LOCALIZING SLD GEOMETRY FOR SCALABLE QUANTUM NATURAL GRADIENT UNDER CIRCUIT NOISE

Anonymous authors
Paper under double-blind review

ABSTRACT

Variational quantum circuits generally produce mixed states in the presence of circuit noise. Quantum natural gradient (QNG) accounts for this mixed-state geometry using the symmetric-logarithmic-derivative (SLD) quantum Fisher information matrix, whose direct estimation can exceed practical circuit-shot budgets. Tractable QNG approximations developed for pure states do not directly provide this geometry, while mixed-state approaches often use more accessible surrogates, such as the Hilbert–Schmidt metric. We propose an approach to approximate the SLD geometry using local reductions of the noisy state. We define graph-local physical windows of radius r and restrict the local SLD score space within each window to centered Pauli strings of weight at most k_S. We assemble the local SLD metrics across fixed parameter blocks to obtain a block-local approximation. Increasing r or k_S yields nested approximation hierarchies, and the complete centered Pauli score space recovers the exact reduced-state SLD metric when no score-space regularization is used. For finite-shot implementation, we estimate the score-Gram and response matrices using local Pauli and parameter-shift measurements without reconstructing the reduced states. We derive finite-shot metric-estimation error bounds and show that, under fixed-locality assumptions, a sufficient metric-estimation cost scales as $O(n \log n)$ for fixed accuracy and confidence. We also establish convergence of the regularized finite-shot update to a stationary neighborhood of the noisy objective under standard stochastic-optimization assumptions. Experiments on noisy variational quantum eigensolver (VQE) instances for the Heisenberg model show improved convergence per cumulative shot compared with SGD, Adam, and approximate mixed-state QNG baselines under matched finite-shot budgets.

1 INTRODUCTION

Variational quantum algorithms (VQAs) combine parameterized quantum circuits with classical optimization to solve problems in quantum simulation, optimization, and machine learning (Cerezo et al., 2021; Bharti et al., 2022). Natural gradient based optimization accounts for the geometry of a parameterized model rather than treating its parameters as Euclidean coordinates (Amari, 1998). Quantum natural gradient (QNG) applies this idea to families of quantum states (Stokes et al., 2020). For ideal unitary circuits, this geometry is described by the Fubini–Study metric. Noisy circuits generally produce mixed states, for which the symmetric-logarithmic-derivative (SLD) quantum Fisher geometry is used (Braunstein & Caves, 1994; Petz & Sudár, 1996). Estimating the full SLD quantum Fisher information matrix (QFIM), however, may require substantially more circuit measurements than estimating the gradient. Current QNG methods either focus on pure-state geometry or replace the mixed-state SLD metric with a more accessible surrogate. In this paper, we take a different approach that restricts the SLD variational problem itself to local physical windows and score spaces. This yields local approximations to the mixed-state SLD geometry that can be estimated from local Pauli measurements.

We localize the SLD variational problem to radius-r windows in the circuit interaction graph and, within each window, restrict the local score space to centered Pauli strings of weight at most k_S. Thus, r controls the physical scope, while k_S controls the local score space; increasing either gives a nested metric hierarchy. We characterize the localization error through an exact SLD-score residual and show that the complete centered Pauli score space recovers the exact reduced-state SLD metric

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

before regularization. We then restrict each local metric to a fixed parameter block and assemble the resulting blocks into a block-diagonal preconditioner. The parameter partition determines the metric couplings that are retained.

For finite-shot implementation, we estimate each local metric from local Pauli and parameter-shift measurements without reconstructing the reduced state. We derive concentration and stability bounds for the resulting regularized estimator. When r, k_S, and the score-space regularization are fixed and the physical windows and parameter blocks are uniformly bounded with $n_B = O(n)$, a sufficient metric-estimation shot cost is $O\left(\frac{n}{\varepsilon^2}\log\frac{n}{\delta}\right)$ for operator-norm error at most ε with probability at least $1-\delta$. For dense interaction graphs, where fixed-radius windows may grow with the system size, this bound no longer applies, although fixed k_S still restricts the score-space to bounded weight Pauli strings. We also establish convergence of the regularized finite-shot update to a stationary neighborhood of the noisy objective under standard stochastic-optimization assumptions.

Numerically, we compare the block-local approximation with the full SLD-QFIM on a smaller noisy Heisenberg VQE instance where the full metric is tractable, and study finite-shot optimization on a larger instance. At the higher tested noise strengths, the block-local metric is closer to the full SLD-QFIM than the tested mixed-state surrogates. Under matched cumulative shot budgets, block-local SLD-QNG achieves the lowest mean energy over most of the shared budget range compared to SGD, Adam, Koczor HS-QNG, and WA-QNG. Although our experiments use parameter-independent gate-local Pauli–Lindblad noise, the geometric construction applies to general differentiable mixed-state families satisfying the SLD support condition and is not tied to a particular noise model. The method optimizes the noisy objective directly rather than removing the noise. This can be integrated with other approaches to error mitigation (Naved et al., 2026).

2 Related Work

Quantum natural gradient (QNG) leverages the geometry of the parametrized quantum circuit for optimization (Stokes et al., 2020). For ideal unitary circuits, this geometry is described by the Fubini–Study metric. Stokes et al. (2020) introduced a block-diagonal approximation to the pure-state Fubini–Study metric. Random natural gradient replaces the quantum metric with a classical Fisher metric obtained from random measurements (Kolotouros & Wallden, 2024). For noisy and nonunitary circuits, Koczor & Benjamin (2022) proposed approximations to mixed-state quantum Fisher geometry based on Hilbert–Schmidt overlaps. Weighted Approximate QNG (WA-QNG) incorporates the objective Hamiltonian by combining weighted Hilbert–Schmidt metrics of local reduced states; its experiments use exact expectation values and metric tensors in noise-free circuits (Shi et al., 2026). The measurement cost of metric-aware variational algorithms has also been analyzed (van Straaten & Koczor, 2021), while simultaneous-perturbation methods provide reduced-cost QFIM estimates (Gacon et al., 2021). These methods reduce the cost of estimating or applying a metric but do not introduce the physical-locality and score-space hierarchies considered here.

Local recovery of quantum Fisher information from finite subsystems and restricted operations has also been studied in quantum metrology (Płodzień & Chwedeńczuk, 2026). In our setting, the SLD variational problem is restricted to graph-local physical windows and local score spaces. This gives explicit physical-locality and score-space hierarchies for natural-gradient optimization, together with a finite-shot estimator. The resulting local Pauli measurements are related to target-aware observable-estimation and classical-shadow methods (Huang et al., 2020; 2021; Hadfield et al., 2022).

3 Background and Setup

3.1 Circuit model and noisy variational family

We consider an n-qubit variational circuit with p trainable parameters $\overrightarrow{\theta} \in \mathbb{R}^p$, composed of fixed gates and trainable Pauli rotations

$$U_j(\theta_j) = \exp(i\theta_j G_j), \qquad G_j^2 = I,$$

where $G_j \in \{I, X, Y, Z\}^{\otimes n}$ is the Pauli-string generator of the j-th trainable gate. In the presence of noise, the circuit is described by the completely positive, trace-preserving (CPTP) map $\mathcal{E}_\eta(\overrightarrow{\theta})$, with output $\rho_\eta(\overrightarrow{\theta}) = \mathcal{E}_\eta(\overrightarrow{\theta})(\rho_{\text{in}})$, where η denotes the fixed noise parameters. We assume that the noise channels have no explicit dependence on $\overrightarrow{\theta}$, so the standard parameter-shift rule remains valid (Schuld et al., 2019). The resulting state $\rho_\eta(\overrightarrow{\theta})$ is generally mixed even when the input state ρ_{in} is pure. For a cost Hamiltonian H, the noisy objective is given by:

$$C_\eta(\overrightarrow{\theta}) = \text{tr}\left(H\rho_\eta(\overrightarrow{\theta})\right), \tag{1}$$

with gradient $g_\eta = \nabla C_\eta(\overrightarrow{\theta})$.

3.2 Mixed-state SLD geometry

Since the noisy circuit output is generally mixed, we use the SLD quantum Fisher geometry (Braunstein & Caves, 1994; Petz & Sudár, 1996). For fixed noise parameters η, we write $\rho(\overrightarrow{\theta}) = \rho_\eta(\overrightarrow{\theta})$ or simply ρ when the dependence on $\overrightarrow{\theta}$ is clear. Let $T_j = \partial_j\rho(\overrightarrow{\theta})$ denote the tangent operator for parameter θ_j. For the objective in Equation (1), $(g_\eta)_j = \text{tr}\,(HT_j)$. The SLD superoperator associated with ρ acts on an operator L as

$$\mathcal{R}_\rho(L) = \frac{1}{2}(\rho L + L\rho).$$

The SLD operator for parameter θ_j is a Hermitian operator S_j satisfying

$$T_j = \mathcal{R}_\rho(S_j) = \frac{1}{2}(\rho S_j + S_j\rho). \tag{2}$$

We use the standard SLD support condition $T_j \in \text{Ran}(\mathcal{R}_\rho)$, so the SLD equation is solvable. For full-rank states this is automatic and the rank-deficient case is discussed in Appendix A.1. We write $S_j = \mathcal{R}_\rho^+(T_j)$, where $+$ denotes the Moore–Penrose pseudoinverse. The corresponding SLD-QFIM is given by

$$F_{j\ell}^{\text{SLD}} = \frac{1}{2}\text{tr}\,(\rho\,\{S_j, S_\ell\}) = \text{tr}\,(T_jS_\ell)\,.$$

The equivalence of the two forms is shown in Appendix A.2.

For a real parameter direction $v \in \mathbb{R}^p$, write $T(v) = \sum_j v_jT_j$ and $S(v) = \sum_j v_jS_j$. The ρ-weighted bilinear form between Hermitian operators L_1 and L_2 is defined by

$$\langle L_1, L_2\rangle_\rho = \text{tr}\,(L_1\mathcal{R}_\rho(L_2)) = \frac{1}{2}\text{tr}\,(\rho(L_1L_2 + L_2L_1))\,, \qquad \|L_1\|_\rho^2 = \langle L_1, L_1\rangle_\rho.$$

When ρ is rank deficient, $\|\cdot\|_\rho$ is a seminorm. By linearity of $\mathcal{R}_\rho$ and Equation (2), we have $T(v) = \mathcal{R}_\rho(S(v))$. Then

$$v^\top F^{\text{SLD}}v = \text{tr}\,(T(v)S(v)) = \text{tr}\,(\mathcal{R}_\rho(S(v))S(v)) = \|S(v)\|_\rho^2 \geq 0.$$

The same quantity can also be written variationally.

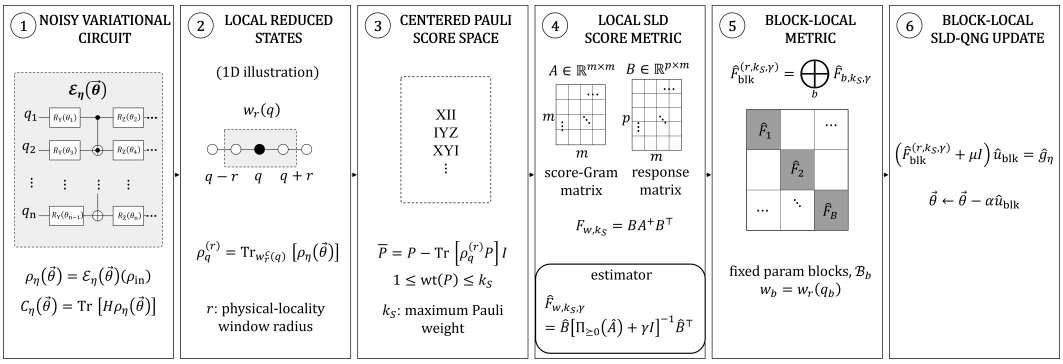
Lemma 3.1 (SLD variational principle)**.** *For every real parameter direction v,*

$$v^\top F^{\text{SLD}}v = \max_{L=L^\dagger}\left\{2tr\,(T(v)L) - tr\left(\rho L^2\right)\right\}.$$

The proof is given in Appendix A.3.

4 Local reduced-state SLD geometry

The exact global SLD-QFIM can become expensive to construct as the system size grows. We restrict the SLD variational problem in Lemma 3.1 to operators supported on small physical windows. This gives the SLD geometry of the corresponding reduced states (Płodzień & Chwedeńczuk, 2026). Within each window, we keep a finite set of centered Pauli scores and retain only the parameter couplings within each block. The resulting block-local SLD-QNG construction is summarized in Figure 1.



LOCALITY–COST TRADEOFF

Increasing r	*Increasing* k_S
more reduced-state information	richer local score space
larger physical window	more Pauli observables

Figure 1: Overview of the proposed block-local SLD-QNG construction. Radius r controls the physical reduced-state window, k_S the local Pauli score space, γ the score-space ridge, and μ the final parameter-space ridge.

4.1 LOCAL REDUCED-STATE HIERARCHY

Let $\Gamma = (V, E)$ denote the interaction graph of the circuit, with $V = \{1, \dots, n\}$ and $\{q, q'\} \in E$ whenever a two-qubit gate acts on qubits q and q'. For $q \in V$ and $r \in \mathbb{Z}_{\geq 0}$, define the radius-$r$ physical window by

$$w_r(q) = \{q' \in V : \mathrm{dist}_\Gamma(q, q') \leq r\}.$$

These windows satisfy $w_r(q) \subseteq w_{r+1}(q)$. For a one-dimensional nearest-neighbor circuit, this reduces to $w_r(q) = \{q - r, \dots, q + r\} \cap \{1, \dots, n\}$. Let $w_r^c(q) = V \setminus w_r(q)$. The reduced state and tangents are given by

$$\rho_q^{(r)} = \mathrm{tr}_{w_r^c(q)}(\rho), \qquad T_{j,q}^{(r)} = \mathrm{tr}_{w_r^c(q)}(T_j).$$

The reduced tangents satisfy $T_{j,q}^{(r)} \in \mathrm{Ran}(\mathcal{R}_{\rho_q^{(r)}})$ whenever $T_j \in \mathrm{Ran}(\mathcal{R}_\rho)$ holds (see Lemma B.1). The corresponding reduced-state SLD operators are $S_{j,q}^{(r)} = \mathcal{R}^+_{\rho_q^{(r)}}\left(T_{j,q}^{(r)}\right)$, with SLD-QFIM entries $\left[F_q^{(r)}\right]_{j\ell} = \mathrm{tr}\left(T_{j,q}^{(r)} S_{\ell,q}^{(r)}\right)$. Although $\rho_q^{(r)}$ acts only on $w_r(q)$, it can depend on all p circuit parameters. So $F_q^{(r)} \in \mathbb{R}^{p \times p}$. The restriction to parameter blocks is introduced later. Since the SLD quantum Fisher information is monotone under CPTP maps (Petz, 1996), the nested physical windows give the following hierarchy.

Theorem 4.1 (Nested local-SLD hierarchy)**.** *For every qubit q and radius $r \in \mathbb{Z}_{\geq 0}$,*

$$0 \preceq F_q^{(r)} \preceq F_q^{(r+1)} \preceq F^{\mathrm{SLD}}.$$

Larger windows retain monotonically more of the SLD geometry, but the radius needed for a given approximation accuracy depends on the circuit and the resulting noisy state.

Let $\mathcal{A}_{\mathrm{global}}$ be the real space of Hermitian operators on the full system, and define the local subspace

$$\mathcal{A}_{q,r} = \left\{ L_{w_r(q)} \otimes I_{w_r^c(q)} : L_{w_r(q)} = L_{w_r(q)}^\dagger \right\}.$$

Theorem 4.2 (Exact localization residual)**.** *For every parameter direction v,*

$$v^\top \left(F^{\mathrm{SLD}} - F_q^{(r)}\right) v = \min_{L \in \mathcal{A}_{q,r}} \|S(v) - L\|_\rho^2.$$

The localization error in direction v is the smallest residual obtained by approximating $S(v)$ with an operator supported on the physical window. Proofs of both results are given in Appendices B.2 and B.3.

4.2 LOCAL SCORE REPRESENTATION

Fix a local window $w = w_r(q)$, and write $\rho_w = \rho_q^{(r)}$, $T_{j,w} = T_{j,q}^{(r)}$ and $S_{j,w} = S_{j,q}^{(r)}$. Let $\{P_a\}_{a=1}^{m_w}$ be a set of nonidentity Pauli strings supported on w. Since ρ_w is a density operator and $\mathrm{tr}(\rho_w) = 1$, we have $\mathrm{tr}(T_{j,w}) = \partial_j \mathrm{tr}(\rho_w) = 0$, so an identity component does not affect the tangent response. The local SLD equation also gives $\mathrm{tr}(\rho_w S_{j,w}) = \mathrm{tr}(T_{j,w}) = 0$. We therefore center the Pauli strings so that they lie in the zero-ρ_w-mean subspace containing the local SLDs:

$$\overline{P}_a = P_a - \mathrm{tr}(\rho_w P_a)\, I. \tag{3}$$

The centered operators are Hermitian and define the restricted local score space

$$\mathcal{V}_w = \mathrm{span}_{\mathbb{R}} \left\{ \overline{P}_1, \ldots, \overline{P}_{m_w} \right\}. \tag{4}$$

For $T_w(v) = \sum_j v_j T_{j,w}$, applying Lemma 3.1 to the reduced family on w and restricting L to $\mathcal{V}_w$ gives

$$v^\top F_{\mathcal{V}_w} v = \max_{L \in \mathcal{V}_w} \left\{ 2\mathrm{tr}\left(T_w(v) L\right) - \mathrm{tr}\left(\rho_w L^2\right) \right\}.$$

Define the response and score-Gram matrices by

$$B_{ja} = \mathrm{tr}\left(T_{j,w} \overline{P}_a\right) = \mathrm{tr}\left(T_{j,w} P_a\right), \qquad A_{aa'} = \frac{1}{2} \mathrm{tr}\left(\rho_w \left\{ \overline{P}_a, \overline{P}_{a'} \right\}\right). \tag{5}$$

Theorem 4.3 (Restricted local SLD)**.** *The SLD variational problem restricted to $\mathcal{V}_w$ yields the metric*

$$F_{\mathcal{V}_w} = B A^+ B^\top, \qquad 0 \preceq F_{\mathcal{V}_w} \preceq F_w^{\mathrm{SLD}}.$$

Here A^+ is the Moore–Penrose pseudoinverse of the score-Gram matrix A. The matrix F_w^{SLD} denotes the exact SLD-QFIM of the reduced family on w. In particular, $F_w^{\mathrm{SLD}} = F_q^{(r)}$ when $w = w_r(q)$.

Thus the restricted metric can be computed from A and B without explicitly computing the local SLD operators, as in moment-matrix Fisher bounds (Gessner et al., 2020). The proof is given in Appendix B.5 and uses the range compatibility shown in Lemma B.2.

Full Pauli score space. If $\{P_a\}_{a=1}^{m_w}$ contains all nonidentity Pauli strings on w, each local SLD can be expanded in the corresponding centered Pauli operators. The restricted metric then equals the exact reduced SLD metric:

$$B A^+ B^\top = F_w^{\mathrm{SLD}}.$$

The proof is given in Appendix B.6. On $|w|$ qubits, the full Pauli score space contains $4^{|w|} - 1$ nonidentity Pauli labels. For an interior radius-one window in one dimension, $|w| = 3$, giving 63 labels.

Score-space truncation. To control the score-space dimension, we retain only Pauli strings of weight at most k_S. Let $\mathcal{S}_{w,k_S} = \{P : \mathrm{supp}(P) \subseteq w, 1 \le \mathrm{wt}(P) \le k_S\}$ denote this set, where $\mathrm{supp}(P)$ and $\mathrm{wt}(P)$ are the support and Pauli weight of P. Centering the retained Pauli strings as in Equation (3) gives the score space

$$\mathcal{V}_{w,k_S} = \mathrm{span}_{\mathbb{R}} \left\{ \overline{P} : P \in \mathcal{S}_{w,k_S} \right\}.$$

The number of retained score labels is $|\mathcal{S}_{w,k_S}| = \sum_{s=1}^{\min\{k_S,|w|\}} \binom{|w|}{s} 3^s$. Let $F_{w,k_S} = F_{\mathcal{V}_{w,k_S}}$, with A_{k_S} and B_{k_S} denoting its score-Gram and response matrices. Since $\mathcal{V}_{w,k_S} \subseteq \mathcal{V}_{w,k_S+1}$, the metrics increase in Loewner order and recover F_w^{SLD} once $k_S \ge |w|$; see Proposition B.3.

4.3 BLOCK-LOCAL GEOMETRY

We next impose a separate approximation in parameter space. The parameter blocks determine which metric couplings are retained, while r determines the physical window used to construct each block. Let $\{\mathcal{B}_b\}_{b=1}^{n_B}$ be a disjoint partition of the trainable parameters, and assign each block an anchor qubit

q_b. We keep this partition fixed as r varies. For each block, let $w_b = w_r(q_b)$. The block-local metric is given by:

$$F_{\mathrm{blk}}^{(r,k_S)} = \bigoplus_{b=1}^{n_B} [F_{w_b,k_S}]_{\mathcal{B}_b\mathcal{B}_b} \succeq 0.$$

For the depth-d hardware-efficient ansatz used in our experiments, we take $n_B = n$, $q_b = b$, and

$$\mathcal{B}_b = \{\theta_{Y,b,\ell}, \theta_{Z,b,\ell}\}_{\ell=1}^{d}. \tag{6}$$

Thus $|\mathcal{B}_b| = 2d$. Other fixed parameter partitions are compatible with the construction, but we do not study their selection here.

To separate physical localization from parameter blocking, define the global and localized block-diagonal metrics, respectively, by

$$F_{\mathrm{BD}}^{\mathrm{global}} = \bigoplus_{b=1}^{n_B} \left[F^{\mathrm{SLD}}\right]_{\mathcal{B}_b\mathcal{B}_b}, \qquad F_{\mathrm{blk}}^{(r)} = \bigoplus_{b=1}^{n_B} \left[F_{q_b}^{(r)}\right]_{\mathcal{B}_b\mathcal{B}_b}.$$

For these exact local blocks,

$$F^{\mathrm{SLD}} - F_{\mathrm{blk}}^{(r)} = \underbrace{F^{\mathrm{SLD}} - F_{\mathrm{BD}}^{\mathrm{global}}}_{E_{\mathrm{cross}}} + \underbrace{F_{\mathrm{BD}}^{\mathrm{global}} - F_{\mathrm{blk}}^{(r)}}_{E_{\mathrm{loc}}^{(r)}}.$$

Here E_{cross} contains the couplings removed by the parameter blocks, while $E_{\mathrm{loc}}^{(r)} \succeq 0$ is the error caused by using reduced physical windows. By Proposition B.5, $E_{\mathrm{loc}}^{(r+1)} \preceq E_{\mathrm{loc}}^{(r)}$, whereas E_{cross} is independent of r. Similarly, applying Proposition B.3 to each block gives $F_{\mathrm{blk}}^{(r,k_S)} \preceq F_{\mathrm{blk}}^{(r,k_S+1)} \preceq F_{\mathrm{blk}}^{(r)}$. Thus increasing r cannot increase the physical-localization error, while increasing k_S captures more of the local SLD geometry within each physical window.

5 Block-local SLD-QNG and finite-shot implementation

We now describe the finite-shot estimation of the block-local metric and the resulting QNG update. For block $\mathcal{B}_b$ with $w_b = w_r(q_b)$, let A_{b,k_S} denote the score-Gram matrix and B_{b,k_S} the response matrix restricted to the parameters in $\mathcal{B}_b$. The unregularized and regularized block metrics are

$$F_{b,k_S} = [F_{w_b,k_S}]_{\mathcal{B}_b\mathcal{B}_b} = B_{b,k_S} A_{b,k_S}^{+} B_{b,k_S}^{\top}, \quad F_{b,k_S,\gamma} = B_{b,k_S} \left(A_{b,k_S} + \gamma I\right)^{-1} B_{b,k_S}^{\top}, \quad \gamma > 0.$$

The ridge γ stabilizes the score-space inversion, and its bias relative to the pseudoinverse metric is bounded in Proposition B.4.

We estimate A_{b,k_S} from local product-Pauli measurements and B_{b,k_S} from parameter-shift measurements, and denote the resulting estimates by $\widehat{A}_{b,k_S}$ and $\widehat{B}_{b,k_S}$. Although the population score-Gram matrix satisfies $A_{b,k_S} \succeq 0$, its finite-shot estimate need not be positive semidefinite. We therefore use

$$\widehat{F}_{b,k_S,\gamma} = \widehat{B}_{b,k_S} \left[\Pi_{\succeq 0}\left(\widehat{A}_{b,k_S}\right) + \gamma I\right]^{-1} \widehat{B}_{b,k_S}^{\top}. \tag{7}$$

Here $\Pi_{\succeq 0}$ denotes projection onto the positive-semidefinite matrices. We assemble the population and estimated block-local metrics as

$$F_{\mathrm{blk}}^{(r,k_S,\gamma)} = \bigoplus_{b=1}^{n_B} F_{b,k_S,\gamma}, \qquad \widehat{F}_{\mathrm{blk}}^{(r,k_S,\gamma)} = \bigoplus_{b=1}^{n_B} \widehat{F}_{b,k_S,\gamma}.$$

The finite-shot direction and parameter update are

$$\left(\widehat{F}_{\mathrm{blk}}^{(r,k_S,\gamma)} + \mu I\right) \widehat{u}_{\mathrm{blk}} = \widehat{g}_\eta, \qquad \overrightarrow{\theta} \leftarrow \overrightarrow{\theta} - \alpha \widehat{u}_{\mathrm{blk}}.$$

Here $\widehat{u}_{\mathrm{blk}}$ is the finite-shot block-local preconditioned direction, $\alpha > 0$ is the step size and $\mu > 0$ stabilizes the parameter update.

We refresh the metric every R_{ref} iterations and reuse its most recent value between refreshes. Compatible measurement data may be reused across overlapping windows and between response and gradient estimation. Further details are given in Appendix E. Algorithm 1 summarizes the procedure.

Finite-shot error. At a metric-refresh iteration, let $\epsilon_{A,b} = \left\| \widehat{A}_{b,k_S} - A_{b,k_S} \right\|_{\mathrm{op}}$, $\epsilon_{B,b} = \left\| \widehat{B}_{b,k_S} - B_{b,k_S} \right\|_{\mathrm{op}}$ and $\beta_b = \|B_{b,k_S}\|_{\mathrm{op}}$. The resulting block error satisfies

$$\left\| \widehat{F}_{b,k_S,\gamma} - F_{b,k_S,\gamma} \right\|_{\mathrm{op}} \leq \frac{2\beta_b \epsilon_{B,b} + \epsilon_{B,b}^2}{\gamma} + \frac{2\beta_b^2 \epsilon_{A,b}}{\gamma^2}.$$

The proof and concentration bounds are given in Appendix E.

Resource scaling. Since the block-local metric is a direct sum, its operator-norm error is the largest block error. Suppose r, k_S, and $\gamma > 0$ are fixed independently of n, $\max_b |w_r(q_b)| = O(1)$, the parameter blocks have bounded size and $n_B = O(n)$. Then, for any $\varepsilon, \delta \in (0,1)$, there exists a local product-Pauli measurement schedule with shot cost $N_{\mathrm{geom}} = O\left(\frac{n}{\varepsilon^2} \log \frac{n}{\delta}\right)$ that estimates the regularized block-local metric to operator-norm error at most ε with probability at least $1 - \delta$. The hidden constant depends on the fixed local window and score-space sizes, the block size and powers of γ^{-1}. The full statement and proof are given in Corollary E.4.

If the metric is refreshed every R_{ref} iterations, a conservative average shot cost per iteration is $N_{\mathrm{iter}}^{\mathrm{SLD}} = N_{\mathrm{grad}} + \frac{N_{\mathrm{geom}}}{R_{\mathrm{ref}}}$, where N_{grad} is the cost of a gradient-only iteration. Compatible measurement reuse can further reduce this cost (see Appendix E.6).

Convergence. For fixed noise parameters η, let $\widehat{F}_{\mathrm{blk}}^t$ denote the most recently refreshed block-local metric at iteration t, and set $M_t = \widehat{F}_{\mathrm{blk}}^t + \mu I$. Write the gradient estimate as $\widehat{g}_\eta^t = \nabla C_\eta(\overrightarrow{\theta}^t) + \xi_t$. Since $\widehat{F}_{\mathrm{blk}}^t \succeq 0$, we have $M_t \succeq \mu I$. The score ridge and the bounded empirical Pauli responses also give

$$M_t \preceq \Lambda_{\max} I, \qquad \Lambda_{\max} = \mu + \frac{4}{\gamma} \max_b \left\{ |\mathcal{B}_b| |\mathcal{S}_{w_b,k_S}| \right\}.$$

Theorem 5.1 (Convergence to a stationary neighborhood)**.** *Assume that C_η is L_C-smooth and bounded below by $C_{\inf}$. Suppose that for $t = 0, \ldots, t_{\max} - 1$,*

$$\mathbb{E}\left[\xi_t \mid \overrightarrow{\theta}^t, M_t \right] = 0, \qquad \mathbb{E}\left[\|\xi_t\|_2^2 \mid \overrightarrow{\theta}^t, M_t \right] \leq \sigma_{g_\eta}^2.$$

If $0 < \alpha \leq \mu / L_C$, then

$$\min_{0 \leq t < t_{\max}} \mathbb{E}\left[\left\| \nabla C_\eta(\overrightarrow{\theta}^t) \right\|_2^2 \right] \leq \frac{2\Lambda_{\max}\left(C_\eta(\overrightarrow{\theta}^0) - C_{\inf} \right)}{\alpha t_{\max}} + \frac{L_C \alpha \Lambda_{\max}}{\mu^2} \sigma_{g_\eta}^2.$$

The theorem gives convergence to a stationary neighborhood of the noisy objective C_η in expected squared gradient norm. With exact gradients, the variance term vanishes. The proof and other details are given in Appendix F.

6 Numerical Experiments

Experimental setup. We study a VQE (Peruzzo et al., 2014) for the ground-state energy of the one-dimensional Heisenberg model,

$$H = \sum_{i=0}^{n-2} \left(X_i X_{i+1} + Y_i Y_{i+1} + Z_i Z_{i+1} \right),$$

with open boundary conditions. The variational circuit is a nearest-neighbor hardware-efficient ansatz (HEA) with R_Y and R_Z rotations and CNOT entangling gates, repeated for d layers. We use parameter independent gate-local Pauli–Lindblad noise with η denoting the common Pauli-generator rate (details in Appendix G.1). For block-local SLD-QNG, the rotation parameters associated with each qubit form a fixed parameter block. Block $\mathcal{B}_b$ is assigned the physical window $w_r(q_b)$ and a centered Pauli score space with maximum weight k_S. Unless stated otherwise, block-local SLD-QNG uses $r = 1$ and $k_S = 3$; because a radius-one window contains at most three qubits, this gives the complete Pauli score space within each window. We evaluate both population-level approximation to the full SLD-QFIM and finite-shot optimization performance. We compare against Koczor HS-QNG (Koczor & Benjamin, 2022) and WA-QNG (Shi et al., 2026), with SGD and Adam additionally included in the finite-shot optimization experiments.

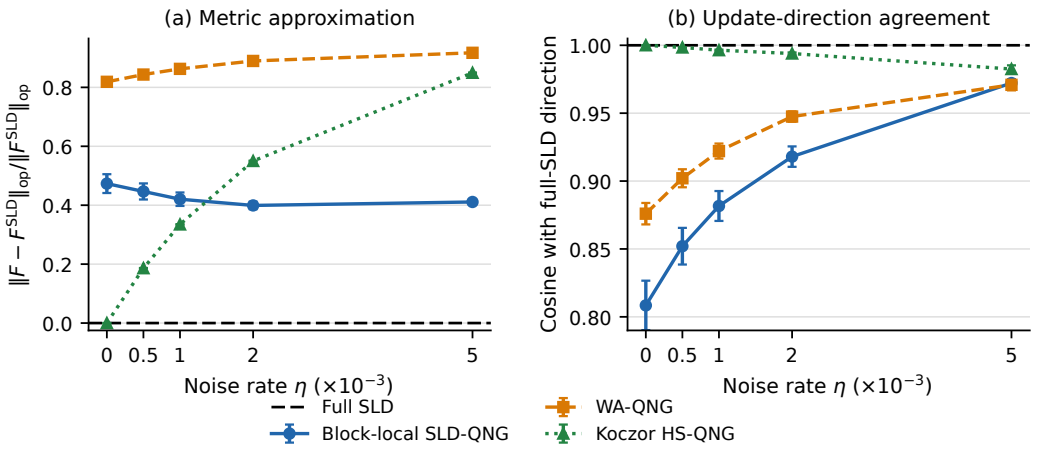


Figure 2: Approximation quality for the 6-qubit, depth-4 circuit over 10 random parameter initializations. (a) Relative operator-norm error w.r.t the full SLD-QFIM. (b) Cosine similarity with the regularized full-SLD update direction. The dashed black lines mark the exact full-SLD reference: zero error in (a) and unit cosine similarity in (b). Error bars show the standard error of the mean.

Approximation quality. At $n = 6$ and $d = 4$, we construct the full mixed-state SLD-QFIM exactly and evaluate population-level approximation errors over 10 random parameter initializations. We measure metric error by $\left\|F - F^{\mathrm{SLD}}\right\|_{\mathrm{op}} / \left\|F^{\mathrm{SLD}}\right\|_{\mathrm{op}}$, where F denotes the approximate metric, and direction agreement by the cosine similarity with the regularized full-SLD direction. At $\eta = 0$, the rescaled Hilbert–Schmidt metric used by Koczor HS-QNG coincides with the full SLD-QFIM, while the block-local metric retains localization and parameter-blocking error. As noise makes the state mixed, the Hilbert–Schmidt metric no longer coincides with the SLD-QFIM. Block-local SLD-QNG has the smallest metric error among the tested methods for $\eta \geq 0.002$ under both uniform and tested nonuniform rates; see Figures 2 and 5. At $\eta = 0.005$, the errors are 0.4110 ± 0.0045, 0.9173 ± 0.0026, and 0.8494 ± 0.0009 for block-local SLD-QNG, WA-QNG, and Koczor HS-QNG, respectively. Their update-direction cosines are 0.9722 ± 0.0024, 0.9709 ± 0.0041, and 0.9825 ± 0.0027. The lower operator-norm error shows that the block-local metric is closer to the full SLD-QFIM, while all three methods produce directions closely aligned with the full-SLD update. At the same noise strength, the cross-block residual E_{cross} has relative operator norm 0.1543 ± 0.0086. Figure 3 separates the two local approximation mechanisms. The physical-localization residual decreases from 0.4461 at $r = 0$ to 0.3783 at $r = 1$ and 0.3552 at $r = 2$. At $r = 1$, the score-space residual decreases from 0.2151 for $k_S = 1$ to 0.0844 for $k_S = 2$ and vanishes for $k_S = 3$, which spans the complete Pauli score space on a radius-one window. This monotonic behavior is consistent with the hierarchies in Propositions B.3 and B.5. Figure 3(c) shows the associated increase in the number of circuit–measurement settings for the conservative independent-block construction.

Finite-shot optimization. We use a 12-qubit, depth-6 circuit with common gate-local Pauli–Lindblad generator rate $\eta = 0.0005$. All methods optimize the same noisy objective and are compared at matched cumulative shot budgets. All metric-based methods refresh their metric every $R_{\mathrm{ref}} = 10$ iterations. Energy, gradient, and metric-estimation shots are all included in the cumulative shot count.

As shown in Figure 4(a), block-local SLD-QNG reaches low energies in fewer iterations. Koczor HS-QNG catches up significantly later, while WA-QNG lies between block-local SLD-QNG and the first-order methods. WA-QNG performs well despite its larger SLD-metric error; this is consistent with its use of a Hamiltonian-aware local geometry rather than a direct approximation to the full SLD-QFIM (Shi et al., 2026). The cumulative-shot comparison in Figure 4(b) also accounts for metric-estimation cost. Block-local SLD-QNG maintains the lowest mean energy over most of the common shot range. Koczor HS-QNG reaches a similar energy per iteration but requires substantially more shots. Additional noise-strength and transverse-field Ising controls are reported in Appendix G.7.

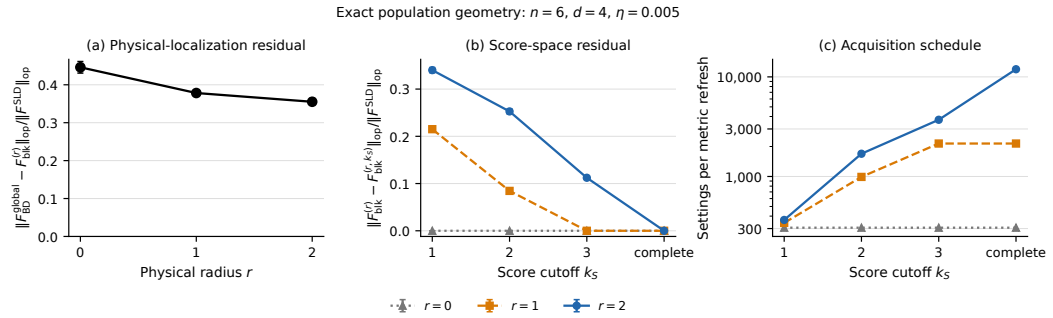


Figure 3: Effects of physical localization and score-space truncation for the 6-qubit, depth-4 circuit at $\eta = 0.005$ over 10 parameter initializations. (a) Physical-localization residual versus r, using the exact reduced-state SLD metric within each window. (b) Score-space residual versus k_S for $r = 0, 1, 2$. (c) Measurement settings per metric refresh for a conservative independent-block product-Pauli schedule. Compatible settings across overlapping windows are not merged. Error bars show the standard error of the mean.

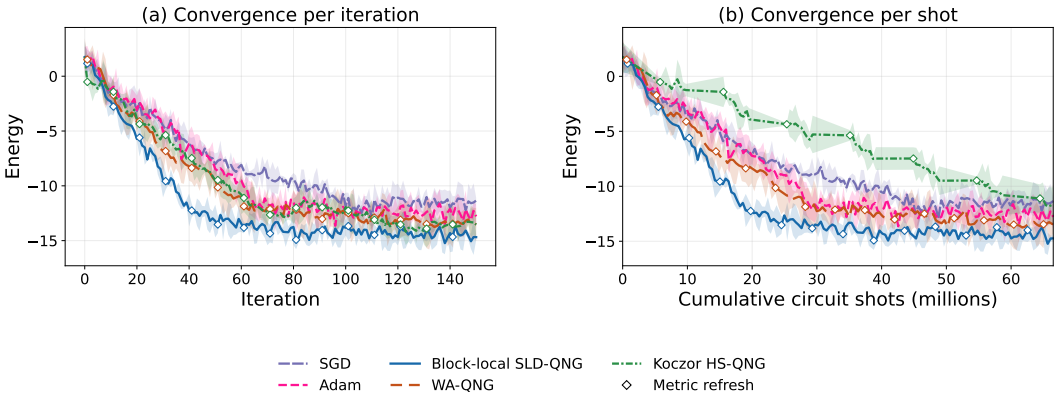


Figure 4: Finite-shot VQE optimization for the 12-qubit, depth-6 Heisenberg instance over 8 random seeds. (a) Energy versus iteration. (b) Energy versus cumulative executed circuit shots, restricted to the largest budget reached by all methods. Diamond symbols mark metric refreshes and shaded regions show pointwise 95% confidence intervals.

7 Discussion

Our results show that localizing the SLD variational problem can retain useful mixed-state geometry while reducing measurement cost. Increasing the physical radius r or score-space cutoff k_S reduces the corresponding approximation error. At the higher tested noise strengths, the block-local metric is closer to the full SLD-QFIM than the tested surrogates. Under matched cumulative shot budgets, block-local SLD-QNG also reaches the lowest mean energy over most of the shared budget range. Metric error alone does not determine optimization performance, since the update also depends on the gradient, regularization, and inverse metric.

The $O(n \log n)$ metric-estimation bound requires bounded physical windows and parameter blocks. The finite-shot bound uses a conservative independent-block schedule, while compatible measurement reuse can reduce the practical cost. Parameter blocking still removes cross-block couplings, even when the local score space is complete. The convergence guarantee applies to the noisy objective and requires the gradient estimator to be conditionally unbiased given the empirical metric. Adaptive choices of r and k_S, ansatz- or problem-dependent parameter blocks, and integration with error-mitigation methods remain open directions.

AI USE STATEMENT

Generative AI tools were used for language editing, code debugging and limited implementation assistance, checking mathematical derivations, and limited feedback on experimental methodology. They were also used to prepare initial drafts of the Introduction and Discussion, which were substantially revised by the authors. All AI-assisted material was reviewed and verified by the authors, who take full responsibility for the content of this work.

ETHICS STATEMENT

This work does not involve human subjects, personal data, or sensitive datasets, and we do not identify ethical concerns specific to this study.

REPRODUCIBILITY STATEMENT

Proofs of all theoretical results are given in Appendices A, B, E and F. The finite-shot estimators, concentration bounds, and measurement-reuse details are provided in Appendix E, with the optimization procedure summarized in Appendix D. The noise model, experimental configurations, baseline definitions, hyperparameters and shot accounting used in the numerical experiments are provided in Appendices G.1 to G.3 and G.5 to G.7. Code will be released upon publication.

REFERENCES

Shun-Ichi Amari. Natural gradient works efficiently in learning. *Neural computation*, 10(2):251–276, 1998.

Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S Kottmann, Tim Menke, et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*, 94(1):015004, 2022.

Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018.

Samuel L Braunstein and Carlton M Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439, 1994.

Marco Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 2021.

Julien Gacon, Christa Zoufal, Giuseppe Carleo, and Stefan Woerner. Simultaneous perturbation stochastic approximation of the quantum fisher information. *Quantum*, 5:567, 2021.

Manuel Gessner, Augusto Smerzi, and Luca Pezzè. Multiparameter squeezing for optimal quantum enhancements in sensor networks. *Nature communications*, 11(1):3817, 2020.

Saeed Ghadimi and Guanghui Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM journal on optimization*, 23(4):2341–2368, 2013.

Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows. *Communications in Mathematical Physics*, 391(3):951–967, 2022.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 2020.

Hsin-Yuan Huang, Richard Kueng, and John Preskill. Efficient estimation of pauli observables by derandomization. *Physical review letters*, 127(3):030503, 2021.

Ali Javadi-Abhari, Matthew Treinish, Kevin Krsulich, Christopher J Wood, Jake Lishman, Julien Gacon, Simon Martiel, Paul D Nation, Lev S Bishop, Andrew W Cross, et al. Quantum computing with qiskit. *arXiv preprint arXiv:2405.08810*, 2024.

Bálint Koczor and Simon C Benjamin. Quantum natural gradient generalized to noisy and nonunitary circuits. *Physical Review A*, 106(6):062416, 2022.

Ioannis Kolotouros and Petros Wallden. Random natural gradient. *Quantum*, 8:1503, 2024.

Md Mobasshir Arshed Naved, Wenbo Xie, Wojciech Szpankowski, and Ananth Grama. Robust integrated learning and pauli noise mitigation for parametrized quantum circuits. *Advances in Neural Information Processing Systems*, 38:14224–14270, 2026.

Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5(1):4213, 2014.

Dénes Petz. Monotone metrics on matrix spaces. *Linear algebra and its applications*, 244:81–96, 1996.

Dénes Petz and Csaba Sudár. Geometries of quantum states. *Journal of Mathematical Physics*, 37(6): 2662–2673, 1996.

Marcin Płodzień and Jan Chwedeńczuk. Operator spreading and recoverability of local quantum fisher information in a $u(1)$-broken spin chain. *arXiv preprint arXiv:2605.02774*, 2026.

Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3):032331, 2019.

Chenyu Shi, Vedran Dunjko, and Hao Wang. Weighted approximate quantum natural gradient for variational quantum eigensolver. *Quantum Science and Technology*, 11(1):015060, 2026.

James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, 2020.

Ewout Van Den Berg, Zlatko K Minev, Abhinav Kandala, and Kristan Temme. Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors. *Nature physics*, 19 (8):1116–1121, 2023.

Barnaby van Straaten and Bálint Koczor. Measurement cost of metric-aware variational quantum algorithms. *PRX Quantum*, 2(3):030324, 2021.

A SLD SUPEROPERATOR DETAILS

A.1 RANK-DEFICIENT SLD SUPPORT CONDITION

Let

$$\rho = \sum_m p_m \left|m\right\rangle \left\langle m\right|.$$

In the eigenbasis of ρ,

$$\begin{aligned}\mathcal{R}_\rho(\left|m\right\rangle \left\langle m'\right|) &= \frac{1}{2}\left(\rho \left|m\right\rangle \left\langle m'\right| + \left|m\right\rangle \left\langle m'\right| \rho\right) \\ &= \frac{1}{2}\left(p_m \left|m\right\rangle \left\langle m'\right| + p_{m'} \left|m\right\rangle \left\langle m'\right|\right) \\ &= \frac{p_m + p_{m'}}{2} \left|m\right\rangle \left\langle m'\right|. \end{aligned} \tag{8}$$

Therefore the Moore–Penrose pseudoinverse acts as

$$\mathcal{R}_\rho^+(\left|m\right\rangle \left\langle m'\right|) = \begin{cases} \dfrac{2}{p_m + p_{m'}} \left|m\right\rangle \left\langle m'\right|, & p_m + p_{m'} > 0, \\ 0, & p_m = p_{m'} = 0. \end{cases} \tag{9}$$

Hence

$$\mathrm{Ker}(\mathcal{R}_\rho) = \mathrm{span}\left\{\left|m\right\rangle \left\langle m'\right| : p_m = p_{m'} = 0\right\}.$$

Let Π_0 be the projector onto $\mathrm{Ker}(\rho)$. From Equation (8), components supported entirely in $\mathrm{Ker}(\rho)$ cannot be produced by $\mathcal{R}_\rho$. So a Hermitian operator T belongs to $\mathrm{Ran}(\mathcal{R}_\rho)$ exactly when

$$\Pi_0 T \Pi_0 = 0. \tag{10}$$

For such T,

$$\mathcal{R}_\rho \mathcal{R}_\rho^+(T) = T.$$

The Hermitian solutions $L = L^\dagger$ of $\mathcal{R}_\rho(L) = T$ are

$$L = \mathcal{R}_\rho^+(T) + K, \qquad K = K^\dagger, \quad \mathcal{R}_\rho(K) = 0.$$

The pseudoinverse selects the minimum-Hilbert–Schmidt-norm representative. For full-rank ρ, $\mathrm{Ker}(\rho) = \{0\}$, so every Hermitian T lies in $\mathrm{Ran}(\mathcal{R}_\rho)$.

A.2 EQUIVALENCE OF SLD-QFIM FORMULAS

Using the SLD equation in Equation (2),

$$\begin{aligned}\mathrm{tr}\left(T_j S_\ell\right) &= \frac{1}{2}\mathrm{tr}\left((\rho S_j + S_j \rho) S_\ell\right) \\ &= \frac{1}{2}\mathrm{tr}\left(\rho S_j S_\ell\right) + \frac{1}{2}\mathrm{tr}\left(S_j \rho S_\ell\right) \\ &= \frac{1}{2}\mathrm{tr}\left(\rho S_j S_\ell\right) + \frac{1}{2}\mathrm{tr}\left(\rho S_\ell S_j\right) \\ &= \frac{1}{2}\mathrm{tr}\left(\rho\{S_j, S_\ell\}\right), \end{aligned}$$

where the third equality uses cyclicity of trace. Therefore

$$F_{j\ell}^{\mathrm{SLD}} = \mathrm{tr}\left(T_j S_\ell\right) = \mathrm{tr}\left(T_j \mathcal{R}_\rho^+(T_\ell)\right).$$

Using the eigenbasis representation of $\mathcal{R}_\rho^+$ in Equation (9),

$$\langle m|\mathcal{R}_\rho^+(T_\ell)|m'\rangle = \begin{cases} \frac{2\langle m|T_\ell|m'\rangle}{p_m + p_{m'}}, & p_m + p_{m'} > 0, \\ 0, & p_m = p_{m'} = 0. \end{cases}$$

Substituting this into $F_{j\ell}^{\mathrm{SLD}} = \mathrm{tr}\left(T_j \mathcal{R}_\rho^+(T_\ell)\right)$ gives

$$\begin{aligned} F_{j\ell}^{\mathrm{SLD}} &= \sum_{m,m'} \langle m|T_j|m'\rangle\langle m'|\mathcal{R}_\rho^+(T_\ell)|m\rangle \\ &= 2 \sum_{\substack{m,m' \\ p_m+p_{m'}>0}} \frac{\langle m|T_j|m'\rangle\langle m'|T_\ell|m\rangle}{p_m + p_{m'}}. \end{aligned}$$

A.3 Proof of the SLD variational principle

Proof. Since $S(v) = \mathcal{R}_\rho^+(T(v))$ and $T(v) \in \mathrm{Ran}(\mathcal{R}_\rho)$,

$$T(v) = \mathcal{R}_\rho(S(v)).$$

By Equation (8), $\mathcal{R}_\rho$ is self-adjoint with respect to the Hilbert–Schmidt inner product. Therefore,

$$\begin{aligned} \mathrm{tr}\,(T(v)L) &= \mathrm{tr}\,(\mathcal{R}_\rho(S(v))L) \\ &= \mathrm{tr}\,(S(v)\mathcal{R}_\rho(L)) \\ &= \langle S(v), L\rangle_\rho. \end{aligned}$$

For Hermitian L,

$$\begin{aligned} \|L\|_\rho^2 &= \mathrm{tr}\,(L\mathcal{R}_\rho(L)) \\ &= \mathrm{tr}\left(\rho L^2\right). \end{aligned}$$

Hence

$$\begin{aligned} 2\mathrm{tr}\,(T(v)L) - \mathrm{tr}\left(\rho L^2\right) &= 2\langle S(v), L\rangle_\rho - \|L\|_\rho^2 \\ &= \|S(v)\|_\rho^2 - \left(\|L\|_\rho^2 - 2\langle S(v), L\rangle_\rho + \|S(v)\|_\rho^2\right) \\ &= \|S(v)\|_\rho^2 - \|L - S(v)\|_\rho^2. \end{aligned} \tag{11}$$

The eigenvalues of $\mathcal{R}_\rho$ in Equation (8) are nonnegative, so $\mathcal{R}_\rho$ is positive semidefinite. Hence

$$\|L - S(v)\|_\rho^2 = 0 \quad \Longleftrightarrow \quad L - S(v) \in \mathrm{Ker}(\mathcal{R}_\rho).$$

By Equation (11), the objective is maximized exactly when $L - S(v) \in \mathrm{Ker}(\mathcal{R}_\rho)$. Therefore the maximizers are

$$L = S(v) + K, \qquad K = K^\dagger, \quad \mathcal{R}_\rho(K) = 0.$$

The maximum value is given by,

$$\max_{L=L^\dagger} \left\{2\mathrm{tr}\,(T(v)L) - \mathrm{tr}\left(\rho L^2\right)\right\} = \|S(v)\|_\rho^2 = v^\top F^{\mathrm{SLD}} v.$$

□

B Proofs and auxiliary results for local reduced-state SLD geometry

B.1 Reduced-state support compatibility

Lemma B.1 (Reduced-state support compatibility). *Let $T = T^\dagger \in \mathrm{Ran}(\mathcal{R}_\rho)$ and, for a subsystem w, define*

$$\rho_w = tr_{w^c}(\rho), \qquad T_w = tr_{w^c}(T).$$

Then

$$T_w \in \mathrm{Ran}(\mathcal{R}_{\rho_w}).$$

Proof. Since $T \in \mathrm{Ran}(\mathcal{R}_\rho)$, there exists a Hermitian operator S such that

$$T = \mathcal{R}_\rho(S) = \frac{1}{2}(\rho S + S\rho).$$

Let P_w be the projector onto $\mathrm{Ker}(\rho_w)$ and define

$$Q_w = P_w \otimes I_{w^c}.$$

By the property of partial trace,

$$\mathrm{tr}\,(Q_w\rho) = \mathrm{tr}\,((P_w \otimes I_{w^c})\rho) = \mathrm{tr}\,(P_w\rho_w) = 0,$$

where the last equality follows because P_w projects onto $\mathrm{Ker}(\rho_w)$. Since $\rho \succeq 0$ and Q_w is a projector, we get

$$\mathrm{tr}\,(Q_w\rho) = \mathrm{tr}\left(\rho^{1/2}Q_w\rho^{1/2}\right) = \mathrm{tr}\left(\left(Q_w\rho^{1/2}\right)^\dagger\left(Q_w\rho^{1/2}\right)\right) = 0.$$

Hence $Q_w\rho^{1/2} = 0$ and therefore $Q_w\rho = \rho Q_w = 0$.

Using the expression for T,

$$Q_wTQ_w = \frac{1}{2}\left(Q_w\rho SQ_w + Q_wS\rho Q_w\right) = 0.$$

Taking the partial trace over w^c gives

$$\mathrm{tr}_{w^c}\left(Q_wTQ_w\right) = P_wT_wP_w = 0.$$

By Equation (10), $P_wT_wP_w = 0$ implies

$$T_w \in \mathrm{Ran}(\mathcal{R}_{\rho_w}).$$

□

B.2 Proof of Theorem 4.1

Proof. For a parameter direction v, write $T_q^{(r)}(v) = \sum_j v_j T_{j,q}^{(r)}$. By definition of the reduced state and reduced tangents,

$$\rho_q^{(r)} = \mathrm{tr}_{w_r^c(q)}\left(\rho\right), \qquad T_q^{(r)}(v) = \mathrm{tr}_{w_r^c(q)}\left(T(v)\right).$$

Let $L_r = L_r^\dagger$ be a Hermitian operator acting on $w_r(q)$. The partial-trace identities give

$$\mathrm{tr}\left(T_q^{(r)}(v)L_r\right) = \mathrm{tr}\left(T(v)\left(L_r \otimes I_{w_r^c(q)}\right)\right),$$
$$\mathrm{tr}\left(\rho_q^{(r)}L_r^2\right) = \mathrm{tr}\left(\rho\left(L_r^2 \otimes I_{w_r^c(q)}\right)\right).$$

The reduced problem can therefore be written as an optimization over embedded local Hermitian operators:

$$v^\top F_q^{(r)} v = \max_{L \in \mathcal{A}_{q,r}} \left\{2\mathrm{tr}\left(T(v)L\right) - \mathrm{tr}\left(\rho L^2\right)\right\}. \tag{12}$$

The operator spaces satisfy

$$\mathcal{A}_{q,r} \subseteq \mathcal{A}_{q,r+1} \subseteq \mathcal{A}_{\mathrm{global}}.$$

Each space contains $L = 0$, so the corresponding maximum is nonnegative. As the spaces grow with r, the maximum cannot decrease and remains bounded by the global SLD maximum. So, for every parameter direction v,

$$0 \le v^\top F_q^{(r)} v \le v^\top F_q^{(r+1)} v \le v^\top F^{\mathrm{SLD}} v.$$

Hence

$$0 \preceq F_q^{(r)} \preceq F_q^{(r+1)} \preceq F^{\mathrm{SLD}}.$$

□

B.3 Proof of Theorem 4.2

Proof. By Equation (12),

$$v^\top F_q^{(r)} v = \max_{L \in \mathcal{A}_{q,r}} \left\{ 2\mathrm{tr}\left(T(v)L\right) - \mathrm{tr}\left(\rho L^2\right) \right\}.$$

Using Equation (11),

$$\begin{aligned} v^\top F_q^{(r)} v &= \max_{L \in \mathcal{A}_{q,r}} \left\{ \|S(v)\|_\rho^2 - \|S(v) - L\|_\rho^2 \right\} \\ &= \|S(v)\|_\rho^2 - \min_{L \in \mathcal{A}_{q,r}} \|S(v) - L\|_\rho^2 . \end{aligned}$$

Since

$$v^\top F^{\mathrm{SLD}} v = \|S(v)\|_\rho^2 ,$$

we obtain

$$v^\top \left(F^{\mathrm{SLD}} - F_q^{(r)} \right) v = \min_{L \in \mathcal{A}_{q,r}} \|S(v) - L\|_\rho^2 .$$

□

B.4 Range compatibility

Optimizing the score coefficients in the proof of Theorem 4.3 gives the stationarity equation $Ac = B^\top v$. Since A may be singular, this equation has a solution only if $B^\top v \in \mathrm{Ran}(A)$. The next lemma shows that this holds for every parameter direction v.

Lemma B.2 (Range compatibility)**.** *For the matrices A and B defined in Equation* (5)*,*

$$\mathrm{Ran}(B^\top) \subseteq \mathrm{Ran}(A).$$

Equivalently, $B^\top v \in \mathrm{Ran}(A)$ for every parameter direction v.

Proof. Let $c = (c_1, \ldots, c_{m_w})^\top \in \mathbb{R}^{m_w}$ satisfy $Ac = 0$, and define

$$L' = \sum_{a=1}^{m_w} c_a \overline{P}_a .$$

Using the definition of L' and Equation (5),

$$\begin{aligned} \|L'\|_{\rho_w}^2 &= \frac{1}{2} \mathrm{tr}\left(\rho_w \left\{L', L'\right\}\right) \\ &= \frac{1}{2} \sum_{a,a'} c_a c_{a'} \mathrm{tr}\left(\rho_w \left\{\overline{P}_a, \overline{P}_{a'}\right\}\right) \\ &= \sum_{a,a'} c_a A_{aa'} c_{a'} \\ &= c^\top A c = 0. \end{aligned}$$

By Lemma B.1, each reduced tangent $T_{j,w}$ admits a local SLD operator $S_{j,w}$ satisfying

$$T_{j,w} = \mathcal{R}_{\rho_w}(S_{j,w}).$$

Now,

$$\begin{aligned} (Bc)_j &= \sum_a c_a B_{ja} \\ &= \mathrm{tr}\left(T_{j,w} L'\right) \\ &= \langle S_{j,w}, L' \rangle_{\rho_w}. \end{aligned}$$

By Cauchy–Schwarz in the ρ_w-weighted seminorm,

$$|\langle S_{j,w}, L' \rangle_{\rho_w}|^2 \leq \|S_{j,w}\|_{\rho_w}^2 \, \|L'\|_{\rho_w}^2 = 0,$$

and hence $(Bc)_j = 0$ for every j. Thus

$$\mathrm{Ker}(A) \subseteq \mathrm{Ker}(B).$$

Taking orthogonal complements and using $A = A^\top$ gives

$$\mathrm{Ran}(B^\top) = \mathrm{Ker}(B)^\perp \subseteq \mathrm{Ker}(A)^\perp = \mathrm{Ran}(A).$$

□

B.5 Proof of Theorem 4.3

This follows the response–covariance structure of moment-matrix Fisher bounds (Gessner et al., 2020), with singular A handled using Lemma B.2.

Proof. For a parameter direction v, write $T_w(v) = \sum_j v_j T_{j,w}$. By Equation (4), any restricted score $L' \in \mathcal{V}_w$ can be written as

$$L' = \sum_{a=1}^{m_w} c_a \overline{P}_a, \qquad c = (c_1, \ldots, c_{m_w})^\top \in \mathbb{R}^{m_w}.$$

Using the definitions of A and B,

$$\mathrm{tr}\,(T_w(v) L') = \sum_{j,a} v_j c_a \mathrm{tr}\left(T_{j,w} \overline{P}_a\right) = v^\top B c,$$

where centering does not change the response because $\mathrm{tr}\,(T_{j,w}) = 0$. Similarly,

$$\mathrm{tr}\left(\rho_w (L')^2\right) = \frac{1}{2} \sum_{a,a'} c_a c_{a'} \mathrm{tr}\left(\rho_w \{\overline{P}_a, \overline{P}_{a'}\}\right) = c^\top A c.$$

Therefore the restricted variational objective is

$$2\mathrm{tr}\,(T_w(v) L') - \mathrm{tr}\left(\rho_w (L')^2\right) = 2 v^\top B c - c^\top A c.$$

Since A is a Gram matrix, $A = A^\top \succeq 0$, so the objective is concave in c. Its gradient is

$$\nabla_c \left(2 v^\top B c - c^\top A c\right) = 2 B^\top v - 2 A c.$$

The stationary points therefore satisfy

$$A c = B^\top v.$$

If A were invertible, this equation would give $c = A^{-1} B^\top v$. Since A may be singular, we instead use the Moore–Penrose pseudoinverse. By Lemma B.2, $B^\top v \in \mathrm{Ran}(A)$, so

$$c^\star = A^+ B^\top v, \qquad A c^\star = A A^+ B^\top v = B^\top v.$$

Thus $c^\star$ satisfies the stationarity equation. Since the objective is concave, it is a global maximizer.

Substituting $c^\star$ into the variational objective gives

$$\max_{c \in \mathbb{R}^{m_w}} \left\{2 v^\top B c - c^\top A c\right\} = 2 v^\top B c^\star - (c^\star)^\top A c^\star = v^\top B A^+ B^\top v.$$

By definition of the restricted metric,

$$v^\top F_{\mathcal{V}_w} v = \max_{c \in \mathbb{R}^{m_w}} \left\{2 v^\top B c - c^\top A c\right\}.$$

Therefore, for every parameter direction v,

$$v^\top F_{\mathcal{V}_w} v = v^\top B A^+ B^\top v,$$

and hence

$$F_{\mathcal{V}_w} = B A^+ B^\top.$$

Since $\mathcal{V}_w$ is a subspace of the Hermitian operators on w, the argument in Appendix B.2 gives, for every v,

$$0 \leq v^\top F_{\mathcal{V}_w} v \leq v^\top F_w^{\mathrm{SLD}} v.$$

Therefore,

$$0 \preceq F_{\mathcal{V}_w} \preceq F_w^{\mathrm{SLD}}.$$

□

B.6 FULL PAULI SCORE SPACE

Let $\{P_a\}_{a=1}^{m_w}$ be the set of all nonidentity Pauli strings on w with $m_w = 4^{|w|} - 1$. For a parameter direction v, define

$$S_w(v) = \sum_j v_j S_{j,w} \qquad\qquad T_w(v) = \sum_j v_j T_{j,w}.$$

By linearity of $\mathcal{R}_{\rho_w}$,

$$\begin{aligned}\mathcal{R}_{\rho_w}(S_w(v)) &= \sum_j v_j \mathcal{R}_{\rho_w}(S_{j,w}) \\ &= \sum_j v_j T_{j,w} \\ &= T_w(v).\end{aligned}$$

So $S_w(v)$ is an SLD operator for the reduced tangent $T_w(v)$. Taking the trace of this equation gives

$$\mathrm{tr}\left(\rho_w S_w(v)\right) = \mathrm{tr}\left(T_w(v)\right) = 0.$$

Since the identity and all nonidentity Pauli strings form a basis for Hermitian operators on w, we can write

$$S_w(v) = c_0 I + \sum_{a=1}^{m_w} c_a P_a, \qquad c_0, c_a \in \mathbb{R}.$$

Using $\mathrm{tr}\left(\rho_w S_w(v)\right) = 0$ gives

$$c_0 = -\sum_{a=1}^{m_w} c_a \mathrm{tr}\left(\rho_w P_a\right).$$

Therefore,

$$\begin{aligned}S_w(v) &= \sum_{a=1}^{m_w} c_a \left(P_a - \mathrm{tr}\left(\rho_w P_a\right) I\right) \\ &= \sum_{a=1}^{m_w} c_a \overline{P}_a \in \mathcal{V}_w.\end{aligned}$$

By Lemma 3.1, $S_w(v)$ attains the exact reduced-state maximum in direction v. Since $S_w(v) \in \mathcal{V}_w$, the same maximum is attained after restricting the score space to $\mathcal{V}_w$. Hence, for every v,

$$v^\top F_{\mathcal{V}_w} v = v^\top F_w^{\mathrm{SLD}} v.$$

Therefore,

$$F_{\mathcal{V}_w} = F_w^{\mathrm{SLD}}.$$

Using $F_{\mathcal{V}_w} = BA^+B^\top$ from Theorem 4.3, we obtain

$$BA^+B^\top = F_w^{\mathrm{SLD}}.$$

B.7 NESTED SCORE-SPACE HIERARCHY

Proposition B.3 (Nested score-space hierarchy)**.** *For a fixed reduced state family on w,*

$$0 \preceq F_{w,1} \preceq F_{w,2} \preceq \cdots \preceq F_{w,|w|} = F_w^{\mathrm{SLD}}.$$

Proof. The centered score spaces satisfy

$$\mathcal{V}_{w,1} \subseteq \mathcal{V}_{w,2} \subseteq \cdots \subseteq \mathcal{V}_{w,|w|}.$$

The same variational-set argument used in Appendix B.2 gives, for every parameter direction v,

$$0 \le v^\top F_{w,1} v \le v^\top F_{w,2} v \le \cdots \le v^\top F_{w,|w|} v.$$

For any $k_S \ge |w|$, the retained Pauli set contains all nonidentity Pauli strings on w. By Appendix B.6,

$$F_{w,k_S} = F_{w,|w|} = F_w^{\mathrm{SLD}}.$$

□

B.8 SCORE-SPACE RIDGE BIAS

For a fixed window w, define the score-regularized metric by

$$F_{w,k_S,\gamma} = B_{k_S} \left(A_{k_S} + \gamma I\right)^{-1} B_{k_S}^\top, \qquad \gamma > 0.$$

Proposition B.4 (Score-space ridge bias)**.** *Assume A_{k_S} has at least one positive eigenvalue, and let*

$$\lambda_{k_S}^+ = \lambda_{\min}^+(A_{k_S})$$

denote its smallest strictly positive eigenvalue. Then

$$0 \preceq B_{k_S} A_{k_S}^+ B_{k_S}^\top - F_{w,k_S,\gamma},$$

and

$$\left\| B_{k_S} A_{k_S}^+ B_{k_S}^\top - F_{w,k_S,\gamma} \right\|_{\mathrm{op}} \leq \frac{\gamma}{\lambda_{k_S}^+ (\lambda_{k_S}^+ + \gamma)} \left\| B_{k_S} \right\|_{\mathrm{op}}^2 .$$

Proof. Let Π_A denote the orthogonal projector onto $\mathrm{Ran}(A_{k_S})$. By Lemma B.2,

$$B_{k_S} = B_{k_S} \Pi_A.$$

Therefore

$$\begin{aligned} &B_{k_S} A_{k_S}^+ B_{k_S}^\top - B_{k_S} (A_{k_S} + \gamma I)^{-1} B_{k_S}^\top \\ &\quad = B_{k_S} \Pi_A \left[A_{k_S}^+ - (A_{k_S} + \gamma I)^{-1} \right] \Pi_A B_{k_S}^\top. \end{aligned}$$

On $\mathrm{Ran}(A_{k_S})$, every eigenvalue satisfies $\lambda \geq \lambda_{k_S}^+ > 0$, and

$$\frac{1}{\lambda} - \frac{1}{\lambda + \gamma} = \frac{\gamma}{\lambda(\lambda + \gamma)} \geq 0.$$

Hence

$$0 \preceq B_{k_S} A_{k_S}^+ B_{k_S}^\top - F_{w,k_S,\gamma}.$$

Moreover,

$$\left\| \Pi_A \left[A_{k_S}^+ - (A_{k_S} + \gamma I)^{-1} \right] \Pi_A \right\|_{\mathrm{op}} \leq \frac{\gamma}{\lambda_{k_S}^+ (\lambda_{k_S}^+ + \gamma)}.$$

It follows that

$$\begin{aligned} \left\| B_{k_S} A_{k_S}^+ B_{k_S}^\top - F_{w,k_S,\gamma} \right\|_{\mathrm{op}} &\leq \left\| B_{k_S} \right\|_{\mathrm{op}}^2 \left\| \Pi_A \left[A_{k_S}^+ - (A_{k_S} + \gamma I)^{-1} \right] \Pi_A \right\|_{\mathrm{op}} \\ &\leq \frac{\gamma}{\lambda_{k_S}^+ (\lambda_{k_S}^+ + \gamma)} \left\| B_{k_S} \right\|_{\mathrm{op}}^2 . \end{aligned}$$

□

B.9 BLOCK-LOCAL RADIUS HIERARCHY

Proposition B.5 (Block-local radius hierarchy)**.** *For fixed parameter blocks,*

$$0 \preceq F_{\mathrm{blk}}^{(r)} \preceq F_{\mathrm{blk}}^{(r+1)} \preceq F_{\mathrm{BD}}^{\mathrm{global}}.$$

Proof. By Theorem 4.1, restricting the metric to each parameter block gives, for every b,

$$0 \preceq \left[F_{q_b}^{(r)} \right]_{\mathcal{B}_b \mathcal{B}_b} \preceq \left[F_{q_b}^{(r+1)} \right]_{\mathcal{B}_b \mathcal{B}_b} \preceq \left[F^{\mathrm{SLD}} \right]_{\mathcal{B}_b \mathcal{B}_b} .$$

Taking the direct sum over $b = 1, \ldots, n_B$ gives

$$0 \preceq F_{\mathrm{blk}}^{(r)} \preceq F_{\mathrm{blk}}^{(r+1)} \preceq F_{\mathrm{BD}}^{\mathrm{global}}.$$

□

C Worked Example of the Local-SLD Construction

Consider a five-qubit chain with depth $d = 2$. Each layer applies trainable Y and Z rotations on every qubit, followed by fixed nearest-neighbor CNOT gates. The CNOT gates are part of the noisy circuit but do not carry trainable parameters.

For each qubit b, the parameter block contains its four trainable rotation parameters from both layers:

$$\mathcal{B}_b = \{\theta_{Y,b,1}, \theta_{Z,b,1}, \theta_{Y,b,2}, \theta_{Z,b,2}\}, \qquad |\mathcal{B}_b| = 4.$$

There are five parameter blocks and $p = 20$ trainable parameters. The block-local preconditioner retains parameter couplings within each block and discards couplings between different blocks.

For radius r, let $w_b = w_r(b)$ denote the physical window associated with block $\mathcal{B}_b$. Now, consider the central block $\mathcal{B}_3$. With radius $r = 1$, its associated physical window is

$$w_3 = w_1(3) = \{2, 3, 4\}.$$

The corresponding reduced state and tangent operators are

$$\rho_{w_3} = \mathrm{tr}_{\{1,5\}}(\rho), \qquad T_{j,w_3} = \mathrm{tr}_{\{1,5\}}(T_j), \qquad j \in \mathcal{B}_3. \tag{13}$$

The four operators T_{j,w_3} describe how the same three-qubit reduced state changes under the four parameters in $\mathcal{B}_3$. The effects of the fixed entangling gates and circuit noise are already included in ρ and T_j. Although entangling gates can spread the effect of a parameter beyond its original qubit, the reduction in Equation (13) retains only the part visible on w_3.

For the three-qubit window w_3, let

$$\mathcal{S}_{w_3,k_S} = \{P : \mathrm{supp}(P) \subseteq w_3, 1 \le \mathrm{wt}(P) \le k_S\},$$
$$\mathcal{V}_{w_3,k_S} = \mathrm{span}_{\mathbb{R}}\left\{\overline{P} : P \in \mathcal{S}_{w_3,k_S}\right\}, \qquad \overline{P} = P - \mathrm{tr}(\rho_{w_3} P)\, I.$$

The number of retained Pauli score operators is

$$|\mathcal{S}_{w_3,k_S}| = \sum_{s=1}^{\min\{k_S,3\}} \binom{3}{s} 3^s.$$

Hence

$$|\mathcal{S}_{w_3,1}| = 9, \qquad |\mathcal{S}_{w_3,2}| = 36, \qquad |\mathcal{S}_{w_3,3}| = 63.$$

Here, $k_S = 1$ retains the weight-one Pauli strings, while $k_S = 2$ retains all Pauli strings of weights one and two. At $k_S = 3$, all $4^3 - 1 = 63$ nonidentity Pauli strings on w_3 are retained, giving the full local score space.

Define the retained Pauli strings by $\{P_a\}_{a=1}^{|\mathcal{S}_{w_3,k_S}|}$, with

$$\overline{P}_a = P_a - \mathrm{tr}(\rho_{w_3} P_a)\, I.$$

To construct the metric block for $\mathcal{B}_3$, we use only the response rows associated with its four parameters. The score-Gram and response matrices are

$$(A_{3,k_S})_{aa'} = \frac{1}{2}\mathrm{tr}\left(\rho_{w_3}\left\{\overline{P}_a, \overline{P}_{a'}\right\}\right),$$
$$(B_{3,k_S})_{ja} = \mathrm{tr}\left(T_{j,w_3}\overline{P}_a\right) = \mathrm{tr}\left(T_{j,w_3} P_a\right), \qquad j \in \mathcal{B}_3.$$

For the choice $k_S = 3$,

$$A_{3,3} \in \mathbb{R}^{63\times 63}, \qquad B_{3,3} \in \mathbb{R}^{4\times 63}.$$

Although A_3 has 63^2 entries, these entries are not measured independently. They are reconstructed from local Pauli moments, as discussed in Appendix E.

The unregularized metric block is

$$F_{3,k_S} = B_{3,k_S} A_{3,k_S}^{+} B_{3,k_S}^{\top}.$$

Let $F_3^{(1)}$ denote the exact reduced-state SLD-QFIM on $w_1(3)$. At $k_S = 3$, the score space is complete, so

$$F_{3,3} = B_{3,3} A_{3,3}^{+} B_{3,3}^{\top} = \left[F_3^{(1)}\right]_{\mathcal{B}_3\mathcal{B}_3}.$$

The regularized construction replaces the pseudoinverse by a ridge solve:

$$F_{3,k_S,\gamma} = B_{3,k_S} \left(A_{3,k_S} + \gamma I\right)^{-1} B_{3,k_S}^{\top}, \qquad \gamma > 0.$$

The final 4×4 matrix is generally dense. Writing the parameters in the order

$$(\theta_{Y,3,1}, \theta_{Z,3,1}, \theta_{Y,3,2}, \theta_{Z,3,2}),$$

we can write

$$F_{3,k_S,\gamma} = \begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{14} \\ f_{21} & f_{22} & f_{23} & f_{24} \\ f_{31} & f_{32} & f_{33} & f_{34} \\ f_{41} & f_{42} & f_{43} & f_{44} \end{pmatrix},$$

where $f_{ij} = f_{ji}$. Thus couplings between different parameters in $\mathcal{B}_3$, such as the coupling between $\theta_{Y,3,1}$ and $\theta_{Z,3,2}$, are retained. The block approximation does not diagonalize the parameters within $\mathcal{B}_3$.

Effect of k_S. Keeping the same physical window $w_3 = \{2, 3, 4\}$ but choosing $k_S = 2$ gives

$$A_{3,2} \in \mathbb{R}^{36\times 36}, \qquad B_{3,2} \in \mathbb{R}^{4\times 36}.$$

The unregularized restricted metrics satisfy

$$0 \preceq F_{3,1} \preceq F_{3,2} \preceq F_{3,3} = \left[F_3^{(1)}\right]_{\mathcal{B}_3\mathcal{B}_3}.$$

Thus r and k_S control different approximations. The radius r determines which physical qubits are retained, while k_S determines which Pauli weights are retained within that window.

Assembly of the block metric. For each block, write

$$F_{b,k_S,\gamma} = \left[F_{w_b,k_S,\gamma}\right]_{\mathcal{B}_b\mathcal{B}_b}.$$

The full block-local metric is then

$$F_{\mathrm{blk}}^{(r,k_S,\gamma)} = \bigoplus_{b=1}^{5} F_{b,k_S,\gamma}.$$

For the open five-qubit chain with $r = 1$, the five windows are

$$w_1 = \{1, 2\}, \quad w_2 = \{1, 2, 3\}, \quad w_3 = \{2, 3, 4\}, \quad w_4 = \{3, 4, 5\}, \quad w_5 = \{4, 5\}.$$

The interior windows contain three qubits and have $4^3 - 1 = 63$ complete nonidentity Pauli scores. The two boundary windows contain two qubits and have $4^2 - 1 = 15$. Since a Pauli string on w_b cannot have weight greater than $|w_b|$, the nominal cutoff $k_S = 3$ is automatically complete on the boundary windows.

Writing $F_b = F_{b,k_S,\gamma}$ for brevity, the preconditioner has the form

$$F_{\mathrm{blk}}^{(r,k_S,\gamma)} = \begin{pmatrix} F_1 & 0 & 0 & 0 & 0 \\ 0 & F_2 & 0 & 0 & 0 \\ 0 & 0 & F_3 & 0 & 0 \\ 0 & 0 & 0 & F_4 & 0 \\ 0 & 0 & 0 & 0 & F_5 \end{pmatrix}.$$

Each F_b is dense within its parameter block, while couplings between different blocks are discarded. For example, the coupling between $\theta_{Y,3,1}$ and $\theta_{Z,3,2}$ is retained because both parameters belong to $\mathcal{B}_3$. Here $w_3 \cap w_4 = \{3, 4\}$. Even though these windows overlap, the coupling between $\theta_{Y,3,1} \in \mathcal{B}_3$ and $\theta_{Z,4,2} \in \mathcal{B}_4$ is set to zero because the parameters belong to different blocks. Increasing r retains more SLD geometry within each block but does not restore couplings between blocks.

Finally, the block-local metric is used to precondition the noisy-objective gradient:

$$u_{\mathrm{blk}} = \left(F_{\mathrm{blk}}^{(r,k_S,\gamma)} + \mu I\right)^{-1} g_\eta, \qquad \mu > 0.$$

Because the metric is block diagonal, the solve can equivalently be performed independently within each parameter block $\mathcal{B}_b$.

D BLOCK-LOCAL SLD-QNG ALGORITHM

Algorithm 1 summarizes the finite-shot optimization procedure, including metric refreshes and reuse between refresh iterations.

Algorithm 1 Block-local SLD QNG

Require: Noisy circuit $\mathcal{E}_\eta(\overrightarrow{\theta})$, noisy objective C_η, initial parameters $\overrightarrow{\theta}^0$, parameter blocks $\{\mathcal{B}_b\}_{b=1}^{n_B}$, anchor qubits $\{q_b\}_{b=1}^{n_B}$, radius r, score cutoff k_S, score-space ridge γ, parameter ridge μ, step size α, refresh period R_{ref}, iteration limit $t_{\max}$

```
Set w_b = w_r(q_b) for every b
for t = 0, ..., t_max − 1 do
    Estimate the finite-shot gradient ĝ_η^t
    if t = 0 or t mod R_ref = 0 then
        for b = 1, ..., n_B do
            Estimate Â^t_{b,k_S} and B̂^t_{b,k_S}, and construct F̂^t_{b,k_S,γ} using Equation (7)
        end for
        Assemble F̂^t_blk = ⊕_b F̂^t_{b,k_S,γ}
    else
        Reuse F̂^t_blk = F̂^{t−1}_blk
    end if
    Solve (F̂^t_blk + μI) û^t_blk = ĝ^t_η
    Update θ⃗^{t+1} = θ⃗^t − α û^t_blk
end for
```

E FINITE-SHOT ESTIMATION

E.1 LOCAL SCORE-GRAM AND RESPONSE ESTIMATORS

For block $\mathcal{B}_b$, let $w_b = w_r(q_b)$ and we write the retained Pauli set as

$$\mathcal{S}_{w_b,k_S} = \{P_a\}_{a=1}^{m_b}, \qquad m_b = |\mathcal{S}_{w_b,k_S}|, \qquad p_b = |\mathcal{B}_b|.$$

For each retained Pauli string P_a, let

$$\nu_a = \mathrm{tr}\left(\rho_{w_b} P_a\right), \qquad \overline{P}_a = P_a - \nu_a I.$$

The score-Gram and response matrices for block $\mathcal{B}_b$ are given by

$$(A_{b,k_S})_{aa'} = \frac{1}{2}\mathrm{tr}\left(\rho_{w_b}\left\{\overline{P}_a, \overline{P}_{a'}\right\}\right), \quad a, a' = 1, \ldots, m_b,$$
$$(B_{b,k_S})_{ja} = \mathrm{tr}\left(T_{j,w_b} P_a\right), \quad j \in \mathcal{B}_b, \quad a = 1, \ldots, m_b.$$

Let

$$G_{aa'} = \frac{1}{2}\mathrm{tr}\left(\rho_{w_b}\{P_a, P_{a'}\}\right)$$

denote the uncentered Pauli moment. Substituting $\overline{P}_a = P_a - \nu_a I$ into the definition of A_{b,k_S} gives

$$\begin{aligned}(A_{b,k_S})_{aa'} &= \frac{1}{2}\mathrm{tr}\left(\rho_{w_b}\{P_a - \nu_a I, P_{a'} - \nu_{a'} I\}\right)\\ &= G_{aa'} - \nu_{a'}\mathrm{tr}\left(\rho_{w_b} P_a\right) - \nu_a \mathrm{tr}\left(\rho_{w_b} P_{a'}\right) + \nu_a \nu_{a'}\mathrm{tr}\left(\rho_{w_b}\right)\\ &= G_{aa'} - \nu_a \nu_{a'}. \end{aligned} \tag{14}$$

The value of $G_{aa'}$ depends on whether P_a and $P_{a'}$ commute:

$$G_{aa'} = \begin{cases} 0, & \{P_a, P_{a'}\} = 0,\\ \mathrm{tr}\left(\rho_{w_b} P_a P_{a'}\right), & [P_a, P_{a'}] = 0. \end{cases}$$

We split the measurements taken at $\overrightarrow{\theta}$ into two independent halves and set

$$\left(\widehat{A}_{b,k_S}\right)_{aa'} = \widehat{G}_{aa'} - \frac{1}{2}\left(\widehat{\nu}_a^{(1)}\widehat{\nu}_{a'}^{(2)} + \widehat{\nu}_a^{(2)}\widehat{\nu}_{a'}^{(1)}\right). \tag{15}$$

Here $\widehat{\nu}_a^{(1)}$ and $\widehat{\nu}_a^{(2)}$ are obtained from the two halves, while $\widehat{G}_{aa'}$ is estimated from the full measurement pool. We use the same estimate for $\widehat{G}_{aa'}$ and $\widehat{G}_{a'a}$, so $\widehat{A}_{b,k_S}$ is symmetric by construction.

Since partial trace commutes with differentiation,

$$T_{j,w_b} = \partial_{\theta_j}\rho_{w_b}.$$

For the noisy circuit with parameter-independent noise, the response entries are obtained by parameter shift

$$\begin{aligned}(B_{b,k_S})_{ja} &= \partial_{\theta_j}\operatorname{tr}(\rho_{w_b}P_a)\\ &= \nu_{ja}^{+} - \nu_{ja}^{-}, \qquad j \in \mathcal{B}_b,\end{aligned} \tag{16}$$

where $\nu_{ja}^{\pm}$ is the expectation of P_a after shifting θ_j by $\pm\pi/4$, with all other parameters fixed. Replacing these expectations by their finite-shot estimates $\widehat{\nu}_{ja}^{\pm}$ gives

$$\left(\widehat{B}_{b,k_S}\right)_{ja} = \widehat{\nu}_{ja}^{+} - \widehat{\nu}_{ja}^{-}. \tag{17}$$

E.2 Concentration of the score-Gram and response estimators

We now bound the error in the score-Gram estimator. Each Pauli expectation is estimated using all compatible measurement outcomes. Let $s_{A,b}$ be the minimum sample count among the Pauli expectations used to construct $\widehat{A}_{b,k_S}$. For cross-fitting, the samples are split equally between two disjoint subsets.

For anticommuting Pauli pairs, $G_{aa'} = 0$, so only commuting pairs contribute to $G_{aa'}$. Let $\mathcal{Q}_b$ denote the set of distinct nonidentity Pauli observables whose expectations are needed for these terms. Since they are supported on w_b, $|\mathcal{Q}_b| \le 4^{|w_b|} - 1$. Including the two cross-fit estimates of each ν_a, the total number of expectation estimates used in the concentration bound is

$$K_{A,b} = |\mathcal{Q}_b| + 2m_b. \tag{18}$$

Lemma E.1 (Score-Gram concentration)**.** *With probability at least $1 - \delta_{A,b}$, the score-Gram estimation error satisfies*

$$\begin{aligned}\epsilon_{A,b} &= \left\|\widehat{A}_{b,k_S} - A_{b,k_S}\right\|_{\mathrm{op}}\\ &\le (4+\sqrt{2})m_b\sqrt{\frac{\log(2K_{A,b}/\delta_{A,b})}{s_{A,b}}}.\end{aligned}$$

Proof. Each of the $K_{A,b}$ estimates in Equation (18) is a sample mean of independent measurement outcomes in $[-1,1]$. If $\overline{Z}_s = s^{-1}\sum_{i=1}^{s} Z_i$ denotes such a sample mean, then for every $t > 0$ Hoeffding's inequality (Hoeffding, 1963) gives

$$\Pr\left(\left|\overline{Z}_s - \mathbb{E}\left[\overline{Z}_s\right]\right| \ge t\right) \le 2\exp\left(-\frac{st^2}{2}\right).$$

To apply a union bound over the $K_{A,b}$ estimates, we bound the failure probability of each estimate by $\delta_{A,b}/K_{A,b}$. Since $\widehat{G}_{aa'}$ uses at least $s_{A,b}$ samples and each cross-fit estimate uses at least $s_{A,b}/2$ samples, we choose $t_{0,b}$ and $t_{1,b}$ so that

$$2\exp\left(-\frac{s_{A,b}t_{0,b}^2}{2}\right) = \frac{\delta_{A,b}}{K_{A,b}}, \qquad 2\exp\left(-\frac{s_{A,b}t_{1,b}^2}{4}\right) = \frac{\delta_{A,b}}{K_{A,b}}.$$

By solving we get

$$t_{0,b} = \sqrt{\frac{2\log\left(2K_{A,b}/\delta_{A,b}\right)}{s_{A,b}}}, \qquad t_{1,b} = 2\sqrt{\frac{\log\left(2K_{A,b}/\delta_{A,b}\right)}{s_{A,b}}}.$$

A union bound over the $K_{A,b}$ estimates then shows that, with probability at least $1-\delta_{A,b}$, all of these estimates satisfy

$$|\widehat{G}_{aa'} - G_{aa'}| \le t_{0,b}, \qquad \left|\widehat{\nu}_a^{(h)} - \nu_a\right| \le t_{1,b},$$

where $h \in \{1,2\}$ labels the two cross-fit subsets.

Because Pauli expectations lie in $[-1,1]$,

$$\begin{aligned}\left|\widehat{\nu}_a^{(1)}\widehat{\nu}_{a'}^{(2)} - \nu_a\nu_{a'}\right| &= \left|\left(\widehat{\nu}_a^{(1)} - \nu_a\right)\widehat{\nu}_{a'}^{(2)} + \nu_a\left(\widehat{\nu}_{a'}^{(2)} - \nu_{a'}\right)\right| \\ &\le \left|\widehat{\nu}_a^{(1)} - \nu_a\right|\left|\widehat{\nu}_{a'}^{(2)}\right| + |\nu_a|\left|\widehat{\nu}_{a'}^{(2)} - \nu_{a'}\right| \\ &\le 2t_{1,b}.\end{aligned}$$

The same bound holds for $\widehat{\nu}_a^{(2)}\widehat{\nu}_{a'}^{(1)}$. Combining these bounds with Equations (14) and (15) gives

$$\begin{aligned}\max_{a,a'}\left|\left(\widehat{A}_{b,k_S} - A_{b,k_S}\right)_{aa'}\right| &\le t_{0,b} + 2t_{1,b} \\ &= (4+\sqrt{2})\sqrt{\frac{\log\left(2K_{A,b}/\delta_{A,b}\right)}{s_{A,b}}}.\end{aligned}$$

Since $\widehat{A}_{b,k_S} - A_{b,k_S}$ is symmetric,

$$\begin{aligned}\left\|\widehat{A}_{b,k_S} - A_{b,k_S}\right\|_{\mathrm{op}} &\le \left\|\widehat{A}_{b,k_S} - A_{b,k_S}\right\|_{\infty} \\ &\le m_b \max_{a,a'}\left|\left(\widehat{A}_{b,k_S} - A_{b,k_S}\right)_{aa'}\right|.\end{aligned}$$

Combining this with the entrywise bound above proves the lemma. □

For the response estimator, let $s_{B,b}$ be the minimum sample count among the shifted Pauli expectations used to construct $\widehat{B}_{b,k_S}$.

Lemma E.2 (Response-matrix concentration)**.** *Assume that the measurement outcomes from the $+$ and $-$ shifted circuits are independent. With probability at least $1-\delta_{B,b}$, the response-matrix estimation error satisfies*

$$\begin{aligned}\epsilon_{B,b} &= \left\|\widehat{B}_{b,k_S} - B_{b,k_S}\right\|_{\mathrm{op}} \\ &\le 2\sqrt{\frac{p_b m_b \log\left(2p_b m_b/\delta_{B,b}\right)}{s_{B,b}}}.\end{aligned}$$

Proof. For a fixed pair (j,a), let

$$\Delta_{ja} = \left(\widehat{B}_{b,k_S} - B_{b,k_S}\right)_{ja}.$$

By Equations (16) and (17),

$$\Delta_{ja} = \left(\widehat{\nu}_{ja}^{+} - \nu_{ja}^{+}\right) - \left(\widehat{\nu}_{ja}^{-} - \nu_{ja}^{-}\right).$$

Each shifted expectation uses at least $s_{B,b}$ independent samples. Since the $+$ and $-$ shifted estimates are independent, Hoeffding's inequality gives, for every $t > 0$,

$$\Pr\left(|\Delta_{ja}| \ge t\right) \le 2\exp\left(-\frac{s_{B,b}t^2}{4}\right).$$

To apply a union bound over the $p_b m_b$ entries, we bound the failure probability of each entry by $\delta_{B,b}/(p_b m_b)$. We choose $t_{B,b}$ so that

$$2\exp\left(-\frac{s_{B,b}t_{B,b}^2}{4}\right) = \frac{\delta_{B,b}}{p_b m_b}.$$

By solving we get,

$$t_{B,b} = 2\sqrt{\frac{\log\left(2p_b m_b/\delta_{B,b}\right)}{s_{B,b}}}.$$

The union bound then shows that, with probability at least $1-\delta_{B,b}$,

$$\max_{j,a}|\Delta_{ja}| \le t_{B,b}.$$

Finally,

$$\begin{aligned}\left\|\widehat{B}_{b,k_S} - B_{b,k_S}\right\|_{\mathrm{op}} &\le \left\|\widehat{B}_{b,k_S} - B_{b,k_S}\right\|_F \\ &\le \sqrt{p_b m_b}\,\max_{j,a}|\Delta_{ja}|.\end{aligned}$$

Substituting the bound on $\max_{j,a}|\Delta_{ja}|$ proves the lemma. □

E.3 Local-metric stability

Let $\beta_b = \|B_{b,k_S}\|_{\mathrm{op}}$. By Equation (16),

$$|(B_{b,k_S})_{ja}| = \left|\nu_{ja}^+ - \nu_{ja}^-\right| \le |\nu_{ja}^+| + |\nu_{ja}^-| \le 2.$$

Since B_{b,k_S} has $p_b m_b$ entries,

$$\begin{aligned}\beta_b = \|B_{b,k_S}\|_{\mathrm{op}} &\le \|B_{b,k_S}\|_F \\ &\le \sqrt{p_b m_b}\max_{j,a}|(B_{b,k_S})_{ja}| \\ &\le 2\sqrt{p_b m_b}.\end{aligned}$$

Before adding the ridge, we replace $\widehat{A}_{b,k_S}$ by its PSD projection, as in Equation (7).

Proposition E.3 (Finite-shot stability of a local block)**.** *For $\gamma > 0$, assume*

$$\left\|\widehat{A}_{b,k_S} - A_{b,k_S}\right\|_{\mathrm{op}} \le \epsilon_{A,b}, \qquad \left\|\widehat{B}_{b,k_S} - B_{b,k_S}\right\|_{\mathrm{op}} \le \epsilon_{B,b}.$$

Then

$$\left\|\widehat{F}_{b,k_S,\gamma} - F_{b,k_S,\gamma}\right\|_{\mathrm{op}} \le \frac{2\beta_b\epsilon_{B,b} + \epsilon_{B,b}^2}{\gamma} + \frac{2\beta_b^2\epsilon_{A,b}}{\gamma^2}.$$

Proof. Since $A_{b,k_S} \succeq 0$, Weyl's inequality gives

$$\begin{aligned}\lambda_{\min}\left(\widehat{A}_{b,k_S}\right) &\ge \lambda_{\min}\left(A_{b,k_S}\right) - \left\|\widehat{A}_{b,k_S} - A_{b,k_S}\right\|_{\mathrm{op}} \\ &\ge -\epsilon_{A,b}.\end{aligned}$$

The PSD projection replaces the negative eigenvalues of $\widehat{A}_{b,k_S}$ by zero. Therefore,

$$\left\|\Pi_{\succeq 0}\left(\widehat{A}_{b,k_S}\right) - \widehat{A}_{b,k_S}\right\|_{\mathrm{op}} = \max\left\{0, -\lambda_{\min}\left(\widehat{A}_{b,k_S}\right)\right\} \le \epsilon_{A,b}.$$

Adding and subtracting $\widehat{A}_{b,k_S}$ and applying the triangle inequality gives

$$\begin{aligned}\left\|\Pi_{\succeq 0}\left(\widehat{A}_{b,k_S}\right) - A_{b,k_S}\right\|_{\mathrm{op}} &\le \left\|\Pi_{\succeq 0}\left(\widehat{A}_{b,k_S}\right) - \widehat{A}_{b,k_S}\right\|_{\mathrm{op}} + \left\|\widehat{A}_{b,k_S} - A_{b,k_S}\right\|_{\mathrm{op}} \\ &\le 2\epsilon_{A,b}. \qquad (19)\end{aligned}$$

Let

$$M_b = A_{b,k_S} + \gamma I, \qquad \widehat{M}_b = \Pi_{\succeq 0}\left(\widehat{A}_{b,k_S}\right) + \gamma I.$$

Both matrices satisfy $M_b \succeq \gamma I$ and $\widehat{M}_b \succeq \gamma I$. So their smallest eigenvalues are at least γ, and

$$\left\|M_b^{-1}\right\|_{\mathrm{op}} \le \frac{1}{\gamma}, \qquad \left\|\widehat{M}_b^{-1}\right\|_{\mathrm{op}} \le \frac{1}{\gamma}.$$

Using the resolvent identity we get,

$$\widehat{M}_b^{-1} - M_b^{-1} = \widehat{M}_b^{-1}\left(M_b - \widehat{M}_b\right) M_b^{-1}.$$

Also,

$$\widehat{M}_b - M_b = \Pi_{\succeq 0}\left(\widehat{A}_{b,k_S}\right) - A_{b,k_S},$$

so Equation (19) gives $\left\|\widehat{M}_b - M_b\right\|_{\mathrm{op}} \le 2\epsilon_{A,b}$. Combining these bounds gives

$$\begin{aligned}\left\|\widehat{M}_b^{-1} - M_b^{-1}\right\|_{\mathrm{op}} &\le \left\|\widehat{M}_b^{-1}\right\|_{\mathrm{op}} \left\|\widehat{M}_b - M_b\right\|_{\mathrm{op}} \left\|M_b^{-1}\right\|_{\mathrm{op}} \\ &\le \frac{2\epsilon_{A,b}}{\gamma^2}.\end{aligned}$$

The population and estimated block metrics are

$$F_{b,k_S,\gamma} = B_{b,k_S} M_b^{-1} B_{b,k_S}^\top,$$
$$\widehat{F}_{b,k_S,\gamma} = \widehat{B}_{b,k_S} \widehat{M}_b^{-1} \widehat{B}_{b,k_S}^\top.$$

To bound the estimation error, we expand $\widehat{F}_{b,k_S,\gamma} - F_{b,k_S,\gamma}$ by adding and subtracting $B_{b,k_S}\widehat{M}_b^{-1}\widehat{B}_{b,k_S}^\top$ and $B_{b,k_S}\widehat{M}_b^{-1} B_{b,k_S}^\top$

$$\begin{aligned}\widehat{F}_{b,k_S,\gamma} - F_{b,k_S,\gamma} &= \left(\widehat{B}_{b,k_S} - B_{b,k_S}\right)\widehat{M}_b^{-1}\widehat{B}_{b,k_S}^\top \\ &\quad + B_{b,k_S}\widehat{M}_b^{-1}\left(\widehat{B}_{b,k_S} - B_{b,k_S}\right)^\top \\ &\quad + B_{b,k_S}\left(\widehat{M}_b^{-1} - M_b^{-1}\right) B_{b,k_S}^\top.\end{aligned}$$

The triangle inequality gives

$$\begin{aligned}\left\|\widehat{B}_{b,k_S}\right\|_{\mathrm{op}} &\le \left\|B_{b,k_S}\right\|_{\mathrm{op}} + \left\|\widehat{B}_{b,k_S} - B_{b,k_S}\right\|_{\mathrm{op}} \\ &\le \beta_b + \epsilon_{B,b}.\end{aligned}$$

Applying the norm bounds to the three terms above yields

$$\begin{aligned}\left\|\widehat{F}_{b,k_S,\gamma} - F_{b,k_S,\gamma}\right\|_{\mathrm{op}} &\le \frac{\epsilon_{B,b}\left(\beta_b + \epsilon_{B,b}\right)}{\gamma} + \frac{\beta_b \epsilon_{B,b}}{\gamma} + \frac{2\beta_b^2 \epsilon_{A,b}}{\gamma^2} \\ &= \frac{2\beta_b\epsilon_{B,b} + \epsilon_{B,b}^2}{\gamma} + \frac{2\beta_b^2\epsilon_{A,b}}{\gamma^2}.\end{aligned}$$

This proves the proposition. □

E.4 Local-block sample complexity

Fix a target block-metric error $\varepsilon > 0$. The local metric stability bound in Proposition E.3 contains one term involving $\epsilon_{A,b}$ and two involving $\epsilon_{B,b}$. For $\beta_b > 0$, it is sufficient to bound each of the three terms in Proposition E.3 by $\varepsilon/3$. This gives

$$\epsilon_{A,b} \le \frac{\varepsilon\gamma^2}{6\beta_b^2}, \qquad \epsilon_{B,b} \le \min\left\{\frac{\varepsilon\gamma}{6\beta_b}, \sqrt{\frac{\varepsilon\gamma}{3}}\right\}.$$

By Lemma E.1, the first condition is satisfied if

$$(4+\sqrt{2})m_b\sqrt{\frac{\log\left(2K_{A,b}/\delta_{A,b}\right)}{s_{A,b}}} \le \frac{\varepsilon\gamma^2}{6\beta_b^2}.$$

Solving for $s_{A,b}$ gives

$$s_{A,b} \ge 36(4+\sqrt{2})^2\frac{m_b^2\beta_b^4}{\varepsilon^2\gamma^4}\log\left(\frac{2K_{A,b}}{\delta_{A,b}}\right).$$

Similarly, Lemma E.2 shows that the condition on $\epsilon_{B,b}$ is satisfied if

$$2\sqrt{\frac{p_b m_b \log\left(2p_b m_b/\delta_{B,b}\right)}{s_{B,b}}} \le \min\left\{\frac{\varepsilon\gamma}{6\beta_b}, \sqrt{\frac{\varepsilon\gamma}{3}}\right\}.$$

Solving for $s_{B,b}$ gives

$$s_{B,b} \ge p_b m_b \log\left(\frac{2p_b m_b}{\delta_{B,b}}\right)\max\left\{\frac{144\beta_b^2}{\varepsilon^2\gamma^2}, \frac{12}{\varepsilon\gamma}\right\}.$$

Up to numerical constants, the sufficient sample counts can be written as

$$s_{A,b} = O\left[\frac{m_b^2\beta_b^4}{\varepsilon^2\gamma^4}\log\left(\frac{2K_{A,b}}{\delta_{A,b}}\right)\right], \tag{20}$$

$$s_{B,b} = O\left[p_b m_b \log\left(\frac{2p_b m_b}{\delta_{B,b}}\right)\max\left\{\frac{\beta_b^2}{\varepsilon^2\gamma^2}, \frac{1}{\varepsilon\gamma}\right\}\right]. \tag{21}$$

If $\beta_b = 0$, the stability bound reduces to $\epsilon_{B,b}^2/\gamma$. So no accuracy condition on $\widehat{A}_{b,k_S}$ is needed, and only the $1/(\varepsilon\gamma)$ term in Equation (21) remains.

The quantities $s_{A,b}$ and $s_{B,b}$ are the minimum sample counts for the required Pauli expectations, rather than total circuit-shot counts. A conservative independent-block construction uses all $3^{|w_b|}$ product-Pauli settings on the unshifted circuit for A_{b,k_S} and on both shifted circuits for each of the p_b parameters for B_{b,k_S}. For this construction, the total number of circuit shots required for block b satisfies

$$N_b^{\mathrm{ind}} \le 3^{|w_b|}\left(s_{A,b} + 2p_b s_{B,b}\right). \tag{22}$$

This schedule is sufficient but not necessarily optimal, since compatible measurements can be reused across observables and overlapping blocks.

E.5 Fixed-locality scaling

The local bounds extend to the full block-local metric through its direct-sum structure:

$$\left\|\widehat{F}_{\mathrm{blk}}^{(r,k_S,\gamma)} - F_{\mathrm{blk}}^{(r,k_S,\gamma)}\right\|_{\mathrm{op}} = \max_{1\le b\le n_B}\left\|\widehat{F}_{b,k_S,\gamma} - F_{b,k_S,\gamma}\right\|_{\mathrm{op}}. \tag{23}$$

Here, the dependence of each block metric on r is implicit through $w_b = w_r(q_b)$.

Corollary E.4 (Fixed-locality finite-shot scaling)**.** *Suppose r, k_S, and $\gamma > 0$ are fixed independently of n, and assume*

$$\max_b |w_r(q_b)| = O(1), \qquad \max_b |\mathcal{B}_b| = O(1), \qquad n_B = O(n).$$

Then, for every $\varepsilon, \delta \in (0,1)$, there exists a local product-Pauli measurement schedule such that

$$\Pr\left[\left\|\widehat{F}_{\mathrm{blk}}^{(r,k_S,\gamma)} - F_{\mathrm{blk}}^{(r,k_S,\gamma)}\right\|_{\mathrm{op}} \le \varepsilon\right] \ge 1-\delta.$$

A sufficient number of circuit shots is

$$N_{\mathrm{geom}} = O\left[n\log\left(\frac{n}{\delta}\right)\left(\frac{1}{\varepsilon^2\gamma^4} + \frac{1}{\varepsilon^2\gamma^2} + \frac{1}{\varepsilon\gamma}\right)\right]. \tag{24}$$

For fixed $\gamma > 0$ and $\varepsilon \in (0,1)$, this reduces to

$$N_{\mathrm{geom}} = O\left(\frac{n}{\varepsilon^2}\log\frac{n}{\delta}\right). \tag{25}$$

Proof. The bounded-window assumption and fixed k_S give

$$m_b = \sum_{\ell=1}^{\min\{k_S,|w_b|\}} \binom{|w_b|}{\ell} 3^\ell = O(1).$$

The bounded-block assumption gives

$$p_b = |\mathcal{B}_b| = O(1),$$

and therefore

$$\beta_b \le 2\sqrt{p_b m_b} = O(1).$$

Moreover, Equation (18) and $|\mathcal{Q}_b| \le 4^{|w_b|} - 1$ imply

$$K_{A,b} = O(1).$$

Thus m_b, p_b, β_b, and $K_{A,b}$ are all $O(1)$, and a complete product-Pauli schedule also contains at most $3^{|w_b|} = O(1)$ measurement settings.

For every block, choose

$$\delta_{A,b} = \delta_{B,b} = \frac{\delta}{2n_B}.$$

Since $K_{A,b}$, p_b, and m_b are all $O(1)$,

$$\log\left(\frac{2K_{A,b}}{\delta_{A,b}}\right) = \log\left(\frac{4K_{A,b}n_B}{\delta}\right) = O\left(\log\frac{n_B}{\delta}\right),$$
$$\log\left(\frac{2p_b m_b}{\delta_{B,b}}\right) = \log\left(\frac{4p_b m_b n_B}{\delta}\right) = O\left(\log\frac{n_B}{\delta}\right).$$

Substituting these bounds into Equations (20) and (21) and using Equation (22) gives

$$N_b^{\mathrm{ind}} = O\left[\log\left(\frac{n_B}{\delta}\right)\left(\frac{1}{\varepsilon^2\gamma^4} + \frac{1}{\varepsilon^2\gamma^2} + \frac{1}{\varepsilon\gamma}\right)\right]$$

for every block.

The probability that any score-Gram or response concentration event fails is at most

$$\sum_{b=1}^{n_B} (\delta_{A,b} + \delta_{B,b}) = \delta.$$

Hence, with probability at least $1-\delta$, every block has estimation error at most ε. By Equation (23), the full block-local metric has the same error bound.

For the independent-block construction,

$$\begin{aligned} N_{\mathrm{geom}} &= \sum_{b=1}^{n_B} N_b^{\mathrm{ind}} \\ &= O\left[n_B \log\left(\frac{n_B}{\delta}\right)\left(\frac{1}{\varepsilon^2\gamma^4} + \frac{1}{\varepsilon^2\gamma^2} + \frac{1}{\varepsilon\gamma}\right)\right] \\ &= O\left[n \log\left(\frac{n}{\delta}\right)\left(\frac{1}{\varepsilon^2\gamma^4} + \frac{1}{\varepsilon^2\gamma^2} + \frac{1}{\varepsilon\gamma}\right)\right], \end{aligned}$$

where the last line uses $n_B = O(n)$. This proves Equation (24).

Finally, because $\gamma > 0$ is fixed, the powers of γ^{-1} are constants. Since $\varepsilon \in (0,1)$, $1/\varepsilon \le 1/\varepsilon^2$. Therefore, Equation (24) reduces to Equation (25). □

This result controls the estimation error relative to the regularized metric $F_{\mathrm{blk}}^{(r,k_S,\gamma)}$. Comparison with the unregularized metric $F_{\mathrm{blk}}^{(r,k_S)}$ also requires the ridge-bias bound in Proposition B.4.

Remark E.5 (Dense interaction graphs). The corollary assumes that the window sizes remain bounded independently of n. This holds for bounded-degree interaction graphs at fixed r. For an all-to-all interaction graph, however, $w_r(q) = V$ for every $r \geq 1$, so the assumption fails and the fixed-locality finite-shot bound does not apply.

Keeping k_S fixed still limits the score dimension to

$$m_b = \sum_{\ell=1}^{k_S} \binom{n}{\ell} 3^\ell = O(n^{k_S}),$$

instead of $4^n - 1$. Thus the score-space construction remains well defined, but the $O(n \log n)$ metric-estimation bound at fixed accuracy and confidence no longer follows from Corollary E.4.

E.6 Measurement reuse

Reuse across score-Gram matrices. The scaling bound in Appendix E.5 assumes that each block is measured independently. For the score-Gram matrices, a single product-Pauli measurement of the full n-qubit system can provide data for several local windows. Let $\chi = (\chi_1, \ldots, \chi_n)$, with $\chi_j \in \{X, Y, Z\}$, denote the Pauli basis measured on each qubit. The outcomes from one such setting can be used to estimate any local Pauli observable whose nonidentity factors agree with χ, and therefore can contribute to entries of A_{b,k_S} for multiple blocks.

On a metric-refresh iteration, let $s_b(\chi)$ denote the number of shots required by block b from setting χ, with $s_b(\chi) = 0$ when that setting is not needed. Since the same outcomes can be reused by all compatible blocks, it is sufficient to take $\max_b s_b(\chi)$ shots for setting χ. Hence,

$$\sum_{\chi} \max_b s_b(\chi) \leq \sum_{\chi} \sum_{b=1}^{n_B} s_b(\chi).$$

For example, suppose two overlapping windows require measurements X_1Z_2 and Z_2Y_3. Both are compatible with the full-system setting $\chi = (X, Z, Y, \ldots)$, so the same circuit shots provide data for both windows. In contrast, X_1Z_2 and X_2Y_3 cannot be obtained from the same setting because they require different measurement bases on qubit 2. Thus the same unshifted measurements can be reused across blocks.

Reuse between response and gradient estimation. Both the response matrix and the gradient are estimated from parameter-shifted circuits. On a metric-refresh iteration, measurements used to estimate B_{b,k_S} may therefore also contribute to the gradient when the required Pauli measurement settings are compatible. These forms of reuse can reduce the practical shot cost relative to measuring all quantities independently.

Relation to target-aware Pauli measurement. The measurement scheme above is tailored to the Pauli observables required by the retained local score spaces. This is related to target-aware strategies such as derandomized and locally biased classical shadows (Huang et al., 2021; Hadfield et al., 2022). Rather than sampling local Pauli bases uniformly, we use the settings required by the score-Gram and response estimators and reuse compatible measurements across targets. More optimized derandomized or biased schedules may further reduce the measurement cost when the retained target set is sparse.

F Convergence analysis

The convergence analysis follows standard smooth nonconvex stochastic optimization arguments (Ghadimi & Lan, 2013; Bottou et al., 2018), adapted here to the regularized block-local preconditioner.

Throughout this section, the noise parameters η are fixed. For $t = 0, \ldots, t_{\max} - 1$, let

$$M_t = \widehat{F}^t_{\mathrm{blk}} + \mu I, \qquad g^t_\eta = \nabla C_\eta(\overrightarrow{\theta}^t), \qquad \widehat{g}^t_\eta = g^t_\eta + \xi_t.$$

The QNG update is

$$\overrightarrow{\theta}^{t+1} = \overrightarrow{\theta}^{t} - \alpha M_t^{-1} \widehat{g}_\eta^t.$$

We assume

$$\mathbb{E}\left[\xi_t \mid \overrightarrow{\theta}^{t}, M_t\right] = 0, \qquad \mathbb{E}\left[\|\xi_t\|_2^2 \mid \overrightarrow{\theta}^{t}, M_t\right] \le \sigma_{g_\eta}^2. \tag{26}$$

Equation (26) is guaranteed by an independent gradient-estimation batch and remains valid under any measurement-reuse scheme that preserves conditional unbiasedness.

F.1 Spectral bounds

Lemma F.1 (Uniform spectral bounds)**.** *For every optimization iteration t,*

$$\mu I \preceq M_t \preceq \Lambda_{\max} I,$$

where

$$\Lambda_{\max} = \mu + \frac{4}{\gamma} \max_b \{|\mathcal{B}_b||\mathcal{S}_{w_b,k_S}|\}.$$

Proof. Consider an iteration t at which the metric is refreshed. For each block, define

$$\widetilde{A}_{b,k_S}^t = \Pi_{\succeq 0}\left(\widehat{A}_{b,k_S}^t\right).$$

Then

$$\widehat{F}_{b,k_S,\gamma}^t = \widehat{B}_{b,k_S}^t \left(\widetilde{A}_{b,k_S}^t + \gamma I\right)^{-1} \widehat{B}_{b,k_S}^{t\top}.$$

Since $\widetilde{A}_{b,k_S}^t \succeq 0$ and $\gamma > 0$,

$$\widetilde{A}_{b,k_S}^t + \gamma I \succeq \gamma I \succ 0.$$

Inversion reverses the Loewner order for positive-definite matrices, so

$$0 \prec \left(\widetilde{A}_{b,k_S}^t + \gamma I\right)^{-1} \preceq \frac{1}{\gamma} I.$$

For any vector x we can write,

$$x^\top \widehat{F}_{b,k_S,\gamma}^t x = \left(\widehat{B}_{b,k_S}^{t\top} x\right)^\top \left(\widetilde{A}_{b,k_S}^t + \gamma I\right)^{-1} \left(\widehat{B}_{b,k_S}^{t\top} x\right) \ge 0.$$

Hence $\widehat{F}_{b,k_S,\gamma}^t \succeq 0$.

Now we have,

$$\begin{aligned}\left\|\widehat{F}_{b,k_S,\gamma}^t\right\|_{\mathrm{op}} &\le \left\|\widehat{B}_{b,k_S}^t\right\|_{\mathrm{op}}^2 \left\|\left(\widetilde{A}_{b,k_S}^t + \gamma I\right)^{-1}\right\|_{\mathrm{op}} \\ &\le \frac{\left\|\widehat{B}_{b,k_S}^t\right\|_{\mathrm{op}}^2}{\gamma}.\end{aligned} \tag{27}$$

Let $p_b = |\mathcal{B}_b|$ and $m_b = |\mathcal{S}_{w_b,k_S}|$. Each entry of $\widehat{B}_{b,k_S}^t$ is the difference of two empirical Pauli means, each lying in $[-1, 1]$:

$$\left|\left(\widehat{B}_{b,k_S}^t\right)_{ja}\right| \le 2.$$

Therefore,

$$\left\|\widehat{B}_{b,k_S}^t\right\|_{\mathrm{op}}^2 \le \left\|\widehat{B}_{b,k_S}^t\right\|_F^2 \le 4 p_b m_b.$$

Combining this with Equation (27) gives

$$\left\|\widehat{F}^t_{b,k_S,\gamma}\right\|_{\mathrm{op}} \le \frac{4p_b m_b}{\gamma}. \tag{28}$$

The empirical metric is a direct sum,

$$\widehat{F}^t_{\mathrm{blk}} = \bigoplus_{b=1}^{n_B} \widehat{F}^t_{b,k_S,\gamma}.$$

Direct sums preserve positive semidefiniteness, and the operator norm of a block-diagonal matrix is the maximum operator norm of its blocks. Therefore,

$$\begin{aligned} \widehat{F}^t_{\mathrm{blk}} &\succeq 0, \\ \left\|\widehat{F}^t_{\mathrm{blk}}\right\|_{\mathrm{op}} &= \max_b \left\|\widehat{F}^t_{b,k_S,\gamma}\right\|_{\mathrm{op}} \\ &\le \frac{4}{\gamma} \max_b \{p_b m_b\}. \end{aligned}$$

Because $\widehat{F}^t_{\mathrm{blk}}$ is positive semidefinite, the operator-norm bound implies

$$0 \preceq \widehat{F}^t_{\mathrm{blk}} \preceq \frac{4}{\gamma} \max_b \{p_b m_b\} I.$$

Adding μI yields

$$\mu I \preceq M_t = \widehat{F}^t_{\mathrm{blk}} + \mu I \preceq \left(\mu + \frac{4}{\gamma} \max_b \{p_b m_b\}\right) I.$$

By the definitions of p_b and m_b, this upper bound is $\Lambda_{\max} I$.

The bound above holds whenever the metric is refreshed. At non-refresh iterations, the algorithm reuses the most recently refreshed metric, so the same spectral bound holds at every optimization iteration t. □

F.2 Expected one-step descent

Lemma F.2 (Expected one-step descent)**.** *Assume that C_η is L_C-smooth and that Equation (26) holds. If $0 < \alpha \le \mu/L_C$, then*

$$\mathbb{E}\left[C_\eta(\overrightarrow{\theta}^{t+1}) \mid \overrightarrow{\theta}^t, M_t\right] \le C_\eta(\overrightarrow{\theta}^t) - \frac{\alpha}{2}(g^t_\eta)^\top M_t^{-1} g^t_\eta + \frac{L_C \alpha^2}{2\mu^2}\sigma^2_{g_\eta}. \tag{29}$$

Proof. Condition on $(\overrightarrow{\theta}^t, M_t)$ and write

$$g^t_\eta = \nabla C_\eta(\overrightarrow{\theta}^t), \qquad \Delta_t = \overrightarrow{\theta}^{t+1} - \overrightarrow{\theta}^t = -\alpha M_t^{-1}(g^t_\eta + \xi_t).$$

By L_C-smoothness,

$$C_\eta(\overrightarrow{\theta}^{t+1}) \le C_\eta(\overrightarrow{\theta}^t) + (g^t_\eta)^\top \Delta_t + \frac{L_C}{2}\|\Delta_t\|_2^2. \tag{30}$$

For the linear term,

$$\begin{aligned} \mathbb{E}\left[(g^t_\eta)^\top \Delta_t \mid \overrightarrow{\theta}^t, M_t\right] &= -\alpha (g^t_\eta)^\top M_t^{-1} \mathbb{E}\left[g^t_\eta + \xi_t \mid \overrightarrow{\theta}^t, M_t\right] \\ &= -\alpha (g^t_\eta)^\top M_t^{-1} g^t_\eta, \end{aligned} \tag{31}$$

where the second equality uses $\mathbb{E}\left[\xi_t \mid \overrightarrow{\theta}^t, M_t\right] = 0$.

For the quadratic term,

$$\|\Delta_t\|_2^2 = \alpha^2 (g^t_\eta + \xi_t)^\top M_t^{-2} (g^t_\eta + \xi_t).$$

Taking conditional expectation and using $\mathbb{E}\left[\xi_t \mid \overrightarrow{\theta}^t, M_t\right] = 0$ gives

$$\mathbb{E}\left[\|\Delta_t\|_2^2 \mid \overrightarrow{\theta}^t, M_t\right] = \alpha^2 (g_\eta^t)^\top M_t^{-2} g_\eta^t + \alpha^2 \mathbb{E}\left[\xi_t^\top M_t^{-2} \xi_t \mid \overrightarrow{\theta}^t, M_t\right]. \tag{32}$$

By Lemma F.1, $M_t \succeq \mu I$. Hence every eigenvalue of M_t is at least μ, which implies

$$M_t^{-2} \preceq \frac{1}{\mu} M_t^{-1}.$$

Therefore,

$$(g_\eta^t)^\top M_t^{-2} g_\eta^t \leq \frac{1}{\mu} (g_\eta^t)^\top M_t^{-1} g_\eta^t. \tag{33}$$

The same spectral bound gives

$$\left\|M_t^{-1}\right\|_{\mathrm{op}} \leq \frac{1}{\mu}.$$

Consequently,

$$\begin{aligned}\mathbb{E}\left[\xi_t^\top M_t^{-2} \xi_t \mid \overrightarrow{\theta}^t, M_t\right] &\leq \frac{1}{\mu^2} \mathbb{E}\left[\|\xi_t\|_2^2 \mid \overrightarrow{\theta}^t, M_t\right] \\ &\leq \frac{\sigma_{g_\eta}^2}{\mu^2}.\end{aligned} \tag{34}$$

Substituting Equations (31), (33) and (34) into Equation (30) yields

$$\begin{aligned}\mathbb{E}\left[C_\eta(\overrightarrow{\theta}^{t+1}) \mid \overrightarrow{\theta}^t, M_t\right] \leq C_\eta(\overrightarrow{\theta}^t) - \alpha \left(1 - \frac{L_C \alpha}{2\mu}\right) (g_\eta^t)^\top M_t^{-1} g_\eta^t \\ + \frac{L_C \alpha^2}{2\mu^2} \sigma_{g_\eta}^2.\end{aligned}$$

Since $\alpha \leq \mu / L_C$,

$$1 - \frac{L_C \alpha}{2\mu} \geq \frac{1}{2}.$$

This proves Equation (29). □

F.3 Proof of Theorem 5.1

Proof. Taking total expectation in Equation (29) gives

$$\mathbb{E}\left[C_\eta(\overrightarrow{\theta}^t) - C_\eta(\overrightarrow{\theta}^{t+1})\right] \geq \frac{\alpha}{2} \mathbb{E}\left[(g_\eta^t)^\top M_t^{-1} g_\eta^t\right] - \frac{L_C \alpha^2}{2\mu^2} \sigma_{g_\eta}^2. \tag{35}$$

By Lemma F.1,

$$M_t \preceq \Lambda_{\max} I.$$

Since $M_t \succ 0$, inversion reverses the Loewner order:

$$M_t^{-1} \succeq \frac{1}{\Lambda_{\max}} I.$$

Hence

$$(g_\eta^t)^\top M_t^{-1} g_\eta^t \geq \frac{1}{\Lambda_{\max}} \left\|g_\eta^t\right\|_2^2. \tag{36}$$

Substituting Equation (36) into Equation (35) and summing over $t = 0, \ldots, t_{\max} - 1$ gives

$$\frac{\alpha}{2\Lambda_{\max}} \sum_{t=0}^{t_{\max}-1} \mathbb{E}\left[\left\|g_\eta^t\right\|_2^2\right] \leq C_\eta(\overrightarrow{\theta}^0) - \mathbb{E}\left[C_\eta(\overrightarrow{\theta}^{t_{\max}})\right] + \frac{L_C \alpha^2 t_{\max}}{2\mu^2}\sigma_{g_\eta}^2.$$

Since C_η is bounded below by $C_{\inf}$,

$$\mathbb{E}\left[C_\eta(\overrightarrow{\theta}^{t_{\max}})\right] \geq C_{\inf}.$$

Therefore,

$$\frac{\alpha}{2\Lambda_{\max}} \sum_{t=0}^{t_{\max}-1} \mathbb{E}\left[\left\|g_\eta^t\right\|_2^2\right] \leq C_\eta(\overrightarrow{\theta}^0) - C_{\inf} + \frac{L_C \alpha^2 t_{\max}}{2\mu^2}\sigma_{g_\eta}^2.$$

Dividing by $\alpha t_{\max}/(2\Lambda_{\max})$ yields

$$\frac{1}{t_{\max}} \sum_{t=0}^{t_{\max}-1} \mathbb{E}\left[\left\|g_\eta^t\right\|_2^2\right] \leq \frac{2\Lambda_{\max}\left(C_\eta(\overrightarrow{\theta}^0) - C_{\inf}\right)}{\alpha t_{\max}} + \frac{L_C \alpha \Lambda_{\max}}{\mu^2}\sigma_{g_\eta}^2.$$

Finally,

$$\min_{0 \leq t < t_{\max}} \mathbb{E}\left[\left\|g_\eta^t\right\|_2^2\right] \leq \frac{1}{t_{\max}} \sum_{t=0}^{t_{\max}-1} \mathbb{E}\left[\left\|g_\eta^t\right\|_2^2\right].$$

Using $g_\eta^t = \nabla C_\eta(\overrightarrow{\theta}^t)$ proves Theorem 5.1. □

G Numerical Details

G.1 Pauli–Lindblad Noise Model

The local-SLD method does not assume a particular noise model. In our experiments, we use a gate-local Pauli–Lindblad model based on Van Den Berg et al. (2023). After each m-qubit gate, $m \in \{1, 2\}$, we apply the channel generated by

$$\mathcal{L}_\eta^{(m)}(\rho) = \sum_{P \in \mathcal{P}_m \setminus \{I^{\otimes m}\}} \lambda_P \left(P\rho P - \rho\right), \qquad \mathcal{P}_m = \{I, X, Y, Z\}^{\otimes m},$$

where $\mathcal{L}_\eta^{(m)}$ is the Pauli–Lindblad generator (Liouvillian) and $\lambda_P \geq 0$ is the rate associated with Pauli generator P. The noise acts only on the qubits involved in the preceding gate. Thus a one-qubit gate has three nonidentity Pauli generators, while a two-qubit gate has 15.

For the uniform-rate model, we set $\lambda_P = \eta$ for every nonidentity generator. The additional nonuniform-rate comparison is described in Appendix G.4.

The individual Pauli generators commute as superoperators, so the full channel can be implemented by composing the corresponding stochastic Pauli channels. For each generator,

$$\exp\left(\mathcal{L}_P\right)(\rho) = (1 - p_P)\rho + p_P P\rho P, \qquad p_P = \frac{1 - e^{-2\lambda_P}}{2},$$

where

$$\mathcal{L}_P(\rho) = \lambda_P \left(P\rho P - \rho\right).$$

G.2 Experimental Configurations

The approximation-quality experiments use a 6-qubit, depth-4 circuit. The full mixed-state SLD-QFIM is computed directly and used as the reference. The population metrics are averaged over 10 random parameter initializations and contain no measurement shot noise, and the regularized update directions use $\mu = 0.1$. Unless otherwise stated, block-local SLD-QNG uses $r = 1$ and $k_S = 3$. The measurement counts in Figure 3(c) are conservative product-Pauli setting counts per metric refresh, not circuit shot counts.

The main finite-shot experiment uses a 12-qubit, depth-6 circuit with $p = 144$ parameters and noise strength $\eta = 0.0005$. Because gate-local noise is applied after every gate, the larger and deeper circuit accumulates more noise. So we use a generator rate smaller than the 6-qubit approximation case, which considers rates up to $\eta = 0.005$. The metric is refreshed every $R_{\mathrm{ref}} = 10$ iterations. All methods optimize the same noisy objective and use the same initial parameters. Gradient evaluations use the same shot allocation across methods. At block-local metric-refresh iterations, the shifted response measurements also provide the gradient, so no additional gradient measurements are needed. Further details are given in Appendix G.6.

All numerical experiments were implemented in Python using Qiskit (v1.4.3), with noisy and finite shot circuit simulation performed using Qiskit Aer (v0.17.1) (Javadi-Abhari et al., 2024). All experiments were run on a multi-core CPU computing cluster.

G.3 Baseline Metrics

We apply both baseline metric constructions to the noisy mixed states considered here as exact population quantities in Figures 2 and 5 and under finite-shot estimation in the optimization experiment in Figure 4. WA-QNG was originally evaluated numerically with exact metric quantities in noise-free circuits.

Hilbert–Schmidt QNG. For the Koczor–Benjamin baseline (Koczor & Benjamin, 2022), we use the rescaled global Hilbert–Schmidt metric

$$F_{ij}^{\mathrm{HS}} = 2\mathrm{tr}\left(\left(\partial_{\theta_i}\rho_\eta\right)\left(\partial_{\theta_j}\rho_\eta\right)\right). \tag{37}$$

At $\eta = 0$, the unitary circuit used in our experiments produces a pure global state, for which F^{HS} coincides with the SLD-QFIM. For mixed states, F^{HS} is generally a surrogate rather than the exact SLD-QFIM. We evaluate it directly on the noisy state without virtual-distillation or other noise-mitigation procedures.

WA-QNG. WA-QNG (Shi et al., 2026) constructs a Hamiltonian-weighted metric from Hilbert–Schmidt metrics of reduced states. Writing

$$H = \sum_m h_m H_m, \qquad \rho_m = \mathrm{tr}_{\mathrm{supp}(H_m)^c}\left(\rho_\eta\right),$$

we use

$$F_{ij}^{\mathrm{WA}} = \frac{2}{\sum_m h_m^2} \sum_m h_m^2 \mathrm{tr}\left(\left(\partial_{\theta_i}\rho_m\right)\left(\partial_{\theta_j}\rho_m\right)\right). \tag{38}$$

In the implementation, Hamiltonian terms with the same support are evaluated using the same reduced state.

G.4 Nonuniform random Pauli rates

We repeat the population-level comparison using fixed nonuniform Pauli–Lindblad rates. For each local Pauli generator, we draw

$$a_P \sim \mathrm{LogNormal}(0, 1)$$

and rescale the resulting rates so that the mean generator rate on every gate support is η. The random and uniform models have the same total generator rate. The same relative rate profiles are used throughout the comparison.

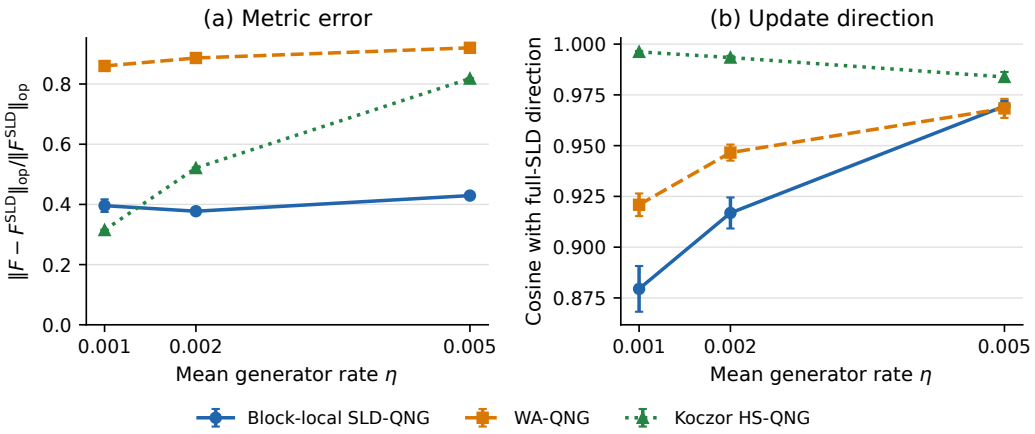


Figure 5: Approximation quality under nonuniform log-normal Pauli–Lindblad rates for the 6-qubit, depth-4 circuit. For each value of η, results are averaged over 3 random choices of the Pauli-generator rates and 10 random parameter initializations. (a) Relative operator-norm error with respect to the full SLD-QFIM. (b) Cosine similarity with the regularized full-SLD update direction. Error bars show the standard error over parameter initializations after averaging over the 3 rate choices.

At $\eta = 0.002$ and 0.005, block-local SLD-QNG has the smallest metric error among the tested approximations. At $\eta = 0.001$, Koczor HS-QNG has the smallest error and also gives the highest direction cosine throughout the tested range. Together with Figure 2, this shows that the lower metric error of block-local SLD-QNG at the two larger rates persists under the tested nonuniform Pauli-generator rate profiles.

G.5 Optimization hyperparameters

The optimization hyperparameters used in the main finite-shot experiment are listed in Table 1. They were selected in preliminary runs and then held fixed across all evaluation seeds and no seed-specific tuning was performed. All metric-based methods use $R_{\mathrm{ref}} = 10$. Adam uses $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon_{\mathrm{Adam}} = 10^{-8}$.

Table 1: Hyperparameters for the 12-qubit, depth-6 finite-shot experiment. A dash indicates that the parameter is not used.

Method	Step size α	Ridge μ	Score ridge γ
SGD	0.3	–	–
Adam	0.25	–	–
Block-local SLD-QNG	0.2	0.1	0.01
WA-QNG	0.2	0.3	–
Koczor HS-QNG	0.2	0.3	–

G.6 Finite-shot accounting

We count one circuit shot for each state preparation and measurement. Two-copy overlap measurements count as two circuit shots per sample.

The 12-qubit, depth-6 experiment has $p = 144$ parameters. Each parameter-shift gradient circuit uses $N_{\mathrm{ps}} = 1536$ shots, and each iteration uses $N_{\mathrm{energy}} = 1024$ shots to estimate the energy. An ordinary iteration costs

$$\begin{aligned} N_{\mathrm{grad}} &= 2pN_{\mathrm{ps}} + N_{\mathrm{energy}} \\ &= 2(144)(1536) + 1024 = 443392 \end{aligned}$$

circuit shots.

Block-local SLD-QNG uses radius-one windows, containing 2 qubits at the boundaries and 3 qubits in the interior, with $k_S = 3$. To estimate the score-Gram matrices, we use 189 distinct Pauli measurement settings after merging identical settings across the 12 overlapping windows. This is the total number of settings shared by all blocks. For comparison, treating the windows independently would give $2(9) + 10(27) = 288$ settings. Repeating each shared setting for 128 shots gives

$$N_A = 189(128) = 24192$$

circuit shots.

At metric-refresh iterations, we use the same shifted product-Pauli measurements to estimate both the response matrices and the energy gradient. At other iterations, the gradient is estimated by the usual parameter-shift rule. For each refresh, we use two shifts for each of the 12 parameters in each of the 12 blocks, with 2560 shots per shifted block parameter. The total cost is

$$N_{g,B} = 2(12)(12)(2560) = 737280$$

circuit shots. Because these measurements also provide the gradient, no separate 442368 shot gradient estimate is needed at a refresh iteration. The total cost of a block-local refresh iteration is therefore

$$\begin{aligned} N_{\mathrm{ref}}^{\mathrm{blk}} &= N_{\mathrm{energy}} + N_A + N_{g,B} \\ &= 1024 + 24192 + 737280 \\ &= 762496 \end{aligned}$$

circuit shots. Relative to an ordinary iteration, the additional geometry cost is

$$\begin{aligned} N_{\mathrm{geom}}^{\mathrm{blk}} &= N_{\mathrm{ref}}^{\mathrm{blk}} - N_{\mathrm{grad}} \\ &= 762496 - 443392 \\ &= 319104. \end{aligned}$$

Using the same measurements for the response matrices and gradient is an implementation choice rather than a requirement of the block-local construction. The convergence result directly covers the variant using an independent gradient batch. The measurement-reuse variant is evaluated empirically.

For WA-QNG, the reduced-state tangents are evaluated on the two-qubit supports of the Heisenberg terms. We collect 576 full-system product-Pauli shadow snapshots for each shifted parameter value and reuse the same snapshots across all Hamiltonian supports. Hence

$$N_{\mathrm{geom}}^{\mathrm{WA}} = 2p(576) = 165888.$$

Koczor HS-QNG estimates the dense Hilbert–Schmidt metric of the full 12-qubit state. Its $p(p+1)/2$ distinct symmetric entries each require 4 shifted-state overlaps. We use 64 samples per overlap and count the two state preparations in each overlap sample as two circuit shots. Therefore,

$$N_{\mathrm{geom}}^{\mathrm{HS}} = \frac{p(p+1)}{2}(4)(64)(2) = 5345280.$$

This is approximately 16.75 times the additional geometry cost of block-local SLD-QNG. Koczor HS-QNG estimates all $p(p+1)/2 = 10440$ symmetric entries through separate overlap measurements, whereas the block-local estimator reuses local Pauli data across overlapping windows.

The metric is refreshed every $R_{\mathrm{ref}} = 10$ iterations. Over 150 iterations, including the initial energy evaluation, the cumulative cost is

$$N_{\mathrm{total}}^{(m)} = 1024 + 150N_{\mathrm{grad}} + 15N_{\mathrm{geom}}^{(m)}.$$

These are the circuit shots used on the horizontal axis of the finite-shot convergence plot.

Acquisition-setting scaling. For the one-dimensional nearest-neighbor ansatz, at fixed depth, radius and score cutoff, the conservative per-refresh setting count is $O(n)$ because each of the $O(n)$ parameter blocks requires a bounded number of local product-Pauli settings. Consider the radius-1 construction with complete local Pauli score spaces ($k_S = 3$, since each window contains at most

Table 2: Circuit shots in the 12-qubit, depth-6 experiment. The geometry column gives the additional shots beyond an ordinary gradient-and-energy iteration.

Method	Support	Additional geometry	Refresh iteration
SGD	–	0	443392
Adam	–	0	443392
Block-local SLD-QNG	$w_1(q_b)$	319104	762496
WA-QNG (shadow)	$\mathrm{supp}(H_m)$	165888	609280
Koczor HS-QNG	global	5345280	5788672

three qubits). The two boundary windows require 9 settings each, while the $n-2$ interior windows require 27 settings each. With $2d$ parameters per block, one unshifted score-Gram acquisition and two shifted response acquisitions per parameter give

$$\begin{aligned} N_{\mathrm{set}} &= (1+4d)\,[2(9)+(n-2)27] \\ &= (1+4d)(27n-36). \end{aligned}$$

For $d=6$, this becomes $N_{\mathrm{set}} = 675n - 900$. At $n=12$, the bracketed term is 288, matching the independent-block score-Gram count above. This is a conservative schedule. The implemented acquisition uses the measurement sharing accounted for above. Combining the $O(n)$ setting count with the shot repetitions required by the concentration analysis gives the $O\left(n\varepsilon^{-2}\log(n/\delta)\right)$ shot bound in Corollary E.4.

G.7 Additional finite-shot controls

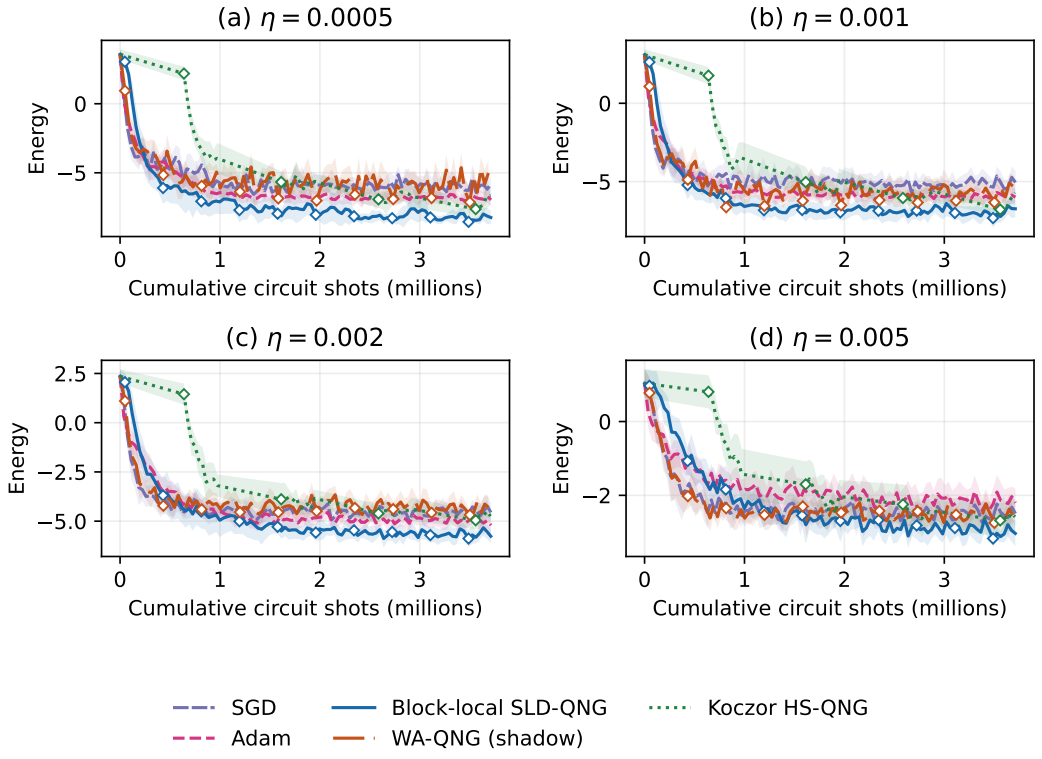


Figure 6: Finite-shot Heisenberg optimization at four Pauli–Lindblad generator rates. Curves show means over 10 seeds, and shaded regions show two-sided 95% confidence intervals. Open diamonds mark metric refreshes. Each panel is restricted to the cumulative-shot range shared by all methods.

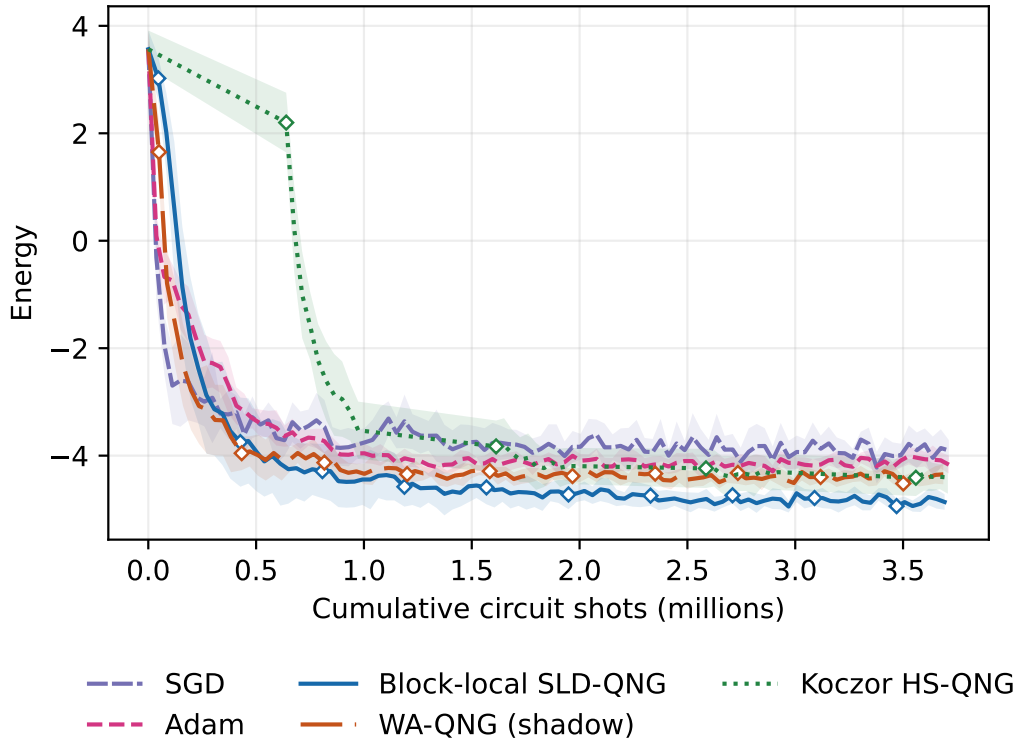


Figure 7: Finite-shot optimization for the 6-qubit TFIM instance as a function of cumulative circuit shots. Curves show means over 10 seeds, shaded regions show two-sided 95% confidence intervals, and open diamonds mark metric refreshes. The plot is restricted to the budget shared by all methods.

We further test the finite-shot behavior across a range of noise strengths and on a different Hamiltonian. These experiments use the circuit-shot counting convention in Appendix G.6. All energy, gradient, and metric-estimation measurements are included in the cumulative shot count. Unless stated otherwise, we use the same optimization hyperparameters as in Table 1.

Noise-strength sweep. We perform finite-shot Heisenberg optimization with the 6-qubit, depth-4 circuit at $\eta \in \{0.0005, 0.001, 0.002, 0.005\}$.

Figure 6 shows that block-local SLD-QNG reaches lower mean energies over much of the shared shot range, although the size of the advantage varies with noise strength. At $\eta = 0.005$, several methods reach similar energy plateaus. These results show that its finite-shot performance is not tied to a single noise rate, but the improvement is not uniform across the tested range.

Transverse-field Ising model. We also consider the transverse-field Ising Hamiltonian

$$H_{\mathrm{TFIM}} = \sum_{i=0}^{n-2} Z_i Z_{i+1} + 0.5 \sum_{i=0}^{n-1} X_i$$

with open boundary conditions. We use the same 6-qubit, depth-4 circuit at $\eta = 0.0005$ and the same circuit-shot accounting convention.

Figure 7 shows that block-local SLD-QNG achieves lower mean energy than SGD over most of the shared shot range and remains competitive with Adam, Koczor HS-QNG, and WA-QNG.