Universal Approximation in Quantum Circuits

Variational Quantum Algorithms · Module 8 — Expressivity Vs Trainability

Node 8.3

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Expressivity vs Trainability **Node:** 8.3

The classical neural network literature has a famous theorem: **a feedforward network with one hidden layer of sufficient width can approximate any continuous function** (Cybenko 1989, Hornik 1991). This is the *universal approximation theorem* and it is what justifies neural networks at the foundational level.

Quantum variational circuits have analogous universality results, and this node walks through them. The key questions:

1. **Is there a quantum analogue of universal approximation?** (Yes, several.)
2. **What does "universal" mean in the quantum setting?** (More than one thing.)
3. **What does universality cost in terms of circuit depth, parameter count, etc.?** (Exponentially much, in general.)
4. **How does universality relate to barren plateaus?** (Negatively — universal ansätze plateau.)

The punchline: quantum universal approximation theorems exist, but they are *existence statements* about ansätze that are exponentially deep or exponentially wide. The practically useful ansätze are not universal — they are restricted, problem-specific, and that restriction is what makes them trainable.

What "Universal" Could Mean For A Quantum Circuit

There are multiple senses in which a parameterized circuit could be "universal."

1. State universality. The ansatz can prepare any state in the Hilbert space:

$$\forall|\phi\rangle \in \mathcal{H},\ \exists\boldsymbol{\theta} : |\psi(\boldsymbol{\theta})\rangle = |\phi\rangle.$$

2. Unitary universality. The ansatz can implement any unitary on the Hilbert space:

$$\forall U \in U(2^n),\ \exists\boldsymbol{\theta} : V(\boldsymbol{\theta}) = U.$$

State universality is weaker (you only need to reach states starting from $|0\rangle^{\otimes n}$). Unitary universality is stronger (you need to be able to act on arbitrary inputs).

3. Function approximation universality. The map $\boldsymbol{\theta} \to \langle\psi(\boldsymbol{\theta})|H|\psi(\boldsymbol{\theta})\rangle$ (or similar observable expectations) can approximate any continuous function on the parameter space. This is the closest analogue of classical universal approximation.

4. Distribution universality. The ansatz can sample from any distribution over computational basis states. Relevant for quantum generative models.

5. Operator universality. The ansatz can prepare any Hermitian observable as a linear combination of its parameters. Relevant for quantum machine learning kernels.

These are *not* equivalent. A circuit can be universal in one sense and not in another. The literature usually means **state universality** unless otherwise specified.

State Universality: Existence

Theorem (state universality). A parameterized circuit on n qubits with at least $2 \cdot 2^n - 2$ real parameters can be state-universal — that is, there exists a circuit construction with this many parameters that can reach every state in the projective Hilbert space.

Proof idea. The projective Hilbert space $\mathbb{CP}^{2^n-1}$ has real dimension $2 \cdot 2^n - 2$. A smooth, surjective map from a parameter space to this manifold must have parameter space dimension at least $2 \cdot 2^n - 2$. With this many parameters, an explicit construction (e.g., the **Schmidt decomposition based circuit** of Iten et al. 2016) achieves universality.

Cost: $2 \cdot 2^n - 2$ parameters means **exponential growth in qubit count**. For $n = 10$, 2046 parameters; for $n = 20$, 2 million; for $n = 50$, 10^{15}.

State universality is theoretically important but **practically infeasible** for any $n > 10$ or so. Universal ansätze are not used in practice; they are reference points to anchor the theory.

Unitary Universality: Existence

Theorem (unitary universality). A parameterized circuit on n qubits can implement any unitary in $U(2^n)$ if and only if its parameter count is at least $(2^n)^2 - 1 = 4^n - 1$.

Cost: $4^n - 1$ parameters — even worse than state universality. Doubly exponential in qubit count.

Unitary universality is **even more theoretical** than state universality. The only reason it's worth knowing about is that it's the natural notion when the ansatz needs to act on arbitrary inputs (e.g., quantum kernels, quantum data processing).

Practical Universality: What "Enough" Looks Like

The interesting question is not "what is the minimum parameter count for universality" but **"how many parameters give an ansatz expressive enough to solve the problem at hand?"**

For chemistry (VQE on small molecules): - Hartree-Fock state plus a UCCSD ansatz with $O(N^4)$ parameters (for N electrons) is enough to reach chemical accuracy on most small molecules. This is *much* less than full universality.

For optimization (QAOA): - A QAOA ansatz with $O(p)$ parameters at depth p is enough to solve many MaxCut and constraint-satisfaction problems with provable approximation guarantees. The required p is small (often $p = 1$ or 2) for many problems, vastly less than universality.

For machine learning (data-encoded models): - A re-uploading ansatz with $O(L)$ data-encoding layers can express functions of growing complexity. Schuld-Sweke-Meyer (2021) showed that L layers gives Fourier series of degree up to L.

The pattern: **practical ansätze trade universality for trainability**. The right amount of expressivity is just enough to contain the answer, and not more.

The Schuld-Sweke-Meyer Theorem

A particularly clean universality result for quantum machine learning models.

Setup. A quantum model is a function $f(x;\boldsymbol{\theta}) = \langle 0|U^\dagger(x,\boldsymbol{\theta})MU(x,\boldsymbol{\theta})|0\rangle$, where x is a data input encoded into the circuit, $\boldsymbol{\theta}$ are trainable parameters, and M is an observable.

Theorem (Schuld-Sweke-Meyer 2021). If the data x is encoded via Pauli rotations e^{-ixH}, and the encoding is repeated L times with trainable circuits in between, then $f(x)$ is a **truncated Fourier series** in x of degree $\le L$. The Fourier coefficients are determined by the trainable circuits; in particular, all Fourier modes up to degree L can be reached by suitable choice of $\boldsymbol{\theta}$.

Implication: with enough re-uploading layers, the model can approximate any continuous function of x on a compact domain (Stone-Weierstrass for trigonometric polynomials). This is a quantum universal approximation theorem in the function-approximation sense.

The catch: the Fourier coefficients are constrained by the structure of the circuit. The model is universal in the sense that *some* coefficients can be chosen freely, but not all combinations are achievable. There are dependencies that the proof doesn't explicitly characterize.

This result is the cleanest connection between quantum circuits and classical function approximation theory, and it's the theoretical foundation for **data re-uploading** as a quantum machine learning architecture.

Universality And Barren Plateaus

The deep tension: **the most universal ansätze are exactly the ansätze most prone to barren plateaus**.

Why: universality (or near-universality) means the ansatz reaches a Haar-like distribution over states. Module 4 showed that Haar-2-design ansätze have exponentially small gradient variance — barren plateau.

Quantitative statement: for an ansatz that is an ϵ-approximate state-2-design,

$$\mathrm{Var}[\partial_j C] \le \frac{1}{2^n - 1}(\|H\|_2^2 + \epsilon \cdot \mathrm{poly}(n)).$$

As $\epsilon \to 0$ (true 2-design), the bound goes to $1/(2^n - 1)$ — exponentially small.

Conclusion: a fully universal ansatz cannot be trained by gradient descent on a generic Hamiltonian. This is a *theorem*, not a heuristic. The only escape is to use a non-universal (structured, restricted) ansatz.

This is the central theoretical message of Module 8: **universality and trainability are in fundamental opposition**, and any practical VQA strategy must explicitly choose where to land on the spectrum.

The Practical Take

Universal approximation theorems for quantum circuits exist, but they:

1. **Require exponentially many parameters** (state univ: $2 \cdot 2^n - 2$, unitary univ: $4^n - 1$).
2. **Imply barren plateaus** (universal = Haar-like = no gradients).
3. **Are not how successful VQAs actually work** (chemistry uses problem-inspired, optimization uses problem-restricted).

The honest summary: universal approximation is a theoretical curiosity for quantum circuits, not a practical guide. **The practically useful theorems are about restricted ansätze**: how much function approximation can a *specific structured* ansatz do, given its symmetries and depth?

This is the territory of nodes 8.4 (trainability bounds for restricted ansätze) and 8.5 (the explicit tradeoff curve).

Universal Approximation In Classical ML, For Comparison

The classical analogue is much friendlier:

- A neural net with one hidden layer of width $O(\text{poly})$ is a universal approximator (Cybenko, Hornik). The width grows polynomially in the function complexity, not exponentially.
- Backpropagation works regardless of width — no analogue of barren plateaus that scales with width.
- In practice, classical NNs are *vastly* over-parameterized, and that over-parameterization is *helpful*, not harmful (Belkin-Hsu-Ma-Mandal "double descent").

In quantum circuits, **the relationship is inverted**: over-parameterization is harmful, and the right amount of structure is essential. This is one of the fundamental qualitative differences between quantum and classical machine learning, and it shapes everything about how VQAs are designed.

Structural Role

This node is the **theoretical reference point** of Module 8. It establishes what universal approximation means in the quantum setting, explains why universal ansätze are not what we want, and motivates the move to restricted ansätze that is the focus of 8.4 and 8.5.

It also provides the connection to classical machine learning theory (and the cautionary tale about how the classical results do *not* translate directly).

Relations to Other Concepts

- **Cybenko 1989, Hornik 1991** — classical universal approximation.
- **Iten et al. 2016** — explicit universal state preparation circuits.
- **Schuld-Sweke-Meyer 2021** — Fourier series structure of data re-uploading.
- **McClean et al. 2018** — barren plateau theorem (the cost of universality).
- **Module 4 (Barren plateaus)** — the trainability cost.

Worked Example — Counting Parameters For A 4-Qubit Universal Ansatz

For 4 qubits, the projective Hilbert space has real dimension $2 \cdot 2^4 - 2 = 30$.

State-universal ansatz: at least 30 parameters. **Unitary-universal ansatz:** at least $4^4 - 1 = 255$ parameters.

A Hardware-Efficient Ansatz with single-qubit R_X, R_Y, R_Z on each qubit (12 parameters per layer) plus 4 nearest-neighbor CNOTs would need at least $\lceil 30/12 \rceil = 3$ layers for state universality (if the CNOTs effectively connect the parameters). In practice, the SJK expressibility approaches Haar at around 3-5 layers, consistent with this dimension counting.

For unitary universality, you'd need much more: at least 21 layers. In practice, even 21 layers doesn't quite achieve unitary universality for HEA because the circuit's connectivity restricts the reachable unitaries.

The dimension counting is a *lower bound* — actual universality requires the right topology of parameter dependence, which not every counting-correct circuit provides.

Visualization

Pending. A chart showing parameter count vs qubit count for state universality, unitary universality, and "practical" UCCSD/HEA. State universality grows as 2^{n+1}, unitary as 4^n, UCCSD as $O(n^4)$, HEA polynomial in n per layer. Caption: "Universality is exponential. Practical ansätze are polynomial."

Exercises

1. For a 1-qubit ansatz $R_Y(\theta_1)R_X(\theta_2)R_Y(\theta_3)$, is it state-universal? Unitary-universal? Justify.
2. The Schuld-Sweke-Meyer theorem says L data-encoding layers give Fourier degree L. For a function $f(x) = \cos(3x) + 0.5\cos(5x)$, what is the minimum L you need to express it exactly?
3. (VQA) For the LiH molecule (10 qubits in a minimal basis), estimate the parameter count of (a) a state-universal ansatz, (b) a UCCSD ansatz, (c) a Hardware-Efficient Ansatz with 5 layers. Comment on the gap.

Research Extensions

- **Quantitative universality vs trainability tradeoffs** — explicit curves for specific ansatz families.
- **Constrained universality** — universality on a symmetry-restricted subspace (e.g., particle-number sectors).
- **Practical universality with depth annealing** — starting with restricted ansätze and growing toward universality.
- **Universal approximation under noise** — how hardware errors affect the universality results.

Cross-links to Other Courses

- **Approximation Theory** — universal approximation theorems, density results.
- **Functional Analysis** — Stone-Weierstrass theorem, denseness in continuous function spaces.
- **Module 4 (Barren Plateaus)** — the trainability cost of universality.
- **Module 2 (Ansatz Design)** — the practical alternative to universality.

Variational Quantum Algorithms · Module 8 — Expressivity Vs Trainability · Node 8.3
Concepts: hilbert-space, parameterized-circuits, computational-complexity
Prerequisites: 8.1, 8.2
Enables: 8.5

hbar.university — own.your.cognition