STAT 4830 — Project Scoping Meeting Prep (Team Internal Document)

Purpose. This document prepares the team for the first project scoping meeting with Prof. Damek Davis. It summarizes three candidate project directions using the instructor's required meeting structure, while providing beginner-friendly explanations so all team members share a common understanding.

Common Theme Across All Topics

All three candidate projects investigate **optimization instability**—situations where optimization algorithms behave unpredictably due to imperfect feedback. The imperfections differ by topic (noise, delay, or drift), but the underlying question is the same: *when do standard optimization guarantees fail to describe practical behavior?*

Option A — Zeroth-Order Reinforcement Learning

Beginner Overview

In reinforcement learning (RL), we aim to maximize expected reward, but cannot directly compute gradients because the environment is a black box. Policy Gradient (PG) methods estimate gradients indirectly using sampled trajectories, which introduces high variance. Evolutionary Strategies (ES) avoid gradients entirely by perturbing parameters and moving toward directions that improve reward on average.

Elevator Pitch (2 sentences)

We study optimization instability in reinforcement learning when gradient estimates are noisy or unreliable, focusing on regimes where policy gradient methods fail.
We compare policy gradients with zeroth-order evolutionary strategies under controlled noise to characterize stability and failure modes rather than benchmark performance.

Self-Critique

1. **Technical risk:** Observed instability may depend on the chosen environment.
2. **Key assumption:** Injected reward noise reflects realistic sources of RL instability.

Proof of Life (Week 1)

Train a small policy using both PG and ES on a simple environment (e.g., CartPole). Sweep reward noise variance and learning rates, and compare stability across random seeds.

Option B — Test-Time / Online Optimization Instability

Beginner Overview

In deployment, data distributions often shift over time. A natural idea is to update models during inference (test-time training or online learning). However, these updates can destabilize models even when training loss decreases, revealing a mismatch between optimization objectives and real-world performance.

Elevator Pitch (2 sentences)

We investigate when test-time or online optimization becomes unstable under distribution shift, even when loss appears to improve.
Using controlled drifting distributions, we analyze how step size, update frequency, and drift speed interact to produce collapse or divergence.

Self-Critique

1. **Key assumption:** Distribution drift can be parameterized smoothly.
2. **Implementation challenge:** Designing drift that is realistic but not trivial.

Proof of Life (Week 1)

Train a classifier on a base distribution. Introduce gradual distribution shift at test time while applying online updates, and track divergence between loss and accuracy.

Option C — Delayed Online Convex Optimization

Beginner Overview

In many real systems, feedback is delayed. In delayed online convex optimization, updates rely on stale gradients. While theory provides regret guarantees, these guarantees may not describe the actual trajectory of the iterates, which can oscillate or diverge.

Elevator Pitch (2 sentences)

We study stability of online convex optimization under delayed feedback, focusing on the mismatch between regret guarantees and the behavior of iterates.
Through synthetic convex problems, we empirically map convergence, oscillation, and divergence regimes.

Self-Critique

1. **Technical risk:** The convex setting may appear overly simple.
2. **Key assumption:** Regret is an insufficient proxy for stability of iterates.

Proof of Life (Week 1)

Run delayed gradient descent on a quadratic loss. Vary delay length and learning rate, and visualize iterate trajectories to identify instability.

Key Concepts (Shared Across Topics)

1 **Instability:** Sensitivity of optimization trajectories to small perturbations.

2 **Noise:** Random fluctuations in gradient or reward signals.

3 **Drift:** Systematic change in the objective over time.

4 **Delay:** Feedback arriving after decisions are made.

One-Sentence Summary. Across all options, we study optimization algorithms under degraded feedback to understand when and why theoretical guarantees fail to capture practical behavior.