Quantum Natural Gradient Descent

Luca Previdi

Vanilla GD[1]

GOAL

Minimize a function f over a set of parameters $\boldsymbol{\theta}$

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \eta \nabla f(\boldsymbol{\theta}^{(k)})$$

VGD step

$$\equiv$$

$$\boldsymbol{\theta}^{(k+1)} = \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d} \left(\langle \boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}, \nabla f(\boldsymbol{\theta}^{(k)}) \rangle + \frac{1}{2\eta} \boxed{||\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}||_2^2} \right)$$

Perform derivatives

$$\mathbf{0} = \nabla f(\boldsymbol{\theta}^{(k)}) + \frac{1}{\eta}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})$$

Assume to have a quantum circuit which evaluates f

Evaluate the gradient through the **shift rule**

$$\partial_i f(\boldsymbol{\theta}) = \frac{f(\theta_i + \varepsilon) - f(\theta_i - \varepsilon)}{2\varepsilon}$$

[1] Kerenidis, Iordanis, and Anupam Prakash. "Quantum gradient descent for linear systems and least squares." *Physical Review A* 101.2 (2020): 022316.

Quantum Information Geometry[2]

Distance between probabilities

$$d(p,q) = \arccos(\langle \sqrt{p}, \sqrt{q} \rangle)$$

$$d^2(p_\theta, p_{\theta+\mathrm{d}\theta}) = \frac{1}{4} \sum_{(i,j)\in[d]^2} I_{ij}(\theta)\mathrm{d}\theta^i \mathrm{d}\theta^j$$

Fisher Information Matrix

$$I_{ij}(\theta) = \sum_{x\in[N]} p_\theta(x) \frac{\partial \log p_\theta(x)}{\partial \theta^i} \frac{\partial \log p_\theta(x)}{\partial \theta^j}$$

Distance between pure states

$$d(P_\psi, P_\phi) = \arccos(|\langle \psi, \phi \rangle|)$$

$$d^2(P_{\psi_\theta}, P_{\psi_{\theta+\mathrm{d}\theta}}) = \sum_{(i,j)\in[d]^2} g_{ij}(\theta)\,\mathrm{d}\theta^i \mathrm{d}\theta^j$$

Quantum Geometric Tensor

$$g_{ij}(\theta) = \mathrm{Re}[G_{ij}(\theta)]$$

$$G_{ij}(\theta) = \left\langle \frac{\partial \psi_\theta}{\partial \theta^i}, \frac{\partial \psi_\theta}{\partial \theta^j} \right\rangle - \left\langle \frac{\partial \psi_\theta}{\partial \theta^i}, \psi_\theta \right\rangle \left\langle \psi_\theta, \frac{\partial \psi_\theta}{\partial \theta^j} \right\rangle$$

[2] Helstrom, Carl W. "Minimum mean-squared error of estimates in quantum statistics." *Physics letters A* 25.2 (1967): 101-102.

Quantum Natural Gradient[3]

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}\in\mathbb{R}^d}{\arg\min}\left(\langle \boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}, \nabla f(\boldsymbol{\theta}^{(k)})\rangle + \frac{1}{2\eta}||\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}||_2^2\right)$$

$$\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}\in\mathbb{R}^d}{\arg\min}\left(\langle \boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}, \nabla f(\boldsymbol{\theta}^{(k)})\rangle + \frac{1}{2\eta}||\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}||_{g(\boldsymbol{\theta}^{(k)})}^2\right)$$

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \eta g^{-1}(\boldsymbol{\theta}^{(k)})\nabla f(\boldsymbol{\theta}^{(k)})$$

QNG step

Natural norm

$$\|\boldsymbol{\theta}\|_{g(\boldsymbol{\theta})}^2 = \langle \boldsymbol{\theta}, g(\boldsymbol{\theta})\boldsymbol{\theta}\rangle$$

$$g^+ = g^T(gg^T)^{-1}$$

Pseudoinverse

[3] Stokes, James, et al. "Quantum natural gradient." *Quantum* 4 (2020): 269.

Quantum Fisher Information Matrix[4]

$\rho_{\boldsymbol{\lambda}}$ Set of quantum states labelled by continuos parameters

$$p_{\boldsymbol{\lambda}}(x) = \mathrm{Tr}\left[\rho_{\boldsymbol{\lambda}} \Pi_x\right]$$ Born rule

Fisher Information

$$I_{\mu\nu} = \sum_x p_{\boldsymbol{\lambda}}(x) \partial_\mu \log p_{\boldsymbol{\lambda}}(x) \partial_\nu \log p_{\boldsymbol{\lambda}}(x)$$

Quantum Fisher Information Matrix

$$Q_{\mu\nu}(\boldsymbol{\lambda}) = \mathrm{Tr}\left[\rho_{\boldsymbol{\lambda}} \frac{L_\mu L_\nu + L_\nu L_\mu}{2}\right]$$

$$\partial_\mu \rho_{\boldsymbol{\lambda}} = \frac{L_\mu \rho_{\boldsymbol{\lambda}} + \rho_{\boldsymbol{\lambda}} L_\mu}{2}$$

Quantum Cramer-Rao Bound

$$N\, V(\hat{\boldsymbol{\lambda}}) \succeq I\left(\boldsymbol{\lambda}, \{\Pi_x\}\right)^{-1} \succeq Q(\boldsymbol{\lambda})^{-1}$$

[4] Albarelli, Francesco, et al. "A perspective on multiparameter quantum metrology: From theoretical tools to applications in quantum imaging." *Physics Letters A* 384.12 (2020): 126311.

Quantum Fisher Information Matrix

$$N\,V(\hat{\boldsymbol{\lambda}}) \succeq I(\boldsymbol{\lambda}, \{\Pi_x\})^{-1} \succeq Q(\boldsymbol{\lambda})^{-1}$$

$$\lim_{N\to\infty} NV(\boldsymbol{\lambda}) = I(\boldsymbol{\lambda})^{-1}$$

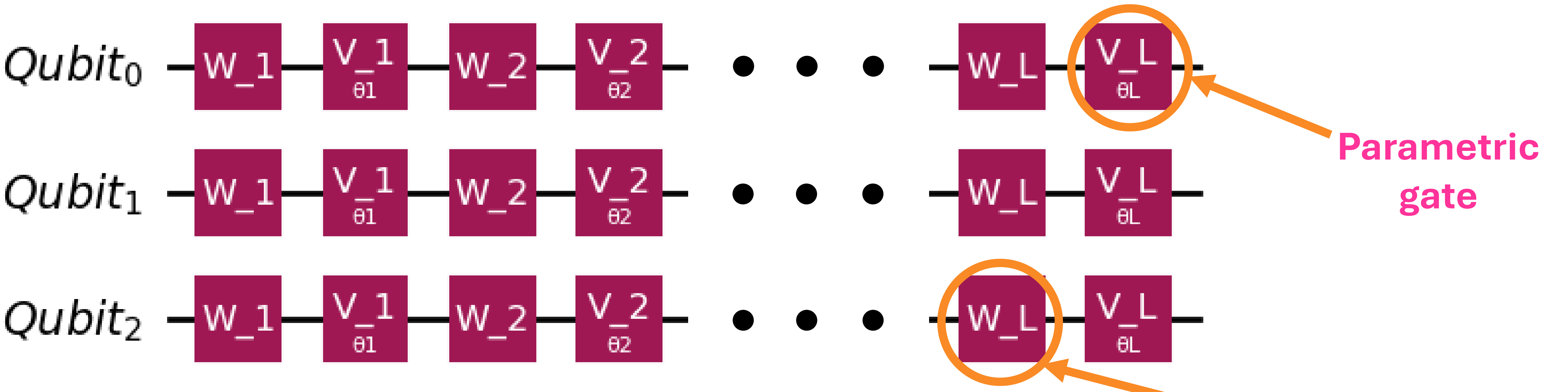
? ? ?

BUT

$$\rho_{\boldsymbol{\lambda}} = |\psi_{\boldsymbol{\lambda}}\rangle\langle\psi_{\boldsymbol{\lambda}}|$$

$Q(\boldsymbol{\lambda}) = \max_{\Pi_x} I(\boldsymbol{\lambda}, \Pi_x)$ **iff** $\langle\psi_{\boldsymbol{\lambda}}|L_\mu L_\nu|\psi_{\boldsymbol{\lambda}}\rangle = 0$

Approximate QFIM Evaluation



$U_L(\theta) := V_L(\boldsymbol{\theta}_L)W_L \cdots V_1(\boldsymbol{\theta}_1)W_1$ Total evolution

$U_{[l_1:l_2]} := V_{l_2}W_{l_2} \cdots V_{l_1}W_{l_1}$ Layers evolution

$\psi_l := U_{[1:l]}|0\rangle$ State after the l-th layer

Approximate QFIM Evaluation

$$\begin{array}{c|cccc} & \boldsymbol{\theta}_1 & \boldsymbol{\theta}_2 & \cdots & \boldsymbol{\theta}_L \\ \hline \boldsymbol{\theta}_1 & \boxed{G^{(1)}} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\theta}_2 & \mathbf{0} & \boxed{G^{(2)}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\theta}_L & \mathbf{0} & \mathbf{0} & \cdots & \boxed{G^{(L)}} \end{array}$$

Block diagonal approximation

Ignore correlations between different layers

$$\partial_i V_l(\boldsymbol{\theta}_l) = -\mathrm{i} K_i V_l(\boldsymbol{\theta}_l)$$ Generators

$$\partial_i K_j = 0 \quad i \neq j \quad \Rightarrow \quad [K_i, K_j] = 0$$

Commuting generators

$$\boxed{g_{ij}^{(l)} = \mathrm{Re}[G_{ij}^{(l)}] = G_{ij}^{(l)} = \langle \psi_l | K_i K_j | \psi_l \rangle - \langle \psi_l | K_i | \psi_l \rangle \langle \psi_l | K_j | \psi_l \rangle}$$

Quantum Imaginary Time Evolution[5]

$$\frac{\partial \left|\psi(t)\right\rangle}{\partial t} = -(\hat{\mathcal{H}} - E_t)\left|\psi(t)\right\rangle$$ Wick rotated Schrödinger equation

$$\Rightarrow \quad \left|\psi(t)\right\rangle = \frac{e^{-\hat{\mathcal{H}}t}}{\sqrt{\left\langle\psi(0)\right| e^{-\hat{\mathcal{H}}t}\left|\psi(0)\right\rangle}}\left|\psi(0)\right\rangle \quad \Rightarrow \quad t \to \infty$$

$\left|\psi(t)\right\rangle$ is the ground state

McLachlan's variational principle[6]

$$\delta \left\| \frac{\partial \left|\psi(\boldsymbol{\theta}^{(t)})\right\rangle}{\partial t} + (\hat{\mathcal{H}} - E_\tau)\left|\psi(\boldsymbol{\theta}^{(t)})\right\rangle \right\|_{\ell^2}^2 = 0 \xrightarrow[\text{Derivatives}]{} g_{ij}(\boldsymbol{\theta}^{(t)})\frac{\partial \theta_j^{(t)}}{\partial t} = -\mathrm{Re}\left\{\left\langle \frac{\partial \psi(\boldsymbol{\theta}^{(t)})}{\partial \theta_i} \middle| \hat{\mathcal{H}}\left|\psi(\boldsymbol{\theta}^{(t)})\right\rangle \right\rangle\right\}$$

Same as QNG!

With: $f(\boldsymbol{\theta}) = -\frac{1}{2}\left\langle\psi(\boldsymbol{\theta})\middle|\hat{\mathcal{H}}\middle|\psi(\boldsymbol{\theta})\right\rangle$

[5] McArdle, Sam, et al. "Variational ansatz-based quantum simulation of imaginary time evolution." *Quantum Information* (2019): 75.
[6] McLachlan, Andrew D. "A variational solution of the time-dependent Schrodinger equation." *Molecular Physics* (1964): 39-44.

OH Molecule

$$H = H_0 - \vec{\mu}_e \cdot \vec{E} - \vec{\mu}_b \cdot \vec{B}$$

Lara's basis

$|J, M, \bar{\Omega}, \varepsilon\rangle$ $J = 3/2$ total angular momentum M projection on x-axis

$\bar{\Omega}$ projection on internuclear axis $\varepsilon = \{e, f\}$ symmetry

$|\vec{B}| > 100\ \mathrm{G}$

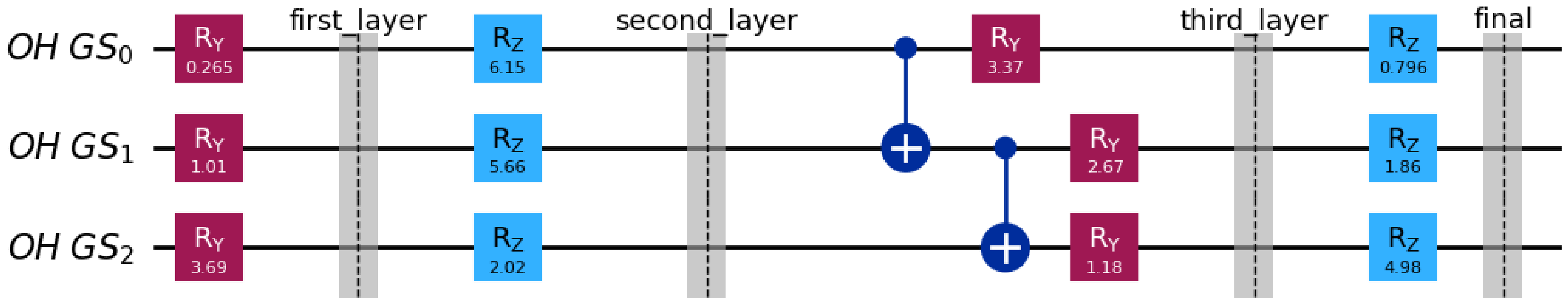
$|\vec{E}| > 1\ \mathrm{kV/cm}$ or $T < 5\mathrm{mK}$

$$\boxed{H = -\tilde{\Delta}\, T_{300} - \frac{4}{5}\tilde{B}(2\, T_{030} + T_{003}) + \frac{2}{5}\tilde{E}_z(2\, T_{130} + T_{103}) - \frac{2}{5}\tilde{E}_y(\sqrt{3}\, T_{101} + T_{111} + T_{122})}$$

with: $T_{ijk} = \sigma_i \otimes \sigma_j \otimes \sigma_k$

[7] Lara, Manuel, Benjamin L. Lev, and John L. Bohn. "Loss of molecules in magneto-electrostatic traps due to nonadiabatic transitions." *Physical Review A—Atomic, Molecular, and Optical Physics* 78.3 (2008): 033433.

Encoding

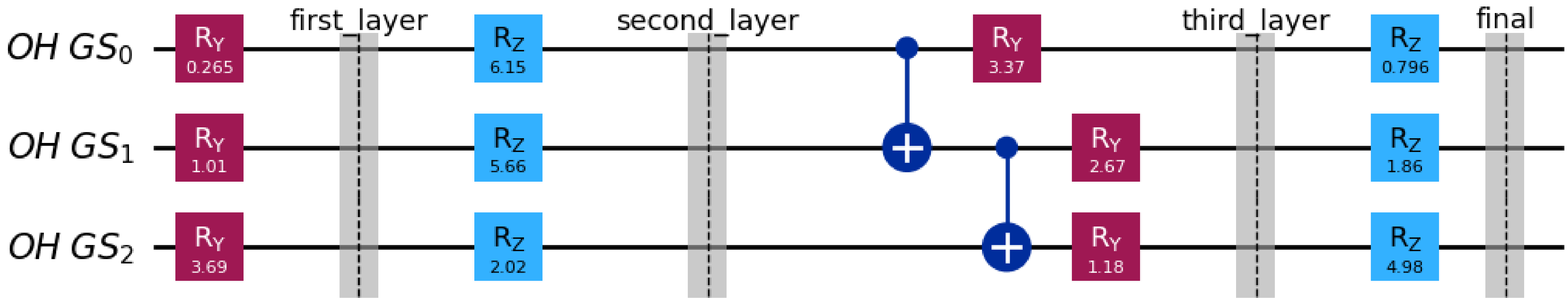


$$V_l(\boldsymbol{\theta}_l) = \bigotimes_{k=1}^{n} R_{P_{l,k}}(\boldsymbol{\theta}_{l,k}) \quad \text{with} \quad R_{P_{l,k}}(\boldsymbol{\theta}_{l,k}) = \exp\left[-\mathrm{i}\frac{\boldsymbol{\theta}_{l,k}}{2} P_{l,k}\right]$$

$$\boldsymbol{\theta}_{l,k} \in [0, 2\pi) \qquad P_{l,k} \in \{\sigma_x, \sigma_y, \sigma_z\}$$

Parametric gates are **Pauli Rotations**

Encoding



Generators are Pauli matrices $K_i = \frac{1}{2}\mathbb{1}^{[1,i)} \otimes P_{l,i} \otimes \mathbb{1}^{(i,n]}$ ⇨ $[K_i, K_j] = 0$

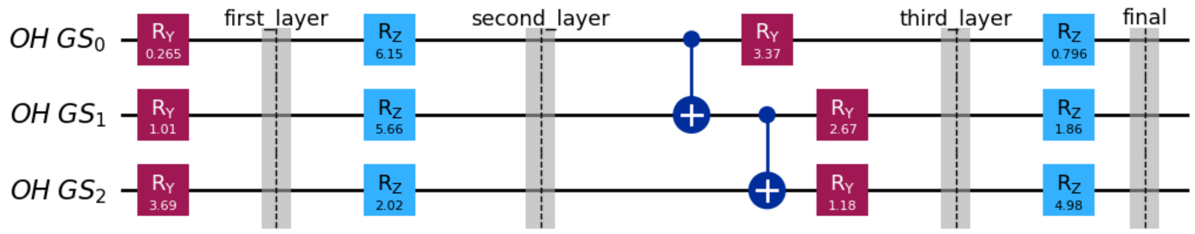
Need to evaluate $\langle \psi_l | \hat{A} | \psi_l \rangle$ in $S_l = \{P_{l,i} \mid 1 \leq i \leq n\} \cup \{P_{l,i}P_{l,j} \mid 1 \leq i < j \leq n\}$

All the operators in S_l commute ⇨ Just need 1 measurement

Qiskit Implementation

Random initial values of the parameters

Parameters encoding



QFIM evaluation

$$\langle\psi_l|K_iK_j|\psi_l\rangle - \langle\psi_l|K_i|\psi_l\rangle\langle\psi_l|K_j|\psi_l\rangle$$

Gradient evaluation

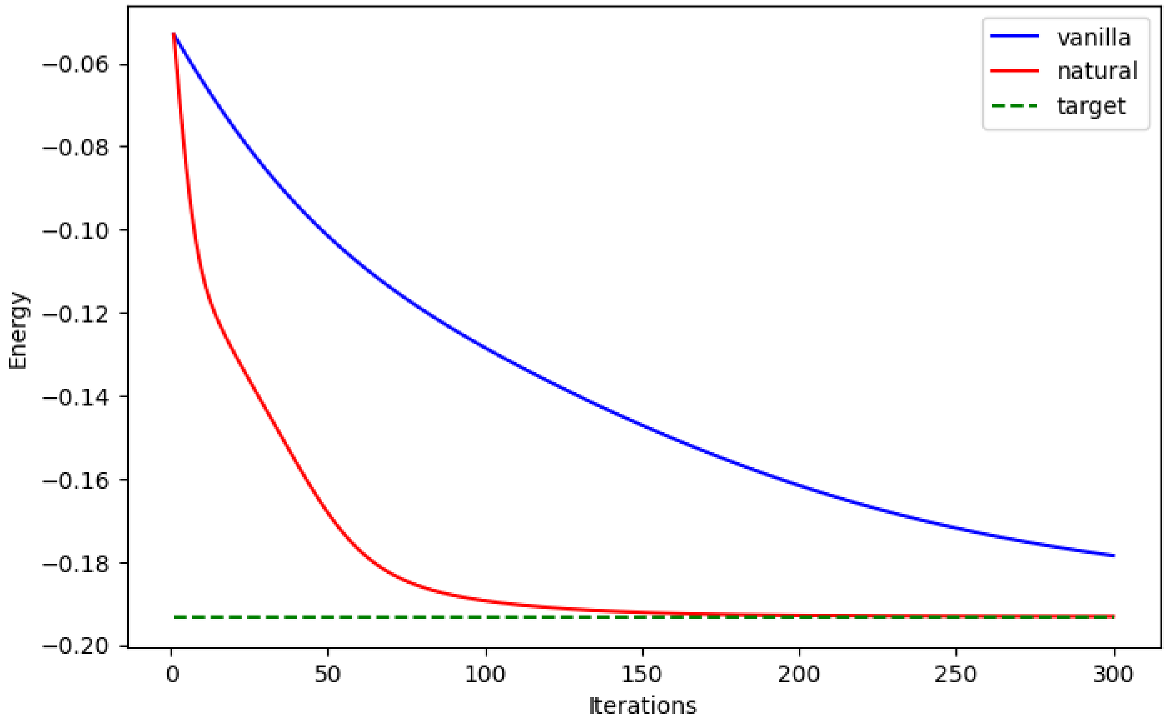
$$\partial_i f(\boldsymbol{\theta}) = \frac{f(\theta_i + \varepsilon) - f(\theta_i - \varepsilon)}{2\varepsilon}$$

Update values

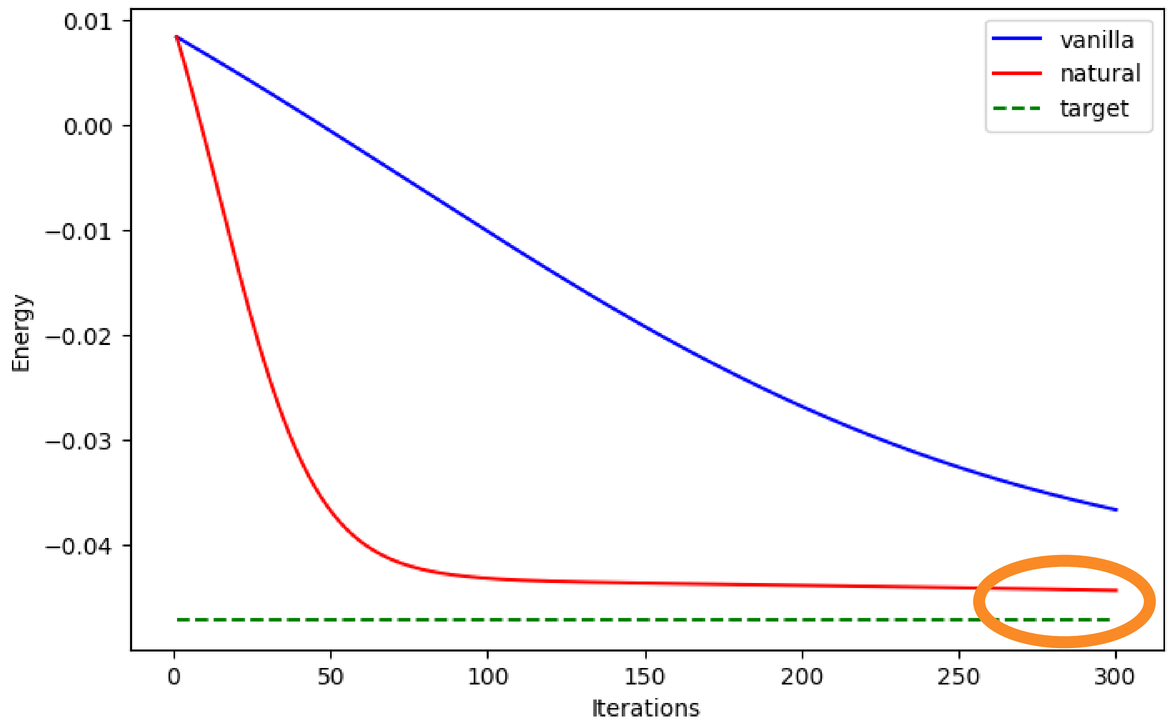
$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \eta g^{-1}(\boldsymbol{\theta}^{(k)})\boldsymbol{\nabla} f(\boldsymbol{\theta}^{(k)})$$

Results

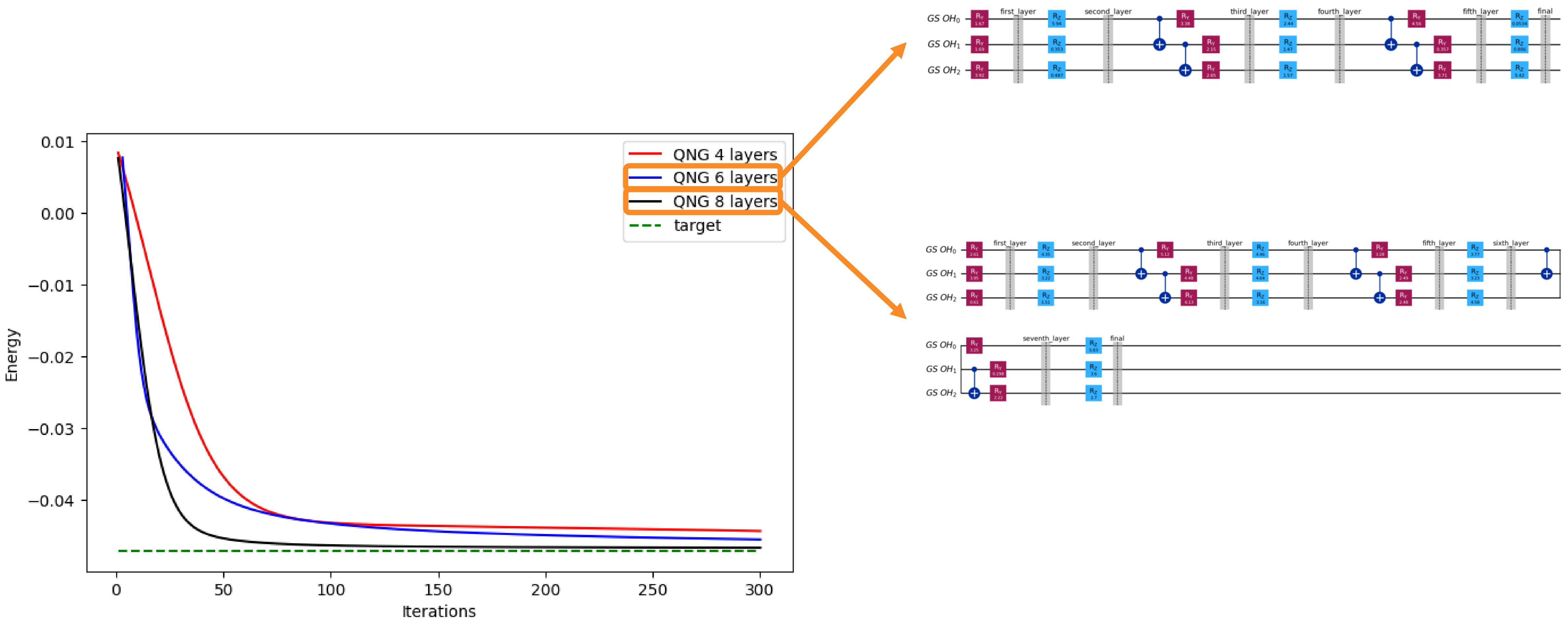
B = 0.1 E = 5e5 θ = π/4



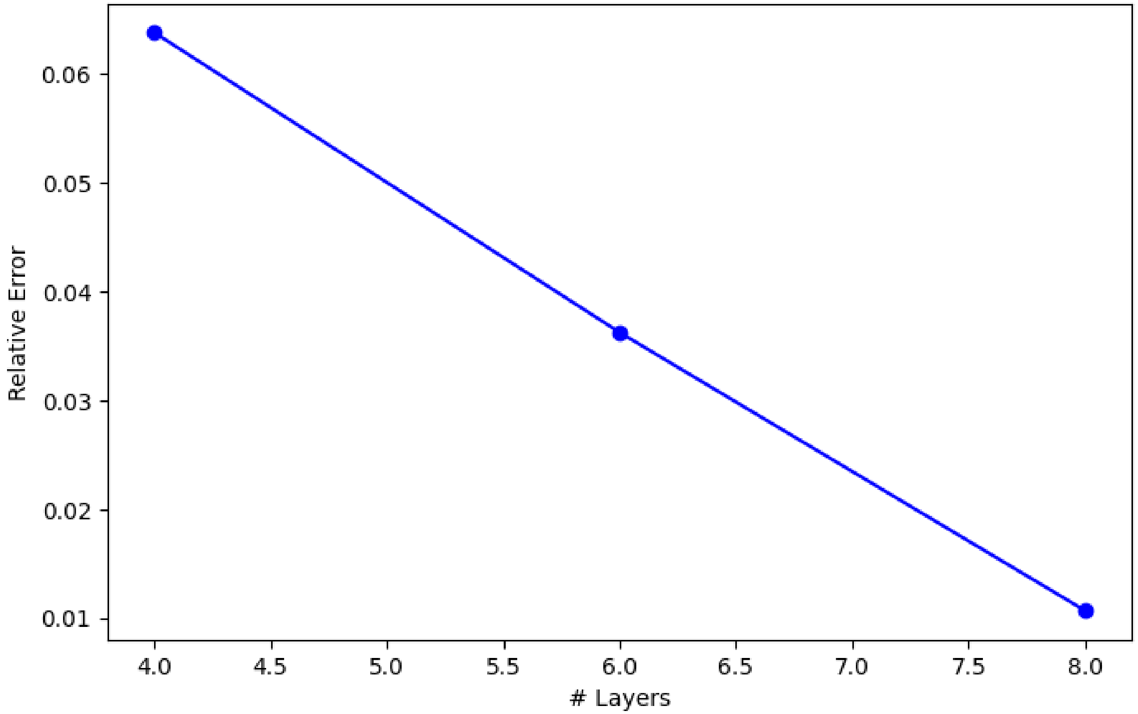
B = 0.01 E = 1e5 θ = 2π/3



Results


QNG 4 layers
QNG 6 layers
QNG 8 layers
target
Energy
Iterations
0.01
0.00
−0.01
−0.02
−0.03
−0.04
0
50
100
150
200
250
300
GS OH0
GS OH1
GS OH2
first_layer
second_layer
third_layer
fourth_layer
fifth_layer
sixth_layer
seventh_layer
final

Results



Relative error seems to scale linearly with the number of layers

But requires furher investigations...

Conclusions

We introduced the Quantum Gemetric Tensor and related it to the Covariance Matrix

We showed that the Quantum Natural Gradient performs better than the standard Vanilla Gradient

We showed that the convergence can be ensured through adding more layers

We pointed out a particular behavior of the relative error wrt the number of layers (in a particular regime)

Thanks for listening

References

[1] Kerenidis, Iordanis, and Anupam Prakash. "Quantum gradient descent for linear systems and least squares." *Physical Review A* 101.2 (2020): 022316.

[2] Helstrom, Carl W. "Minimum mean-squared error of estimates in quantum statistics." *Physics letters A* 25.2 (1967): 101-102.

[3] Stokes, James, et al. "Quantum natural gradient." *Quantum* 4 (2020): 269.

[4] Albarelli, Francesco, et al. "A perspective on multiparameter quantum metrology: From theoretical tools to applications in quantum imaging." *Physics Letters A* 384.12 (2020): 126311.

[5] McArdle, Sam, et al. "Variational ansatz-based quantum simulation of imaginary time evolution." *Quantum Information* (2019): 75.

[6] McLachlan, Andrew D. "A variational solution of the time-dependent Schrodinger equation." *Molecular Physics* (1964): 39-44.

[7] Lara, Manuel, Benjamin L. Lev, and John L. Bohn. "Loss of molecules in magneto-electrostatic traps due to nonadiabatic transitions." *Physical Review A—Atomic, Molecular, and Optical Physics* 78.3 (2008): 033433.