Pontryagin Minimum Principle

Variational Quantum Algorithms · Module 7 — Control Theoretic View

Node 7.5

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Control-Theoretic View **Node:** 7.5

The **Pontryagin minimum principle** (PMP) is the central theorem of classical optimal control theory. It gives **necessary conditions for a control trajectory to be optimal** — a generalization of the calculus-of-variations Euler-Lagrange equations to problems with control inputs.

For VQAs, the PMP has both theoretical and practical importance:

1. **Theoretically**, it tells you what an optimum *must* look like — every optimal VQA parameter setting satisfies a specific structural condition derived from the PMP.
2. **Practically**, it gives a *different* way to derive gradient conditions and update rules. The PMP-derived gradient is equivalent to the parameter-shift gradient, but the derivation is cleaner and connects to the broader optimal control literature.
3. **Conceptually**, the PMP makes the "two-point boundary value" structure of VQA optimization explicit, which clarifies why some VQAs are easy to optimize and others are not.

This node walks through the PMP, applies it to VQAs, and discusses the bang-bang structure it predicts for QAOA-like ansätze.

The Statement Of The Pontryagin Minimum Principle

Setup. Consider the optimal control problem

$$\min_{\mathbf{u}} J = \phi(\mathbf{x}(T)) + \int_0^T L(\mathbf{x}, \mathbf{u}, t)\, dt, \quad \dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, t), \quad \mathbf{x}(0) = \mathbf{x}_0.$$

Define the control Hamiltonian:

$$H(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}, t) = L(\mathbf{x}, \mathbf{u}, t) + \boldsymbol{\lambda}^T f(\mathbf{x}, \mathbf{u}, t).$$

Here $\boldsymbol{\lambda}(t) \in \mathbb{R}^n$ is the **costate** (or "adjoint") variable — a Lagrange multiplier for the dynamics constraint.

Theorem (Pontryagin et al. 1962). If $\mathbf{u}^*(t)$ is optimal and $\mathbf{x}^*(t)$ is the corresponding state trajectory, there exists a costate $\boldsymbol{\lambda}^*(t)$ such that:

1. **State equation:** $\dot{\mathbf{x}}^* = \partial H / \partial \boldsymbol{\lambda}$.
2. **Costate equation:** $\dot{\boldsymbol{\lambda}}^* = -\partial H / \partial \mathbf{x}$.
3. **Minimum condition:** $\mathbf{u}^*(t) = \arg\min_{\mathbf{u}} H(\mathbf{x}^*(t), \mathbf{u}, \boldsymbol{\lambda}^*(t), t)$.
4. **Boundary conditions:** $\mathbf{x}^*(0) = \mathbf{x}_0$ (initial), $\boldsymbol{\lambda}^*(T) = \partial \phi / \partial \mathbf{x}|_{\mathbf{x}(T)}$ (terminal).

Interpretation: the PMP says optimal trajectories are characterized by a *two-point boundary value problem* in the state and costate, and at every instant the control minimizes the Hamiltonian. The costate plays the role of "marginal value of an extra unit of state" — it's the sensitivity of the cost-to-go with respect to the current state.

Why The PMP Is Useful

A few reasons:

1. It is necessary for optimality. Any optimal trajectory satisfies the PMP. So if you find a trajectory that does *not*, it's not optimal.

2. It reduces an infinite-dimensional problem to ODEs. The original problem is over function spaces (the control $\mathbf{u}(t)$), but the PMP gives a finite-dimensional system (state, costate, and a pointwise minimization).

3. It gives structural information. Even without solving, you can read off properties: when the Hamiltonian is linear in the control, the optimum is **bang-bang** (control jumps between extreme values); when it's quadratic, the optimum is **smooth** (linear feedback).

4. It generalizes to constraints, inequality constraints, free terminal time, and stochastic versions.

For VQAs, the PMP gives a non-obvious viewpoint on parameter optimization that is sometimes more natural than gradient descent.

The PMP Applied To VQAs

For a VQA, the optimal control problem is:

- **State:** $|\psi(t)\rangle$, evolving via discretized Schrödinger equation.
- **Control:** parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)$, one per gate (or one per gate-time interval).
- **Dynamics:** unitary evolution under the gate Hamiltonians.
- **Cost:** terminal expectation $C = \langle\psi(T)|H|\psi(T)\rangle$, no running cost.

Discretized PMP for VQA:

The "time" is the position in the circuit. The state $|\psi_k\rangle$ at step k is the result of applying the first k gates. The costate $|\lambda_k\rangle$ is propagated *backward* from $|\lambda_T\rangle = H|\psi_T\rangle$ using the adjoint dynamics.

The **gradient with respect to** θ_k is then

$$\frac{\partial C}{\partial \theta_k} = -i\langle\lambda_k|\frac{\partial U_k}{\partial \theta_k}U_k^\dagger|\psi_k\rangle + \text{c.c.},$$

which is exactly the form one gets from the parameter-shift rule. **The PMP-derived gradient and the parameter-shift gradient are the same expression in different notation.**

Why this matters: the costate-based view shows that the gradient computation is doing **forward propagation of the state** plus **backward propagation of the costate**, with a contraction at each step. This is structurally identical to backpropagation in classical neural networks. The "VQA backprop" picture is the PMP applied to circuits.

This is not new — Schuld et al. and others have made the connection explicit — but the PMP formulation is the *most general* statement of the underlying structure.

Bang-Bang Control And QAOA

A particularly important consequence of the PMP: when the control Hamiltonian is **linear in the control variable**, the optimum is **bang-bang** — the control switches between its extreme values.

For QAOA: the ansatz is

$$|\psi(\boldsymbol{\beta}, \boldsymbol{\gamma})\rangle = \prod_k e^{-i\beta_k H_M} e^{-i\gamma_k H_C} |+\rangle^{\otimes n},$$

where H_M is the mixer and H_C is the cost. This is a *piecewise-constant control*: at each step, the control selects either H_M or H_C, and the only choice is the duration.

PMP analysis. The control Hamiltonian for the optimal control problem is

$$H_{\text{ctrl}}(\boldsymbol{\theta}) = i\boldsymbol{\lambda}^\dagger \cdot (H_M\beta + H_C\gamma)|\psi\rangle + \text{c.c.}$$

This is *linear* in the durations β_k, γ_k. So the optimum durations are at the *boundary* of their allowed range — which is the bang-bang result.

Implication: the optimal QAOA (in the unconstrained-duration limit) uses only "extreme" values for β_k, γ_k — it switches abruptly between the cost and mixer Hamiltonians. Real QAOA has constrained durations (typically $[0, \pi]$), which softens the bang-bang prediction but the underlying structure is preserved.

This was first observed by Brandao et al. (2018) and subsequently studied as the **bang-bang QAOA** literature. The connection to Pontryagin gives the theoretical foundation.

Singular Arcs And QAOA Difficulty

A subtlety: when the control Hamiltonian is *exactly* zero in some region (the coefficient of the control vanishes identically), the bang-bang result fails — you have a **singular arc** where the optimal control is determined by higher-order conditions.

For VQAs: singular arcs correspond to **flat regions of the cost landscape** — the gradient (which is the linear coefficient of the control in the Hamiltonian) is identically zero. **Barren plateaus are exactly singular arcs in the optimal control formulation.**

This gives a *third* characterization of barren plateaus, complementary to: - Module 4's measure-theoretic view (Haar 2-design integration). - Module 8's Lie-algebraic view (small gradient variance from large DLA). - Now: **the plateau is a region where the PMP minimum condition is degenerate** — the control variable doesn't appear linearly in the Hamiltonian, so the standard gradient information is gone.

Implication for optimizer design: in a singular arc, you cannot use the PMP-derived gradient. You need either second-order conditions (Hessian) or non-gradient methods (population optimizers, Module 6.7-6.8). This is the optimal-control justification for Module 6's hybrid methods.

The Dual Variable Interpretation

A useful interpretation of the costate $\boldsymbol{\lambda}(t)$: it is the **dual variable** for the dynamics constraint. In Lagrangian terms, $\boldsymbol{\lambda}^T f(\mathbf{x}, \mathbf{u})$ is the penalty for violating $\dot{\mathbf{x}} = f$.

For VQAs, this means the costate $|\lambda_k\rangle$ is the "sensitivity of the final cost to perturbations of the state at step k." Knowing the costate at every step tells you exactly how much each intermediate quantum state matters for the final outcome.

Practical use: the costate identifies the *bottleneck* gates in a circuit — the gates whose perturbations have the largest effect on the final cost. These are the gates that should be parameterized most carefully and updated most aggressively. Gates with small costate norm are not contributing much and can be left alone or removed (this is the basis of **adaptive ansatz** methods like ADAPT-VQE).

Numerical Methods For PMP

Solving the PMP numerically is hard because it's a two-point boundary value problem (state at $t = 0$, costate at $t = T$).

Standard methods:

1. **Shooting.** Guess $\boldsymbol{\lambda}(0)$, integrate forward, check if the terminal condition is satisfied, adjust the guess.
2. **Multiple shooting.** Break the trajectory into segments, with a guess for the state and costate at each segment boundary.
3. **Collocation.** Discretize state and costate as polynomials over time intervals; solve the resulting algebraic system.
4. **Indirect methods generally.** Solve the PMP system directly as a system of nonlinear equations.

Compared to gradient descent. Gradient descent on the parameters is the *direct method* — discretize, then optimize — and is much simpler. The PMP-based methods are theoretically cleaner but numerically harder.

For most VQAs, gradient descent dominates. The PMP is used for *theoretical analysis* (proving things about optimal trajectories), not for the actual optimization.

When The PMP Gives New Information

A practical question: does the PMP ever give VQA practitioners *new* information they couldn't get from parameter-shift gradient descent?

Yes, in three specific situations:

1. Bang-bang structure prediction. Knowing that the optimum is bang-bang lets you initialize at the extreme values, which often gives better convergence than uniform initialization.

2. Singular arc detection. If you compute the costate norm and it's small, you're in a singular region (barren plateau). This is a fast diagnostic for plateaus.

3. Adaptive ansatz design. The costate identifies which gates contribute most. Adaptive ansatz methods (ADAPT-VQE) use this to grow the ansatz one gate at a time.

These are not earth-shattering, but they are concrete benefits. They justify the theoretical investment in understanding the PMP for VQA practitioners.

Stochastic And Quantum Pontryagin

A few extensions of the PMP relevant to VQAs:

1. Stochastic PMP. Generalizes to systems with noise, replacing the deterministic costate with a stochastic process. Used in stochastic optimal control. For VQAs with shot noise, this is the right framework, but the analysis is more complex.

2. Quantum PMP. Bonnard, Sugny, and others have developed PMP variants for quantum systems with explicit unitary constraints. These give time-optimal results for specific quantum gates (e.g., Khaneja-Brockett-Glaser).

3. Chattering (Fuller's phenomenon). In some optimal control problems, the optimal trajectory has *infinitely many* bang-bang switches in finite time. This is a pathological case but it appears in some quantum optimal control problems and is worth being aware of.

Structural Role

This node is the **theoretical capstone** of Module 7's optimal control thread. It introduces the Pontryagin minimum principle, applies it to VQAs, identifies the costate-gradient connection, and explains the bang-bang and singular-arc structures that appear in QAOA-like ansätze.

It directly connects to node 7.6 (feedback control), which uses the PMP framework to design closed-loop optimizer architectures.

Relations to Other Concepts

- **Pontryagin et al. (1962)** — *The Mathematical Theory of Optimal Processes*, the original.
- **Brandão et al. (2018)** — bang-bang QAOA.
- **Bonnard-Sugny** — quantum optimal control via PMP.
- **Module 6.1 (Gradient Descent)** — the practical use of PMP-derived gradients.
- **Module 4 (Barren Plateaus)** — singular arcs as plateaus.

Worked Example — PMP For A 1-Qubit State Preparation

Problem. Steer a 1-qubit state from $|0\rangle$ to $|1\rangle$ using the control Hamiltonian $H(t) = u(t)X$, with $|u(t)| \leq 1$.

Setup. - State: $|\psi(t)\rangle = c_0(t)|0\rangle + c_1(t)|1\rangle$. - Cost: terminal fidelity to $|1\rangle$, $J = -|c_1(T)|^2$. - Dynamics: $i|\dot{\psi}\rangle = u(t)X|\psi\rangle$.

PMP analysis. The control Hamiltonian is linear in u, so the optimum is bang-bang: $u^*(t) = \pm 1$.

Solving. With $u = +1$, the dynamics rotates the state at unit angular velocity around the X axis. To go from $|0\rangle$ to $|1\rangle$, you need to rotate by π. The minimum time is $T^* = \pi$.

Costate. The costate is $|\lambda(t)\rangle = \pm 2\mathrm{Im}[c_1^*]|1\rangle - 2\mathrm{Re}[c_1^*]|0\rangle$ or similar — geometrically, it's the gradient of $-|c_1|^2$ with respect to the state.

Connection to a single-parameter VQA. The minimum time π corresponds to the parameter $\theta = \pi$ in $R_X(\theta)$. The "bang-bang" optimum corresponds to "use the maximum allowed parameter value." This is trivial in this 1-parameter case but generalizes.

For multi-parameter ansätze, the bang-bang structure says: in each gate's parameter, the optimum tends to be at the extremes of the allowed range. This is *not* true in general for VQAs (because the cost is non-convex), but it is the right intuition in many cases.

Visualization

Pending. A diagram of a bang-bang control trajectory: the control $u(t)$ jumps between $+1$ and -1, the state evolves smoothly, the costate is computed by backward propagation. Caption: "Optimum, costate, control: PMP in three lines."

Exercises

1. For an LQR problem, derive the optimal feedback law from the PMP. How does it match the Riccati result?
2. The PMP says the costate satisfies $\dot{\boldsymbol{\lambda}} = -\partial H_{\text{ctrl}}/\partial \mathbf{x}$. For a VQA with terminal cost only, derive the costate dynamics.
3. (VQA) For a QAOA at depth $p = 2$ with a 4-vertex MaxCut, write the discretized PMP system. How many costate variables are there?

Research Extensions

- **PMP-aware VQA optimizers** — using costate information to drive parameter updates beyond simple gradient descent.
- **Singular arc detection in practice** — fast tests for whether you're in a barren plateau using costate norms.
- **Bang-bang VQA ansatz design** — explicitly constructing ansätze that have bang-bang optima for tractable problems.
- **Stochastic PMP for shot-noise VQAs** — formal extension to noisy gradient estimates.

Cross-links to Other Courses

- **Calculus of Variations** — Euler-Lagrange equations, Hamilton's principle.
- **Optimal Control (Lewis-Vrabie-Syrmos)** — the textbook treatment.
- **Quantum Optimal Control (D'Alessandro)** — the quantum-specific variant.
- **Module 7.4 (Optimal Control Theory)** — the parent framework.
- **Module 6 (Optimization Dynamics)** — the practical optimizer side.

Variational Quantum Algorithms · Module 7 — Control Theoretic View · Node 7.5
Concepts: control-theory, constrained-optimization, variational-principle
Prerequisites: 7.4
Enables: 7.6

hbar.university — own.your.cognition