Local and Global Minima

Variational Quantum Algorithms · Module 3 — Cost Landscapes

Node 3.3

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Cost Landscapes **Node:** 3.3

This node zooms in on the most important topological feature of a VQA landscape — its **minima** — and dissects what they mean for optimization.

The mantra "we want to find the global minimum" is correct in the abstract but operationally misleading in VQA practice. For most realistic VQA instances, the global minimum is **unreachable** by local optimization, and the practical question is **which local minimum can the optimizer find, and how good is it?**

Definitions

A **critical point** $\boldsymbol{\theta}^*$ of $C : \mathbb{T}^P \to \mathbb{R}$ satisfies $\nabla C(\boldsymbol{\theta}^*) = 0$.

A critical point is a **local minimum** if there exists a neighborhood U of $\boldsymbol{\theta}^*$ such that $C(\boldsymbol{\theta}^*) \leq C(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in U$.

A critical point is a **strict local minimum** if the inequality is strict for $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$.

A **global minimum** is the lowest local minimum on the entire domain.

The **basin of attraction** of a local minimum $\boldsymbol{\theta}^*$ is the set of starting points $\boldsymbol{\theta}_0$ from which gradient descent (with appropriately small step size) converges to $\boldsymbol{\theta}^*$.

How Many Minima?

A typical VQA cost function on P parameters has **many** local minima — usually more than P, often growing polynomially or even exponentially in P.

This is true for two related reasons:

1. Symmetry. Ansatz symmetries create equivalence classes of minima. If the ansatz has a permutation symmetry over k parameters, there are at least $k!$ symmetry-equivalent minima for any non-symmetric configuration. Permutation-symmetric HEAs at moderate qubit count can have hundreds of equivalent minima.

2. Compositional cancellation. Layers in a deep ansatz can compose to (approximately) the identity in many distinct ways, creating local minima at parameter configurations whose effective state is similar. $R_X(\theta_1)R_X(\theta_2)$ has a 1-parameter family of (θ_1, θ_2) configurations that all give the same single-qubit unitary, and any cost function on this combined parameterization inherits a 1-parameter degeneracy.

The number of minima grows roughly with the number of "ways" the ansatz can prepare each state, which for expressive ansätze is large.

How Good Are The Local Minima?

A more important question than "how many?" is **how good are they?**

For UCCSD-style chemically-motivated ansätze, the result is often striking: **almost all local minima are within chemical accuracy of the global minimum.** That is, any optimizer that finds *any* local minimum is essentially "done." This is one of the key practical advantages of problem-inspired ansätze.

For HEA-style hardware-motivated ansätze, the result is much worse: local minima can range from "barely better than random" to "near the global minimum," with the typical value far above the global minimum and the gap growing with depth. HEAs do have many local minima, but most of them are mediocre.

This is the **quality of local optima** problem, and it is one of the main reasons HEAs are usually outperformed by problem-inspired ansätze on chemistry instances even when both ansätze are equally trainable. The HEA *can* find a low-cost point, but the typical low-cost point it lands in is much higher than the typical low-cost point of UCCSD.

Why You (Usually) Can't Find The Global Minimum

For high-dimensional non-convex problems, finding the global minimum is generally NP-hard. For VQA problems specifically, several additional obstacles apply:

1. Exponentially many local minima. Even if you could count them, you can't afford to visit them all.

2. Saddles between basins. Local optimizers cannot cross saddles upward. To get from one basin to another, you need either restart-based exploration or a global method.

3. Noisy evaluations. You don't actually know whether a point is a local minimum or just a region with small gradient. Distinguishing requires more shots.

4. Limited query budget. Even global methods need many queries to explore a high-dimensional space, and each query is expensive.

In practice, VQA practitioners do **multi-start optimization**: run gradient descent (or another local method) from many random initializations, take the best result. This is not guaranteed to find the global minimum but typically finds a good one. Algorithms like HOPSO and CMA-ES are essentially principled multi-start strategies that share information between restarts.

The "Reachability vs Findability" Distinction

A critical conceptual move: **a state being reachable on the manifold is not the same as a state being findable by the optimizer.**

For an ansatz that contains the true ground state in its image: - **Reachable**: there exists $\boldsymbol{\theta}^*$ such that $|\psi(\boldsymbol{\theta}^*)\rangle$ equals the ground state. - **Findable**: starting from a typical initialization, the optimizer converges to $\boldsymbol{\theta}^*$ (or to a parameter giving a state with energy near the ground state) within the available query budget.

The first is a property of the ansatz manifold (Module 2.9). The second is a property of the *combination* of the ansatz, the cost function, the optimizer, and the initialization.

A barren plateau is the extreme case of "reachable but not findable" — the manifold contains the answer, but the cost gradient is so small everywhere that no local optimizer can detect a descent direction.

This is also exactly the conceptual move that node 1.9 (separation of objective and dynamics) demands: reachability is an objective property, findability is a composition property.

What An Optimizer Actually Settles On

In practice, when a VQA optimizer "converges," it has found one of the following:

1. **A local minimum** of the noisy cost function (which may or may not be a local minimum of the noiseless cost).
2. **A noise-equilibrium point** where the gradient signal is below the noise floor and the optimizer's updates cancel out on average.
3. **A point in a long flat region** where the gradient is small but not zero, and the optimizer is moving slowly.

Distinguishing these is hard. A common heuristic is to look at the variance of the cost over the last K iterations: if it is dominated by the optimizer's step size, you are at a noise equilibrium; if it is dominated by shot noise, you are at a stable point.

Practical Implications

Three operational rules of thumb that follow from the above:

1. Always do multi-start. Single-start optimization wastes the structure of the landscape — you can pay the cost of multiple optimizations in parallel and at least know which basin you've found.

2. Don't trust convergence as quality. A "converged" run might be in a mediocre local minimum. Always compare the converged value against a known reference (Hartree-Fock, MP2, classical CCSD).

3. Pay attention to optimizer-specific basin preferences. Different optimizers find different local minima from the same starting point, and their preferences can be exploited (or accidentally encode bias).

These are not theorems. They are tactical observations from practice. Most published VQA results follow them implicitly. Most failed VQA experiments did not.

Structural Role

This node sharpens the topology vocabulary of node 3.2 by zooming in on minima specifically. It is also the conceptual setup for Module 4 (barren plateaus): the failure mode introduced there is "all local minima become unreachable in expectation," which is a more extreme form of the problem characterized here.

Relations to Other Concepts

- **Loss landscape geometry of neural networks** — many of the same phenomena (many minima, mostly equivalent for overparameterized networks; basin connectivity).
- **Random energy models** — statistical physics models for random landscapes that predict the distribution of local minima energies.
- **Combinatorial optimization** — global vs local minima as a problem class.
- **Multi-start optimization** — the standard practical technique for landscape exploration.

Worked Example — Counting Minima In An HEA

Consider a 2-qubit HEA with 1 layer: $|\psi(\boldsymbol{\theta})\rangle = \text{CNOT} \cdot (R_Y(\theta_1) \otimes R_Y(\theta_2)) \cdot \text{CNOT} \cdot (R_Y(\theta_3) \otimes R_Y(\theta_4))|00\rangle$.

For a Hamiltonian $H = Z_0 + Z_1$, the cost function $C(\boldsymbol{\theta})$ is a 4D function. Numerical exploration shows:

- **At least 16 local minima** in the fundamental cell $[-\pi, \pi]^4$, arising from sign-flip symmetries: $(\theta_i \to -\theta_i + 2\pi)$ gives an equivalent state up to a global phase, and 4 parameters $\times$ 2 sign choices $= 2^4 = 16$.
- **The global minimum** is at $(0, 0, \pi, \pi)$ or its sign-flip equivalents, where both qubits are flipped to $|11\rangle$ and $\langle Z_0 + Z_1\rangle = -2$.
- **All 16 minima have the same energy** (because they all map to the same state up to a global phase).
- **Saddles between them** correspond to half-flipped configurations with energy 0 or near-0.

So this 4-parameter ansatz has 16 equivalent global minima and a complicated structure of saddles between them. A gradient-descent optimizer starting from a random init will find one of the 16 — which one depends on the basin of attraction the init falls into — but since they're all equivalent, "finding the global minimum" is trivial here.

For a more interesting Hamiltonian (e.g., $H = Z_0 + Z_1 + 0.3X_0X_1$), the symmetry is broken slightly and the 16 minima split into 16 *near-equivalent* minima at slightly different energies. Then the choice of optimizer starts to matter: a finer optimizer will find a slightly lower one, but they're all close.

This is the pattern: **HEAs generate many minima by symmetry, and which one you find usually doesn't matter much — until you turn off the symmetry, at which point the structure suddenly becomes important.**

Visualization

Pending. A 2D slice of the 4D HEA landscape, showing 4 of the 16 minima (one quadrant of the sign-flip group). Color: cost. Annotations: basins, saddles, the global value. Caption: "4 of 16 equivalent minima."

Exercises

1. For the cost function $C(\theta_1, \theta_2) = -\cos\theta_1 - \cos\theta_2$ on $[-\pi, \pi]^2$, find all local minima. How many are there? How many are global?
2. Show that the cost function in Exercise 1 has the same number of local maxima as local minima, by Morse-theoretic reasoning.
3. (VQA) Describe a multi-start optimization protocol for a 20-parameter VQA. How many restarts would you do? How would you decide when to stop?

Research Extensions

- **Counting critical points of variational ansätze** — analytic upper and lower bounds.
- **Energy distribution of local minima** — what is the typical gap between the lowest local minimum found and the global minimum?
- **Basin volume estimation** — for a given local minimum, how big is its basin? Determines how easy it is to find by random initialization.
- **Saddle-aware optimizers** — methods that explicitly seek out and cross saddles.

Cross-links to Other Courses

- **Optimization Theory** — local vs global minima, convergence guarantees.
- **Statistical Physics** — random landscape models, Kac-Rice formula for critical point counting.
- **Module 4 (Barren Plateaus)** — the "all minima unreachable" failure mode.
- **Module 6 (Optimization Dynamics)** — how different optimizers handle multi-minima landscapes.

Variational Quantum Algorithms · Module 3 — Cost Landscapes · Node 3.3
Concepts: optimization-landscape, gradient-descent, cost-function, convergence
Prerequisites: 3.1, 3.2
Enables: 3.5

hbar.university — own.your.cognition