The Ansatz as a Manifold

Variational Quantum Algorithms · Module 2 — Parameterized Circuits

Node 2.9

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Parameterized Circuits **Node:** 2.9

This is the **closing node** of Module 2 and its most geometric. Everything in the module so far has been described in coordinates: parameter vectors, Pauli strings, gate sequences, bitstrings, shot counts. This node steps back from coordinates and looks at the underlying geometric object: **the set of states the ansatz can prepare is a smooth submanifold of complex projective space, and many of the hardest optimization phenomena in VQA are properties of that manifold rather than properties of any particular parameterization.**

This is the geometric viewpoint that makes Module 3 (cost landscapes), Module 4 (barren plateaus), and Module 6 (optimization dynamics) feel coherent rather than ad-hoc. If you internalize this node, the rest of the course is a series of statements about a single geometric object that you can mentally visualize.

The Manifold

For an ansatz $U(\boldsymbol{\theta}) : \mathbb{R}^P \to U(2^n)$ acting on $|0\rangle^{\otimes n}$, the **ansatz manifold** is the image:

$$\mathcal{M} = \{|\psi(\boldsymbol{\theta})\rangle : \boldsymbol{\theta} \in \mathbb{R}^P\} \subset \mathbb{CP}^{2^n-1}.$$

We work in complex projective space $\mathbb{CP}^{2^n-1}$ rather than the unit sphere of $\mathbb{C}^{2^n}$ because global phases don't matter physically — $|\psi\rangle$ and $e^{i\alpha}|\psi\rangle$ are the same state.

Three structural facts about this set:

1. Dimension. The manifold has intrinsic dimension at most P, the parameter count. It can have lower dimension if some parameters are redundant (the parameterization is *singular* somewhere) or if the ansatz contains an exact symmetry that closes off a direction.

2. Smoothness. For circuit ansätze built from analytic gate families (rotations, controlled rotations), the parameterization $\boldsymbol{\theta} \to |\psi(\boldsymbol{\theta})\rangle$ is real-analytic almost everywhere. The manifold is smooth except at isolated singular points.

3. Restriction. $\mathcal{M}$ is a strict, generally low-dimensional subset of $\mathbb{CP}^{2^n-1}$, which itself has dimension $2(2^n - 1)$ — exponential in qubit count. For $n = 10$ qubits, the ambient projective space has real dimension 2046, while a typical 30-parameter ansatz manifold has dimension 30 (or less). **The ansatz traces out a thin 30-dimensional sheet inside a 2046-dimensional space.** Optimization is the act of moving along this sheet looking for a low-energy point.

Tangent Vectors and the Quantum Fisher Information

At a point $|\psi(\boldsymbol{\theta})\rangle$ on the manifold, the tangent space is spanned by the directional derivatives

$$|\partial_i\psi\rangle := \frac{\partial}{\partial\theta_i}|\psi(\boldsymbol{\theta})\rangle.$$

These vectors live in the ambient $\mathbb{C}^{2^n}$ but project (after removing the component parallel to $|\psi\rangle$) into the tangent space of $\mathcal{M}$ at $|\psi(\boldsymbol{\theta})\rangle$.

The natural metric on the manifold is the **Fubini-Study metric**, which on the tangent space takes the form

$$g_{ij}(\boldsymbol{\theta}) = \mathrm{Re}\langle\partial_i\psi|\partial_j\psi\rangle - \langle\partial_i\psi|\psi\rangle\langle\psi|\partial_j\psi\rangle.$$

In quantum information this is known as the **quantum Fisher information matrix** (up to a factor of 4) and it measures how "distinguishable" two infinitesimally separated parameter values are in terms of physical quantum measurements.

The crucial point: this metric is **not** the Euclidean metric δ_{ij} on parameter space. Different parameter directions span tangent vectors of very different lengths in the ambient state space. A "step of size ϵ in parameter space" can correspond to a tiny physical change in the state (if the parameter is in a redundant or weakly-coupled direction) or a huge physical change (if the parameter directly steers the state).

This mismatch is the reason **vanilla gradient descent in parameter space is not optimal** for VQAs. The natural object to descend is the cost gradient pulled back through the inverse metric:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \eta\, g^{-1}(\boldsymbol{\theta}_k)\,\nabla C(\boldsymbol{\theta}_k).$$

This is the **quantum natural gradient** (Stokes et al. 2020) and it is treated in detail in Module 6. For now, the important takeaway is: **the natural geometry of the ansatz lives on the manifold, not in parameter coordinates.**

Singularities

A point $\boldsymbol{\theta}_0$ is **singular** if the tangent vectors $\{|\partial_i\psi(\boldsymbol{\theta}_0)\rangle\}_{i=1}^P$ do not span a P-dimensional subspace — that is, if some directional derivative collapses to zero, or two directions become parallel.

At a singular point: - The local dimension of the manifold drops below P. - The metric g_{ij} becomes singular (determinant zero). - The natural gradient is undefined (the inverse metric blows up). - Multiple distinct $\boldsymbol{\theta}$-paths can lead to the same $|\psi\rangle$, creating a degeneracy.

These singularities are not abstract pathologies — they show up in real ansätze. Typical examples: - **Identity points**: a layer with all rotation angles set to 0 contributes the identity matrix and removes itself from the parameterization. - **Symmetry-equivalent parameters**: two parameters whose effects on the state are related by an exact group action. - **Compositional cancellations**: $R_X(\theta)R_X(-\theta) = I$ creates a 1-parameter family of $(\theta, -\theta)$ configurations all mapping to the same state.

The barren plateau phenomenon (Module 4) is *not* the same as a singularity — it is a global statistical phenomenon — but it can be exacerbated by the manifold passing through near-singular configurations.

Why This Matters For Optimization

Three operationally important consequences of the manifold view:

1. Step size in parameter space $\neq$ step size in state space. A learning rate η that is appropriate in one region of parameter space may be too aggressive or too conservative in another, depending on local manifold geometry. This is why adaptive learning rates (Adam, AdaGrad) often help VQAs — they implicitly normalize for changing local geometry.

2. Curvature drives convergence rate. Just as in classical optimization, the local curvature of the manifold (and the cost surface on it) determines how quickly an optimizer can converge. Highly curved regions need small steps; flat regions allow large steps. Vanilla gradient descent ignores this; second-order methods (natural gradient, Newton-Raphson, BFGS) try to exploit it.

3. The manifold is the same regardless of parameterization. Two ansätze that produce the same set of reachable states have the same manifold — even if their parameterizations are very different. **The geometry is intrinsic; the coordinates are convention.** This is the fundamental reason that the *choice of parameterization* matters even when it doesn't change what states can be reached: it changes the relationship between parameter directions and tangent directions, and therefore changes which optimizers will be efficient.

The Picture for HEAs and UCCSDs

Two quick intuitive snapshots:

HEA manifold. Roughly Haar-uniform on the symmetry sector reachable by the ansatz. As depth grows, the manifold fills out more and more of the unit sphere in Hilbert space. At the depth where it approximates a 2-design, the manifold is "everywhere dense" in a useful sense — and that's exactly when barren plateaus appear. The manifold is too uniform; almost no direction is special.

UCCSD manifold. Heavily structured by the Hartree-Fock reference and the chosen excitations. Sits in a low-dimensional symmetry sector (fixed particle number, spin). Has natural starting point (the HF state, where all parameters are zero) and natural directions of motion (each excitation amplitude). The manifold is small but well-shaped, and the cost function has identifiable minima rather than uniform plateaus.

The contrast: **HEA gives you a featureless manifold; UCCSD gives you a featured one.** The featured manifold is harder to derive but easier to optimize.

Structural Role

This is the **conceptual capstone** of Module 2. It pulls Module 2's coordinate-level facts (gates, parameters, depths, encodings, shots) up into a coordinate-free geometric picture.

Forwards: - Module 3 (Cost Landscapes) treats the cost function $C : \mathcal{M} \to \mathbb{R}$ as a function on the manifold and studies its critical points and topology. - Module 4 (Barren Plateaus) is a statement about the typical curvature of the cost function on the manifold under random parameterization. - Module 6 (Optimization Dynamics) asks how to update parameters in a way that respects manifold geometry rather than parameter coordinates. - Module 8 (Expressivity vs Trainability) is the formal version of "featureless vs featured manifold."

This is the node where Module 2 stops being about plumbing and starts being about geometry.

Relations to Other Concepts

- **Information geometry** — the general theory of metric structures on parameterized probability/quantum models. Amari and Nagaoka are the standard references.
- **Riemannian manifolds** — the mathematical setting for the natural gradient and second-order methods.
- **Complex projective space** $\mathbb{CP}^{2^n-1}$ — the ambient space in which the ansatz manifold lives.
- **Fubini-Study metric** — the canonical metric on $\mathbb{CP}^n$, and the source of the quantum Fisher information.
- **Stiefel manifold** — alternative parameterization for orthonormal frames, occasionally used in tensor-network analogues of ansatz design.

Worked Example — Two Manifolds With The Same Reach

Consider two single-qubit ansätze:

Ansatz A: $|\psi_A(\theta, \phi)\rangle = R_Z(\phi) R_Y(\theta)|0\rangle$. Two parameters. Reaches every pure single-qubit state. The manifold is the **full Bloch sphere**.

Ansatz B: $|\psi_B(\alpha, \beta, \gamma)\rangle = R_Y(\gamma) R_Z(\beta) R_Y(\alpha)|0\rangle$. Three parameters. Also reaches every pure single-qubit state. The manifold is **also the full Bloch sphere**.

Same manifold. Different parameterizations.

What differs: - **Local metric.** At a generic point, the metric g_{ij} for ansatz A is full rank 2×2. For ansatz B it is rank-2 in a 3×3 matrix — there is one redundant direction at every point. - **Optimization behavior.** A gradient-descent run on ansatz A walks across the Bloch sphere in 2D. The same algorithm on ansatz B has an extra "null direction" (the redundant one) that the optimizer can drift along without changing the state. - **Parameter-space distance.** Two states close together on the Bloch sphere can correspond to parameter vectors that are far apart in either ansatz — but the relationship is different in the two parameterizations.

Both manifolds are the same set of physical states. **The cost landscape differences come entirely from how the parameters cover the manifold.** This is one of the cleanest demonstrations that the choice of parameterization matters even when the reachable set does not.

Visualization

Pending. A diagram showing the Bloch sphere with two coordinate grids overlaid: one from ansatz A (latitude/longitude-like, smooth), one from ansatz B (a 3D parameter space being projected, with a "redundant fiber" at each point shown as a small circle). Caption: "Same manifold, different coordinates, different gradient flows."

Exercises

1. For the 2-qubit ansatz $U(\theta_1, \theta_2) = (R_Y(\theta_1) \otimes R_Y(\theta_2))|00\rangle$, compute the tangent vectors $|\partial_1\psi\rangle$ and $|\partial_2\psi\rangle$. Are they orthogonal? What is the dimension of the resulting manifold?
2. Show that the parameterization $R_X(\theta_1) R_X(\theta_2)|0\rangle$ has a 1-dimensional kernel everywhere — i.e., the linear combination $|\partial_1\psi\rangle - |\partial_2\psi\rangle = 0$. What does this imply about the dimension of the manifold?
3. (VQA) Sketch an argument for why the natural gradient $g^{-1}\nabla C$ is invariant under reparameterizations of the ansatz, while the Euclidean gradient ∇C is not.

Research Extensions

- **Quantum information geometry of variational models** — formalizing the manifold view across ansatz families.
- **Geodesic optimization on ansatz manifolds** — using exact or approximate Riemannian optimizer steps that follow geodesics rather than straight lines in parameter space.
- **Manifold curvature and barren plateaus** — connecting global geometric properties to optimization difficulty.
- **Riemannian regularization** — adding curvature-aware penalties that act intrinsically rather than in parameter coordinates.

Cross-links to Other Courses

- **Differential Geometry** — manifolds, tangent spaces, metrics, curvature.
- **Information Geometry** — Fisher metric, dual connections, exponential families.
- **Module 3 (Cost Landscapes)** — the cost function as a scalar field on the manifold.
- **Module 6 (Optimization Dynamics)** — the natural gradient as the manifold-aware update rule.

Variational Quantum Algorithms · Module 2 — Parameterized Circuits · Node 2.9
Concepts: inner-product, hilbert-space, parameterized-circuits, dimensionality
Prerequisites: 2.1, 2.2, 2.3, 2.6

hbar.university — own.your.cognition