PAC global optimization for VQE in low-curvature geometric regimes

Benjamin Asch

ETH Zürich

We give noise-robust, Probably Approximately Correct (PAC) guarantees of global ε-optimality for the Variational Quantum Eigensolver under explicit geometric conditions. For periodic ansatzes with bounded generators—yielding a globally Lipschitz cost landscape on a toroidal parameter space—we assume that the low-energy region containing the global minimum is a Morse–Bott submanifold whose normal Hessian has rank $r = O(\log p)$ for p parameters, and which satisfies polynomial fiber regularity with respect to coordinate-aligned, embedded flats. This low-curvature-dimensional structure serves as a model for regimes in which only a small number of directions control energy variation, and is consistent with mechanisms such as strong parameter tying together with locality in specific multiscale and tied shallow architectures.

Under this assumption, the sample complexity required to find an ε-optimal region with confidence $1 - \delta$ scales with the curvature dimension r rather than the ambient dimension p. With probability at least $1-\delta$, the algorithm outputs a region in which all points are ε-optimal, and at least one lies within a bounded neighborhood of the global minimum. The resulting complexity is quasi-polynomial in p and ε^{-1} and logarithmic in δ^{-1}. This identifies a geometric regime in which high-probability global optimization remains feasible despite shot noise.

Keywords: Variational quantum eigensolver, Lipschitz optimization, randomized sampling, PAC guarantees, quantum optimization

1 Introduction

Variational quantum algorithms (VQAs) [1] remain among the promising near-term approaches for leveraging noisy intermediate-scale quantum (NISQ) devices [2–4]. They operate by parameterizing a family of quantum circuits and optimizing an expected energy or cost functional. Among the most prominent examples, the Variational Quantum Eigensolver (VQE) [5, 6] has achieved empirical success across quantum chemistry and condensed-matter applications [7, 8]. Despite this progress, the theoretical understanding of variational optimization remains incomplete. The induced landscapes are generally nonconvex and high-dimensional, often shaped by symmetries, degeneracies, and noise that complicate the geometry of global minima [9, 10].

Several analyses capture restricted regimes. When the ansatz map is locally surjective on the relevant manifold, gradient flow can converge to the global minimum given infinite precision and time [11], though this condition is rarely met by hardware-efficient circuits. Conversely, in sufficiently expressive families, gradients can vanish exponentially with system size, producing the barren-plateau phenomenon that renders local descent ineffective [12, 13]. More recently, it has been shown that shallow alternating-layer circuits with guiding states and small initialization admit a linearized training interpretation, enabling convergence and generalization guarantees across input families [14], albeit under a warm-start requirement. Global or gradient-free optimizers—Bayesian, evolutionary, or heuristic [15, 16]—may perform well empirically, but typically lack rigorous non-asymptotic guarantees and scale poorly with the number of parameters. Classical

Benjamin Asch: benjamintasch@berkeley.edu

Lipschitz-based global optimization methods [17] produce correct solutions but rely on uniform partitions of the ambient space, yielding exponential dependence on the parameter dimension. None of these approaches explain why the strong low-dimensional structure empirically observed in VQE landscapes can dramatically reduce the effective difficulty of the optimization.

Our starting point is the geometry of the landscape itself. For a periodic ansatz with bounded generators and a bounded Hamiltonian H, the resulting cost is globally Lipschitz and has uniformly bounded curvature. These regularities arise directly from physically motivated architectures. Rather than viewing VQE optimization as stochastic descent over a rugged surface, we treat it as probabilistic elimination carried out over random geodesic flats of a smooth toroidal manifold.

Low-energy regions in VQE landscapes have been observed to concentrate on lower-dimensional manifolds or "dents" [18–20]. We later describe how strong parameter tying together with architecturally induced locality provides one mechanism that plausibly yields reduced normal Hessian rank, and we highlight other architectures with similar structure. Motivated by this geometry, we introduce *Adaptive Lipschitz Elimination on the Torus* (ALeT), a modification of classical Lipschitz-elimination methods [17] designed to run in quasipolynomial PAC time under this landscape structure. ALeT removes provably suboptimal regions using conservative Lipschitz certificates and finite-shot confidence intervals, and is efficient when the normal Hessian rank satisfies $r = O(\log p)$. With probability at least $1 - \delta$, the returned region is ε-optimal: every point it contains obeys $C(\theta) \leq C^* + \varepsilon$. We emphasize that the guarantees established here are conditional on explicit geometric regularity assumptions (See Section 2); we do not claim tractability of VQE optimization in general. Furthermore, the solution is ε-optimal for the variational landscape, not necessarily for the exact ground-state energy of the Hamiltonian.

This framework unifies geometry and probability into a common perspective on variational optimization. Efficiency arises not from surjectivity or near-convexity, but from structural properties—periodicity, Lipschitz continuity, low-dimensional curvature, and fiber regularity—that are often naturally present in physically motivated settings. Section 2 formalizes the setup and assumptions. Section 3 identifies architectural regimes admitting low normal Hessian rank. Section 4 introduces ALeT and establishes its PAC guarantees. Section 5 discusses limitations, implications, and directions for future work.

Contributions. This work provides a conditional tractability result for global optimization of VQE landscapes. Our main contributions are:

- Structural regime. We formalize a geometric regime—a Morse–Bott low-energy manifold with low normal curvature dimension and polynomial fiber-regularity—under which approximate global optimization becomes feasible with high probability even in the presence of sampling noise.

- Noise-robust global optimization. We introduce ALeT, a Lipschitz-based elimination method operating on random $(r+1)$-dimensional flats, and prove PAC-style guarantees showing it returns an ε-optimal parameter region using only noisy cost evaluations.

- Quasi-polynomial complexity. Under the above structural assumptions, ALeT attains sample complexity $\mathrm{poly}(p, \varepsilon^{-1})^{O(r)}$, which becomes quasi-polynomial whenever the normal curvature rank satisfies $r = O(\log p)$.

- Geometric interpretation. Our analysis indicates that tractability is governed not by global smoothness or convexity surrogates, but by the *effective geometric dimension* of curvature in the low-energy region, providing a precise analogue of structural landscape assumptions in classical nonconvex optimization.

2 Setup

We work with standard product ansatzes on a toroidal parameter space. This section fixes notation and records the conditional geometric and noise assumptions under which the main results hold.

2.1 Parameter space and cost landscape

The variational circuit is

$$U(\theta) \;=\; \prod_{j=1}^{p} e^{-i\theta_j A_j}\,, \qquad \theta \in \mathbb{T}^p := (\mathbb{R}/\gamma\mathbb{Z})^p,$$

acting on a fixed reference state $|0\rangle$. The cost function is

$$C(\theta) \;=\; \langle 0|\, U(\theta)^{\dagger} H U(\theta)\, |0\rangle,$$

for a Hamiltonian H. We equip $\mathbb{T}^p$ with the geodesic distance

$$\mathrm{dist}_{\mathbb{T}^p}(\theta,\theta') \;:=\; \min_{k\in\mathbb{Z}^p} \|\theta-\theta'+\gamma k\|_2,$$

where $\gamma := 2\pi q$ is the periodicity of the torus. We implicitly identify $\mathbb{T}^p$ with its fundamental domain $[-q\pi, q\pi)^p$ whenever convenient. Closed metric balls are denoted $\mathbb{B}(x,\mathfrak{r})$, and an $\mathfrak{r}$-net of $S \subseteq \mathbb{T}^p$ is a finite set of centers whose $\mathfrak{r}$-balls cover S. Hausdorff $\mathcal{H}^k$ is the k-dimensional volume on k-rectifiable sets.

2.2 Periodic generators and Lipschitz regularity

The toroidal structure is induced by periodic dependence of $U(\theta)$ on each parameter. The only structural input is periodicity and a uniform bound on H.

Assumption 2.1 (Periodic bounded generators)**.** *Each gate depends on its parameter only through $e^{-i\theta_j A_j}$, with A_j having rational spectrum $\lambda_k \in \frac{1}{q}\mathbb{Z}$ for some $q\in\mathbb{N}$. Then $\theta \mapsto U(\theta)$ is γ-periodic in each coordinate with $\gamma = 2\pi q$. Throughout, we will (*w.l.o.g.*) take $q = 1$, and we regard θ modulo γ as living on $\mathbb{T}^p$.*

Assumption 2.2 (Bounded Hamiltonian)**.** *The Hamiltonian satisfies $\|H\|_2 \le \Lambda$ for some known $\Lambda > 0$.*

Under these conditions the cost is globally Lipschitz on $(\mathbb{T}^p, \mathrm{dist}_{\mathbb{T}^p})$.

Proposition 2.3 (Global Lipschitz continuity)**.** *Under Assumptions 2.1 and 2.2,*

$$|C(\theta)-C(\theta')| \;\le\; 2\Lambda \sqrt{\sum_j \|A_j\|^2}\, \|\theta-\theta'\|_2 \qquad \text{for all } \theta,\theta' \in \mathbb{T}^p.$$

In particular, C is L-Lipschitz with $L \le 2\Lambda\sqrt{\sum_j \|A_j\|^2}$.

The proof is a standard commutator bound for product ansatzes and is included in App. A for completeness. There we also introduce the *conjugated generators*

$$U_{>j}(\theta) \;:=\; \prod_{k=j+1}^{p} e^{-i\theta_k A_k}, \qquad \widetilde{A}_j(\theta) \;:=\; U_{>j}^{\dagger}(\theta) A_j U_{>j}(\theta),$$

which control both gradients and Hessians in later sections.

2.3 Shot noise model

Algorithmically we never observe $C(\theta)$ exactly but only via finite-shot estimates. We adopt a minimal, homoscedastic model; variance-adaptive generalizations would only change constants.

Assumption 2.4 (Finite-shot measurements)**.** *Each evaluation at θ returns an unbiased estimator $\widehat{C}(\theta)$ computed from n_{shots} independent quantum measurements, with outcomes bounded in $[-R/2, R/2]$ for a known $R > 0$.*

By Hoeffding's inequality we immediately obtain per-point confidence intervals.

Lemma 2.5 (Energy concentration)**.** *Under Assumption 2.4, for any fixed* θ *and* $\alpha \in (0,1)$*,*

$$\mathbb{P}\Big(|\widehat{C}(\theta) - C(\theta)| \le \mathrm{rad}(n_{\mathrm{shots}}, \alpha)\Big) \;\ge\; 1-\alpha,$$

where

$$\mathrm{rad}(n_{\mathrm{shots}}, \alpha) := R\sqrt{\frac{\ln(2/\alpha)}{2n_{\mathrm{shots}}}}.$$

These radii are fed directly into the elimination certificates in Section 4.

2.4 Dent manifold and Morse–Bott structure

The PAC guarantees are localized to a geometrically regular low-energy region containing the global minimum. We model this region as a Morse–Bott critical submanifold.

Definition 2.1 (Dent manifold)**.** A subset $M \subset \mathbb{T}^p$ is a dent manifold for C if,

1. *Global minimum containment:* there exists $\theta^\star \in M$ such that $C(\theta^\star) \le C(\theta)$ for all $\theta \in \mathbb{T}^p$.
2. *Criticality on* M*:* $\nabla C(\theta) = 0$ for all $\theta \in M$.
3. *Constant normal rank:* there exists $r \ge 1$ s.t. $\mathrm{rank}\big(\nabla^2 C(\theta)\big) = \mathrm{rank}\big(\nabla^2 C(\theta)|_{N_\theta M}\big) = r$ for all $\theta \in M$.
4. *Positive curvature in normal directions:* $\nabla^2 C(\theta)|_{N_\theta M} \succ 0$ for all $\theta \in M$.

Here $N_\theta M$ is the normal space to M at θ.

Intuitively, M is a smooth, low-dimensional "valley floor" on which the cost is flat. This is the precise version of the "dent" regime motivated in the introduction. Parameter symmetries, layer tying, and gauge freedoms all naturally generate such Morse–Bott regions. Approximate Morse–Bott would suffice; we state the exact version for clarity and ease of analysis.

2.5 Fiber regularity of the dent

The randomized slicing analysis in Section 4 requires that M not concentrate an exponential fraction of its volume on a vanishing set of fibers. We encode this by a polynomial fiber-regularity condition.

Fix a coordinate-aligned, linear $(r+1)$-plane $V \subset \mathbb{R}^p$ with orthogonal complement $V^\perp$ of dimension $m-1 = p-r-1$, and let $\mathbb{T}_{V^\perp}$ denote the subtorus corresponding to $V^\perp$. For $b \in \mathbb{T}_{V^\perp}$, write

$$X(b) \;:=\; \mathcal{H}^1(M \cap E(b,V)),$$

where $\mathcal{H}^1$ denotes 1-dimensional Hausdorff measure (so $X(b) = 0$ whenever the intersection is empty or 0-dimensional), and $E(b,V) \subset \mathbb{T}^p$ is the geodesic flat through b parallel to V. Let $\mu := \mathbb{E}_b[X(b)]$ denote the expectation with respect to normalized Haar measure on $\mathbb{T}_{V^\perp}$.

Assumption 2.6 (Polynomial fiber regularity)**.** *We say* M *has polynomial fiber regularity (with respect to* V*) if there exists a constant* $A_{p,m} \ge 1$*, depending at most polynomially on* (p,m)*, such that*

$$X(b) \;\le\; A_{p,m}\,\mu \qquad \text{for Haar-almost every } b \in \mathbb{T}_{V^\perp}.$$

Therefore no typical slice exceeds the mean slice length by more than a polynomial factor in (p,m).

Example 2.1 (Tubular dent intersection implies fiber regularity)**.** Suppose that, for $(b,t) \in \mathbb{T}_{V^\perp} \times \mathbb{R}$, the set $M \cap E(b,V)$ is contained in a tube

$$M \cap E(b,V) \;\subseteq\; \{(b,t) : b \in B,\; -\rho(b) \le t \le \rho(b)\},$$

where $B \subset \mathbb{T}_{V^\perp}$ is measurable and

$$0 < \rho_{\min} \leq \rho(b) \leq \rho_{\max} \qquad (b \in B), \qquad \frac{\rho_{\max}}{\rho_{\min}} \leq \mathrm{poly}(p, m),$$

and where the base has non-negligible measure,

$$\frac{\mathcal{H}^{m-1}(B)}{(2\pi)^{m-1}} \geq \frac{1}{\mathrm{poly}(p, m)}.$$

Then M satisfies Assumption 2.6 with $A_{p,m} = \mathrm{poly}(p, m)$.

Proof. For $b \in B$, the slice $M \cap E(b, V)$ coincides with an interval of length $X(b)$, and for $b \notin B$ we have $X(b) = 0$. Thus

$$\mu = \mathbb{E}_b X(b) = \frac{1}{(2\pi)^{m-1}} \int_{\mathbb{T}_{V^\perp}} X(b)\, db \geq \frac{2\rho_{\min}}{\mathrm{poly}(p, m)}.$$

For $b \in B$,

$$X(b) \leq 2\rho_{\max} \leq \frac{\rho_{\max}}{\rho_{\min}} \frac{(2\pi)^{m-1}}{\mathcal{H}^{m-1}(B)}\, \mu = A_{p,m}\, \mu,$$

and for $b \notin B$ the inequality is trivial since $X(b) = 0$. Hence Assumption 2.6 holds with the stated $A_{p,m}$. □

Remark 2.1. Assumption 2.6 is a structural regularity condition placed on the dent manifold M. It rules out the pathological situation where M concentrates an exponential fraction of its volume on a vanishing set of fibers of the projection $\pi_{V^\perp}$. The assumption is not derived from the curvature or positive-reach hypotheses; it is an additional geometric requirement that determines whether randomized slicing can be used to certify non-emptiness with polynomial overhead.

In highly structured VQE landscapes, empirical studies of dent-like minima and symmetry-reduced low-energy regions suggest that the mass of near-optimal sets is typically spread over many fibers rather than collapsing onto a negligible subset. Assumption 2.6 should be viewed as a conditional geometric hypothesis, such that whenever a VQE landscape exhibits a Morse–Bott minimum whose low-energy region is not exponentially concentrated along any coordinate fiber, the guarantees of Section 4 go through.

Establishing verifiable sufficient conditions for fiber regularity–possibly using curvature measures [21, 22] or projection identities of Federer-Kubota type [23]–is an interesting direction but not required for our results. Though harder to motivate, a simple and weaker second-moment bound also suffices for the Paley-Zygmund step of Section 4.

3 Low Hessian Rank

The efficiency of ALeT depends on curvature concentrating in a small number of normal directions. Crucially, the relevant Hessian is the *normal Hessian along the dent manifold M*, not the ambient Hessian at generic parameter values.

In this section we describe architectural conditions under which the normal curvature rank satisfies $r = O(\log p)$. Parameter tying alone does not guarantee low curvature rank unless such tying is so strong so as to induce a rank $O(\log p)$ tying map (chain rule); rather, tying must interact with locality and multiscale structure to collapse the span of *backpropagated generator templates* that control the curvature near M. The following discussion and architectural examples serve as motivation, illustrating mechanisms that plausibly yield low effective curvature rank, but we do not prove low-rank bounds from architectural assumptions in the current work.

3.1 Tying, locality, and the effective curvature dimension

Recall from App. B that the Hessian in a direction $v \in \mathbb{R}^p$ can be written as

$$\partial_v^2 C(\theta) = -\langle [K_\theta(v), [H, K_\theta(v)]] \rangle + \langle i[H, \dot{K}] \rangle,$$

where $K_\theta(v) = \sum_j v_j \tilde{A}_j(\theta)$ is a linear combination of the *backpropagated/conjugated generators* $\tilde{A}_j(\theta) = U^\dagger_{>j}(\theta) A_j U_{>j}(\theta)$, so that the curvature rank at θ is controlled by the dimension of

$$\mathcal{K}(\theta) := \mathrm{span}\{\tilde{A}_j(\theta) : 1 \leq j \leq p\},$$

and its restriction to the normal bundle $N_\theta M$. If $\dim \mathcal{K}(\theta) = r_{\mathrm{eff}}$, then $\mathrm{rank}(\nabla^2 C(\theta)|_{N_\theta M}) \leq r_{\mathrm{eff}}$.

Parameter tying enters only indirectly: tying groups parameters so that many $\tilde{A}_j(\theta)$ coincide *once propagated through the circuit*. Whether this collapse is strong enough to force $r_{\mathrm{eff}} = O(\log p)$ depends on the circuit architecture.

Definition 3.1 (Operator-template collapse)**.** We say an ansatz exhibits operator-template collapse of order r near M if there exists a neighborhood $\mathcal{U}$ of M such that

$$\dim \mathcal{K}(\theta) \leq r \qquad (\theta \in \mathcal{U}).$$

This is the actual structural condition needed for low curvature rank. It is a restriction on the *local effective degrees of freedom of the generator algebra*.

3.2 Tied parameter classes and template collapse

Let the p parameters be partitioned into k tied classes, with $k = O(\log p)$. Write $T : \mathbb{R}^p \to \mathbb{R}^k$ for the induced tying map. Tying alone restricts directions in parameter space, but, does not, except under the aforementioned condition, by itself limit $\mathcal{K}(\theta)$.

However, when tying is combined with *locality* or *multiscale structure*, the backpropagated operators $\tilde{A}_j(\theta)$ do not proliferate independently, but rather remain confined to a light cone of bounded width or a constant-size template at each scale. The effect is that many distinct parameters yield identical or linearly dependent $\tilde{A}_j(\theta)$ once pushed through the circuit.

Consider a depth-D architecture on a bounded-degree lattice, with at most $O(1)$ tied classes per layer. If the backward light cone of each gate has size $O(1)$ (independent of p), then for θ in a neighborhood of any Morse–Bott minimum, $\dim \mathcal{K}(\theta) = O(D)$.

Locality ensures that backpropagation maps each tied class to one of $O(D)$ operator templates, and tangential directions along M contribute no curvature. Choosing depth $D = \Theta(\log p)$ gives $\dim \mathcal{K}(\theta) = O(\log p)$, so the normal Hessian rank is at most $O(\log p)$.

3.3 Architectural regimes achieving $r = O(\log p)$

Hardware-efficient layers with tied gate types. [7, 24] In D alternating layers of local rotations and entanglers on a bounded-degree graph, tying all gates of the same type in each layer gives $O(1)$ tied classes per layer.

$$r_{\mathrm{eff}} = O(D).$$

With $D = \Theta(\log p)$, $r_{\mathrm{eff}} = O(\log p)$.

Shallow multiscale ansatzes (TTN/MERA). [25–27] In tree or MERA layouts, the number of scales is $\Theta(\log p)$, and tying parameters within each scale gives $O(1)$ tied classes per scale. Each scale induces only a constant number of backpropagated templates, suggesting $r_{\mathrm{eff}} = O(\log p)$ in a neighborhood of the optimum.

4 Adaptive Lipschitz Elimination on the Torus (ALeT)

ALeT adapts the Lipschitz-elimination of LIPO [17] to the geometric setting of VQE landscapes on the torus. The algorithm differs in two key respects. First, instead of working in the full parameter space $\mathbb{T}^p$, it restricts search to random $(r+1)$-dimensional geodesic flats $E \subset \mathbb{T}^p$ that intersect the Morse–Bott dent M with positive probability. Second, it replaces exact function evaluations with finite-shot estimates and embeds confidence intervals directly into the elimination rule. These modifications make elimination feasible in the low–normal-rank regime $r = O(\log p)$ and yield PAC guarantees under shot noise.

Let $M \subset \mathbb{T}^p$ be the dent manifold with intrinsic dimension m and normal Hessian rank $r = p{-}m$. Along M, the cost is locally flat ($C|_M \equiv C^\star$ on connected components), while the normal directions exhibit strictly positive curvature.

4.1 Randomized slicing of the dent

We parametrize flats via a linear $(r{+}1)$-plane $V \in \mathrm{Gr}(r{+}1, \mathbb{R}^p)$ and a translation $v \in \mathbb{T}^p$. Writing $w : \mathbb{R}^p \to \mathbb{T}^p$ for the coordinate-wise mod-2π map, define the geodesic flat

$$E(v, V) \;:=\; w(v + V) \;\subset\; \mathbb{T}^p.$$

For each such flat we consider the one-dimensional slice length of the intersection with the dent,

$$X(v) \;:=\; \mathcal{H}^1\big(M \cap E(v, V)\big) \;\in [0, \infty).$$

The following identity expresses the average slice length in terms of the m-dimensional volume of M, $\mathcal{H}^m(M)$, and the average transversality of M to $V^\perp$, $K(V; M)$.

Proposition 4.1 (Translation average)**.** *Fix a coordinate-aligned $(r{+}1)$-plane $V \subset \mathbb{R}^p$ and sample $v \sim \mathrm{Unif}(\mathbb{T}^p)$. Then*

$$\mathbb{E}_v X(v) \;=\; (2\pi)^{-(m-1)}\, K(V; M)\, \mathcal{H}^m(M),$$

where

$$K(V; M) := \frac{1}{\mathcal{H}^m(M)} \int_M J_{m-1}\big(D\pi_{V^\perp}|_{T_x M}\big)\, d\mathcal{H}^m(x)$$

and J_{m-1} is the $(m{-}1)$-Jacobian of $\pi_{V^\perp}$ restricted to $T_x M$.

Proof. Decompose $\mathbb{R}^p = V \oplus V^\perp$ so that $\mathbb{T}^p \simeq \mathbb{T}^{r+1} \times \mathbb{T}^{m-1}$, and let $\pi_{V^\perp} : \mathbb{T}^p \to \mathbb{T}^{m-1}$ be the projection. For $v = (a, b)$ with $a \in \mathbb{T}^{r+1}$ and $b \in \mathbb{T}^{m-1}$, the flat $E(v, V)$ is precisely the fiber $\pi_{V^\perp}^{-1}(b)$. Fubini and the coarea formula applied to $\pi_{V^\perp}|_M$ give

$$\int_{\mathbb{T}^p} X(v)\, d\mathcal{H}^p(v) = (2\pi)^{r+1} \int_M J_{m-1}\big(D\pi_{V^\perp}|_{T_x M}\big)\, d\mathcal{H}^m(x),$$

and dividing by $\mathcal{H}^p(\mathbb{T}^p) = (2\pi)^p$ yields the stated expectation. □

Under Assumption 2.6, the slice lengths are not allowed to concentrate exponentially on a negligible set of fibers.

Lemma 4.2 (Second moment under fiber regularity)**.** *Under Assumption 2.6,*

$$\mathbb{E}_v X(v)^2 \;\le\; \frac{A_{p,m}}{(2\pi)^{2(m-1)}}\, K(V; M)^2\, \mathcal{H}^m(M)^2.$$

Proof. Assumption 2.6 states that $X(v) \le A_{p,m}\mu$ almost everywhere, where $\mu = \mathbb{E}_v X(v)$. Thus $X(v)^2 \le A_{p,m}\mu\, X(v)$ and

$$\mathbb{E}_v X(v)^2 \;\le\; A_{p,m}\mu^2.$$

Substituting μ from Proposition 4.1 gives the claim. □

Combining the first and second moments via Paley–Zygmund yields a uniform lower bound on the probability that a random flat intersects M.

Theorem 4.3 (Hit probability for a random flat)**.** *Let $v \sim \mathrm{Unif}(\mathbb{T}^p)$ and $X(v)$ as above. Then*

$$\Pr_v\big[M \cap E(v, V) \neq \varnothing\big] \;\ge\; \frac{\big(\mathbb{E}_v X(v)\big)^2}{\mathbb{E}_v X(v)^2} \;\ge\; A_{p,m}^{-1}.$$

Remark 4.1. For a generic choice of $(r{+}1)$-plane V, standard transversality arguments imply $K(V; M) > 0$ for any fixed Morse–Bott M; we restrict to such V. Assumption 2.6 then ensures that the second moment of $X(v)$ is controlled polynomially in (p, m). Without such a regularity condition, one can still use positive reach and translative intersection formulas [21, 22, 28], but the resulting bounds generally deteriorate exponentially in m when a large fraction of the mass of M is concentrated in a small set of fibers.

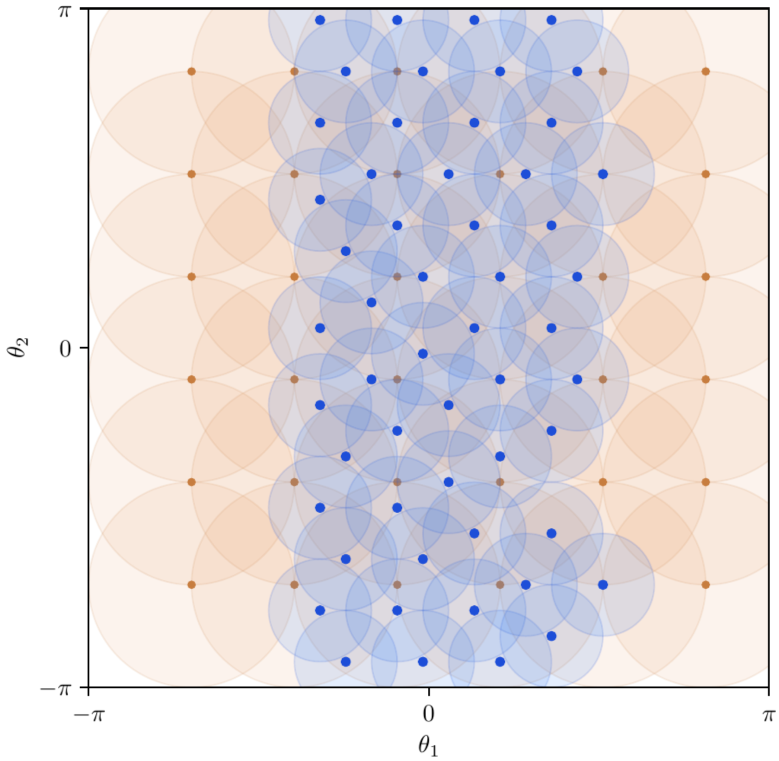

Figure 1: Two steps of ALeT: (orange) an initial $\mathfrak{r}$-net eliminates exterior balls via the EC; (blue) a refined $\mathfrak{r}/2$-net covers the surviving set.

4.2 Noisy elimination certificates on a flat

Fix a flat $E = E(v, V)$ that intersects M, and endow E with its geodesic metric inherited from $\mathbb{T}^p$. Let $L > 0$ be any valid global Lipschitz constant for C on $\mathbb{T}^p$. For a subset $S \subseteq E$ and radius $\mathfrak{r} > 0$, write $\mathcal{C}_{\mathfrak{r}}(S)$ for an $\mathfrak{r}$-net of S.

On round t, ALeT maintains a survivor set $S_t \subseteq E$ and an $\mathfrak{r}_t$-net $\mathcal{C}_t := \mathcal{C}_{\mathfrak{r}_t}(S_t)$, with $\mathfrak{r}_{t+1} = \mathfrak{r}_t/\Delta$ for some fixed $\Delta > 1$. Each query at $\theta \in \mathcal{C}_t$ returns an unbiased estimator $\widehat{C}_t(\theta)$ from n_{shots} independent shots, bounded in $[-R/2, R/2]$ by Assumption 2.4. For any confidence level $\alpha \in (0, 1)$ define

$$\mathrm{rad}(\alpha) := R\sqrt{\frac{\ln(2/\alpha)}{2\, n_{\text{shots}}}}.$$

With $N_t := |\mathcal{C}_t|$, we set

$$\alpha_{t,\theta} := \frac{6\delta_{\text{noise}}}{\pi^2\, t^2\, N_t}, \qquad \theta \in \mathcal{C}_t,$$

and form

$$\mathrm{LCB}_t(\theta) := \widehat{C}_t(\theta) - \mathrm{rad}(\alpha_{t,\theta}), \qquad \mathrm{UCB}_t^\star := \min_{\phi \in \mathcal{C}_t}\big(\widehat{C}_t(\phi) + \mathrm{rad}(\alpha_{t,\phi})\big).$$

The *noisy elimination certificate* at round t is

$$\mathrm{LCB}_t(\theta) \;>\; \mathrm{UCB}_t^\star + L\,\mathfrak{r}_t. \tag{4.1}$$

If (4.1) holds, then for every $x \in B_E(\theta, \mathfrak{r}_t)$ we have $C(x) > \mathrm{UCB}_t^\star \geq \min_{\phi\in\mathcal{C}_t} C(\phi)$, hence the entire ball is certifiably suboptimal with respect to confidence $\alpha_{t,\theta}$ and can be removed. Our choice of α_t is not strictly necessary, but simplifies analysis. In practice, it may often be useful to maintain a constant $\mathrm{rad}(\alpha_t)$ and include this in the chosen ε; for such cases, note that the growth of required n_{shots} at timestep t, n_t, is not prohibitive under the adaptive scheduler.

4.3 ALeT on a fixed flat

Algorithm 1 ALeT on a fixed flat $E = E(v, V) \subset \mathbb{T}^p$

Inputs: final radius $\mathfrak{r}_{\text{fin}}$, shrink factor $\Delta > 1$, Lipschitz bound L, noise budget $\delta_{\text{noise}} \in (0, 1)$.
Initialize $S_0 \leftarrow E$, choose $\mathfrak{r}_0$, and set $\mathcal{C}_0 \leftarrow \mathcal{C}_{\mathfrak{r}_0}(S_0)$.
for $t = 0, 1, 2, \ldots$ **do**
    Query $\widehat{C}_t(\theta)$ for all $\theta \in \mathcal{C}_t$.
    Compute $\mathrm{LCB}_t(\theta)$ and $\mathrm{UCB}_t^\star$ as above.
    Elimination Step:

$$\mathcal{C}_t^{\text{keep}} \leftarrow \Big\{\theta \in \mathcal{C}_t :\ \mathrm{LCB}_t(\theta) \leq \mathrm{UCB}_t^\star + L\mathfrak{r}_t\Big\}.$$

    Set $S_{t+1} \leftarrow \bigcup_{\theta \in \mathcal{C}_t^{\text{keep}}} B_E(\theta, \mathfrak{r}_t)$.
    if $\mathfrak{r}_t \leq \mathfrak{r}_{\text{fin}}$ **then**
        break
    end if
    $\mathfrak{r}_{t+1} \leftarrow \mathfrak{r}_t/\Delta, \quad \mathcal{C}_{t+1} \leftarrow \mathcal{C}_{\mathfrak{r}_{t+1}}(S_{t+1})$.
end for
Output: S_T as an ε-optimal region on E.

4.4 Correctness guarantees on a single flat

We first note that survivor sets are nested.

Theorem 4.4 (Monotonicity on a flat)**.** *For Algorithm 1 one has $S_{t+1} \subseteq S_t$ for all t.*

Proof. By construction, S_{t+1} is the union of balls around $\mathcal{C}_t^{\text{keep}} \subseteq \mathcal{C}_t$, so every kept ball at round $t+1$ is contained in a ball that survived round t. □

The main guarantee is an ε-optimality statement on E with high probability over the measurement noise. Let $\mathfrak{r}_T$ be the final radius reached and define

$$\varepsilon_T := 3L\,\mathfrak{r}_T + 4\,\mathrm{rad}(\alpha_{T,\theta}).$$

Lemma 4.5 (Survival of the minimizer on a flat)**.** *Let $E \subset \mathbb{T}^p$ be a fixed flat and let $x^\star \in E$ be a minimizer of C over E. On the event $\mathcal{G}$ (all confidence intervals hold), the point $x^\star$ lies in every survivor set S_t and, for each round t, there exists a kept center $\theta_t \in \mathcal{C}_t^{\text{keep}}$ such that*

$$\mathrm{dist}_E(\theta_t, x^\star) \leq \mathfrak{r}_t.$$

Proof. We argue by induction on t. For $t = 0$ we have $S_0 = E$, so $x^\star \in S_0$. Assume $x^\star \in S_t$. Because $\mathcal{C}_t$ is an $\mathfrak{r}_t$-net of S_t, there exists $\zeta_t \in \mathcal{C}_t$ with $\mathrm{dist}_E(\zeta_t, x^\star) \leq \mathfrak{r}_t$.

We first show that ζ_t cannot be eliminated on $\mathcal{G}$. Suppose, for contradiction, that $\zeta_t \notin \mathcal{C}_t^{\text{keep}}$, i.e. the elimination certificate (4.1) holds at ζ_t:

$$\mathrm{LCB}_t(\zeta_t) \ > \ \mathrm{UCB}_t^\star + L\,\mathfrak{r}_t.$$

On $\mathcal{G}$, we have $C(\zeta_t) \geq \mathrm{LCB}_t(\zeta_t)$ and $C(\zeta_t) \leq C(x^\star) + L\mathfrak{r}_t$ by Lipschitz continuity and $\mathrm{dist}_E(\zeta_t, x^\star) \leq \mathfrak{r}_t$. Thus

$$C(x^\star) + L\mathfrak{r}_t \ \geq \ C(\zeta_t) \ \geq \ \mathrm{LCB}_t(\zeta_t) \ > \ \mathrm{UCB}_t^\star + L\,\mathfrak{r}_t,$$

which implies $C(x^\star) > \mathrm{UCB}_t^\star$. On the other hand, for any $\phi \in \mathcal{C}_t$, the event $\mathcal{G}$ gives $C(\phi) \leq \widehat{C}_t(\phi) + \mathrm{rad}(\alpha_{t,\phi})$, so

$$\mathrm{UCB}_t^\star = \min_{\phi \in \mathcal{C}_t} \big(\widehat{C}_t(\phi) + \mathrm{rad}(\alpha_{t,\phi})\big) \ \geq \ \min_{\phi \in \mathcal{C}_t} C(\phi) \ \geq \ C(x^\star),$$

since $x^\star$ minimizes C over $E \supseteq \mathcal{C}_t$. This contradicts $C(x^\star) > \mathrm{UCB}_t^\star$. Hence $\zeta_t \in \mathcal{C}_t^{\text{keep}}$.

By definition of S_{t+1},

$$S_{t+1} = \bigcup_{\theta \in \mathcal{C}_t^{\mathrm{keep}}} B_E(\theta, \mathfrak{r}_t),$$

and $x^\star \in B_E(\zeta_t, \mathfrak{r}_t)$ by construction, so $x^\star \in S_{t+1}$. Taking $\theta_t := \zeta_t$ completes the induction and the proof. □

Theorem 4.6 (ε-optimality with confidence)**.** *Let $C^\star := \min_{\theta \in E} C(\theta)$. With probability at least $1 - \delta_{\mathrm{noise}}$ (over all measurement noise), every $x \in S_T$ returned by Algorithm 1 satisfies*

$$C(x) \;\le\; C^\star + \varepsilon_T.$$

Proof. Let $\mathcal{G}$ be the event that every queried θ at every round has its true value within the corresponding confidence interval, i.e.

$$|\,\widehat{C}_t(\theta) - C(\theta)\,| \le \mathrm{rad}(\alpha_{t,\theta}) \quad \text{for all } t \text{ and all } \theta \in \mathcal{C}_t.$$

By a union bound and the choice of $\alpha_{t,\theta}$,

$$\Pr(\mathcal{G}^c) \;\le\; \sum_{t \ge 1} \sum_{\theta \in \mathcal{C}_t} \alpha_{t,\theta} \;\le\; \sum_{t \ge 1} \frac{6\delta_{\mathrm{noise}}}{\pi^2 t^2} \;\le\; \delta_{\mathrm{noise}}.$$

Hence $\Pr(\mathcal{G}) \ge 1 - \delta_{\mathrm{noise}}$.

Fix a round t and work on the event $\mathcal{G}$. Let $R_t := \max_{\phi \in \mathcal{C}_t} \mathrm{rad}(\alpha_{t,\phi})$; note that in our construction $\mathrm{rad}(\alpha_{t,\phi})$ does not depend on ϕ, so $\mathrm{rad}(\alpha_{t,\phi}) = R_t$ for all $\phi \in \mathcal{C}_t$.

Take any kept center $\theta \in \mathcal{C}_t^{\mathrm{keep}}$. Since θ is not eliminated, the elimination certificate (4.1) fails at θ, so

$$\mathrm{LCB}_t(\theta) \;=\; \widehat{C}_t(\theta) - \mathrm{rad}(\alpha_{t,\theta}) \;\le\; \mathrm{UCB}_t^\star + L\,\mathfrak{r}_t.$$

Rearranging and using $\mathrm{rad}(\alpha_{t,\theta}) = R_t$ gives

$$\widehat{C}_t(\theta) \;\le\; \mathrm{UCB}_t^\star + L\,\mathfrak{r}_t + R_t.$$

On $\mathcal{G}$ we have $C(\theta) \le \widehat{C}_t(\theta) + \mathrm{rad}(\alpha_{t,\theta}) = \widehat{C}_t(\theta) + R_t$, hence

$$C(\theta) \;\le\; \mathrm{UCB}_t^\star + L\,\mathfrak{r}_t + 2R_t. \tag{4.2}$$

Let $x^\star \in E$ be a minimizer of C over E, so $C(x^\star) = C^\star$. Because $\mathcal{C}_t$ is an $\mathfrak{r}_t$-net of S_t and $x^\star \in S_t$, there exists $\zeta \in \mathcal{C}_t$ such that $\mathrm{dist}_E(\zeta, x^\star) \le \mathfrak{r}_t$. Lipschitz continuity then implies

$$C(\zeta) \;\le\; C(x^\star) + L\,\mathfrak{r}_t \;=\; C^\star + L\,\mathfrak{r}_t.$$

On $\mathcal{G}$,

$$\widehat{C}_t(\zeta) + \mathrm{rad}(\alpha_{t,\zeta}) \;\le\; C(\zeta) + 2\mathrm{rad}(\alpha_{t,\zeta}) \;\le\; C^\star + L\,\mathfrak{r}_t + 2R_t.$$

By definition of $\mathrm{UCB}_t^\star$,

$$\mathrm{UCB}_t^\star \;=\; \min_{\phi \in \mathcal{C}_t} \big(\widehat{C}_t(\phi) + \mathrm{rad}(\alpha_{t,\phi})\big) \;\le\; \widehat{C}_t(\zeta) + \mathrm{rad}(\alpha_{t,\zeta}),$$

so combining the last two displays yields

$$\mathrm{UCB}_t^\star \;\le\; C^\star + L\,\mathfrak{r}_t + 2R_t. \tag{4.3}$$

Substituting (4.3) into (4.2) gives, for every $\theta \in \mathcal{C}_t^{\mathrm{keep}}$,

$$C(\theta) \;\le\; \big(C^\star + L\,\mathfrak{r}_t + 2R_t\big) + L\,\mathfrak{r}_t + 2R_t \;=\; C^\star + 2L\,\mathfrak{r}_t + 4R_t.$$

At the final round T, every $x \in S_T$ lies within distance at most $\mathfrak{r}_T$ of some kept center $\theta_x \in \mathcal{C}_T^{\mathrm{keep}}$ by construction of the net. Using the Lipschitz property once more,

$$C(x) \;\le\; C(\theta_x) + L\,\mathfrak{r}_T \;\le\; C^\star + 3L\,\mathfrak{r}_T + 4R_T,$$

where $R_T = \max_{\phi \in \mathcal{C}_T} \mathrm{rad}(\alpha_{T,\phi}) = \mathrm{rad}(\alpha_{T,\theta^\star})$ for any $\theta^\star \in \mathcal{C}_T$.

By the definition of ε_T, $\varepsilon_T = 3L\,\mathfrak{r}_T + 4\,\mathrm{rad}(\alpha_{T,\theta}) = 3L\,\mathfrak{r}_T + 4R_T$, so we obtain

$$C(x) \;\le\; C^\star + \varepsilon_T$$

for all $x \in S_T$ on the event $\mathcal{G}$. Since $\Pr(\mathcal{G}) \ge 1 - \delta_{\mathrm{noise}}$, the theorem follows. □

We now quantify the number of oracle calls required to reach a prescribed final radius on a single flat.

Theorem 4.7 (Sample complexity on a single flat). *Let $E = E(v, V) \subset \mathbb{T}^p$ be an $(r{+}1)$-dimensional geodesic flat and assume $C|_E$ is L-Lipschitz. Let $\varepsilon_{\mathrm{eff}} := \varepsilon_T - 4\mathrm{rad}_T > 0$ be the effective accuracy at the final radius*

$$\mathfrak{r}_T = \frac{\varepsilon_{\mathrm{eff}}}{3L}, \qquad \mathrm{rad}_T := \mathrm{rad}(n_{\mathrm{shots}}, \alpha_{T,\theta}).$$

Let $\kappa_{r+1} := \mathcal{H}^{r+1}(B_1(0))$ denote the volume of the $(r{+}1)$-dimensional Euclidean unit ball. Then there exists a universal constant $C_0 > 0$ such that

$$N_{\mathrm{flat}} \;\le\; C_0\, \kappa_{r+1}^{-1} \left(\frac{6\pi L}{\varepsilon_{\mathrm{eff}}}\right)^{r+1},$$

where N_{flat} is the total number of oracle queries made by Algorithm 1 on E.

Proof. Each connected component of E is a flat $(r{+}1)$-torus of diameter at most 2π in the induced metric. Let $d := r{+}1$. For any $\rho \in (0, 2\pi]$, a standard volume argument yields

$$N_{\mathrm{cov}}(E, \rho) \;\le\; C_1 \frac{\mathcal{H}^d(E)}{\mathcal{H}^d(B_\rho)} \;=\; C_1\, \kappa_d^{-1} \Big(\frac{2\pi}{\rho}\Big)^d, \tag{4.4}$$

for some universal $C_1 > 0$. At the last elimination round, $\mathcal{C}_T$ is an $\mathfrak{r}_T$-net of a subset of E, so

$$|\mathcal{C}_T| \;\le\; N_{\mathrm{cov}}(E, \mathfrak{r}_T) \;\le\; C_1\, \kappa_d^{-1} \Big(\frac{2\pi}{\mathfrak{r}_T}\Big)^d.$$

Because $\mathfrak{r}_{t+1} = \mathfrak{r}_t/\Delta$, the covering numbers $|\mathcal{C}_t|$ grow geometrically in t, and the total number of queries satisfies

$$N_{\mathrm{flat}} = \sum_{t=0}^{T} |\mathcal{C}_t| \;\le\; C_2\, |\mathcal{C}_T| \;\le\; C_0\, \kappa_d^{-1} \Big(\frac{2\pi}{\mathfrak{r}_T}\Big)^d$$

for a universal constant $C_2 > 0$ and $C_0 := C_1 C_2$. Substituting $\mathfrak{r}_T = \varepsilon_{\mathrm{eff}}/(3L)$ yields the stated bound. □

The dependence on dimension can be made explicit via the asymptotics of κ_d.

Corollary 4.8 (Quasi-polynomial scaling for $r = O(\log p)$). *Under the hypotheses of Theorem 4.7, one has*

$$\log N_{\mathrm{flat}} \;=\; O\!\left(d \log d + d \log \frac{L}{\varepsilon_{\mathrm{eff}}}\right), \qquad d = r{+}1.$$

In particular, if $r = O(\log p)$, then $d = O(\log p)$ and

$$N_{\mathrm{flat}} \;\le\; \exp\!\Big(O\big(\log p\, [\log\log p + \log(L/\varepsilon_{\mathrm{eff}})\,]\big)\Big),$$

i.e. quasi-polynomial in both p and $1/\varepsilon_{\mathrm{eff}}$.

Proof. The unit-ball volume satisfies

$$\kappa_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)} = \exp\!\big(-\tfrac{d}{2}\log d + O(d)\big),$$

so $\log \kappa_d^{-1} = O(d \log d)$ as $d \to \infty$. Taking logs in Theorem 4.7 gives

$$\log N_{\mathrm{flat}} \;\le\; O\big(d \log d\big) + d \log \frac{L}{\varepsilon_{\mathrm{eff}}},$$

which is the first claim. If $r = O(\log p)$ then $d = O(\log p)$ and $\log d = O(\log\log p)$, yielding the stated quasi-polynomial bound. □

4.5 Global PAC guarantee over multiple flats

Algorithm 1 provides an ε-optimal region on a *fixed* flat E that intersects M, with failure probability controlled by the measurement noise budget. To obtain a global guarantee on $\mathbb{T}^p$, we run ALeT on multiple independent random flats and use Theorem 4.3 to control the probability that no flat ever meets the dent manifold. We then select a subset of survivor sets via a simple data-dependent rule based on empirical costs.

We explicitly distinguish two sources of failure:

- *Intersection failure.* None of the sampled flats intersects M.
- *Noise failure.* At least one confidence interval used by ALeT on some flat is violated.

Let $\delta_{\rm int}, \delta_{\rm noise} \in (0,1)$ be budgets for these two events and write

$$\delta \;:=\; \delta_{\rm int} + \delta_{\rm noise}.$$

Algorithm over flats. Sample ℓ independent translations $v^{(1)}, \dots, v^{(\ell)} \sim \mathrm{Unif}(\mathbb{T}^p)$ and set $E^{(j)} := E(v^{(j)}, V)$ for a fixed coordinate-aligned $(r+1)$-plane V. On each $E^{(j)}$ run Algorithm 1 with *per-flat* noise budget $\delta_{\rm noise}/\ell$ and final resolution $\mathfrak{r}_T > 0$, obtaining survivor sets $S_T^{(j)} \subseteq E^{(j)}$ and terminal nets $\mathcal{C}_T^{(j)} = \mathcal{C}_{\mathfrak{r}_T}(S_T^{(j)})$. Define the per-flat and global empirical minima

$$\widehat{C}_{\min}^{(j)} := \min_{\theta \in \mathcal{C}_T^{(j)}} \widehat{C}_T(\theta), \qquad \widehat{C}_{\min} := \min_{1 \le j \le \ell} \widehat{C}_{\min}^{(j)}.$$

Let $R_T := \mathrm{rad}(n_{\rm shots}, \alpha_{T,\theta})$ be the terminal confidence radius (common to all centers at round T), and define the selection band

$$\beta_T \;:=\; L\,\mathfrak{r}_T + 2R_T.$$

We keep exactly those flats that contain at least one "near-minimal" center:

$$J_{\rm keep} := \Big\{ j \in \{1, \dots, \ell\} : \; \exists\, \theta \in \mathcal{C}_T^{(j)} \text{ with } \widehat{C}_T(\theta) \le \widehat{C}_{\min} + \beta_T \Big\}.$$

Finally, we output the union

$$S_{\rm out} \;:=\; \bigcup_{j \in J_{\rm keep}} S_T^{(j)}.$$

Remark 4.2. The above selection rule is fully implementable: it uses only the terminal empirical estimates $\widehat{C}_T(\theta)$ and the known radii R_T and $\mathfrak{r}_T$. It ensures that any flat whose true minimum is within $O(L\mathfrak{r}_T + R_T)$ of the global minimum will be retained whenever all confidence intervals hold.

We now state the global PAC guarantee, phrased in terms of the terminal radius $\mathfrak{r}_T$ and confidence radius R_T.

Theorem 4.9 (Global PAC guarantee on $\mathbb{T}^p$). *Let $M \subset \mathbb{T}^p$ be a Morse–Bott dent manifold with normal Hessian rank $r = p - m$ and polynomial fiber regularity in the sense of Assumption 2.6, with parameter $A_{p,m} = \mathrm{poly}(p, m)$. Let $L > 0$ be a Lipschitz constant for C on $\mathbb{T}^p$. Fix $\delta \in (0,1)$ and choose $\delta_{\rm int}, \delta_{\rm noise} > 0$ with $\delta_{\rm int} + \delta_{\rm noise} \le \delta$. Run the above multi-flat procedure with*

$$\ell \;\ge\; A_{p,m} \log \frac{1}{\delta_{\rm int}}$$

and per-flat noise budget $\delta_{\rm noise}/\ell$ in each call to Algorithm 1. Suppose the terminal radius $\mathfrak{r}_T$ and confidence radius R_T satisfy

$$\varepsilon \;:=\; 5L\,\mathfrak{r}_T + 8R_T \;>\; 0.$$

Then, with probability at least $1 - \delta$ (over both the random choice of flats and all measurement noise), the output $S_{\rm out}$ obeys:

1. Every $\theta \in S_{\rm out}$ is ε-optimal,

$$C(\theta) \;\le\; C^\star + \varepsilon, \qquad C^\star := \min_{\phi \in \mathbb{T}^p} C(\phi).$$

2. At least one point of $S_{\rm out}$ lies within geodesic distance at most $\mathfrak{r}_T$ of the global minimum set M, i.e. there exists $\theta_{\rm hit} \in S_{\rm out}$ and $\theta^\star \in M$ with $\mathrm{dist}_{\mathbb{T}^p}(\theta_{\rm hit}, \theta^\star) \le \mathfrak{r}_T$.

Moreover, if $r = O(\log p)$ and $A_{p,m} = \mathrm{poly}(p, m)$, the total number of oracle queries scales quasi-polynomially in p and $1/\varepsilon$.

Proof. **Noise event.** For flat $E^{(j)}$, let $\mathcal{G}_j$ denote the event that all confidence intervals used by Algorithm 1 on that flat hold. By the proof of Theorem 4.6, run with noise budget $\delta_{\rm noise}/\ell$, we have $\Pr(\mathcal{G}_j^c) \le \delta_{\rm noise}/\ell$. A union bound gives

$$\Pr\Big(\bigcup_{j=1}^{\ell} \mathcal{G}_j^c\Big) \;\le\; \sum_{j=1}^{\ell} \Pr(\mathcal{G}_j^c) \;\le\; \delta_{\rm noise},$$

so the global noise-good event $\mathcal{G} := \bigcap_{j=1}^{\ell} \mathcal{G}_j$ satisfies $\Pr(\mathcal{G}) \ge 1 - \delta_{\rm noise}$.

Intersection event. By Theorem 4.3, a single random flat intersects M with probability at least $1/A_{p,m}$. For ℓ independent flats, the probability that none intersects M is at most $(1 - 1/A_{p,m})^\ell \le \exp(-\ell/A_{p,m})$. Choosing $\ell \ge A_{p,m}\log(1/\delta_{\rm int})$ ensures that the intersection failure event $\mathcal{W}^c$ ("no flat hits M") has probability at most $\delta_{\rm int}$, so $\Pr(\mathcal{W}) \ge 1 - \delta_{\rm int}$.

Step 1: the hitting flat is kept. Work on $\mathcal{G} \cap \mathcal{W}$. Let $x^\star \in M$ be a global minimizer of C on $\mathbb{T}^p$, so $C(x^\star) = C^\star$. On $\mathcal{W}$ there exists at least one flat $E^{(j_{\rm hit})}$ with $E^{(j_{\rm hit})} \cap M \neq \varnothing$, and in particular $x^\star \in E^{(j_{\rm hit})}$.

For each flat $E^{(j)}$, let $x_\star^{(j)}$ be a minimizer of C over $E^{(j)}$, and set $C_\star^{(j)} := C\big(x_\star^{(j)}\big)$. Then $C_\star^{(j_{\rm hit})} = C^\star$. By Lemma 4.5 applied on each flat under $\mathcal{G}_j$, for every j and every round t there exists a kept center $\zeta_t^{(j)} \in \mathcal{C}_t^{(j),\rm keep}$ with $\mathrm{dist}_{E^{(j)}}(\zeta_t^{(j)}, x_\star^{(j)}) \le \mathfrak{r}_t$. In particular, at $t = T$ we obtain a kept center $\zeta^{(j)} \in \mathcal{C}_T^{(j),\rm keep}$ such that

$$\mathrm{dist}_{E^{(j)}}(\zeta^{(j)}, x_\star^{(j)}) \le \mathfrak{r}_T.$$

By Lipschitz continuity and $\mathcal{G}_j$,

$$C(\zeta^{(j)}) \;\le\; C_\star^{(j)} + L\,\mathfrak{r}_T, \qquad \widehat{C}_T(\zeta^{(j)}) \;\le\; C(\zeta^{(j)}) + R_T \;\le\; C_\star^{(j)} + L\,\mathfrak{r}_T + R_T.$$

Let $\widehat{C}_{\min}^{(j)} := \min_{\theta \in \mathcal{C}_T^{(j)}} \widehat{C}_T(\theta)$ and $\widehat{C}_{\min} := \min_j \widehat{C}_{\min}^{(j)}$. On $\mathcal{G}$ we also have

$$\widehat{C}_T(\theta) \ge C(\theta) - R_T \quad \text{for all } \theta,$$

so

$$\widehat{C}_{\min}^{(j)} \ge C_\star^{(j)} - R_T \quad \text{for all } j, \qquad \widehat{C}_{\min} \ge C^\star - R_T.$$

On the other hand,

$$\widehat{C}_{\min}^{(j_{\rm hit})} \;\le\; \widehat{C}_T(\zeta^{(j_{\rm hit})}) \;\le\; C^\star + L\,\mathfrak{r}_T + R_T,$$

so

$$\widehat{C}_{\min} \;\le\; \widehat{C}_{\min}^{(j_{\rm hit})} \;\le\; C^\star + L\,\mathfrak{r}_T + R_T.$$

Combining,

$$\big|\widehat{C}_{\min} - C^\star\big| \;\le\; L\,\mathfrak{r}_T + R_T.$$

Now consider the hitting flat $j_{\rm hit}$. Its kept center $\zeta^{(j_{\rm hit})}$ satisfies

$$\widehat{C}_T\big(\zeta^{(j_{\rm hit})}\big) \;\le\; C^\star + L\,\mathfrak{r}_T + R_T,$$

while, as just shown, $\widehat{C}_{\min} \ge C^\star - R_T$. Hence

$$\widehat{C}_T\big(\zeta^{(j_{\rm hit})}\big) - \widehat{C}_{\min} \;\le\; (C^\star + L\,\mathfrak{r}_T + R_T) - (C^\star - R_T) \;=\; L\,\mathfrak{r}_T + 2R_T \;=\; \beta_T.$$

By definition of $J_{\rm keep}$, this implies $j_{\rm hit} \in J_{\rm keep}$. In particular, $S_T^{(j_{\rm hit})} \subseteq S_{\rm out}$.

Step 2: flat-wise minima on kept flats are near-global. Let $j \in J_{\text{keep}}$. By definition of J_{keep} there exists a "witness" $\theta^{(j)} \in \mathcal{C}_T^{(j)}$ such that

$$\widehat{C}_T(\theta^{(j)}) \leq \widehat{C}_{\min} + \beta_T.$$

On $\mathcal{G}$, for any θ we have $\widehat{C}_T(\theta) \geq C(\theta) - R_T$, so

$$C(\theta^{(j)}) \leq \widehat{C}_T(\theta^{(j)}) + R_T \leq \widehat{C}_{\min} + \beta_T + R_T.$$

Using $\widehat{C}_{\min} \leq C^\star + L\,\mathfrak{r}_T + R_T$ from above,

$$C(\theta^{(j)}) \leq C^\star + L\,\mathfrak{r}_T + R_T + \beta_T + R_T = C^\star + 2L\,\mathfrak{r}_T + 4R_T.$$

Since $\theta^{(j)} \in E^{(j)}$, the flat-wise minimum obeys

$$C_\star^{(j)} = \min_{x \in E^{(j)}} C(x) \leq C(\theta^{(j)}) \leq C^\star + 2L\,\mathfrak{r}_T + 4R_T.$$

Step 3: global ε-optimality of S_{out}. For each flat $E^{(j)}$, Theorem 4.6 (applied with per-flat budget $\delta_{\text{noise}}/\ell$) guarantees that, on $\mathcal{G}_j$,

$$C(x) \leq C_\star^{(j)} + \varepsilon_T, \qquad x \in S_T^{(j)},$$

where $\varepsilon_T = 3L\,\mathfrak{r}_T + 4R_T$. Therefore, on $\mathcal{G}$ and for any $j \in J_{\text{keep}}$ and any $x \in S_T^{(j)}$,

$$C(x) \leq C_\star^{(j)} + 3L\,\mathfrak{r}_T + 4R_T \leq \big(C^\star + 2L\,\mathfrak{r}_T + 4R_T\big) + 3L\,\mathfrak{r}_T + 4R_T = C^\star + 5L\,\mathfrak{r}_T + 8R_T.$$

By definition, every $\theta \in S_{\text{out}}$ lies in some $S_T^{(j)}$ with $j \in J_{\text{keep}}$, so

$$C(\theta) \leq C^\star + \varepsilon, \qquad \varepsilon := 5L\,\mathfrak{r}_T + 8R_T,$$

which proves item (1).

For item (2), recall from Step 1 that $j_{\text{hit}} \in J_{\text{keep}}$. By Lemma 4.5, the global minimizer $x^\star$ lies in every survivor set $S_t^{(j_{\text{hit}})}$ and, in particular, in $S_T^{(j_{\text{hit}})} \subseteq S_{\text{out}}$. Hence we may take $\theta_{\text{hit}} := x^\star$ and $\theta^\star := x^\star$, yielding $\mathrm{dist}_{\mathbb{T}^p}(\theta_{\text{hit}}, \theta^\star) = 0 \leq \mathfrak{r}_T$.

Step 4: failure probability and sample complexity. We have

$$\Pr\big((\mathcal{G} \cap \mathcal{H})^c\big) \leq \Pr(\mathcal{G}^c) + \Pr(\mathcal{H}^c) \leq \delta_{\text{noise}} + \delta_{\text{int}} \leq \delta,$$

so both properties hold with probability at least $1-\delta$. The stated quasi-polynomial sample complexity follows by combining Corollary 4.8 (for each flat) with the bound $\ell \leq \mathrm{poly}(p,m)\log(1/\delta_{\text{int}})$ when $A_{p,m} = \mathrm{poly}(p,m)$ and $r = O(\log p)$. □

5 Discussion and Outlook

This work develops a geometric–probabilistic framework for global variational optimization on the torus. Starting from the regularity implied by periodic generators and a bounded Hamiltonian (§2), we identified three key structural properties that govern tractability: a Morse–Bott minimum set, concentration of curvature in $r = O(\log p)$ normal directions, and a fiber-regularity condition on the low-energy region. Within this regime, ALeT provides noise-robust PAC guarantees whose dependence on p is quasipolynomial. The central point is that the analysis rests on geometric structure—not on surjectivity conditions or fine-grained access to derivatives.

Scope. The guarantees obtained here are not intended to model generic VQE behavior. They apply to a structured subclass characterized by periodic parameterizations, uniform Lipschitz regularity, and a low-curvature-dimensional geometry near the optimum. In such cases, randomized slicing followed by Lipschitz elimination removes large portions of the search region at controlled cost. The analysis tracks only intersection probabilities and high-confidence elimination, making the regime both mathematically transparent and practically testable.

Relation to hardness. These results are fully consistent with known hardness of ground-state problems [29, 30]. Hardness reductions generate landscapes with rapidly oscillatory or combinatorially encoded structure, outside the Morse–Bott and fiber-regularity conditions used here. In contrast, the architectural mechanisms leading to $r = O(\log p)$ in §3—layer tying and shallow multi-scale patterns—reduce expressivity in exchange for controlled curvature. Our guarantees therefore delineate settings in which global optimization becomes feasible, rather than asserting tractability in the worst case.

Assumptions in context. The global Lipschitz bound (Prop. 2.3) is conservative; in practice, the relevant local constants along survivor sets are often smaller. The normal-rank condition need not hold globally—it is sufficient that the Hessian maintains rank r on neighborhoods of near-optimal level sets. Finally, Assumption 2.6 enters only through the second-moment bound required for the Paley–Zygmund argument; no additional global geometric control is used.

Noise and robustness. The guarantees assume bounded, unbiased, IID shot noise, but the analysis extends directly to variance-adaptive confidence radii (e.g., empirical Bernstein) and to mild correlations. The elimination mechanism depends only on the availability of high-probability deviation bounds and is otherwise agnostic to the particular estimator. This modularity is beneficial in NISQ regimes where noise characteristics may vary across evaluations.

Practical considerations. ALeT requires only the ability to evaluate $\widehat{C}(\theta)$ at prescribed centers and to maintain survivor sets across shrinking nets. Its effectiveness depends on the achievable shot precision: very small target radii r_T may demand large numbers of samples, even when the geometric conditions are favorable. One possible application is to furnish reliable warm-start regions for local optimizers, reducing the chance of initialization in suboptimal basins.

Limitations and future directions. The method is efficient only when the normal Hessian rank satisfies $r = O(\log p)$, which corresponds to architectures with restricted expressivity. The analysis assumes an a priori upper bound on r, though a simple logarithmic search schedule with a budgeted hierarchy of flats could eliminate this requirement. We do not optimize constants; our bounds are asymptotic and may be loose for moderate p. Several natural extensions remain open, including adaptive estimation of local Lipschitz constants and incorporating coarse curvature information without sacrificing the PAC guarantees. Our results can be extended to the setting of approximate Morse–Bott minima, which is left to future work.

Broader outlook. The combination of Lipschitz elimination and randomized slicing applies whenever low-energy regions concentrate on low-complexity sets inside a space of finite volume. Beyond quantum applications, this suggests a portable template for high-dimensional global optimization in regimes where curvature or intrinsic dimension is effectively compressed. We expect that incorporating additional geometric information while preserving the same probabilistic backbone will further expand the scope of problems for which reliable global guarantees are attainable.

Acknowledgments

I am grateful to Christian Arenz for mentorship and many discussions that helped to inspire this project. I also thank Sabina Dragoi for helpful discussions. Early conceptualization of this work was supported by NSF Grant No. 1953745 (REU). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect those of the NSF. Automated tools were used in drafting and editing; all conceptual and technical content is the author's own.

A Lipschitz and Hessian Bounds

We now prove the Lipschitz and Hessian bounds on our cost function described in Section 2. These are standard proofs, but we include them for completeness. Tighter bounds exist.

Proof of Proposition 2.3. Differentiating $U(\theta)$ with respect to θ_j gives us

$$\partial_{\theta_j} U(\theta) = \left(\prod_{k<j} e^{-i\theta_k A_k} \right) \left(-iA_j e^{-i\theta_j A_j} \right) \left(\prod_{k>j} e^{-i\theta_k A_k} \right) = -iU\tilde{A}_j \qquad \widetilde{A}_j(\theta) := U_{>j}^{\dagger} A_j U_{>j},$$

By differentiating through the product ansatz,

$$\partial_{\theta_j} C(\theta) = i \, \langle 0 | [U^{\dagger}(\theta) H U(\theta), \widetilde{A}_j(\theta)] | 0 \rangle,$$

so that $\|\widetilde{A}_j(\theta)\| = \|A_j\|$. Using the unitary invariance of the operator norm and the basic commutator inequality $\|[X, Y]\| \leq 2\|X\| \, \|Y\|$ then gives

$$|\partial_{\theta_j} C(\theta)| \ \leq \ \|[U^{\dagger}(\theta) H U(\theta), \widetilde{A}_j(\theta)]\| \ \leq \ 2\|H\| \, \|A_j\| \ \leq \ 2\Lambda \, \|A_j\|.$$

Hence $\|\nabla C(\theta)\|_2^2 \leq 4\Lambda^2 \sum_j \|A_j\|^2$.

$$\implies |C(\theta) - C(\theta')| \ \leq \ \sup_{\xi \in [\theta, \theta']} \|\nabla C(\xi)\|_2 \, \mathrm{dist}_{\mathbb{T}^p}(\theta, \theta'),$$

establishing global Lipschitz continuity with the claimed constant. □

B Linear Effective Ansatz

In this section, we provide derivation for the linear effective ansatz mentioned earlier. Recall that we have the product ansatz and cost function,

$$U(\theta) = \prod_{j=1}^{p} e^{-i\theta_j A_j}, \qquad C(\theta) = \langle 0 | U(\theta)^{\dagger} H U(\theta) | 0 \rangle.$$

We've also previously defined the following operators

$$U_{>j} := \prod_{k=j+1}^{p} e^{-i\theta_k A_k}, \qquad \tilde{A}_j := U_{>j}^{\dagger} A_j U_{>j}.$$

B.1 Directional Derivative of the Unitary

Differentiating U with respect to θ_j gives

$$\partial_{\theta_j} U(\theta) = \left(\prod_{k<j} e^{-i\theta_k A_k} \right) \left(-iA_j e^{-i\theta_j A_j} \right) \left(\prod_{k>j} e^{-i\theta_k A_k} \right) = -iU(\theta)\tilde{A}_j(\theta).$$

Along a direction $v \in \mathbb{R}^p$,

$$\partial_v U(\theta) := \sum_{j=1}^{p} v_j \partial_{\theta_j} U(\theta) = -iU(\theta) \sum_{j=1}^{p} v_j \tilde{A}_j(\theta).$$

We then define

$$K_\theta(v) := \sum_{j=1}^{p} v_j \tilde{A}_j(\theta).$$

B.2 First and Second-Order Expansions of the Cost

For small t, our first-order expansion of $U(\theta)$ is

$$U(\theta + tv) = U(\theta)\,(I - itK_\theta(v)) + O(t^2).$$

Therefore,

$$\begin{aligned} C(\theta + tv) &= \langle 0|U^\dagger(\theta + tv)HU(\theta + tv)|0\rangle \\ &= \langle 0|U^\dagger HU|0\rangle + it\langle 0|U^\dagger[H, K_\theta(v)]U|0\rangle + O(t^2), \end{aligned}$$

giving us,

$$\partial_v C(\theta) = i\langle [H, K_\theta(v)]\rangle.$$

Performing the same expansion one more time gives

$$\partial_v^2 C(\theta) = -\langle [K_\theta(v), [H, K_\theta(v)]]\rangle + \langle i[H, \dot{K}]\rangle.$$

References

[1] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. "Variational quantum algorithms". Nature Reviews Physics **3**, 625–644 (2021).

[2] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. "Noisy intermediate-scale quantum algorithms". Reviews of Modern Physics**94** (2022).

[3] John Preskill. "Quantum computing in the nisq era and beyond". Quantum **2**, 79 (2018).

[4] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. "A quantum approximate optimization algorithm" (2014). arXiv:1411.4028.

[5] 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** (2014).

[6] Jules Tilly, Hongxiang Chen, Shuxiang Cao, Dario Picozzi, Kanav Setia, Ying Li, Edward Grant, Leonard Wossnig, Ivan Rungger, George H. Booth, and Jonathan Tennyson. "The variational quantum eigensolver: A review of methods and best practices". Physics Reports **986**, 1–128 (2022).

[7] 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**, 242–246 (2017).

[8] Cornelius Hempel, Christine Maier, Jonathan Romero, Jarrod McClean, Thomas Monz, Heng Shen, Petar Jurcevic, Ben P. Lanyon, Peter Love, Ryan Babbush, Alán Aspuru-Guzik, Rainer Blatt, and Christian F. Roos. "Quantum chemistry calculations on a trapped-ion quantum simulator". Physical Review X**8** (2018).

[9] Eric R. Anschuetz. "Critical points in quantum generative models" (2023). arXiv:2109.06957.

[10] Eric R. Anschuetz and Bobak T. Kiani. "Quantum variational algorithms are swamped with traps". Nature Communications**13** (2022).

[11] Marco Wiedmann, Daniel Burgarth, Gunther Dirr, Thomas Schulte-Herbrüggen, Emanuel Malvetti, and Christian Arenz. "On the convergence of the variational quantum eigensolver and quantum optimal control" (2025). arXiv:2509.05295.

[12] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. "Barren plateaus in quantum neural network training landscapes". Nature Communications**9** (2018).

[13] M. Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles. "Cost function dependent barren plateaus in shallow parametrized quantum circuits". Nature Communications**12** (2021).

[14] Tuyen Nguyen and Mária Kieferová. "Theoretical guarantees of variational quantum algorithm with guiding states" (2025). arXiv:2510.06764.

[15] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando de Freitas. "Taking the human out of the loop: A review of bayesian optimization". Proc. IEEE **104**, 148–175 (2016).

[16] Nikolaus Hansen. "The cma evolution strategy: A tutorial" (2023). arXiv:1604.00772.

[17] Cédric Malherbe and Nicolas Vayatis. "Global optimization of lipschitz functions" (2017). arXiv:1703.02628.

[18] Bryan T. Gard, Linghua Zhu, George S. Barron, Nicholas J. Mayhall, Sophia E. Economou, and Edwin Barnes. "Efficient symmetry-preserving state preparation circuits for the variational quantum eigensolver algorithm". npj Quantum Information**6** (2020).

[19] Hamzat A. Akande, Alexandre Perrin, Bruno Senjean, and Matthieu Saubanère. "Variational quantum subspace construction via symmetry-preserving cost functions". Phys. Rev. A **112**, 032623 (2025).

[20] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. "The power of quantum neural networks". Nature Computational Science **1**, 403–409 (2021).

[21] Herbert Federer. "Curvature measures". Transactions of the American Mathematical Society **93**, 418–491 (1959).

[22] Jan Rataj and Martina Zähle. "Curvature measures of singular sets". Springer Monographs in Mathematics. Springer Nature Switzerland AG. (2019).

[23] Martin Lotz, Michael B. McCoy, Ivan Nourdin, Giovanni Peccati, and Joel A. Tropp. "Concentration of the intrinsic volumes of a convex body" (2019). arXiv:1810.12412.

[24] Lorenzo Leone, Salvatore F.E. Oliviero, Lukasz Cincio, and M. Cerezo. "On the practical usefulness of the Hardware Efficient Ansatz". Quantum **8**, 1395 (2024).

[25] G. Vidal. "Class of quantum many-body states that can be efficiently simulated". Physical Review Letters**101** (2008).

[26] Y.-Y. Shi, L.-M. Duan, and G. Vidal. "Classical simulation of quantum many-body systems with a tree tensor network". Physical Review A**74** (2006).

[27] Philipp Seitz, Ismael Medina, Esther Cruz, Qunsheng Huang, and Christian B. Mendl. "Simulating quantum circuits using tree tensor networks". Quantum **7**, 964 (2023).

[28] Shiri Artstein-Avidan, Apostolos Giannopoulos, and Vitali D Milman. "Asymptotic geometric analysis, part i". Mathematical surveys and monographs. American Mathematical Society. (2015).

[29] Sevag Gharibian and Carsten Hecht. "Hardness of approximation for ground state problems" (2024). arXiv:2411.04874.

[30] Julia Kempe, Alexei Kitaev, and Oded Regev. "The complexity of the local hamiltonian problem" (2005). arXiv:quant-ph/0406180.