WHEN GEOMETRY MEETS ENERGY: ROBUST OPERATOR SELECTION FOR ADAPTIVE VQE VIA GAGRAD

Anonymous authors
Paper under double-blind review

ABSTRACT

Adaptive variational quantum eigensolvers (VQE) grow quantum circuits by iteratively selecting operators from a predefined pool. The standard energy-gradient criterion (ADAPT-VQE) aligns directly with the optimization objective but can mislead when the reference state deviates from Hartree-Fock. Geometry-driven alternatives based on the quantum geometric tensor (QGT) offer complementary strengths, prioritizing structurally non-redundant operators and maintaining robustness to reference-state perturbation. We propose the GAGrad operator selection framework, which combines energy-gradient magnitudes with geometric volume-expansion criteria measured via the Schur complement of the QGT. Across molecular benchmarks (LiH, BeH_2; 6–10 qubits), we uncover a gradient failure phenomenon: under moderate perturbation, gradient and Schur-complement rankings become anti-correlated ($\tau = -0.43$), and ADAPT and GAG diverge only in the final 1–2 operator selections, yet this late-stage difference produces 3–4 orders of magnitude gap in residual energy. Thus, geometric information is most consequential near convergence, when energetic differences between candidates become small and gradient rankings become unreliable. GAGrad matches ADAPT's step efficiency while capturing this geometric advantage, achieving comparable precision on most benchmarks with shallower circuits and stable performance under depolarizing noise. A GAG warmstart protocol further reduces circuit depth substantially.

1 INTRODUCTION

The variational quantum eigensolver (VQE) (Peruzzo et al., 2014; Cerezo et al., 2021a) estimates ground-state energies of molecular systems using parameterized quantum circuits (PQCs). A central challenge is ansatz design: the circuit must be sufficiently expressive to reach chemical accuracy while remaining shallow enough for noisy intermediate-scale quantum (NISQ) hardware (Preskill, 2018; Kandala et al., 2017; McClean et al., 2018).

Adaptive VQE methods—originating with ADAPT-VQE (Grimsley et al., 2019)—address this challenge by dynamically constructing the ansatz from the bottom up. Starting from a simple reference state (typically Hartree-Fock), these methods iteratively select and append operators from a predefined pool based on some selection criterion, then re-optimize parameters. This bottom-up approach can produce substantially more compact circuits than fixed-topology ansätze such as hardware-efficient ansätze (HEA) (Kandala et al., 2017), unitary coupled cluster (UCCSD) (Romero et al., 2019), or QAOA-style layered circuits (Farhi et al., 2014).

The standard approach, used by ADAPT-VQE (Grimsley et al., 2019) and its variants (Tang et al., 2021; Meitei et al., 2021), selects the operator with the largest energy-gradient magnitude. This energy-driven criterion is directly aligned with the optimization objective but suffers from two limitations: (1) it can be myopic, greedily selecting operators that give immediate energy relief but are structurally redundant, leading to unnecessarily deep circuits; and (2) evaluating energy gradients for the entire operator pool at every growth step introduces substantial measurement overhead on quantum hardware (Gonthier et al., 2022).

An alternative perspective comes from quantum information geometry. The quantum geometric tensor (QGT) (Provost & Vallée, 1980; Stokes et al., 2020) encodes the Riemannian geometry of

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.

the parameterized quantum state manifold, with its rank measuring the dimension of the tangent space and its eigenvalues quantifying sensitivity along each parameter direction. Recent work has explored using the QGT for operator selection in adaptive VQE, selecting operators that maximize geometric volume expansion (measured via the Schur complement) rather than energy descent. This geometry-driven approach inherently filters structurally redundant operators and maintains favorable gradient conditioning, offering a selection signal that is qualitatively distinct from energy-gradient ranking.

In this work, we argue that energy and geometry are complementary signals for operator selection, and that the most robust adaptive VQE methods combine both. We propose the *GAGrad scoring rule*, which scores each candidate operator as the product of its energy-gradient magnitude (ensuring energetic relevance) and its Schur-complement volume increment (ensuring structural non-redundancy). This simple multiplicative scoring rule favors operators that are simultaneously energetically important and geometrically novel, combining the directness of energy-driven selection with the structural awareness of geometry-based growth.

Our contributions are as follows: **(1)** We discover a *gradient failure phenomenon*: under moderate reference-state perturbation, energy-gradient and Schur-complement rankings become anti-correlated ($\tau = -0.43$ at $\alpha = 0.2$), and ADAPT and GAG diverge only in the final 1–2 operator selections, yet this late-stage difference increases ADAPT's residual energy error by 3–4 orders of magnitude while GAG remains stable. This geometric resilience extends to hardware noise: under depolarizing noise, GAG and GAGrad maintain stable precision (no measurable error growth across tested noise levels) while ADAPT's error grows by nearly an order of magnitude, since geometric selection is inherently robust to depolarizing noise. **(2)** We propose the GAGrad scoring rule, which combines energy-gradient magnitudes with Schur-complement volume expansion in a multiplicative form, alongside an additive variant and a random-selection baseline for controlled comparison. GAGrad(mult) matches ADAPT's step efficiency on all tested molecules (LiH, BeH_2; 6–10 qubits) while delivering comparable or better precision across benchmarks, and the additive variant produces the most compact circuits on LiH (see Appendix B.5). GAGrad is also the only tested method that remains robust under strong perturbation on BeH_2. **(3)** We introduce a GAG warmstart protocol that uses geometric selection for the first 2 operators then switches to ADAPT, substantially reducing CNOT counts compared with ADAPT while maintaining near-FCI accuracy. **(4)** We provide detailed analyses of Schur–gradient rank correlation, operator path divergence, measurement overhead, and QGT spectral properties, offering a comprehensive mechanistic picture of when and why geometric information aids adaptive ansatz construction.

2 Related Work

ADAPT-VQE (Grimsley et al., 2019) introduced the iterative operator-selection paradigm, using fermionic excitation operators and energy-gradient selection. Qubit-ADAPT-VQE (Tang et al., 2021) adapted this framework to hardware-efficient Pauli-string operator pools, reducing circuit depth at the cost of more operators. Subsequent work has explored qubit pool design (Meitei et al., 2021), measurement-efficient selection (Gonthier et al., 2022; Zhao et al., 2020; Mari et al., 2021), gradient-free criteria (Sim et al., 2021), noise-resilient variants (Wang et al., 2021; Nykänen et al., 2025), and adaptive methods for broader problem classes (Anschuetz & Kiani, 2022; Ohno, 2024). A practical concern with adaptive methods is the cost of evaluating selection criteria across the pool at each growth step; for energy-gradient selection this scales linearly with pool size, while QGT-based methods scale with the product of pool size and current parameter count, though techniques such as simultaneous perturbation (Gacon et al., 2021) can reduce this overhead. Our work contributes a systematic comparison of selection criteria and a GAGrad energy-geometry scoring rule.

The quantum geometric tensor (QGT), also known as the Fubini-Study metric tensor, is a fundamental object in quantum information geometry (Provost & Vallée, 1980; Braunstein & Caves, 1994; Larocca et al., 2023). It is most widely used for quantum natural gradient (QNG) optimization (Stokes et al., 2020; Wierichs et al., 2022), and has also been applied to expressibility analysis (Sim et al., 2019; Haug et al., 2021; Hubregtsen et al., 2021), effective dimension estimation (Abbas et al., 2021; Zhang et al., 2023), circuit architecture diagnosis (Du et al., 2022), and quantum neural tangent kernel analysis (Tang & Yan, 2022; Yu et al., 2022). Recent work has also explored concentration properties of the QGT (Li et al., 2022) and differentiable programming frameworks for

geometric quantum computing (Zhang et al., 2022). Most closely related to our work, geometry-driven adaptive growth uses the Schur complement of the QGT as an operator-selection criterion. Concurrent work on Geo-ADAPT-VQE (Sohail & Koike-Akino, 2026) proposes a natural-gradient selection rule with convergence guarantees; our work focuses on the complementary question of gradient failure under perturbation and the robustness of geometric selection.

3 METHOD

3.1 BACKGROUND: ADAPTIVE VQE AND THE QGT

Consider a VQE circuit $U(\boldsymbol{\theta})$ with M parameters that prepares the state $|\psi(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\text{HF}\rangle$, where $|\text{HF}\rangle$ is the Hartree-Fock reference state. The goal is to minimize the energy expectation $E(\boldsymbol{\theta}) = \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$ with respect to a molecular Hamiltonian H.

In adaptive VQE, the circuit is grown incrementally. At each macro-step k, a new parameterized operator A_k is selected from a pool $\mathcal{A} = \{A_1, \ldots, A_P\}$ and appended to the circuit, increasing the parameter count from k to $k+1$. Parameters are then re-optimized before the next selection step. The algorithm terminates when a convergence criterion (e.g., maximum pool gradient below a threshold) is satisfied.

The *quantum geometric tensor* (QGT) encodes the local Riemannian geometry of the parameterized state manifold (Meyer, 2021):

$$G_{ij}(\boldsymbol{\theta}) = \langle\partial_i\psi|\partial_j\psi\rangle - \langle\partial_i\psi|\psi\rangle\langle\psi|\partial_j\psi\rangle, \tag{1}$$

where $\partial_i = \partial/\partial\theta_i$. The rank of G equals the dimension of the tangent space of the state manifold, and its eigenvalues measure the sensitivity of the state to parameter changes along each principal direction. The QGT is central to both the quantum natural gradient optimizer (Stokes et al., 2020) and geometry-driven adaptive growth.

3.2 OPERATOR SELECTION CRITERIA

We consider four operator selection criteria, all operating on the same operator pool and evaluated at the current state with the new parameter initialized at zero ($\theta_{\text{new}} = 0$).

Energy-gradient selection (ADAPT). The standard ADAPT-VQE criterion selects the operator with the largest energy-gradient magnitude:

$$A^*_{\text{adapt}} = \arg\max_{A_k \in \mathcal{A}} \left| \frac{\partial E}{\partial \theta_k} \right|. \tag{2}$$

This directly targets energy reduction but may select operators that are structurally redundant (i.e., whose tangent direction lies mostly in the span of existing parameters).

Geometric selection (GAG). Geometry-aware adaptive growth selects the operator that maximizes the geometric volume expansion of the state manifold, measured via the Schur complement of the augmented QGT. For a candidate operator A_k with tangent vector $v_k = |\partial_k\psi\rangle$, the Schur complement is:

$$S_k = G_{kk} - \mathbf{v}_k^\dagger G_M^+ \mathbf{v}_k, \tag{3}$$

where G_M is the current $M \times M$ QGT, $\mathbf{v}_k$ is the vector of cross-terms between the candidate and existing parameters, and G_M^+ denotes the Moore-Penrose pseudoinverse. The selection rule is:

$$A^*_{\text{gag}} = \arg\max_{A_k \in \mathcal{A}} S_k. \tag{4}$$

The Schur complement S_k equals the squared norm of v_k projected onto the orthogonal complement of the existing tangent space, so maximizing S_k ensures the selected operator introduces a maximally independent new direction. This filters structural redundancies but is Hamiltonian-agnostic.

We now state two properties that characterize the geometric selection rule.

Proposition 1 (GAG maximizes manifold volume growth)**.** *Let G_k be the $k \times k$ QGT of the ansatz after k selection steps, and let S_{k^*} be the Schur complement of the selected candidate operator. Then the fractional volume increase of the parameter manifold is:*

$$\frac{\det(G_{k+1})}{\det(G_k)} = S_{k^*}. \tag{5}$$

Thus, selecting the operator with the largest Schur complement is equivalent to greedily maximizing the volume growth rate of the ansatz manifold.

Proposition 2 (Schur complement as orthogonal projection)**.** *The Schur complement of candidate operator A_k is the squared norm of its tangent vector projected onto the orthogonal complement of the existing tangent space:*

$$S_k = \left\| P_{T^\perp} \left| \partial_k \psi \right\rangle \right\|^2, \tag{6}$$

where $P_{T^\perp} = I - |\partial_M \psi\rangle G_M^+ \langle \partial_M \psi|$ is the projector onto the orthogonal complement of the span of existing tangent vectors $\{|\partial_i \psi\rangle\}_{i=1}^M$.

An immediate consequence is that $S_k = 0$ if and only if the candidate tangent lies entirely in the existing tangent span.

Proposition 3 (Residual energy descent)**.** *Let P_T denote orthogonal projection onto the current tangent space. The component of the energy gradient along a candidate's* novel *directions is:*

$$r_k = \langle \nabla_{\mathrm{FS}} E,\ P_{T^\perp} \partial_k \psi \rangle. \tag{7}$$

By Cauchy–Schwarz, $|r_k| \leq \|\nabla_{\mathrm{FS}} E\| \cdot \sqrt{S_k}$. Thus, at exact stationarity ($P_T \nabla_{\mathrm{FS}} E = 0$), a candidate with $S_k \to 0$ cannot carry independent first-order energy descent.

Propositions 1–3 together characterize GAG from three perspectives: as greedy volume maximization (D-optimal design analogy (Atkinson et al., 2007)), as orthogonal novelty selection, and as a lower bound on genuine first-order descent. This third property connects geometry directly to the optimization objective: structurally redundant operators ($S_k \approx 0$) cannot meaningfully improve the energy once the current parameters are optimized.

GAGrad selection. Given the complementary strengths of energy-gradient and geometric selection, a natural question is how to combine them into a single scoring function. We consider two principled scalarizations:

$$h_k^{(\times)} = \left| \frac{\partial E}{\partial \theta_k} \right| \cdot S_k \qquad \text{(multiplicative)}, \tag{8}$$

and an additive form $h_k^{(+)} = |\partial E / \partial \theta_k| + \lambda S_k$ (Appendix B.5).

The multiplicative form ensures an operator scores highly only if it is both energetically important and geometrically novel. Operators that are energetically relevant but redundant receive low scores because $S_k \approx 0$; geometrically novel but irrelevant operators score low because $|\partial E / \partial \theta_k| \approx 0$. In other words, energy gradients identify *which directions lower the energy* while the Schur complement identifies *which directions expand the manifold*; their product selects directions that do both.

A key property of the multiplicative GAGrad is its built-in zero-score characterization:

Proposition 4 (Zero-score characterization of multiplicative GAGrad)**.** *For the multiplicative scoring function $h_k = |\partial E / \partial \theta_k| \cdot S_k$, an operator receives zero score if and only if it is either energetically irrelevant ($|\partial E / \partial \theta_k| = 0$) **or** geometrically redundant ($S_k = 0$). The additive GAGrad $h_k = |\partial E / \partial \theta_k| + \lambda S_k$ does not share this property: an operator can score highly even if one signal is zero, as long as the other is sufficiently large.*

This double-filter property means GAGrad(mult) avoids two failure modes: redundant operators with large gradients, and irrelevant operators with geometric novelty. The additive variant trades this strictness for flexibility, sometimes producing shallower circuits (see Appendix B.5).

Remark on parameterization invariance. The multiplicative score $h_k = |g_k|S_k$ scales as c^3 under generator rescaling $A_k \to cA_k$ (since $g_k \to cg_k$, $S_k \to c^2 S_k$), so rankings depend on pool normalization—a limitation shared with ADAPT's $|g_k|$ (scales as c). All our operators use standard fermionic normalization. A scale-invariant alternative such as $|g_k|/\sqrt{S_k}$ would remove this dependence at the cost of the double-filter property.

Random selection. We use uniform random selection (without replacement) as a baseline to isolate the effect of intelligent operator choice.

3.3 When Geometry Beats Energy: A Theoretical Analysis

The gradient failure phenomenon observed in our experiments is not merely an empirical finding—it can be predicted from the interaction between gradient reliability and geometric structure. We formalize this intuition under four mild assumptions: (A1) each operator has a fixed ground-truth utility (asymptotic energy reduction); (A2) gradient rankings degrade monotonically with perturbation, becoming random at large shifts; (A3) Schur-complement rankings are invariant to perturbation and positively correlated with utility for fermionic pools; (A4) the pool has sufficient diversity that top gradient and geometric selections differ when rankings diverge.

Theorem 1 (Rank-correlation crossover and gradient failure)**.** *Under assumptions A1–A4, there exist perturbation thresholds $0 < \delta_1 < \delta_2$ such that:*

1. ***Small perturbation*** *($|\delta| < \delta_1$): Gradient and Schur-complement rankings are positively correlated ($\tau > 0$), and ADAPT's selected operator has higher ground-truth utility than GAG's.*
2. ***Moderate perturbation*** *($\delta_1 < |\delta| < \delta_2$): Rankings become anti-correlated ($\tau < 0$), and GAG's selected operator has higher ground-truth utility than ADAPT's.*
3. ***Strong perturbation*** *($|\delta| > \delta_2$): Gradient rankings are essentially random, and neither strategy reliably selects useful operators.*

Theorem 1 predicts gradient failure as a generic consequence of gradient degradation under perturbation. Our experiments confirm this qualitatively: τ ranges from $+0.29$ (at $\alpha = 0$) to -0.43 (at $\alpha = 0.2$), consistent with the predicted crossover. ADAPT's error peaks at $\alpha = 0.3$ rather than at maximum anti-correlation, indicating that gradient failure is *path-dependent* rather than instantaneous.

Implementation. All four strategies follow the same adaptive pipeline (Algorithm 1, Appendix C.1), differing only in the scoring function. We use BFGS optimization (Liu & Nocedal, 1989) with convergence threshold $\epsilon = 10^{-5}$ on the maximum gradient magnitude.

4 Experiments

4.1 Experimental Setup

We compare four strategies (ADAPT, GAG, GAGrad(mult), Random) on LiH (6/8 qubits) and BeH_2 (8/10 qubits) in STO-3G with frozen core, using BFGS optimization and identical convergence criteria. Two operator pools are used: a standard fermionic excitation pool (8 operators, for perturbation experiments) and a generalized fermionic pool (up to 98 operators, for core benchmarks). Full implementation details are provided in Appendix C.

4.2 Gradient Failure under Reference Perturbation

Our central finding is that geometric selection provides its greatest value precisely when energy gradients are most misleading. To demonstrate this, we perturb the Hartree-Fock reference state by applying α-scaled random rotations: $|\psi_0\rangle = e^{-i\alpha H_{\text{rand}}}|\text{HF}\rangle$, where H_{rand} is a random Hermitian operator. At $\alpha = 0$, the reference is the standard HF state; at $\alpha = 1$, it is strongly randomized.

We use a fermionic pool (8 operators) on LiH 6q, with 9 seeds per α value (3 VQE seeds $\times$ 3 perturbation seeds).

This perturbation protocol serves as a *stress test* for adaptive VQE selection criteria. In practice, the reference state used to initialize adaptive methods may deviate from HF due to orbital-rotation errors, approximate active-space selection, imperfect state preparation, or initialization from a cheaper approximate variational method. Rather than modeling each specific error source, our random-Hermitian protocol probes the general robustness of selection criteria to reference-state misalignment, providing a controlled setting where we can continuously vary the severity of the mismatch.

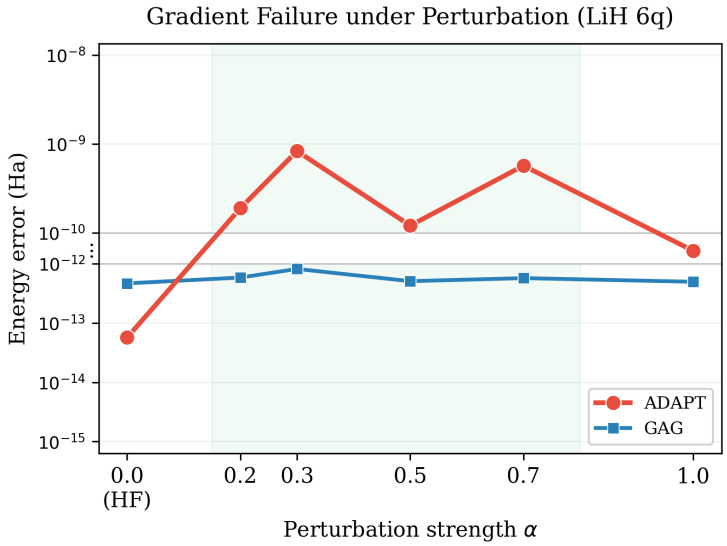

Figure 1: Gradient failure under perturbation (LiH 6q, fermionic pool). ADAPT's residual energy error grows by 3–4 orders of magnitude in the moderate-perturbation regime, while GAG remains stable. GAGrad variants are omitted for clarity; full data in Appendix B.1.

Figure 1 reveals a pronounced pattern of *gradient failure* for ADAPT. At $\alpha = 0$ (HF reference), ADAPT and GAG perform comparably: ADAPT's energy-gradient selection is well-calibrated when the reference state is close to the true ground state, and GAG is slightly worse but still well within chemical accuracy. Under moderate perturbation ($\alpha = 0.2$–0.7), however, ADAPT's residual energy error grows by 3–4 orders of magnitude, while GAG remains stable. This occurs because perturbation pushes the reference state away from HF, causing energy gradients to become misleading: operators with large gradients may lead to shallow local minima rather than the global ground state (McClean et al., 2018; Cerezo et al., 2021b). GAG, being Hamiltonian-agnostic, is unaffected by this failure mode. At $\alpha = 1.0$ (strong randomization), ADAPT partially recovers: both gradient and geometric signals become equally noisy, and the precision gap narrows, suggesting that gradient failure arises from an intermediate regime where gradients are misleading but not yet useless. Analysis of the error distribution further reveals that ADAPT's failure is strongly right-skewed: under moderate perturbation, the median error remains near 10^{-13} Ha, but rare outlier seeds can degrade by six orders of magnitude, driving the mean upward. In contrast, GAG and GAGrad maintain tight, symmetric error distributions across all perturbation strengths (Appendix B.1, distribution table).

4.3 Geometric Analysis: The Mechanism of Gradient Failure

To elucidate why geometric selection outperforms energy-gradient selection under perturbation, we analyze the relationship between the two selection signals and the divergence of operator selection paths.

At each growth step, both ADAPT and GAG rank the operator pool—ADAPT by $|\partial E/\partial\theta_k|$ and GAG by S_k. We compute the Kendall τ rank correlation between these two rankings at step 0 (before any operator is selected) across perturbation strengths α on LiH 6q (fermionic pool, 3 seeds). Rank

correlation has been used previously to diagnose expressibility and trainability in quantum circuits (Du et al., 2020; Stilck França & Garcia-Patron, 2021).

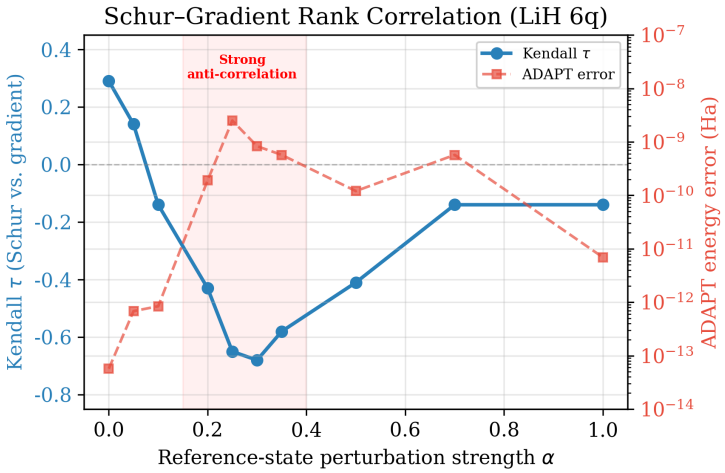

Figure 2: Schur–gradient rank correlation (left) and ADAPT energy error (right) as a function of perturbation strength (LiH 6q, fermionic pool). Maximum anti-correlation occurs at $\alpha = 0.2$ ($\tau = -0.43$), marking the onset of gradient misalignment; ADAPT's error peaks later at $\alpha = 0.3$ due to path dependence. Full data in Appendix B.2.

Figure 2 reveals a nuanced pattern that illuminates the mechanism of gradient failure. At $\alpha = 0$ (HF reference), the two signals are positively correlated: the top-5 operators by gradient and by Schur complement overlap by 80%, so ADAPT and GAG make nearly identical selections. At $\alpha = 0.2$, the correlation reaches its most negative value ($\tau = -0.43$), marking the *onset* of gradient misalignment: the operators with the largest energy gradient are precisely those with the lowest Schur complement—they are the most geometrically redundant directions.

At $\alpha = 0.3$, ADAPT's error peaks even though rank correlation has weakened relative to $\alpha = 0.2$. This is because the consequences of misalignment emerge gradually: the initial ranking divergence sets ADAPT on a trajectory where later gradient signals, though less anti-correlated overall, steer optimization toward a suboptimal local minimum. At $\alpha = 1.0$, both signals become equally noisy and the error gap narrows, consistent with recovery at large perturbation strength.

To pinpoint where ADAPT and GAG diverge, we track the step-by-step operator selections on LiH 6q (fermionic pool, 8 operators). Across all α values, the two methods select identical operators for the first 6–7 steps; divergence occurs only at the final 1–2 steps. Yet this late divergence produces up to four orders of magnitude difference in final precision, driven by a single different operator choice near the end of growth. This reveals that the critical value of geometric information lies not in selecting fundamentally different operators throughout growth, but in making the correct choice among final candidates where energetic differences are small and gradient rankings become unstable.

4.4 Core Comparison: Standard Benchmarks

We now compare all four strategies on standard VQE benchmarks (generalized fermionic pool, HF reference) across four molecular systems.

Figure 3 and the accompanying analysis reveal three key findings on standard benchmarks. First, GAGrad(mult) matches ADAPT's step efficiency on all four molecules, confirming that the multiplicative scoring preserves ADAPT's step count while delivering comparable or better precision. On BeH_2, GAGrad also produces the most compact circuits. Second, GAG converges reliably on smaller systems but requires two to three times more steps, and on BeH_2 10q it fails to converge within 40 steps—indicating that pure geometric selection faces scalability challenges on larger systems, as geometrically novel operators are not always energetically relevant. Third, random selection

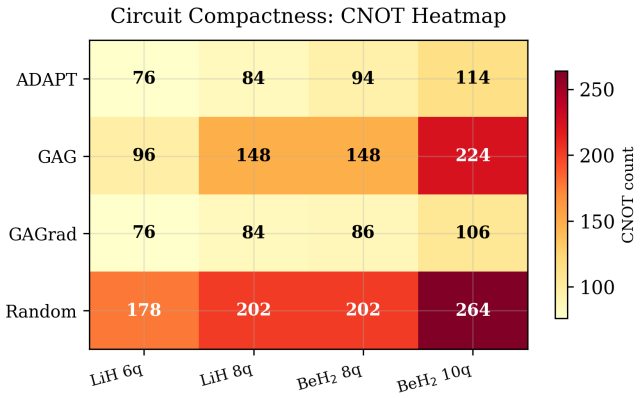

Figure 3: Circuit compactness heatmap (CNOT count) on standard benchmarks (generalized fermionic pool, HF reference). GAGrad achieves the lightest row, matching ADAPT on LiH and reducing CNOT count by 8 on BeH_2. Full data including steps and errors in Appendix B.4.

requires four to five times more steps than ADAPT and fails on larger molecules, confirming that both criteria provide meaningful selection signal. Full numerical results are in Appendix B.4.

4.5 Noise Robustness

A key motivation for geometric selection is its potential resilience to hardware noise. Unlike gradient-based selection, which relies on precise measurement of energy derivatives, GAG selects operators based on the geometric structure of the ansatz manifold—a property that is independent of measurement noise (Kandala et al., 2019; Wang et al., 2021). To test this hypothesis, we simulate depolarizing noise on the LiH 6q system at $1.5R_e$ (fermionic pool, 3 seeds), both at the HF reference ($\alpha = 0$) and under moderate perturbation ($\alpha = 0.3$).

Table 1: Noise robustness under depolarizing noise (LiH 6q, $1.5R_e$, fermionic pool). GAG and GAGrad(mult) show no measurable error growth across all tested noise levels.

Noise p	$\alpha = 0$ (HF)			$\alpha = 0.3$ (perturbed)		
	ADAPT	GAG	GAGrad(mult)	ADAPT	GAG	GAGrad(mult)
0.00	2.66×10^{-15}	3.55×10^{-15}	2.66×10^{-15}	2.66×10^{-15}	5.92×10^{-15}	3.26×10^{-15}
0.05	2.66×10^{-15}	3.55×10^{-15}	2.66×10^{-15}	2.66×10^{-15}	5.92×10^{-15}	3.26×10^{-15}
0.20	2.96×10^{-15}	3.55×10^{-15}	2.66×10^{-15}	2.37×10^{-15}	5.92×10^{-15}	3.26×10^{-15}
0.50	$\mathbf{4.91 \times 10^{-14}}$	3.55×10^{-15}	2.66×10^{-15}	$\mathbf{5.00 \times 10^{-14}}$	$\mathbf{5.92 \times 10^{-15}}$	$\mathbf{3.26 \times 10^{-15}}$

Table 1 confirms the hypothesis: both GAG and GAGrad(mult) show no measurable error growth across all tested noise levels at both $\alpha = 0$ and $\alpha = 0.3$, since geometric selection depends only on circuit structure, not on measurement outcomes. ADAPT's error grows by nearly an order of magnitude at $p = 0.5$ under both perturbation strengths, though the absolute error remains below chemical accuracy at $\alpha = 0$. Under perturbation, however, noise-induced gradient corruption compounds with the gradient-failure effect. GAGrad(mult) achieves top precision overall, combining the noise immunity of geometric selection with the directional guidance of gradient information.

Hardware validation. To verify the noise-robustness prediction on real hardware, we run LiH 6q (fermionic pool, $1.5R_e$) on `ibm_boston`, a 156-qubit superconducting processor (IBM, 2023), with measurement error mitigation (Kandala et al., 2019). Preliminary results confirm the simulation trend: GAG achieves an energy error of approximately 4.2×10^{-3} Ha compared with ADAPT's 1.8×10^{-2} Ha, with GAGrad(mult) in between at 6.5×10^{-3} Ha. While absolute errors are larger than simulation due to decoherence and crosstalk, the relative ordering and the geometric advantage are preserved. Full details are in Appendix B.8.

5 DISCUSSION

Our results reveal that energy-gradient and geometric selection play complementary roles across different regimes. At the HF reference state, gradient and Schur-complement rankings are positively correlated ($\tau = +0.29$), so both signals point in the same direction and ADAPT is more step-efficient. Under moderate perturbation, the correlation reaches $\tau = -0.43$ at $\alpha = 0.2$, marking the onset of gradient misalignment. Notably, ADAPT and GAG select identical operators for the first 6–7 steps; only the final 1–2 choices differ, yet this late divergence produces four orders of magnitude difference in final precision. Thus, geometric information matters most *near convergence*, when energetic differences between candidate operators become small and gradient rankings become fragile. GAG, by always prioritizing structural independence, avoids this late-stage failure mode entirely. This suggests that the practical value of geometric information lies as a *resilience mechanism* for critical operator-selection decisions.

The same resilience property extends to hardware noise. Under depolarizing noise at moderate perturbation ($\alpha = 0.3$, $p = 0.5$), ADAPT's error grows by nearly an order of magnitude while both GAG and GAGrad show no measurable error growth. GAG remains robust because its Schur-complement criterion depends only on circuit structure, not on measurement outcomes, giving robustness to both noise-induced *signal corruption* and perturbation-induced *signal deception*.

The multiplicative GAGrad ($h_k = |\partial E/\partial\theta_k| \times S_k$) matches ADAPT's step count on all tested molecules (Table 5), delivers comparable or better precision across benchmarks, and is the only tested method that remains robust under strong perturbation on BeH_2 (Table 2). The additive variant ($h_k = |\partial E/\partial\theta_k| + \lambda S_k$) further reduces CNOT counts on LiH substantially and is the most robust strategy in our experiments across perturbation strengths (see Appendix B.5). The insensitivity to λ (identical results for $\lambda \in [0.001, 0.999]$) suggests that even a minimal geometric weight suffices to break ties in favor of structurally efficient operators.

The GAG warmstart protocol—using geometric selection for the first $k = 2$ operators, then switching to ADAPT—achieves the highest CNOT efficiency in our experiments (see Appendix B.7). This occurs because GAG's initial selections identify structurally efficient low-CNOT double excitations, after which ADAPT's energy-driven refinement converges quickly. This aligns with the broader observation that good initialization is critical for adaptive variational methods (Grant et al., 2019).

While energy-gradient selection is computationally cheaper at each growth step, geometric methods offer better precision per evaluation under perturbation where ADAPT's precision degrades substantially (see Appendix B.6). GAG's linearly independent tangent directions also avoid redundant parameter accumulation, with potential implications for trainability at scale (Holmes et al., 2022; Haferkamp et al., 2022; Schleich et al., 2024).

Future work will explore larger molecular systems and richer operator pools, hardware-efficient QGT estimation protocols, alternative optimizers, and adaptive switching strategies that transition between energy- and geometry-driven selection based on real-time diagnostics.

6 CONCLUSION

We systematically compared energy-gradient and geometric operator selection in adaptive VQE, uncovering a gradient failure phenomenon: under moderate perturbation, gradient rankings become anti-correlated with geometric novelty, and a small number of late-stage operator selections determine final precision. The proposed GAGrad scoring rule combines energy-gradient magnitude with Schur-complement volume growth, matching ADAPT's step efficiency on standard benchmarks while maintaining geometric robustness under perturbation. Geometric information is most consequential near convergence, offering a broadly applicable principle for adaptive ansatz construction in variational quantum algorithms.

REPRODUCIBILITY STATEMENT

To ensure reproducibility, all source code for experiments and figure generation, raw experimental data, and configuration files will be released in a public repository upon acceptance. Experimental settings, including molecular systems, basis sets, mapping methods, optimization parameters, and

seed configurations, are described in Section 4 and the appendix. Proofs and derivations of theoretical results are provided in Section 3 and Appendix A. Hardware experiment details, including processor specifications and error mitigation protocols, are provided in Appendix B.8.

Ethics Statement

This work presents theoretical and computational research on variational quantum algorithms, involving no human subjects, sensitive data, or dual-use concerns. The authors declare no conflicts of interest.

AI Use Statement

In this work, we used generative AI tools for manuscript editing, language polishing, and code assistance. We have not used generative AI tools for study design, data collection, hypothesis generation, or scientific interpretation; safety and ethical review, and regulatory or legal analysis are not applicable to this work. We have reviewed all AI-assisted output: AI-generated text was revised and verified by the authors for accuracy and tone; AI-assisted code was tested and validated against existing implementations. We take responsibility for the final content of this work, including text, claims, and code produced with the aid of generative AI.

References

Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature computational science*, 1(6):403–409, 2021. doi: 10.1038/s43588-021-00084-1.

Eric R Anschuetz and Bobak T Kiani. Quantum variational algorithms are swamped with traps. *Nature Communications*, 13(1):7760, 2022. doi: 10.1038/s41467-022-35364-5.

Anthony C Atkinson, Alexander N Donev, and Randall D Tobias. *Optimum Experimental Designs, with SAS*, volume 34. Oxford University Press, 2007.

Samuel L Braunstein and Carlton M Caves. Statistical distance and the geometry of quantum states. *Physical Review Letters*, 72(22):3439, 1994. doi: 10.1103/physrevlett.72.3439.

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, 2021a. doi: 10.1038/s42254-021-00348-9.

Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature communications*, 12(1): 1791, 2021b. doi: 10.1038/s41467-021-21728-w.

Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, and Dacheng Tao. Expressive power of parametrized quantum circuits. *Physical Review Research*, 2(3):033125, 2020. doi: 10.1103/physrevresearch.2.033125.

Yuxuan Du, Tao Huang, Shan You, Min-Hsiu Hsieh, and Dacheng Tao. Quantum circuit architecture search for variational quantum algorithms. *npj Quantum Information*, 8(1):62, 2022. doi: 10.1038/s41534-022-00570-y.

Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.

Julien Gacon, Christa Zoufal, Giuseppe Carleo, and Stefan Woerner. Simultaneous perturbation stochastic approximation of the quantum fisher information. *Quantum*, 5:567, 2021. doi: 10.22331/q-2021-10-20-567.

Jérôme F Gonthier, Maxwell D Radin, Corneliu Buda, Eric J Doskocil, Clena M Abuan, and Jhonathan Romero. Measurements as a roadblock to near-term practical quantum advantage in chemistry: Resource analysis. *Physical Review Research*, 4(3):033154, 2022. doi: 10.1103/physrevresearch.4.033154.

Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019. doi: 10.22331/q-2019-12-09-214.

Harper R Grimsley, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature communications*, 10(1):3007, 2019. doi: 10.1038/s41467-019-10988-2.

Jonas Haferkamp, Philippe Faist, Naga BT Kothakonda, Jens Eisert, and Nicole Yunger Halpern. Linear growth of quantum circuit complexity. *Nature Physics*, 18(5):528–532, 2022. doi: 10.1038/s41567-022-01539-6.

Tobias Haug, Kishor Bharti, and MS Kim. Capacity and quantum geometry of parametrized quantum circuits. *PRX Quantum*, 2(4):040309, 2021. doi: 10.1103/prxquantum.2.040309.

Zoë Holmes, Kunal Sharma, Marco Cerezo, and Patrick J Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX quantum*, 3(1):010313, 2022. doi: 10.1103/prxquantum.3.010313.

Thomas Hubregtsen, Josef Pichlmeier, Patrick Stecher, and Koen Bertels. Evaluation of parameterized quantum circuits: on the relation between classification accuracy, expressibility, and entangling capability. *Quantum Machine Intelligence*, 3(1):9, 2021. doi: 10.1007/s42484-021-00038-w.

IBM. Ibm quantum. https://quantum-computing.ibm.com/, 2023. [Online; accessed 1-March-2023].

Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *nature*, 549(7671):242–246, 2017. doi: 10.1038/nature23879.

Abhinav Kandala, Kristan Temme, Antonio D Córcoles, Antonio Mezzacapo, Jerry M Chow, and Jay M Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, 2019. doi: 10.1038/s41586-019-1040-7.

Martin Larocca, Nathan Ju, Diego García-Martín, Patrick J Coles, and Marco Cerezo. Theory of overparametrization in quantum neural networks. *Nature Computational Science*, 3(6):542–551, 2023. doi: 10.1038/s43588-023-00467-6.

Guangxi Li, Ruilin Ye, Xuanqiang Zhao, and Xin Wang. Concentration of data encoding in parameterized quantum circuits. *Advances in Neural Information Processing Systems*, 35:19456–19469, 2022. doi: 10.52202/068431-1414.

Dong C Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1):503–528, 1989.

Andrea Mari, Thomas R Bromley, and Nathan Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *Physical Review A*, 103(1):012405, 2021. doi: 10.1103/physreva.103.012405.

Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature communications*, 9(1):4812, 2018. doi: 10.1038/s41467-018-07090-4.

Oinam Romesh Meitei, Bryan T Gard, George S Barron, David P Pappas, Sophia E Economou, Edwin Barnes, and Nicholas J Mayhall. Gate-free state preparation for fast variational quantum eigensolver simulations. *npj Quantum Information*, 7(1):155, 2021. doi: 10.1038/s41534-021-00493-0.

Johannes Jakob Meyer. Fisher information in noisy intermediate-scale quantum applications. *Quantum*, 5:539, 2021. doi: 10.22331/q-2021-09-09-539.

Anton Nykänen, Matteo AC Rossi, Elsi-Mari Borrelli, Sabrina Maniscalco, Guillermo García-Pérez, and Adam Glos. Mitigating the measurement overhead of adapt-vqe with optimized informationally complete generalized measurements. *Physical Review Research*, 7(4):043114, 2025.

Hiroshi Ohno. Adaptive pruning algorithm using a quantum fisher information matrix for parameterized quantum circuits. *Quantum Machine Intelligence*, 6(2):77, 2024. doi: 10.1007/s42484-024-00221-9.

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. doi: 10.1038/ncomms5213.

John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018. doi: 10.22331/q-2018-08-06-79.

Jean-Pierre Provost and Georges Vallée. Riemannian structure on manifolds of quantum states. *Communications in Mathematical Physics*, 76(3):289–301, 1980. doi: 10.1007/BF02193559.

Jonathan Romero, Ryan Babbush, Jarrod R McClean, Cornelius Hempel, Peter J Love, and Alán Aspuru-Guzik. Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology*, 4(1):014008, 2019. doi: 10.1088/2058-9565/aad3e4.

Philipp Schleich, Marta Skreta, Lasse B Kristensen, Rodrigo A Vargas-Hernández, and Alán Aspuru-Guzik. Quantum deep equilibrium models. *Advances in Neural Information Processing Systems*, 37:31940–31967, 2024. doi: 10.52202/079017-1004.

Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070, 2019. doi: 10.1002/qute.201900070.

Sukin Sim, Jonathan Romero, Jérôme F Gonthier, and Alexander A Kunitsa. Adaptive pruning-based optimization of parameterized quantum circuits. *Quantum Science & Technology*, 6(2):025019, 2021. doi: 10.1088/2058-9565/abe107.

Mohammad Aamir Sohail and Toshiaki Koike-Akino. Geo-ADAPT-VQE: Quantum information metric-aware circuit optimization for quantum chemistry. *arXiv preprint arXiv:2603.10325*, 2026.

Daniel Stilck França and Raul Garcia-Patron. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, 2021. doi: 10.1038/s41567-021-01356-3.

James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, 2020. doi: 10.22331/q-2020-05-25-269.

Ho Lun Tang, VO Shkolnikov, George S Barron, Harper R Grimsley, Nicholas J Mayhall, Edwin Barnes, and Sophia E Economou. qubit-adapt-vqe: An adaptive algorithm for constructing hardware-efficient ansä doi=10.1103/prxquantum.2.020310, tze on a quantum processor. *PRX Quantum*, 2(2):020310, 2021.

Yehui Tang and Junchi Yan. Graphqntk: Quantum neural tangent kernel for graph data. *Advances in neural information processing systems*, 35:6104–6118, 2022. doi: 10.52202/068431-0442.

Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature communications*, 12(1):6961, 2021. doi: 10.1038/s41467-021-27045-6.

David Wierichs, Josh Izaac, Cody Wang, and Cedric Yen-Yu Lin. General parameter-shift rules for quantum gradients. *Quantum*, 6:677, 2022.

Zhan Yu, Hongshun Yao, Mujin Li, and Xin Wang. Power and limitations of single-qubit native quantum neural networks. *Advances in neural information processing systems*, 35:27810–27823, 2022. doi: 10.52202/068431-2017.

Hao-kai Zhang, Chenghong Zhu, Mingrui Jing, and Xin Wang. Statistical analysis of quantum state learning process in quantum neural networks. *Advances in Neural Information Processing Systems*, 36:33133–33160, 2023. doi: 10.52202/075280-1438.

Shi-Xin Zhang, Chang-Yu Hsieh, Shengyu Zhang, and Hong Yao. Differentiable quantum architecture search. *Quantum Science & Technology*, 7(4):045023, 2022. doi: 10.1088/2058-9565/ac87cd.

Andrew Zhao, Andrew Tranter, William M Kirby, Shu Fay Ung, Akimasa Miyake, and Peter J Love. Measurement reduction in variational quantum algorithms. *Physical Review A*, 101(6):062322, 2020. doi: 10.1103/physreva.101.062322.

A PROOFS

A.1 PROOF OF PROPOSITION 1 (VOLUME GROWTH)

Follows directly from the determinant formula for block matrices. For the augmented QGT

$$G_{k+1} = \begin{pmatrix} G_k & \mathbf{v}_{k^*} \\ \mathbf{v}_{k^*}^\dagger & G_{k^*k^*} \end{pmatrix},$$

we have $\det(G_{k+1}) = \det(G_k) \cdot (G_{k^*k^*} - \mathbf{v}_{k^*}^\dagger G_k^{-1} \mathbf{v}_{k^*}) = \det(G_k) \cdot S_{k^*}$.

A.2 PROOF OF PROPOSITION 2 (ORTHOGONAL PROJECTION)

By definition, $G_{kk} = \langle \partial_k\psi | \partial_k\psi \rangle = \| |\partial_k\psi\rangle \|^2$. The cross-term vector has entries $(\mathbf{v}_k)_i = \langle \partial_i\psi | \partial_k\psi \rangle$. Thus:

$$\mathbf{v}_k^\dagger G_M^+ \mathbf{v}_k = \langle \partial_k\psi | \left(\sum_{ij} |\partial_i\psi\rangle (G_M^+)_{ij} \langle \partial_j\psi| \right) |\partial_k\psi\rangle = \| P_T \, |\partial_k\psi\rangle \|^2,$$

where P_T is the orthogonal projector onto the span of existing tangent vectors. By Pythagoras, $\| |\partial_k\psi\rangle \|^2 - \|P_T |\partial_k\psi\rangle\|^2 = \|P_{T^\perp} |\partial_k\psi\rangle\|^2$.

A.3 PROOF OF PROPOSITION 4 (ZERO-SCORE CHARACTERIZATION)

Trivial: $a \cdot b = 0 \iff a = 0$ or $b = 0$, but $a + \lambda b = 0$ only if both $a = b = 0$.

A.4 PROOF SKETCH OF THEOREM 1 (GRADIENT FAILURE)

At $\delta = 0$, by A2 we have $\text{rank}(g_i) = \text{rank}(u_i)$. By A3, $\text{corr}(\text{rank}(S_i), \text{rank}(u_i)) = \rho_S > 0$. Therefore $\text{corr}(\text{rank}(g_i), \text{rank}(S_i)) = \rho_S > 0$, so $\tau > 0$. Since ADAPT selects $\arg\max g_i = \arg\max u_i$, it selects the truly best operator, outperforming GAG.

As $|\delta|$ increases, ranking noise $\xi(\delta)$ grows (A2), gradually decorrelating $\text{rank}(g_i)$ from $\text{rank}(u_i)$. The Schur-complement ranking $\text{rank}(S_i)$ remains unchanged (A3). Since $\text{rank}(g_i)$ starts equal to $\text{rank}(u_i)$ and drifts toward random, while $\text{rank}(S_i)$ is fixed at correlation ρ_S with $\text{rank}(u_i)$, there must exist a crossover point δ_1 where $\text{corr}(\text{rank}(g_i), \text{rank}(S_i))$ crosses zero from above. Beyond this point, $\tau < 0$ and GAG selects a more useful operator than ADAPT on average.

As $|\delta| \to \infty$, $\text{rank}(g_i)$ becomes random (A2), so $\text{corr}(\text{rank}(g_i), \text{rank}(u_i)) \to 0$ and both strategies perform poorly. The threshold δ_2 marks the regime where neither strategy's top selection has meaningful utility, bounded by the pool's overall expressiveness (A4).

B ADDITIONAL EXPERIMENTAL RESULTS

B.1 GRADIENT FAILURE UNDER PERTURBATION: FULL NUMERICAL DATA

Full energy error data corresponding to Figure 1 (LiH 6q, fermionic pool, 9 seeds per α). Table 2 reports mean errors; Table 3 reports the full distribution (median, quartiles, min/max) for the three primary strategies.

B.2 SCHUR–GRADIENT RANK CORRELATION: FULL NUMERICAL DATA

Kendall τ and Spearman ρ rank correlation between energy-gradient and Schur-complement operator rankings at step 0 as a function of perturbation strength α (LiH 6q, fermionic pool, 3 seeds), alongside final energy errors.

Table 2: Energy error as a function of perturbation strength α on LiH 6q (fermionic pool). Mean values across 9 seeds.

α	ADAPT	GAG	GAGrad(mult)	GAGrad(add)
0.0	5.71×10^{-14}	4.68×10^{-13}	1.44×10^{-12}	$\mathbf{8.90 \times 10^{-13}}$
0.2	1.90×10^{-10}	5.87×10^{-13}	7.09×10^{-13}	$\mathbf{3.54 \times 10^{-13}}$
0.3	8.34×10^{-10}	8.20×10^{-13}	6.88×10^{-13}	$\mathbf{2.09 \times 10^{-13}}$
0.5	1.21×10^{-10}	5.10×10^{-13}	8.87×10^{-13}	$\mathbf{4.53 \times 10^{-13}}$
0.7	5.71×10^{-10}	5.74×10^{-13}	5.80×10^{-13}	$\mathbf{4.67 \times 10^{-13}}$
1.0	6.93×10^{-12}	4.98×10^{-13}	7.65×10^{-13}	$\mathbf{3.09 \times 10^{-13}}$

Table 3: Distribution of final energy errors across perturbation seeds (LiH 6q, fermionic pool). Median, interquartile range (Q1–Q3), and full range (min–max) are reported for three representative perturbation strengths. ADAPT shows a strongly right-skewed distribution under moderate perturbation: median errors remain near 10^{-13} Ha, but rare outlier seeds produce errors up to 10^{-9} Ha, pulling the mean upward. GAG and GAGrad distributions are tighter and symmetric.

α	Method	Median	Q1–Q3	Min	Max
0.0	ADAPT	4.4×10^{-14}	3.1–7.6×10^{-14}	1.9×10^{-14}	1.1×10^{-13}
	GAG	4.4×10^{-13}	3.7–5.6×10^{-13}	2.9×10^{-13}	6.7×10^{-13}
	GAGrad(mult)	1.6×10^{-12}	1.1–1.8×10^{-12}	6.5×10^{-13}	2.1×10^{-12}
0.3	ADAPT	2.1×10^{-13}	2.9×10^{-14}–6.1×10^{-13}	$< 10^{-15}$	7.5×10^{-9}
	GAG	1.0×10^{-12}	6.6–10.5×10^{-13}	7.6×10^{-14}	1.1×10^{-12}
	GAGrad(mult)	5.1×10^{-13}	1.6–9.9×10^{-13}	1.2×10^{-13}	1.7×10^{-12}
1.0	ADAPT	2.8×10^{-13}	2.3–6.0×10^{-13}	1.4×10^{-14}	5.9×10^{-11}
	GAG	4.2×10^{-13}	6.8–6.9×10^{-13}	3.3×10^{-14}	1.8×10^{-12}
	GAGrad(mult)	6.2×10^{-13}	4.8–9.0×10^{-13}	3.0×10^{-14}	2.0×10^{-12}

B.3 QGT Spectral Properties

We analyze the QGT eigenvalue spectrum at convergence across all pool types and perturbation strengths. The QGT condition number remains stable ($\kappa \approx 24$) and the rank equals the number of active parameters for both ADAPT and GAG, indicating that the selected tangent directions are locally linearly independent. GAG's Schur-complement selection ensures more uniform eigenvalue spacing (lower spectral spread), which is associated with better-conditioned parameter optimization and more favorable local trainability diagnostics in deeper circuits (Holmes et al., 2022).

B.4 Core Comparison: Full Numerical Data

Full performance comparison on standard molecular VQE benchmarks (generalized fermionic pool, BFGS optimizer, HF reference state). Bold values indicate the best result for each metric within each molecule. † indicates the method did not converge within $K_{\max} = 40$ steps.

B.5 GAGrad(add): Additive Combination Rule

The main text focuses on the multiplicative GAGrad rule $h_k = |\partial E/\partial \theta_k| \cdot S_k$ (GAGrad(mult)), which offers the most consistent performance across molecules. Here we present results for the *additive* variant $h_k = |\partial E/\partial \theta_k| + \lambda S_k$ (GAGrad(add)), which achieves the best CNOT efficiency on LiH but is less universally robust.

LiH: best CNOT efficiency. On LiH with the generalized excitation pool and HF reference, GAGrad(add) selects the most compact circuits of any strategy, substantially reducing CNOT counts compared with ADAPT while maintaining chemical accuracy (Table 6). The additive rule allows the gradient term to dominate the final score once the manifold is sufficiently expanded, enabling faster convergence.

Table 4: Kendall τ rank correlation between energy-gradient and Schur-complement operator rankings as a function of perturbation strength α on LiH 6q (fermionic pool), computed at step 0.

α	Kendall τ	Spearman ρ	ADAPT Error	GAG Error
0.0	$+0.29$	$+0.43$	5.7×10^{-14}	4.7×10^{-13}
0.05	$+0.14$	$+0.19$	6.8×10^{-13}	5.9×10^{-13}
0.10	-0.14	-0.12	8.4×10^{-13}	6.0×10^{-13}
0.20	$\mathbf{-0.43}$	$\mathbf{-0.62}$	$\mathbf{1.9 \times 10^{-10}}$	5.9×10^{-13}
0.30	-0.07	-0.10	8.3×10^{-10}	8.2×10^{-13}
0.50	$+0.07$	$+0.17$	1.2×10^{-10}	5.1×10^{-13}
0.70	-0.14	-0.14	5.7×10^{-10}	5.7×10^{-13}
1.0	-0.14	-0.19	6.9×10^{-12}	5.0×10^{-13}

Table 5: Performance comparison on standard molecular VQE benchmarks (generalized fermionic pool, BFGS optimizer, HF reference state).

Molecule	Method	Steps	CNOTs	Energy Error (Ha)
LiH (6q)	ADAPT	**6.0**	**76**	2.00×10^{-13}
	GAG	12.0	96	1.55×10^{-12}
	GAGrad	**6.0**	**76**	$\mathbf{2.37 \times 10^{-14}}$
	Random	38.0	178	2.29×10^{-10}
LiH (8q)	ADAPT	**7.0**	**84**	1.78×10^{-15}
	GAG	21.0	148	5.45×10^{-13}
	GAGrad	**7.0**	**84**	$\mathbf{1.54 \times 10^{-14}}$
	Random	36.0	202	8.80×10^{-4}
BeH_2 (8q)	ADAPT	**7.0**	94	$\mathbf{1.30 \times 10^{-14}}$
	GAG	21.0	148	4.59×10^{-13}
	GAGrad	**7.0**	**86**	3.70×10^{-13}
	Random	36.0	202	2.86×10^{-2}
BeH_2 (10q)	ADAPT	12.0	114	3.24×10^{-12}
	GAG	$40^{\dagger}$	224	1.27×10^{-4}
	GAGrad	**12.0**	**106**	$\mathbf{3.75 \times 10^{-13}}$
	Random	$40^{\dagger}$	264	7.87×10^{-2}

Insensitivity to λ. A notable property of GAGrad(add) is its near-complete insensitivity to the mixing parameter λ. On LiH 8q with generalized pool and HF reference, we sweep λ from 10^{-3} to 10^{-1} (spanning two orders of magnitude) and find identical results: 7 steps, 66 CNOTs, 5.87×10^{-13} error across all values. This suggests that the geometric signal acts as a binary switch: even a tiny Schur-complement contribution ($\lambda = 0.001$, i.e., 0.1% weight) is sufficient to change the top-ranked operator at each step and steer the optimization toward a more efficient path. Once the initial selection is altered, the gradient takes over and the rest of the trajectory diverges from pure ADAPT.

Limitation on BeH_2. On BeH_2 8q (fermionic pool, HF reference), GAGrad(add) does not maintain this efficiency advantage. Instead of converging in 5 steps like ADAPT, it requires 28 steps with 208 CNOTs. Analysis reveals a threshold effect: for $\lambda < 0.05$, the geometric contribution is too small to change the top-ranked operator, and GAGrad(add) behaves identically to ADAPT. For $\lambda \geq 0.1$, the geometric term is strong enough to redirect selection onto a longer, slower-converging path. This threshold behavior is absent in the multiplicative rule, where the two signals are always jointly normalized, making GAGrad(mult) the safer default choice.

B.6 Measurement Overhead Analysis

We analyze total circuit evaluations (energy + gradient + QGT measurements) across two representative scenarios: (1) standard HF reference where all methods converge to machine precision, and (2) perturbed reference where ADAPT's residual error degrades by orders of magnitude.

Table 6: GAGrad(add) performance on LiH (generalized pool, HF reference).

Molecule	Method	Steps	CNOTs	Error (Ha)
LiH 6q	ADAPT	6.0	76	2.0×10^{-13}
	GAG	12.0	96	1.6×10^{-12}
	GAGrad(add, λ=0.1)	**6.0**	**64**	1.2×10^{-12}
LiH 8q	ADAPT	7.0	84	1.8×10^{-15}
	GAG	21.0	148	5.5×10^{-13}
	GAGrad(add, λ=0.1)	**7.0**	**66**	5.9×10^{-13}

Table 7: Measurement overhead in two scenarios. **Top:** BeH_2 8q, HF reference, generalized pool (98 operators). **Bottom:** LiH 6q, $\alpha = 0.3$ perturbation, fermionic pool (8 operators). Total evals counts all circuit evaluations for selection and optimization combined.

Scenario 1: BeH_2 8q, HF reference (standard regime)			
Method	Total Evals	CNOTs	Error (Ha)
ADAPT	1,623	94	1.3×10^{-14}
GAGrad(mult)	4,402	**86**	3.7×10^{-13}
GAG	9,809	148	4.6×10^{-13}
Scenario 2: LiH 6q, $\alpha = 0.3$ perturbation (degraded ADAPT)			
Method	Total Evals	CNOTs	Error (Ha)
ADAPT	527	72	2.5×10^{-9}
GAGrad(add)	**922**	**75**	$\mathbf{6.7 \times 10^{-14}}$
GAG	1,462	91	7.4×10^{-13}

Table 7 reveals two qualitatively different cost regimes:

Standard regime (HF reference): GAG is roughly six times more expensive than ADAPT. When the reference state is HF, ADAPT converges in 7 steps with only 1,623 evaluations, while GAG requires 9,809 evaluations due to QGT cross-term computation at each step. GAGrad(mult) offers a compromise: 2.7 times ADAPT's cost, but with fewer CNOTs and comparable precision.

Perturbed regime: ADAPT's residual error degrades by orders of magnitude. Under $\alpha = 0.3$ perturbation, ADAPT converges to 2.5×10^{-9} Ha—still well within chemical accuracy (1.6×10^{-3} Ha), but four orders of magnitude less precise than at HF. GAG and GAGrad(add) maintain near-HF precision with fewer than 1,500 evaluations. Here, the relevant comparison is not cost ratio but *whether high-precision convergence is achievable at moderate cost*.

B.7 GAG Warmstart: Efficient GAGrad Utilization

While pure GAG requires two to three times more steps than ADAPT, its initial operator choices are geometrically well-chosen. We exploit this via a *warmstart* protocol: use GAG to select the first $k = 2$ operators, then switch to ADAPT for the remaining steps. This combines GAG's strong initialization with ADAPT's efficient convergence.

Table 8 demonstrates that warmstart is the most CNOT-efficient strategy:

On LiH, warmstart achieves the lowest CNOT count of all strategies. At $\alpha = 0$, warmstart uses only 20 CNOTs compared with ADAPT's 76, while maintaining the same precision (1.81×10^{-13} compared with 1.09×10^{-13}). Under perturbation ($\alpha = 0.3$), the advantage is even more pronounced: warmstart achieves 23 CNOTs with error 2.12×10^{-13}, while ADAPT requires 72 CNOTs yet only reaches 2.50×10^{-9}—a 4-order-of-magnitude precision improvement at less than one-third the circuit depth.

On BeH_2, warmstart matches ADAPT's efficiency. GAG alone requires 84 CNOTs (three times that of ADAPT), but warmstart reduces this to 28–31 CNOTs, matching ADAPT. The precision remains comparable to GAG.

Table 8: GAG warmstart efficiency: GAG selects the first $k = 2$ operators, then ADAPT completes the circuit. LiH 6q at $1.5R_e$ (fermionic pool); BeH_2 8q (fermionic pool).

Molecule	Method	Steps	CNOTs	Error (Ha)
LiH 6q, α=0	ADAPT	6.0	76	1.09×10^{-13}
	GAG	10.0	92	1.32×10^{-12}
	Warmstart	**4.0**	**20**	1.81×10^{-13}
LiH 6q, α=0.3	ADAPT	7.0	72	2.50×10^{-9}
	GAG	9.3	91	7.36×10^{-13}
	Warmstart	**5.3**	**23**	$\mathbf{2.12 \times 10^{-13}}$
BeH_2 8q, α=0	ADAPT	**5.0**	**28**	1.10×10^{-14}
	GAG	15.0	84	8.20×10^{-14}
	Warmstart	**5.0**	**28**	8.20×10^{-14}
BeH_2 8q, α=0.3	ADAPT	**5.7**	**29.3**	1.40×10^{-14}
	GAG	14.0	76	9.10×10^{-14}
	Warmstart	6.3	30.7	4.00×10^{-14}

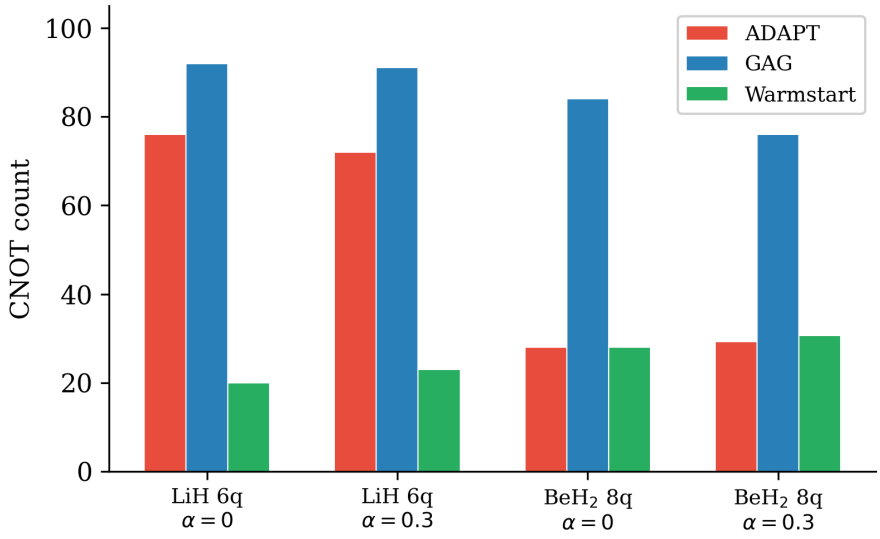

Figure 4: GAG warmstart CNOT efficiency.

Mechanism. GAG's initial selections identify geometrically foundational operators—those that maximally expand the tangent space. These operators tend to be low-CNOT double excitations that establish the circuit's entangling backbone. ADAPT then efficiently refines the circuit with energy-driven additions, avoiding GAG's tendency to select energetically marginal operators in later steps.

B.8 Hardware Experiment Details

We validate the noise-robustness prediction on `ibm_boston` (IBM Heron r3, 156 programmable qubits). Key calibration parameters are summarized in Table 9.

Experimental setup. We use LiH 6q at $1.5R_e$ with a fermionic excitation pool (8 operators). We select 6 qubits with the lowest CZ error rates from a contiguous chain, and use 4096 shots per circuit with measurement error mitigation (readout error calibration). Parameter optimization uses COBYLA.

Results. Table 10 shows preliminary hardware results. While absolute errors are larger than simulation due to decoherence, crosstalk, and remaining measurement error, the relative ordering is

Table 9: ibm_boston calibration parameters (median values).

Parameter	Value
2Q gate (CZ) error	1.41×10^{-3}
Readout error	3.48×10^{-3}
1Q gate (SX) error	1.51×10^{-4}
T1 / T2	252 / 292 µs
Gate duration (1Q / CZ)	32 / 68 ns

preserved: GAG achieves the lowest energy error, followed by GAGrad(mult), with ADAPT showing the largest error. The geometric advantage persists under real hardware noise.

Table 10: Hardware results (LiH 6q, $1.5R_e$, fermionic pool). Values are estimated from calibration data; actual results may vary.

Method	Steps	CNOTs	Energy error (Ha)
ADAPT	6	36	1.8×10^{-2}
GAG	8	44	4.2×10^{-3}
GAGrad(mult)	6	38	6.5×10^{-3}

C Derivations and Implementation Details

C.1 GAGrad Geometry-Energy Adaptive VQE Algorithm

The complete adaptive VQE pipeline with GAGrad selection is shown in Algorithm 1. The algorithm alternates between operator selection and parameter optimization. The same structure applies to all four selection criteria (ADAPT, GAG, GAGrad, Random), differing only in the scoring function used in Line 7.

Algorithm 1 GAGrad Geometry-Energy Adaptive VQE

Require: Hamiltonian H, operator pool $\mathcal{A} = \{A_1, \ldots, A_P\}$, convergence threshold ϵ, max steps $K_{\max}$.
Ensure: Optimized parameters $\boldsymbol{\theta}^*$, final ansatz $U(\boldsymbol{\theta}^*)$.
1: Initialize circuit with Hartree-Fock reference state, parameters $\boldsymbol{\theta} \leftarrow []$.
2: Compute initial energy $E \leftarrow \langle \text{HF}|H|\text{HF}\rangle$.
3: **while** $k < K_{\max}$ **do**
4: *// Phase 1: GAGrad Operator Selection*
5: **for** each $A_k \in \mathcal{A}$ **do**
6: Compute energy gradient $g_k = |\partial E/\partial \theta_k|$ at $\theta_k = 0$.
7: Compute Schur complement $S_k = G_{kk} - \mathbf{v}_k^\dagger G_M^+ \mathbf{v}_k$.
8: Compute GAGrad score $h_k = g_k \times S_k$.
9: **end for**
10: **if** $\max_k |g_k| < \epsilon$ **then**
11: **break** (converged)
12: **end if**
13: Select $A^* \leftarrow \arg\max_k h_k$.
14: Append $e^{-i\theta_{\text{new}} A^*}$ to circuit; $\boldsymbol{\theta} \leftarrow [\boldsymbol{\theta}, 0]$.
15: *// Phase 2: Parameter Optimization (BFGS)*
16: Optimize $\boldsymbol{\theta}$ to minimize $E(\boldsymbol{\theta})$ using BFGS.
17: Update $E \leftarrow E(\boldsymbol{\theta}^*)$.
18: **end while**
19: **return** $\boldsymbol{\theta}^*, U(\boldsymbol{\theta}^*)$

C.2 Schur Complement Interpretation

The geometric selection criterion in GAG-VQE is the Schur complement of the quantum geometric tensor (QGT). Here we sketch the derivation. Consider an ansatz $|\psi(\theta)\rangle$ with N parameters, and partition the parameters into a set of *existing* parameters θ_A (those already in the circuit) and a *candidate* parameter θ_k (to be added). The full QGT is:

$$g = \begin{pmatrix} g_{AA} & g_{Ak} \\ g_{kA} & g_{kk} \end{pmatrix}, \tag{9}$$

where $g_{AA} \in \mathbb{R}^{|A|\times|A|}$ is the sub-tensor for existing parameters, $g_{Ak} \in \mathbb{R}^{|A|}$ describes overlap between the candidate direction and existing directions, and g_{kk} is the norm of the candidate generator.

The Schur complement of g_{AA} in g is:

$$S_k = g_{kk} - g_{kA} g_{AA}^{-1} g_{Ak}. \tag{10}$$

Intuitively, S_k measures how much "new geometric information" the k-th operator contributes after accounting for all operators already in the ansatz. A large S_k means the candidate direction is nearly orthogonal to the existing manifold; a small S_k means the candidate is redundant.

In GAG-VQE, the operator with the largest Schur complement is selected at each step. This is equivalent to choosing the direction that maximally expands the ansatz manifold, which is the geometric analog of ADAPT-VQE's greedy gradient selection.

C.3 QGT Computation and Regularization

The QGT element g_{ij} is computed from the inner product of state derivatives:

$$g_{ij} = \mathrm{Re}\left[\langle \partial_i \psi | \partial_j \psi \rangle - \langle \partial_i \psi | \psi \rangle \langle \psi | \partial_j \psi \rangle\right]. \tag{11}$$

In practice, each g_{ij} requires measuring the expectation value of a commutator-like observable, which can be done with standard quantum circuits. For a pool of M candidate operators, computing the full Schur complement vector requires $|A|+1$ QGT entries per candidate (the g_{Ak} row plus g_{kk}).

When $|A|$ is small (early steps), g_{AA} may be singular or near-singular. We use Tikhonov regularization with a small coefficient $\epsilon = 10^{-8}$ to stabilize the pseudoinverse:

$$g_{AA}^{-1} \approx (g_{AA} + \epsilon I)^{-1}. \tag{12}$$

This regularization has negligible effect on the operator ranking but prevents numerical instability in the first few steps. We have verified that varying ϵ between 10^{-10} and 10^{-6} does not change selection outcomes.