{=iA_i}\underbrace{\Big\langle\frac{\partial\psi}{\partial\theta_j}\Big|\psi\Big\rangle}_{=iA_j} \to -A_iA_j$$

$$+ \underbrace{\Big\langle\psi\Big|\frac{\partial\psi}{\partial\theta_j}\Big\rangle}_{=-iA_j}\underbrace{\Big\langle\psi\Big|\frac{\partial\psi}{\partial\theta_i}\Big\rangle}_{-iA_i} + \Big\langle\frac{\partial\psi}{\partial\theta_j}\Big|\frac{\partial\psi}{\partial\theta_i}\Big\rangle\Big]\dot{\theta}_i\dot{\theta}_j$$

$$\to -A_iA_j$$

$$= 2\,\mathrm{Re}\Big[\underbrace{\langle \frac{\partial\psi}{\partial\theta_i} \Big| \frac{\partial\psi}{\partial\theta_j}\rangle + \langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle\langle\psi|\frac{\partial\psi}{\partial\theta_j}\rangle}_{G_{ij}[\theta(\tau)]}\Big] \cdot \dot\theta_i\,\dot\theta_j$$

$$= 2\,\mathrm{Re}\big[G_{ij}[\theta(\tau)]\big]\,\dot\theta_i\,\dot\theta_j = 2\, g_{ij}[\theta(\tau)]\,\dot\theta_i\,\dot\theta_j .$$

1

2

$$\mathrm{Tr}\Big[\Big(\overbrace{\frac{\partial\rho}{\partial\theta_i}}^{=|\frac{\partial\psi}{\partial\theta_i}\rangle\langle\psi| + |\psi\rangle\langle\frac{\partial\psi}{\partial\theta_i}|}\Big)^{\dagger}\Big(\underbrace{H\overbrace{\rho}^{=|\psi\rangle\langle\psi|} + \rho H}_{\text{in principle } \rho H + H\rho} - 2\langle H\rangle\rho\Big)$$

focus on pure states

$$+ \Big(H\rho + \rho H - 2\langle H\rangle\rho\Big)\Big(\underbrace{\frac{\partial\rho}{\partial\theta_i}}_{=|\frac{\partial\psi}{\partial\theta_i}\rangle\langle\psi| + |\psi\rangle\langle\frac{\partial\psi}{\partial\theta_i}|}\Big)\Big]\,\dot\theta_i =$$

see below (i) + (ii) + (iii) + (iv)

$$\stackrel{?}{=} 2\Big(\underbrace{\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle + \langle\psi|H|\frac{\partial\psi}{\partial\theta_i}\rangle}_{= 2\,\mathrm{Re}\left(\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle\right)}$$

$$- 4\langle H\rangle\Big(\underbrace{\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle + \langle\frac{\partial\psi}{\partial\theta_i}|\psi\rangle}\Big)\Big\}\,\dot\theta_i =$$

$$\overset{=0}{} = 4\,\mathrm{Re}\left(\langle \frac{\partial \psi}{\partial \theta_i} | H | \psi \rangle\right).$$

③

$$\mathrm{Tr}\left[\left(H\rho + \rho H - 2\langle H\rangle \rho\right)\left(H\rho + \rho H - 2\langle H\rangle\rho\right)\right] =$$

$$= \cancel{\langle H\rangle^2} + \langle H^2\rangle - 2\cancel{\langle H\rangle^2} + \langle \underline{H^2}\rangle + \cancel{\langle H\rangle^2} - 2\langle H\rangle^2$$

$$- 2\cancel{\langle H\rangle^2} - 2\cancel{\langle H\rangle^2} + \cancel{4\langle H\rangle^2} =$$

$$= 2\left(\langle H^2\rangle - \langle H\rangle^2\right).$$

Now, we can collect the three terms

$$\boxed{\begin{aligned} L^2 &= ① + ② + ③ = 2\,\mathrm{Re}\left[G_{ij}[\theta(\tau)]\right]\dot{\theta}_i\dot{\theta}_j \\ &+ 4\,\mathrm{Re}\left(\langle \frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle\right)\dot{\theta}_i + 2\left(\langle H^2\rangle - \langle H\rangle^2\right) \end{aligned}}$$

Extra calculation for step (2) above:

(i)

$$\mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^{\dagger} H \rho\right] = \mathrm{Tr}\left[\left(\left|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle\psi| + |\psi\rangle\left\langle\frac{\partial \psi}{\partial \theta_i}\right|\right) H|\psi\rangle\langle\psi|\right]$$

$$= \left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle H\rangle + \left\langle\frac{\partial \psi}{\partial \theta_i}\middle|H\middle|\psi\right\rangle .$$

(ii)

$$\mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^{\dagger} \rho H\right] = \mathrm{Tr}\left[\left(\left|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle\psi| + |\psi\rangle\left\langle\frac{\partial \psi}{\partial \theta_i}\right|\right)|\psi\rangle\langle\psi| H\right] =$$

$$= \left\langle\psi\middle|H\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle + \underbrace{\left\langle\frac{\partial \psi}{\partial \theta_i}\middle|\psi\right\rangle\langle H\rangle}_{= -\left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle H\rangle}$$

(iii)

$$\mathrm{Tr}\left[H\rho \underbrace{\frac{\partial \rho}{\partial \theta_i}}_{\left|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle\psi| + |\psi\rangle\left\langle\frac{\partial \psi}{\partial \theta_i}\right|}\right] = \langle H\rangle\left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle + \left\langle\frac{\partial \psi}{\partial \theta_i}\middle|H\middle|\psi\right\rangle$$

(iv)

$$\mathrm{Tr}\left[\rho H \frac{\partial \rho}{\partial \theta_i}\right] = \left\langle\psi\middle|H\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle + \langle H\rangle\underbrace{\left\langle\frac{\partial \psi}{\partial \theta_i}\middle|\psi\right\rangle}_{= -\left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle}$$

It is now straightforward to derive an equation of motion for the variational parameters $\theta(\tau)$ that minimizes L^2 using the variational principle:

$$L^2 = 2\,\mathrm{Re}[G_{ij}]\,\dot\theta_i\dot\theta_j + 4\,\underbrace{\mathrm{Re}\left(\left\langle \frac{\partial\psi}{\partial\theta_i}\middle| H \middle|\psi\right\rangle\right)}_{=\frac{1}{2}\frac{\partial}{\partial\theta_i}\langle\psi|H|\psi\rangle}\dot\theta_i + 2\left(\langle H^2\rangle - \langle H\rangle^2\right)$$

$$\delta L^2 = 0 \Rightarrow \qquad \frac{\delta L^2}{\delta\dot\theta_i} = 0 \quad \forall i$$

$$\Leftrightarrow 4\,\mathrm{Re}[G_{ij}]\,\dot\theta_j + 4\,\mathrm{Re}\left(\left\langle\frac{\partial\psi}{\partial\theta_i}\middle|H\middle|\psi\right\rangle\right) = 0$$

$$\Leftrightarrow g_{ij}\,\dot\theta_j = -\frac{\partial}{\partial\theta_i}\underbrace{\frac{1}{2}\langle\psi|H|\psi\rangle}_{=C(\theta)\text{ from earlier}} = -\frac{\partial}{\partial\theta_i}C(\theta)$$

$$\Rightarrow \boxed{g[\theta(\tau)]\,\dot\theta(\tau) = -\vec\nabla C[\theta(\tau)]}$$

Same equation as we had derived before starting from $\underset{\delta\theta \in \mathbb{R}^{N_\theta}}{\operatorname{argmin}} \| |\Psi_{\bar{\theta}}\rangle - |\Psi_{\theta+\delta\theta}\rangle\langle\Psi_{\theta+\delta\theta}|\Psi_{\bar{\theta}}\rangle \|^2 .$

Now that we have a classical EOM for the variational parameters, what do we need to address:

① • Measure $g_{ij}[\theta]$ and $\mathrm{Re}\left(\langle \frac{\partial\Psi}{\partial\theta_i} | H | \Psi\rangle\right)$ on quantum computer

Can be done using Hadamard test circuits & direct measurements.

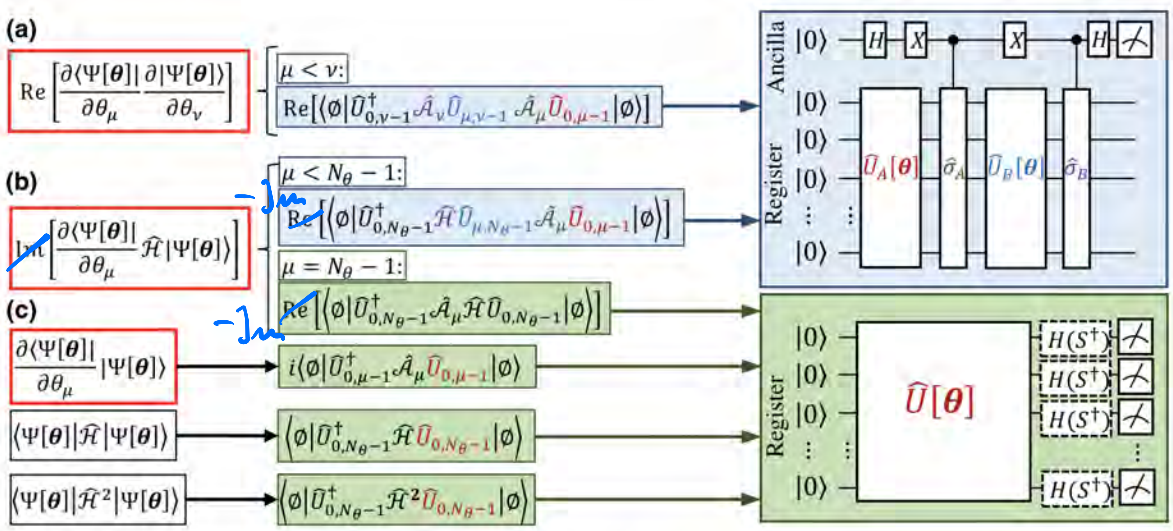


FIG. 2. Quantum-circuit implementation of the AVQDS algorithm. The left column lists the unique terms to be evaluated in Eqs. (5) and (7) of VQDS, with the terms (a)–(c) highlighted in red also involved in the ansatz adaptive procedure in AVQDS. The middle column specifies the expressions when the wave-function ansatz takes the pseudo-Trotter form of Eq. (10) with $\hat{U}_{j,k}[\theta] = \prod_{\mu=j}^{k} e^{-i\theta_\mu\hat{\mathcal{A}}_\mu}$ and $|\Psi_0\rangle = |\emptyset\rangle \equiv \otimes_{j=0}^{N-1}|0\rangle$ for an N-qubit system. Two types of quantum circuits are adopted: a green block for the direct measurement circuit, and a blue block for a generalized Hadamard test circuit [27,45]. The direct measurement circuit includes optional Hadamard gate H or Hadamard-phase gate $HS^\dagger$ when measuring X or Y-Pauli strings present in $\hat{\mathcal{A}}_\mu$, $\hat{\mathcal{A}}^2_\mu$, $\hat{\mathcal{H}}$, and $\hat{\mathcal{H}}^2$. Accord-

Originally for real-time evolution, where $\mathrm{Im}\langle\frac{\partial\Psi}{\partial\theta_i}|H|\Psi\rangle$ occurs.

(2) How to optimally distribute measurements (shots) across different circuits?

- Kozcov: minimize $\mathrm{var}(\|\Delta\vec{\theta}\|_2)$
- alternatives, e.g., minimize variance of MacLachlan distance $\mathrm{var}(L)$.

(3) How to select the ansatz?

- fixed ansatz
- adaptively expanded ansatz

Adaptive ansatz generation: flexible and shown to produce shallow circuits with near optimal scaling of N_θ and N_{CNOT} with system size, when optimal scaling is known

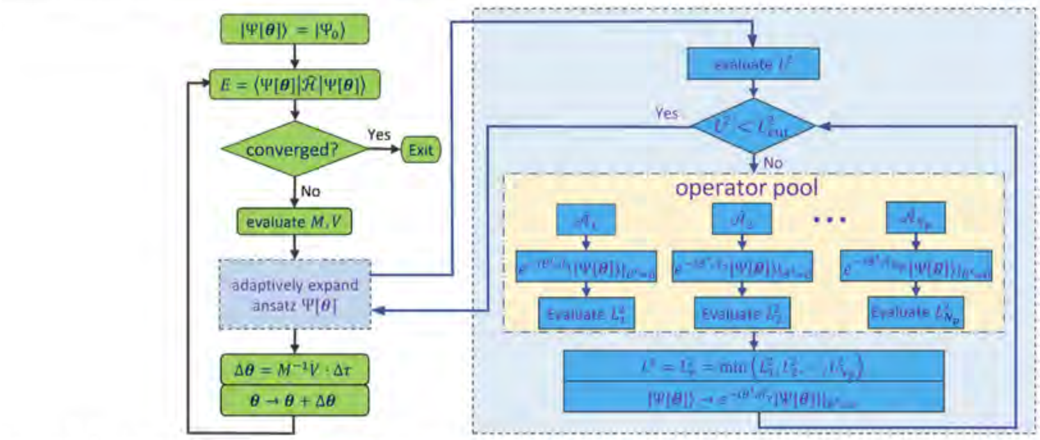


Figure 1. Schematic illustration of variational quantum imaginary time evolution algorithm, with an additional module to adaptively expand the ansatz. The green flowchart on the left shows a typical VQITE calculation. In AVQITE, a module (blue) is introduced to adaptively expand the variational ansatz by selectively appending parametric rotation gates to keep the McLachlan distance L^2 under a threshold L^2_{cut} along the imaginary-time evolution path.

4 How to best deal with inversion of g_{ij}, which is often singular in practice (large condition number)?

- directly solve linear system of equations

$$g\,\dot{\theta} + \vec{\nabla} C(\theta) = 0$$

e.g. using

numpy.linalg.lstsq

`linalg.lstsq(a, b, rcond='warn')` [source]

Return the least-squares solution to a linear matrix equation.

Computes the vector x that approximately solves the equation `a @ x = b`. The equation may be under-, well-, or over-determined (i.e., the number of linearly independent rows of a can be less than, equal to, or greater than its number of linearly independent columns). If a is square and of full rank, then x (but for round-off error) is the "exact" solution of the equation. Else, x minimizes the Euclidean 2-norm $||b - ax||$. If there are multiple minimizing solutions, the one with the smallest 2-norm $||x||$ is returned.

Parameters: **a** : *(M, N) array_like*

"Coefficient" matrix.

b : *{(M,), (M, K)} array_like*

Ordinate or "dependent variable" values. If b is two-dimensional, the least-squares solution is calculated for each of the K columns of b.

rcond : *float, optional*

Cut-off ratio for small singular values of a. For the purposes of rank determination, singular values are treated as zero if they are smaller than *rcond* times the largest singular value of a.

- use Tikhonov regularization

$$g \longrightarrow g + \xi \mathbb{1} \Rightarrow \tilde{g}^{-1} = [g + \xi \mathbb{1}]^{-1}$$

$\xi \ll 1$ (e.g. typically $10^{-1} \leq \xi \leq 10^{-6}$).

Real-time evolution:

One can derive similar classical EOM for $\theta(t)$ from McLachlan's principle applied to real-time dynamics. Here, the exact state evolution is given by the von-Neumann eq.:

$$\frac{\partial \rho(t)}{\partial t} = -i\left[H, \rho(t)\right]$$

Applying the variational principle to minimize the McLachlan distance L^2 between variational & exact time evolution yields:

$$\delta L^2 = 0$$

$$\Rightarrow \delta \left\| \frac{\partial \rho[\theta(t)]}{\partial \theta_i} \dot{\theta}_i + i \underbrace{\left[H, \rho[\theta(t)]\right]}_{H\rho - \rho H} \right\|^2 = 0$$

Rewriting L^2 as:

$$L^2 = \mathrm{Tr}\Big\{\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^\dagger\dot\theta_i - i(\rho H - H\rho)\Big]$$

$\updownarrow$ note!

$$\Big[\Big(\frac{\partial\rho}{\partial\theta_j}\Big)\dot\theta_j + i(H\rho - \rho H)\Big]\Big\}$$

$$= \mathrm{Tr}\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^\dagger\Big(\frac{\partial\rho}{\partial\theta_j}\Big)\Big]\dot\theta_i\,\dot\theta_j$$

$$+ i\,\mathrm{Tr}\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^\dagger(H\rho - \rho H)\Big]\dot\theta_i$$

$$- i\,\mathrm{Tr}\Big[(\rho H - H\rho)\Big(\frac{\partial\rho}{\partial\theta_i}\Big)\Big]\dot\theta_i$$

$$+ \mathrm{Tr}\big[(H\rho - \rho H)(H\rho - \rho H)\big]$$

Focusing on pure states: $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$

$$= \Big|\frac{\partial\psi}{\partial\theta_i}\Big\rangle\langle\psi| + |\psi\rangle\Big\langle\frac{\partial\psi}{\partial\theta_i}\Big|$$

(1) $$\mathrm{Tr}\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^\dagger\Big(\frac{\partial\rho}{\partial\theta_j}\Big)\Big] =$$

$$= 2\,\mathrm{Re}[G_{ij}(\theta)] \quad \text{as before}$$

(2)

$$\mathrm{Tr}\Big[\Big(\overbrace{\frac{\partial\rho}{\partial\theta_i}}^{=|\frac{\partial\psi}{\partial\theta_i}\rangle\langle\psi| + |\psi\rangle\langle\frac{\partial\psi}{\partial\theta_i}|}\Big)^{\dagger}(H\rho-\rho H)\Big]$$

$$-\mathrm{Tr}\Big[(\rho H-H\rho)\Big(\frac{\partial\rho}{\partial\theta_i}\Big)\Big] =$$

$$= \langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle\langle H\rangle + \langle\frac{\partial\psi}{\theta_i}|H|\psi\rangle$$

$$-\langle\psi|H|\frac{\partial\psi}{\partial\theta_i}\rangle - \langle\frac{\partial\psi}{\partial\theta_i}|\psi\rangle\langle H\rangle$$

$$+\Big(\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle\langle H\rangle + \langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle$$

$$-\langle\psi|H|\frac{\partial\psi}{\partial\theta_i}\rangle - \langle\frac{\partial\psi}{\partial\theta_i}|\psi\rangle\langle H\rangle\Big) =$$

$$= 2\langle H\rangle\Big[\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle - \underbrace{\langle\frac{\partial\psi}{\partial\theta_i}|\psi\rangle}_{=-\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle}\Big]$$

$$+2\Big[\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle - \langle\psi|H|\frac{\partial\psi}{\partial\theta_i}\rangle\Big] =$$

$$= 4\langle H\rangle\underbrace{\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle}_{=-iA_i} + 4i\,\mathrm{Im}\Big[\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle\Big]$$

③ irrelevant for EOM

We thus find

$$L^2 = 2\,\mathrm{Re}\left[G_{ij}(\theta)\right]\dot\theta_i\,\dot\theta_j + \left(-4\,\mathrm{Im}\left[\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle\right] - 4\langle H\rangle\,\mathrm{Im}\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle\right)\dot\theta$$

$$+\ \text{const.}$$

$\Rightarrow$ $\delta L^2 = 0$ yields EOM for $\theta(t)$:

$$\Downarrow$$

$$4\,\underbrace{\mathrm{Re}\left[G_{ij}\right]}_{g_{ij}}\dot\theta_j = 4\,\underbrace{\mathrm{Im}\left[\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle + \langle H\rangle\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle\right]}_{=V_i}$$

$$\Rightarrow\ \boxed{g_{ij}[\theta(t)]\,\dot\theta_j(t) = V_i}$$

solve EOM on classical computer

measure on QC