Regularization as Dynamical Modification (and Why It Is Not Bias)

Variational Quantum Algorithms · Module 5 — Regularization

Node 5.5

Position in Knowledge Graph

System: Variational Quantum Algorithms **Structural Domain:** Regularization **Node:** 5.5

This is the **central claim** of the regularization thesis. It is the node that motivates why Module 5 is labelled "an original contribution" rather than "standard background." Everything in nodes 5.1 through 5.4 was standard ML / inverse-problems theory. This node argues that the standard theory **misframes what regularization is doing in the VQA setting**, and proposes a replacement framing.

The thesis claim

> In variational quantum algorithms, regularization should be understood not as injecting a prior into the objective, but as **modifying the dynamics of the optimizer** under stochastic feedback. The objective remains unchanged. Only the trajectory through parameter space is altered.

Source: thesis Chapter 6.2 [@thesis-ch6].

Formal Definition

Let the unregularized parameter update rule be the discrete-time dynamical system

$$\boldsymbol{\theta}_{k+1} = \mathcal{A}\big(\boldsymbol{\theta}_k,\, \hat{C}(\boldsymbol{\theta}_k),\, \mathcal{H}_k\big),$$

where $\mathcal{A}$ is the classical optimizer, $\hat{C}$ is the noisy cost estimate, and $\mathcal{H}_k$ is the optimizer's internal state at iteration k.

A **regularizer** is a modification of this map:

$$\boldsymbol{\theta}_{k+1} = \mathcal{A}_{\text{reg}}\big(\boldsymbol{\theta}_k,\, \hat{C}(\boldsymbol{\theta}_k),\, \mathcal{H}_k;\, \lambda\big),$$

where λ controls the strength.

Crucially, $\mathcal{A}_{\text{reg}}$ is not derived from a modified cost. It is a direct modification of the update rule itself — adding a damping term, smoothing the input signal, bounding the step size, etc. The fixed points of $\mathcal{A}_{\text{reg}}$ are *not* generally those of $\arg\min C + \lambda R$.

Intuition

In a noise-free convex setting, gradient descent on $C(\boldsymbol{\theta}) + \lambda R(\boldsymbol{\theta})$ and gradient descent on C with a damping term added to the update rule are mathematically the same thing in the limit $k \to \infty$. **They converge to the same point.** This is why the standard ML literature can use prior-injection and dynamical-modification interchangeably — at convergence, they agree.

In VQA, you never reach convergence. Shot noise, NISQ-era circuit depth limits, and finite measurement budgets mean the optimizer lives in a permanent transient regime. **The two framings disagree about what is happening in the transient.**

The prior-injection framing says: "we are searching for the regularized minimum." The dynamical-modification framing says: "we are flowing along a stochastic trajectory whose distribution at iteration k is what we care about."

The latter is the right description for VQA. The former is a useful classical limit, but it does not predict what an actual VQA run looks like at iteration 50 with 1000 shots per evaluation.

Why it is not bias

A common objection: "if you modify the dynamics, surely you are introducing bias?"

No. Or rather: not in the sense the objector means.

A *biased estimator* is one whose expected steady-state value is not the true minimum. A *dynamical modification* is one whose **trajectory** is different but whose **steady-state** (in the limit $k \to \infty$ and $\lambda \to 0$) is the same as the unmodified system.

Schematically:

$$\lim_{k \to \infty} \lim_{\lambda \to 0} \mathbb{E}[\boldsymbol{\theta}_k^{(\lambda)}] = \boldsymbol{\theta}^\star$$

where $\boldsymbol{\theta}^\star$ is the true minimizer of C.

The order of limits matters. You have to take $\lambda \to 0$ **before** $k \to \infty$ for the unbiased property to hold. This is the regime in which **the cosine schedule (node 5.8) operates**: λ is annealed to zero as k grows, ensuring that the regularizer fades out before the steady state is reached.

This is the polemic point. A static regularizer (constant λ) does introduce bias — it shifts the steady state. A scheduled regularizer that vanishes at convergence does not introduce bias — it only shapes the path. The thesis argument is that **only the scheduled, dynamical interpretation makes sense in VQA**, because static regularization in a transient regime is just adding bias for no asymptotic benefit.

Structural Role

This is the load-bearing node of Module 5. Nodes 5.6, 5.7, 5.8 are all instances of dynamical modification. Node 5.9 is the empirical evidence that the framing pays off.

It is also the link between Module 5 and Module 6 (Optimization Dynamics). The dynamical-modification framing only makes sense if you have already accepted that the optimizer is a dynamical system and not a function evaluator — see node 6.1 for the foundation.

Relations to Other Concepts

- **Stochastic approximation theory** (Robbins-Monro) — the closest classical analogue. Annealed step sizes are conceptually identical to scheduled regularization.
- **Implicit bias of SGD** — the observation that gradient descent itself produces solutions that are "regularized" without any explicit penalty, depending on the optimizer.
- **Dynamical systems / control theory** — the natural language for describing the modified update map.
- **Effective potential** — viewing $\mathcal{A}_{\text{reg}}$ as gradient flow on a modified energy landscape that depends on k and λ.

Worked Derivations

Damped gradient descent as dynamical modification

Consider gradient descent with a damping term:

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

If λ_k is constant, this is L2-regularized gradient descent and converges to $\arg\min C + \frac{\lambda}{2}\|\boldsymbol{\theta}\|^2$ — a biased steady state.

If $\lambda_k \to 0$ such that $\sum_k \lambda_k = \infty$ but $\sum_k \lambda_k \eta_k < \infty$, the steady state is the unbiased $\arg\min C$. The damping shapes the path but vanishes asymptotically.

Why constant λ is bad in VQA

For constant λ, the steady state of the regularized dynamics is

$$\boldsymbol{\theta}^\star_\lambda \neq \arg\min C.$$

The displacement $\|\boldsymbol{\theta}^\star_\lambda - \arg\min C\|$ is bounded by something proportional to λ/μ, where μ is the local strong-convexity constant. In VQA, μ is small (weakly identified parameters) and λ is non-trivial (you need it for stability), so the displacement is *not negligible*. You are paying a permanent bias for a temporary stability gain. **This is the core argument for scheduled regularization.**

Visualization

Pending. Diagram: two trajectory plots in 2D parameter space — one with constant λ (ends at biased point), one with scheduled λ (ends at the true minimum, but visits a more stable transient region).

Exercises

1. Construct a 1D quadratic example where constant L2 regularization shifts the minimum by an amount you can compute exactly.
2. Show that for the same example, a cosine-scheduled L2 with $\lambda_k = \lambda_0 \cos^2(\pi k/2K)$ reaches the true minimum at iteration K.
3. Argue (without proof) why the dynamical-modification framing is unnecessary in noise-free convex optimization, and necessary in noisy non-convex settings.

Research Extensions

- **Stochastic differential equation models of SGD** — continuous-time limit of the discrete dynamics; turns regularization into a drift term.
- **Mean-field theory of optimizer dynamics** — averaging over noise realizations to predict trajectory distributions.
- **Connection to quantum annealing schedules** — the cosine schedule in 5.8 is closely related to adiabatic quantum schedules.

Cross-links to Other Courses

- **Statistics Module 9** — Stochastic Optimization (Robbins-Monro, annealed step sizes).
- **Math-Intuition** — Dynamical systems, fixed points, and basins of attraction.
- **Quantum Foundations** — Adiabatic theorem (analogous schedule structure).

Variational Quantum Algorithms · Module 5 — Regularization · Node 5.5
Concepts: regularization, dynamical-systems, stochastic-optimization, cost-function
Prerequisites: 3.1, 5.1, 5.4, 6.1
Enables: 5.6, 5.7, 5.8

hbar.university — own.your.cognition