Control with Learning on the Fly: First Toy Problems

PIs: Charlie Fefferman, Clancy Rowley

Grad students: Bernat Guillen
               Sam Otto
               Amlan Sinha
               Melanie Weber

PRINCETON UNIVERSITY
The Toy Problems

- Apply a time-dependent control $u$ to keep the position $q$ of a moving particle close to zero.
- $u(t), q(t) \in \mathbb{R}^1$.
- Dynamics of $q$ depend on $u$, random noise and one unknown parameter $a$. 

The Toy Problems

- Apply a time-dependent control $u$ to keep the position $q$ of a moving particle close to zero.
- $u(t), q(t) \in \mathbb{R}^1$.
- Dynamics of $q$ depend on $u$, random noise and one unknown parameter $a$.
- The particle starts at time zero at position $q(0) = 0$.
- Until time $T_0$, we may observe $q(t)$, but we cannot control it; we must take $u = 0$ for all times $< T_0$.
- Starting at time $T_0$, we are free to apply any control we please (but we’re not allowed to look into the future).
Dynamics of the moving particle

Three toy models:

TM I: \[ dq = (a \, dt + dW) + u \, dt \]

TM II: \[ dq = (aq \, dt + dW) + u \, dt \]

TM III: \[ dq = dW + au \, dt, \]

where

\[ q = \text{position}, \quad u = \text{control}, \quad t = \text{time}, \]
\[ dW = \text{White noise}, \]
\[ a = \text{unknown parameter} \]
Objective

We want to apply a control strategy $u$ to minimize the expected value of

$$S = \int_{T_0}^{T_1} (q^2 + \lambda u^2) \, dt,$$

where $\lambda > 0$ is a known coefficient.

- If $a$ is known, that’s a standard control theory problem.
- If $a$ is unknown, then as time $t$ increases, we learn more and more about $a$ from history up to time $t$. So we are doing control with learning on the fly.
Our toy problems vs. real problems

- Among other differences, real problems are infinite-dimensional, while our toy problems are one-dimensional.
- However, our toy problems already force us to face significant issues. We hope to “climb a ladder” by solving ever harder problems, starting with our toy problems.
How to formulate the problem?

Bayesian control
vs.
Agnostic control

Bayesian: assume a probability density

$$d\text{Prob} = \rho(a) \, da$$

reflecting our prior belief about the unknown $a$. Pick our strategy to minimize the expected value of $S$. 
Agnostic control

What to do if we know *absolutely nothing* about the unknown $a$?
(no prior belief)
How to formulate the problem?
The idea of regret

- We play against an opponent who knows the value of $a$.
- We pick our control strategy.
- Opponent picks optimal strategy for $a$.

\[
S_{\text{our team}}(a) = \text{Expected value of } S \text{ obtained using our strategy, given the value of } a.
\]

\[
S_{\text{opponent}}(a) = \text{Expected value of } S \text{ obtained by opponent, given the value of } a.
\]
The idea of regret

Additive regret

\[ AR(a) = S_{\text{our team}}(a) - S_{\text{opponent}}(a) \geq 0. \]

Multiplicative regret

\[ MR(a) = \frac{S_{\text{our team}}(a)}{S_{\text{opponent}}(a)} \geq 1. \]

Worst-case regret

\[ AR_* = \sup_{a \in \mathbb{R}} AR(a) \]
\[ MR_* = \sup_{a \in \mathbb{R}} MR(a) \]

These depend on our strategy.
Find a control strategy \( u \) that minimizes \( AR_\star \) (or \( MR_\star \)).
A variant ("Fuel Tax")

As before:

▶ We must control the system with no knowledge of $a$.
▶ Our opponent has perfect knowledge of $a$, and plays optimally.

However...
A variant ("Fuel Tax")

Our score for a given value of $a$ is

$$S_{\text{our team}}(a) = \text{Expected value of } \int_{T_0}^{T_1} (q^2 + \lambda u^2) \, dt$$

(which depends on our strategy)
A variant ("Fuel Tax")

Our score for a given value of $a$ is

$$S_{\text{our team}}(a) = \text{Expected value of } \int_{T_0}^{T_1} (q^2 + \lambda u^2) \, dt$$

(which depends on our strategy)

Our opponent’s score for a given value of $a$ is

$$S_{\text{opponent}}(a) = \text{Expected value of } \int_{T_0}^{T_1} (q^2 + \hat{\lambda} u^2) \, dt,$$

with $\hat{\lambda} > \lambda$ ("fuel tax").
A variant ("Fuel Tax")

We want to find a strategy such that

\[ S_{\text{our team}}(a) \leq S_{\text{opponent}}(a), \quad \text{for all } a \in \mathbb{R}, \]

with \( \hat{\lambda} \) as small as possible.

Another variant

Restrict \( a \) to an interval \( I \) (maybe \([0, \infty)\)) but assume no prior belief about where \( a \) is likely to be within \( I \).
Review

We have proposed 3 toy models:

\[ dq = (a \, dt + dW) + u \, dt \]
\[ dq = (aq \, dt + dW) + u \, dt \]
\[ dq = dW + au \, dt \]
We formulated several notions of optimality for control strategies $u$, namely

- Bayesian
- Least additive regret
- Least multiplicative regret
- Fuel tax variant
Results so far

For the toy model

\[ dq = (a \, dt + dW) + u \, dt \]

- We have a complete understanding of the Bayesian problem with prior belief regarding unknown \( a \) given by a normal distribution \( \mathcal{N}(a_0, \sigma) \).
- We have found a strategy that provably minimizes additive regret.
Results so far

It turns out that

- The optimal strategy for additive regret is the limit of the Bayesian strategy for prior belief $\mathcal{N}(a_0, \sigma)$ as $\sigma \to \infty$ for (arbitrary) fixed $a_0$. 
Results so far

It turns out that

- The optimal strategy for additive regret is the limit of the Bayesian strategy for prior belief $\mathcal{N}(a_0, \sigma)$ as $\sigma \to \infty$ for (arbitrary) fixed $a_0$.
- The additive regret $AR(a)$ for the optimal strategy is independent of $a$. 

Results so far

It turns out that

- The optimal strategy for additive regret is the limit of the Bayesian strategy for prior belief $N(a_0, \sigma)$ as $\sigma \to \infty$ for (arbitrary) fixed $a_0$.

- The additive regret $AR(a)$ for the optimal strategy is independent of $a$.

For other models and/or other notions of optimality, we are just getting started.
Thank you!