Almost Optimal Agnostic Control of Unknown Linear Dynamics

J. Carruth, M. Eggl, C. Fefferman, C. Rowley

January 2024

Abstract

We consider a simple control problem in which the underlying dynamics depend on a parameter a that is unknown and must be learned. We study three variants of the control problem: Bayesian control, in which we have a prior belief about a; bounded agnostic control, in which we have no prior belief about a but we assume that a belongs to a bounded set; and fully agnostic control, in which a is allowed to be an arbitrary real number about which we have no prior belief. In the Bayesian variant, a control strategy is optimal if it minimizes a certain expected cost. In the agnostic variants, a control strategy is optimal if it minimizes a quantity called the worst-case regret. For the Bayesian and bounded agnostic variants above, we produce optimal control strategies. For the fully agnostic variant, we produce almost optimal control strategies, i.e., for any $\varepsilon > 0$ we produce a strategy that minimizes the worst-case regret to within a multiplicative factor of $(1 + \varepsilon)$.

The purpose of this note is to announce the results of our companion papers [5, 6]. These papers explore a new flavor of adaptive control theory, which we call "agnostic control"; see also [4, 7, 9, 10]. While our exposition here borrows heavily from the introductions of [5, 6], we think the results benefit from a unified presentation. Moreover, we give here a more detailed overview of the results of [6] than is given in the introduction to that paper.

Many works in adaptive control theory attempt to control a system whose underlying dynamics are initially unknown and must be learned from observation. The goal is then to bound Regret, a quantity defined by comparing our expected cost with that incurred by an opponent who knows the underlying dynamics and plays optimally. Typically one tries to achieve a regret whose order of magnitude is as small as possible after a long time. Adaptive control theory has extensive practical applications; see, e.g., [2, 8, 11, 12] for some examples.

In some applications, we don't have the luxury of waiting for a long time. This is the case, e.g., for a pilot attempting to land an airplane following the sudden loss of a wing, as in [3]. Our goal here is to achieve the absolute minimum

This work was supported by AFOSR grant FA9550-19-1-0005 and by the Joachim Herz Foundation..

possible regret over a fixed, finite time horizon. This objective poses formidable mathematical challenges, even for simple model systems.

We will study a one-dimensional, linear model system whose dynamics depend on a single unknown parameter a. When a is large positive, the system is highly unstable. (There is no "stabilizing gain" for all a.) We will make progressively weaker assumptions about the unknown parameter a—eventually, we will assume that a may be any real number and we won't assume that we are given a Bayesian prior probability distribution for it.

We now give a precise statement of our problem.

The Model System

Our system consists of a particle moving in one dimension, influenced by our control and buffeted by noise. The position of our particle at time t is denoted by $q(t) \in \mathbb{R}$. At each time t, we may specify a "control" $u(t) \in \mathbb{R}$, determined by history up to time t, i.e., by $(q(s))_{s\in[0,t]}$. A "strategy" (aka "policy") is a rule for specifying $u(t)$ in terms of $(q(s))_{s\in[0,t]}$ for each t. We write $\sigma, \sigma', \sigma^*$, etc. to denote strategies. The noise is provided by a standard Brownian motion $(W(t))_{t\geq 0}$.

The particle moves according to the stochastic ODE

$$dq(t) = \big(aq(t) + u(t)\big)dt + dW(t), \qquad q(0) = q_0, \tag{1}$$

where a and q_0 are real parameters. Due to the noise in (1), $q(t)$ and $u(t)$ are random variables; these random variables depend on our strategy σ, and we often write $q^\sigma(t)$, $u^\sigma(t)$ to make that dependence explicit.

Over a time horizon $T > 0$, we incur a Cost, given† by

$$\textsc{Cost}(\sigma, a) = \int_0^T \big\{(q^\sigma(t))^2 + (u^\sigma(t))^2\big\}dt. \tag{2}$$

This quantity is a random variable determined by a, q_0, T and our strategy σ. Here, the starting position q_0 and time horizon T are fixed and known.

We would like to pick our strategy σ to keep our cost as low as possible. We examine several variants of the above control problem, making successively weaker assumptions regarding our knowledge of the parameter a. The first variant is simply the classical case, in which a is a known real number. In the second variant, we assume that the parameter a is unknown, but subject to a given prior probability distribution supported on a bounded interval. In the third variant, we assume that the parameter a belongs to a bounded interval, but is otherwise unknown (in particular, we do *not* assume that we are given a prior belief about a). In the fourth and final variant, we assume that a is unknown and may be any real number (again, we do not assume that we are given a prior belief about a). We refer to the third and fourth variants, in which we are not given a prior belief about a, as *agnostic control.*

† By rescaling, we can consider seemingly different cost functions of the form $\int_0^T (q^2 + \lambda u^2)$ for $\lambda > 0$.

Variant I: Classical Control

We suppose first that the parameter a is known. We write $\text{ECost}(\sigma, a; T, q_0)$, or sometimes $\text{ECost}(\sigma, a)$, to denote the expected Cost incurred by executing a given strategy σ. Our task is to pick σ to minimize $\text{ECost}(\sigma, a; T, q_0)$. As shown in textbooks (e.g., [1]), there is an elementary formula for the optimal strategy, denoted $\sigma_{\text{opt}}(a)$, given by

$$u(t) = -\kappa(T-t, a)q(t),$$

where

$$\kappa(s, a) = \frac{\tanh(s\sqrt{a^2+1})}{\sqrt{a^2+1} - a\tanh(s\sqrt{a^2+1})}.$$

We refer to $\sigma_{\text{opt}}(a)$ as the *optimal known-a strategy.* It will be important later to note that $\sigma_{\text{opt}}(a)$ satisfies the inequality

$$|u(t)| \leq C\max\{a, 1\} \cdot |q(t)| \text{ for an absolute constant } C. \tag{3}$$

Variant II: Bayesian Control

We now suppose that the parameter a is unknown, but is subject to a given prior probability distribution $d\text{Prior}(a)$ supported in an interval $[-a_{\max}, a_{\max}]$. Our goal is then to pick a strategy σ to minimize our expected cost, given by

$$\text{ECost}(\sigma, d\text{Prior}) = \int_{-a_{\max}}^{a_{\max}} \text{ECost}(\sigma, a)\ d\text{Prior}(a). \tag{4}$$

Before presenting rigorous results, we provide a heuristic discussion.

First of all, since $d\text{Prior}$ is supported in $[-a_{\max}, a_{\max}]$, a glance at (3) suggests that our optimal strategy σ will satisfy

$$|u^\sigma(t)| \leq Ca_{\max}|q^\sigma(t)|. \tag{5}$$

In [6], we introduce the notion of a *tame strategy* σ, which satisfies the estimate

$$|u^\sigma(t)| \leq C^\sigma_{\text{TAME}}[|q^\sigma(t)| + 1] \quad (\text{for all } t \in [0, T]) \tag{6}$$

with probability 1, for a constant C^σ_{TAME} called a *tame constant* for σ (note that C_{TAME} may depend on $a_{\max}$). Thus, we expect that the optimal strategy for Bayesian control will be tame.

Next, we note a major simplification. In principle, a strategy σ is a one-parameter family of functions on an infinite-dimensional space, because for each t it specifies $u(t)$ in terms of the path $(q(s))_{s\in[0,t]}$. However, reasoning heuristically, one computes that the posterior probability distribution for the unknown a, given a past history $(q(s))_{s\in[0,t]}$ is determined by the prior $d\text{Prior}(a)$, together with the two observable quantities

$$\zeta_1(t) = \int_0^t q(s)[dq(s) - u(s)ds] \quad \text{and} \quad \zeta_2(t) = \int_0^t (q(s))^2\ ds \geq 0. \tag{7}$$

Therefore, it is natural to suppose that the optimal strategy $\sigma_{\text{Bayes}}(d\text{Prior})$ takes the form

$$u(t) = \tilde{u}(q(t), t, \zeta_1(t), \zeta_2(t)) \tag{8}$$

for a function $\tilde{u}$ on $\mathbb{R} \times [0, T] \times \mathbb{R} \times [0, \infty)$.

So, instead of looking for a one-parameter family of functions on an infinite-dimensional space, we merely have to specify a function $\tilde{u}$ of four variables.

It isn't hard to apply heuristic reasoning to derive a PDE for the function $\tilde{u}$ in (8). To do so, we introduce the *cost-to-go*, $S(q, t, \zeta_1, \zeta_2)$, defined as the expected value of

$$\int_t^T \left\{ (q^\sigma(t))^2 + (u^\sigma(t))^2 \right\} \, ds, \tag{9}$$

conditioned on

$$q(t) = q, \quad \zeta_1(t) = \zeta_1, \quad \zeta_2(t) = \zeta_2, \tag{10}$$

with our strategy σ picked to minimize (9). Clearly,

$$\begin{aligned} &S(q, t, \zeta_1, \zeta_2) \\ &\quad = \min_u \{ (q^2 + u^2) dt + \mathrm{E}[S(q + dq, t + dt, \zeta_1 + d\zeta_1, \zeta_2 + d\zeta_2)] \} + o(dt), \end{aligned} \tag{11}$$

where (q, t, ζ_1, ζ_2) evolves to $(q + dq, t + dt, \zeta_1 + d\zeta_1, \zeta_2 + d\zeta_2)$ if we apply the control u at time t. Here, $\mathrm{E}[\cdots]$ denotes expected value conditioned on (10).

Moreover, the optimal control at time t (given (10)) is precisely the value of u that minimizes the right-hand side of (11); we denote it $\tilde{u}(t)$.

Taylor-expanding the right-hand side of (11), and taking $dt \to 0$, we arrive at the *Bellman equation*

$$\begin{aligned} 0 = &\partial_t S + (\bar{a}(\zeta_1, \zeta_2) q + \tilde{u}) \partial_q S + \bar{a}(\zeta_1, \zeta_2) q^2 \partial_{\zeta_1} S + q^2 \partial_{\zeta_2} S + \frac{1}{2} \partial_q^2 S \\ &+ q \partial_{q \zeta_1} S + \frac{1}{2} q^2 \partial_{\zeta_1}^2 S + (q^2 + \tilde{u}^2), \end{aligned} \tag{12}$$

where $\bar{a}(\zeta_1, \zeta_2)$ is the posterior expected value of a given (10); explicitly,

$$\bar{a}(\zeta_1, \zeta_2) = \frac{\int_{-a_{\max}}^{a_{\max}} a \exp\left(-\frac{a^2}{2} \zeta_2 + a \zeta_1 \right) \, d\text{Prior}(a)}{\int_{-a_{\max}}^{a_{\max}} \exp\left(-\frac{a^2}{2} \zeta_2 + a \zeta_1 \right) \, d\text{Prior}(a)}. \tag{13}$$

Moreover, the minimizer $\tilde{u}$ for the right-hand side of (11) is given by

$$\tilde{u}(q, t, \zeta_1, \zeta_2) = -\frac{1}{2} \partial_q S(q, t, \zeta_1, \zeta_2). \tag{14}$$

Together with (12), we impose the obvious terminal condition

$$S|_{t=T} = 0, \tag{15}$$

and the natural requirement

$$S \geq 0. \tag{16}$$

Our plan to solve for the optimal Bayesian control is thus to solve (12)–(16) for S and $\tilde{u}$, and then set $u^{\sigma_{\text{Bayes}}(d\text{Prior})}(t) := \tilde{u}(q(t), t, \zeta_1(t), \zeta_2(t))$.

We have produced numerical solutions to (12)–(16), but we don't have rigorous proofs of existence or regularity. We proceed by imposing the following assumption.

PDE Assumption. *Equations* (12)–(16) *admit a solution* $S \in C^{2,1}(\mathbb{R} \times [0,T] \times \mathbb{R} \times [0,\infty))$*, satisfying the estimates*

$$|\partial^{\alpha}_{q,t,\zeta_1,\zeta_2} S| \leq K \cdot [1 + |q| + |\zeta_1| + \zeta_2|]^{m_0} \text{ a.e. for } |\alpha| \leq 3, \tag{17}$$

and

$$|\tilde{u}| \leq C_{\text{TAME}} \cdot [1 + |q|] \text{ for all } (q, t, \zeta_1, \zeta_2), \tag{18}$$

for some K*,* m_0*,* C_{TAME}*.*

Assumption (18) asserts that our strategy $\sigma_{\text{Bayes}}(d\text{Prior})$, given by (12)–(16), is a tame strategy, as expected.

Our numerical simulations appear to confirm (17), (18). Accordingly, our *PDE Assumption* seems safe.

We are ready to present our rigorous results on optimal Bayesian control; these are proved in [6].

Theorem 1 (Optimal Bayesian Strategy)**.** *Fix a probability distribution* $d\text{Prior}$*, supported on* $[-a_{\max}, a_{\max}]$*, and suppose our* PDE Assumption *is satisfied. Let* $\sigma = \sigma_{\text{Bayes}}(d\text{Prior})$ *be the strategy obtained by solving* (12)–(16)*. Then*

(A) $\text{ECost}(\sigma, d\text{Prior}) = S(q_0, 0, 0, 0)$*, with* S *as in* (12)–(16)*.*

(B) *Let* σ' *be any other strategy. Then*

$$\text{ECost}(\sigma', d\text{Prior}) \geq \text{ECost}(\sigma, d\text{Prior}),$$

with equality only when we have

$$u^{\sigma'}(t) = u^{\sigma}(t) \text{ for a.e. } t \quad \text{and} \quad q^{\sigma'}(t) = q^{\sigma}(t) \text{ for all } t,$$

with probability 1.

When the competing strategy σ' is assumed to be tame, we can sharpen the above uniqueness assertion (B) to a quantitative result.

Theorem 2 (Quantitative Uniqueness of the Optimal Bayesian Strategy)**.** *Let* $d\text{Prior}$ *and* $\sigma = \sigma_{\text{Bayes}}(d\text{Prior})$ *be as in Theorem 1. Given* $\varepsilon > 0$*, and given a constant* $\hat{C}$*, there exists* $\delta > 0$ *for which the following holds.*

Let σ' *be a tame strategy with tame constant at most* $\hat{C}$*. If*

$$\text{ECost}(\sigma', d\text{Prior}) \leq \text{ECost}(\sigma, d\text{Prior}) + \delta,$$

then the expected value of

$$\int_0^T \{|q^{\sigma}(t) - q^{\sigma'}(t)|^2 + |u^{\sigma}(t) - u^{\sigma'}(t)|^2\} \, dt$$

is less than ε*.*

Theorem 2 plays a crucial rôle in our analysis of agnostic control for bounded a (see [6] for details).

We now discuss an issue arising in the proofs of our results on Bayesian control: We need a rigorous definition of a strategy. Clearly, the phrase "a rule for specifying $u(t)$ in terms of past history" isn't precise.

We want to allow $u(t)$ to depend discontinuously on past history $(q(s))_{s\in[0,t]}$. For instance, we should be allowed to set

$$u(t) = \begin{cases} -q(t) & \text{if } |q(t)| > 1, \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, we had better make sure that we can produce solutions of our stochastic ODE

$$dq = (aq + u)dt + dW. \tag{19}$$

Without the noise dW, we have a standard ODE, and the usual existence and uniqueness theorems for ODE would require Lipschitz continuity of u.

We proceed as follows.

At first we fix a partition

$$0 = t_0 < t_1 < \cdots < t_N = T \tag{20}$$

of the time interval $[0, T]$. We restrict ourselves to strategies σ in which the control $u(t)$ is constant in each interval $[t_\nu, t_{\nu+1})$, and in which, for each ν, $u(t_\nu)$ is determined by $(q(t_\gamma))_{\gamma\leq\nu}$, together with "coin flips" $\vec{\xi} = (\xi_1, \xi_2, \dots) \in \{0, 1\}^{\mathbb{N}}$. We assume that $u(t_\nu)$ is a Borel measurable function of $(q(t_1), \dots, q(t_\nu), \vec{\xi})$, and that for all ν we have

$$|u(t_\nu)| \leq C_{\text{TAME}}[|q(t_\nu)| + 1].$$

We call such a strategy a *tame strategy associated to the partiton* (20) with *tame constant* C_{TAME}. For such strategies, it is easy to define the solutions $q^\sigma(t)$, $u^\sigma(t)$ of our stochastic ODE (19).

Most of our work lies in controlling and optimizing tame strategies associated to a sufficiently fine partition (20). In particular, we prove approximate versions of Theorems 1 and 2 in the setting of such tame strategies.

We then define a tame strategy (not associated to any partition) by considering a sequence $\pi_1, \pi_2, \dots$ of ever-finer partitions of $[0, T]$. To each partition π_n we associate a tame strategy σ_n with a tame constant C_{TAME} independent of n. If the resulting $q^{\sigma_n}(t)$ and $u^{\sigma_n}(t)$ tend to limits, in an appropriate sense, as $n \to \infty$, then we declare those limits $q(t)$, $u(t)$ to arise from a *tame strategy* σ.

Finally, we drop the restriction to tame strategies and consider general strategies. To do so, we consider a sequence $(\sigma_n)_{n=1,2,\dots}$ of tame strategies, *not* assumed to have a tame constant independent of n. If the relevant $q^{\sigma_n}(t)$ and $u^{\sigma_n}(t)$ converge, in a suitable sense, as $n \to \infty$, then we say that the limits $q(t)$ and $u(t)$ arise from a strategy σ.

It isn't hard to pass from tame strategies associated to partitions of $[0, T]$ to general tame strategies, and then to pass from such tame strategies to general strategies. The work in proving Theorems 1 and 2 lies in our close study of tame strategies associated to fine partitions. We refer the reader to [6] for details.

Variant III: Agnostic Control for Bounded a

We now suppose that our parameter a is confined to a bounded interval $[-a_{\max}, a_{\max}]$ but is otherwise unknown. In particular, we don't assume that we are given a Bayesian prior probability distribution $d\text{Prior}(a)$. Consequently, we cannot define a notion of expected cost by formula (4).

Instead, our goal will be to minimize *worst-case regret*, defined by comparing the performance of our strategy with that of the optimal known-a strategy $\sigma_{\text{opt}}(a)$. We will introduce several variants of the notion of regret.

Let us fix a starting position q_0, a time horizon T, and an interval $[-a_{\max}, a_{\max}]$ guaranteed to contain the unknown a. To a given strategy σ, we associate the following functions on $[-a_{\max}, a_{\max}]$:

- Additive Regret, defined as

$$\text{AR}(\sigma, a) = \text{E}\textsc{Cost}(\sigma, a) - \text{E}\textsc{Cost}(\sigma_{\text{opt}}(a), a) \geq 0.$$

- Multiplicative Regret (aka "competitive ratio"), defined as

$$\text{MR}(\sigma, a) = \frac{\text{E}\textsc{Cost}(\sigma, a)}{\text{E}\textsc{Cost}(\sigma_{\text{opt}}(a), a)} \geq 1.$$

- Hybrid Regret, defined in terms of a parameter $\gamma > 0$ by setting

$$\text{HR}_\gamma(\sigma, a) = \frac{\text{E}\textsc{Cost}(\sigma, a)}{\text{E}\textsc{Cost}(\sigma_{\text{opt}}(a), a) + \gamma}.$$

Writing $\textsc{Regret}(\sigma, a)$ to denote any one of the above three functions on $[-a_{\max}, a_{\max}]$, we define the *worst-case regret*

$$\textsc{Regret}^*(\sigma) = \sup \left\{ \textsc{Regret}(\sigma, a) : a \in [-a_{\max}, a_{\max}] \right\}.$$

We seek a strategy σ that minimizes worst-case regret.

The above notions are useful in different regimes. If we expect to pay a large cost, then we care more about multiplicative regret then about additive regret. (If we have to pay 10^9 dollars, we are unimpressed by a savings of 10^5 dollars.) Similarly, if our expected cost is small, then we care more about additive regret then about multiplicative regret. (If we pay only 10^{-5} dollars, we don't care that we might instead pay 10^{-9} dollars.) If we fix γ to be a cost we are willing to neglect, then hybrid regret $\text{HR}_\gamma(\sigma, a)$ provides meaningful information regardless of the order of magnitude of the expected cost.

So far, we have defined three flavors of worst-case regret, and posed the problem of minimizing that regret. The solution to our agnostic control problem is given by the following result, proved in [6].

Theorem 3. *Fix* $[-a_{\max}, a_{\max}]$*,* q_0*,* T *(and* γ *if we use hybrid regret). Suppose our* PDE Assumption *is satsified. Then there exist a probability measure* $d\text{Prior}(a)$*, a finite subset* $E \subset [-a_{\max}, a_{\max}]$*, and a strategy* σ*, for which the following hold.*

(I) σ *is the optimal Bayesian strategy for the prior probability distribution* $d\text{Prior}$*.*

(II) $d\text{Prior}$ *is supported in the finite set* E*.*

(III) E *is precisely the set of points at which the function* $a \mapsto \text{Regret}(\sigma, a)$ *achieves its maximum on the interval* $[-a_{\max}, a_{\max}]$*.*

(IV) $\text{Regret}^*(\sigma) \leq \text{Regret}^*(\sigma')$ *for any other strategy* σ'*.*

So, for optimal agnostic control, we should pretend to believe that the unknown a is confined to a finite set E and governed by the probability distribution $d\text{Prior}$, even though in fact we know nothing about a except that it lies in $[-a_{\max}, a_{\max}]$.

It is easy to see that conditions (I), (II), (III) in Theorem 3 imply condition (IV) (we give the argument later in this Section). The hard part of Theorem 3 is the assertion that there exist $d\text{Prior}$, E, σ satisfying (I), (II), (III); we now give an overview of how this is done.

We first prove an analogous result for the setting in which the unknown a is confined to a finite subset $A \subset [-a_{\max}, a_{\max}]$. Once that's done, we take a sequence of fine nets, e.g.,

$$A_n = [-a_{\max}, a_{\max}] \cap 2^{-n}\mathbb{Z},\ n = 1, 2, 3, \ldots$$

and deduce Theorem 3 by applying our result to the A_n and passing to the limit.

We sketch the ideas for finite A.

First of all, because we allow strategies to depend on coinflips, it's easy to define intermediate or "mixed" strategies between two given strategies σ_0 and σ_1. Given a number $\theta \in [0, 1]$, we play strategy σ_1 with probability θ, and we play instead strategy σ_0 with probability $1 - \theta$. We write σ_θ to denote that mixed strategy. Clearly, we have

$$\text{ECost}(\sigma_\theta, a) = \theta\text{ECost}(\sigma_1, a) + (1 - \theta)\text{ECost}(\sigma_0, a) \text{ for any } a \in \mathbb{R}.$$

Now let $A \subset [-a_{\max}, a_{\max}]$ be finite. We associate to any given strategy σ its *cost vector*, defined as

$$\overrightarrow{\text{ECost}}(\sigma) = (\text{ECost}(\sigma, a))_{a \in A} \in \mathbb{R}^A.$$

Thanks to our discussion of intermediate strategies, the set of all cost vectors of arbitrary strategies is a convex set $\mathcal{K} \subset \mathbb{R}^A$.

For $\varepsilon > 0$, we call a strategy σ_0 *ε-efficient* if there is no competing strategy σ' such that

$$\text{ECost}(\sigma', a) < \text{ECost}(\sigma_0, a) - \varepsilon \text{ for all } a \in A.$$

A simple convexity argument shows that any ε-efficient strategy σ_0 is within ε of optimal for some Bayesian prior probability distribution $(p(a))_{a\in A}$ on A. To see this, we form the convex set $\mathcal{K}_-$, consisting of all vectors $(v_a)_{a\in A} \in \mathbb{R}^A$ such that

$$v_a < \text{ECost}(\sigma_0, a) - \varepsilon \text{ for all } a \in A.$$

Since σ_0 is ε-efficient, the convex sets $\mathcal{K}$ and $\mathcal{K}_-$ are disjoint, hence there is a nonzero linear functional λ on $\mathbb{R}^A$ such that $\lambda(v_-) \leq \lambda(v)$ for all $v_- \in \mathcal{K}_-$, $v \in \mathcal{K}$. From the functional λ we can easily read off a probability distribution $(p(a))_{a\in A}$ on A such that

$$\sum_{a\in A} p(a)\text{ECost}(\sigma_0, a) \leq \sum_{a\in A} p(a)\text{ECost}(\sigma', a) + \varepsilon$$

for every competing strategy σ'.

Thus, as claimed, any ε-efficient strategy is within ε of best possible for Bayesian control for some prior probability distribution on A. Now we are ready for the analogue of Theorem 3 for finite A. The result is as follows.

Lemma 1 (Agnostic Control Lemma)**.** *Let $A \subset [-a_{\max}, a_{\max}]$ be finite, and let $\varepsilon > 0$ be given. Then there exist a subset $A_0 \subset A$, a probability measure μ on A_0, and a strategy σ with the following properties.*

- *σ is the optimal Bayesian strategy for the prior μ.*
- $\text{Regret}(\sigma, a) \leq \text{Regret}(\sigma, a_0) + \varepsilon$ *for all $a \in A$ and $a_0 \in A_0$.*

In particular,

$$|\text{Regret}(\sigma, a_0) - \text{Regret}(\sigma, a_0')| \leq \varepsilon \textit{ for } a_0, a_0' \in A.$$

The proof of the Agnostic Control Lemma proceeds by induction on $\#A$, the number of elements of A. (So it is essential that the Lemma deals only with finite A.)

In the base case $\#A = 1$, we have $A = \{a_0\}$ for some a_0. We take $A_0 = A$, $\mu =$ point mass at a_0, $\sigma =$ optimal known-a strategy for $a = a_0$. The conclusions of the Lemma are obvious.

For the induction step, we fix $k \geq 2$ and suppose our Lemma holds whenever $\#A < k$. We then prove the Lemma for $\#A = k$.

Thus, let $\#A = k$, and let $\varepsilon > 0$. We define suitable small positive numbers

$$\varepsilon_4 \ll \varepsilon_3 \ll \cdots \ll \varepsilon_0 = \varepsilon.$$

For $A' \subset [-a_{\max}, a_{\max}]$ finite, we define

$$\text{Regret}_{\max}(\sigma, A') = \max\{\text{Regret}(\sigma, a) : a \in A'\}$$

for any strategy σ.

Let $\hat{\sigma}$ be a strategy for which $\text{REGRET}_{\max}(\hat{\sigma}, A)$ is within ε_4 of least possible. Then $\hat{\sigma}$ is ε_3-efficient. Indeed, if any competing strategy σ' satisfied

$$\text{ECOST}(\sigma', a) < \text{ECOST}(\hat{\sigma}, a) - \varepsilon_3 \text{ for all } a \in A,$$

then $\text{REGRET}_{\max}(\sigma', A)$ would be smaller than $\text{REGRET}_{\max}(\hat{\sigma}, A)$ by more than ε_4, contradicting the defining property of $\hat{\sigma}$. Since ε_3-efficient strategies are within ε_3 of best possible for some Bayesian prior, there exists a probability distriubtion μ on A such that

$$\text{ECOST}(\hat{\sigma}, \mu) \leq \text{ECOST}(\sigma', \mu) + \varepsilon_3 \tag{21}$$

for any competing strategy σ'.

In particular, let σ be the optimal Bayesian strategy for the prior μ. Then (21) gives

$$\text{ECOST}(\hat{\sigma}, \mu) \leq \text{ECOST}(\sigma, \mu) + \varepsilon_3.$$

Theorem 2* therefore implies that

$$|\text{ECOST}(\hat{\sigma}, a) - \text{ECOST}(\sigma, a)| \leq \varepsilon_3 \text{ for all } a \in A,$$

and therefore

$$|\text{REGRET}_{\max}(\hat{\sigma}, A) - \text{REGRET}_{\max}(\sigma, A)| \leq \varepsilon_2.$$

Together with the defining property of $\hat{\sigma}$, this shows that

$$\text{REGRET}_{\max}(\sigma, A) \leq \text{REGRET}_{\max}(\sigma', A) + 2\varepsilon_2. \tag{22}$$

for any competing strategy σ'.

It may happen that

$$\text{REGRET}(\sigma, a) \geq \text{REGRET}_{\max}(\sigma, A) - \varepsilon_1 \text{ for all } a \in A. \tag{23}$$

In that case, we have

$$\text{REGRET}_{\max}(\sigma, A) - \varepsilon_4 \leq \text{REGRET}(\sigma, a) \leq \text{REGRET}_{\max}(\sigma, A) \text{ for all } a \in A,$$

so the conclusions of our lemma hold for the above μ, σ with $A_0 = A$. Hence, we may assume that (23) is false.

We set

$$A_0 = \{a \in A : \text{REGRET}(\sigma, A) \geq \text{REGRET}_{\max}(\sigma, a) - \varepsilon_1\}.$$

Since (23) is false, we have $\#A_0 < \#A = k$, hence, by our induction hypothesis, Lemma 1 applies to A_0.

Thus, there exist a subset $A_{00} \subset A_0$, a probability measure μ_0 on A_{00}, and a strategy σ_0, such that

*Theorem 2 applies only to tame strategies. In this article, we oversimplify by ignoring that issue. See [6] for a correct discussion.

- σ_0 is the optimal Bayesian strategy for the prior μ_0, and
- $\text{REGRET}(\sigma_0, a) \leq \text{REGRET}(\sigma_0, a_0) + \varepsilon_4$ for all $a \in A_0$, $a_0 \in A_{00}$.

We then show that the conclusions of Lemma 1 hold, with A_{00}, μ_0, σ_0 in place of A_0, μ, σ. This completes our induction on $\#A$, proving Lemma 1.

Once we have established Lemma 1, we can easily pass from the finite sets $A_n = [-a_{\max}, a_{\max}] \cap 2^{-n}\mathbb{Z}$ to the full interval $[-a_{\max}, a_{\max}]$ by a weak compactness argument. This proves conclusions (I), (II), (III) of Theorem 3 except for the finiteness of the set E on which the function

$$[-a_{\max}, a_{\max}] \ni a \mapsto \text{REGRET}(\sigma, a)$$

takes its maximum.

To see that E is finite, we examine the function

$$F : \mathbb{R} \ni a \mapsto \text{REGRET}(\sigma, a).$$

We prove that F is real-analytic and grows exponentially fast as $a \to +\infty$. Consequently, $F|_{[-a_{\max}, a_{\max}]}$ is a nonconstant real-analytic function, which can therefore achieve its maximum at only finitely many points. Thus, (I), (II), and (III) hold with E finite.

It remains only to deduce conclusion (IV) from (I), (II), (III). Let $d\text{Prior}$, σ, E be as in (I), (II), (III) of Theorem 3. Since σ is the optimal Bayesian strategy for $d\text{Prior}$ (by (I)), and since $d\text{Prior}$ is supported on the finite set E (by (II)), we have for any other strategy σ' that

$$\text{ECOST}(\sigma, a_0) \leq \text{ECOST}(\sigma', a_0) \text{ for some } a_0 \in E.$$

In particular, we have

$$\text{REGRET}(\sigma, a_0) \leq \text{REGRET}(\sigma', a_0) \text{ for some } a_0 \in E.$$

Combining this with (III), we see that for any $a \in [-a_{\max}, a_{\max}]$ we have

$$\text{REGRET}(\sigma, a) \leq \text{REGRET}(\sigma', a_0).$$

Therefore (I), (II), (III) of Theorem 3 easily imply (IV).

This concludes our discussion of agnostic control for bounded a; for details, see [6]. Finally, we pass to the most general case.

Variant IV: Fully Agnostic Control

Finally, we make no assumption whatever regarding the unknown a. Our a may be any real number, and we are not given a Bayesian prior distribution for it. If a is large positive, then the system is highly unstable. Our goal is again to minimize worst-case regret, defined as in the previous section, except now the supremum is taken over all $a \in \mathbb{R}$. We confine ourselves to hybrid regret.

We now denote the hybrid regret of a strategy σ by $\mathrm{HR}_\gamma(\sigma, a; q_0, T)$, to make explicit the rôle of the starting position q_0 and time horizon T. Thus, for fixed γ, q_0, T, we are trying to minimize

$$\mathrm{HR}^*_\gamma(\sigma; q_0, T) = \sup_{a\in\mathbb{R}} \mathrm{HR}_\gamma(\sigma; a, q_0, T).$$

We remark that this sup may be infinite.

We strengthen our *PDE Assumption* by assuming also that the constant C_{TAME} in (18) grows at most as a power of $a_{\max}$ when $a_{\max} \gg 1$, i.e., we assume that (12)–(17) hold and that there exists an integer n_0 for which

$$|\tilde{u}| \leq C_0 \cdot [1 + a_{\max}^{n_0}] \cdot [1 + |q|] \text{ for all } (q, t, \zeta_1, \zeta_2) \tag{24}$$

(recall that $a_{\max} > 0$). This seems plausible; we have argued that most likely $C_{\mathrm{TAME}} = O(a_{\max})$ (see (5)).

The main result of our paper [5] is that, with negligible increase in regret, we can reduce matters to agnostic control for bounded a. Specifically, we prove the following Theorem.

Theorem 4. *Fix a time horizon T, a nonzero starting position q_0, and constants C_0, n_0 (to be used in the estimate* (24)*). Then given $\varepsilon > 0$ there exists $a_{\max} > 0$ for which the following holds.*

Let σ be a strategy for the starting position q_0 and time horizon $T + \varepsilon$. Suppose σ satisfies estimate (24) *for $a_{\max}$ and the given C_0, n_0.*

Then there exists a strategy σ_* for the starting position q_0 and time horizon T, satisfying the following estimates.

(A) *For $a \in [-a_{\max}, a_{\max}]$ we have*

$$\begin{aligned}&\mathrm{E}\textsc{Cost}(\sigma_*, a; T, q_0)\\ &\qquad \leq \varepsilon + (1+\varepsilon) \cdot \sup\left\{\mathrm{E}\textsc{Cost}(\sigma, a'; T+\varepsilon, q_0) : |a' - a| \leq \varepsilon|a|\right\}.\end{aligned}$$

(B) *For $a \notin [-a_{\max}, a_{\max}]$ we have*

$$\mathrm{E}\textsc{Cost}(\sigma_*, a; T, q_0) \leq \varepsilon + (1+\varepsilon) \cdot \mathrm{E}\textsc{Cost}(\sigma_{\mathrm{opt}}(a), a; T, q_0).$$

So, if $a \in [-a_{\max}, a_{\max}]$, then σ_* performs almost as well as σ; and if $a \notin [-a_{\max}, a_{\max}]$, then σ_* performs almost as well as the optimal known-a strategy $\sigma_{\mathrm{opt}}(a)$.

Using Theorem 4, we construct strategies σ that come arbitrarily close to minimizing worst-case hybrid regret. Assume that we are given constants γ, T, q_0, C_0, m_0 as in Theorem 4. We let $\varepsilon > 0$ be given and we take $a_{\max}$ to be a large enough positive real number (depending on ε as well as the constants above).

We let σ_0 be the optimal agnostic control strategy for worst-case hybrid regret with starting position q_0 and time horizon $T + \varepsilon$, and with a confined to the interval $[-(1+\varepsilon)a_{\max}, (1+\varepsilon)a_{\max}]$. (Of course, this is Variant III above). We assume (24) holds for σ_0.

Applying Theorem 4 to σ_0, we obtain a strategy σ_{Ag} for time horizon T so that:

- For $a \in [-a_{\max}, a_{\max}]$, the strategy σ_{Ag} performs only slightly worse than the worst-case performance of the strategy σ_0 on the slightly larger interval $[-(1+\varepsilon)a_{\max}, (1+\varepsilon)a_{\max}]$.
- For $a \notin [-a_{\max}, a_{\max}]$, the strategy σ_{Ag} performs only slightly worse than the optimal known-a strategy $\sigma_{\mathrm{opt}}(a)$.

From this, it's easy to deduce that the worst-case hybrid regret of the strategy σ_{Ag} (for fully agnostic control, i.e., with $a \in \mathbb{R}$) is at most $O(\varepsilon)$ percent worse than that of σ_0 (for agnostic control with a confined to $[-(1+\varepsilon)a_{\max}, (1+\varepsilon)a_{\max}]$). The worst-case hybrid regret of the optimal strategy σ_0 on the interval $[-(1+\varepsilon)a_{\max}, (1+\varepsilon)a_{\max}]$ is, of course, bounded above by the worst-case hybrid regret of *any* strategy σ for fully agnostic control (i.e., with $a \in \mathbb{R}$). Consequently, we have

$$\mathrm{HR}^*_\gamma(\sigma_{\mathrm{Ag}}; q_0, T) \leq (1 + C\varepsilon) \cdot \mathrm{HR}^*_\gamma(\sigma; q_0, T+\varepsilon)$$

for any competing strategy σ.

Thus, building on our solution for the control problem in Variant III, we have produced an almost optimal strategy for fully agnostic control. For a more detailed overview of the proof of Theorem 4, we refer the reader to the introduction of [5].

A Future Direction

In [6], we discuss several unsolved problems suggested by our work in [5, 6]. Here, we discuss one of those unsolved problems in more detail. Specifically, we speculate briefly on a particular model problem in which we don't know *a priori* what our control does.

Consider a particle governed by the stochastic ODE

$$dq(t) = au(t)dt + dW(t), \qquad q(0) = 0. \tag{25}$$

As usual, $q(t)$ denotes position, $u(t)$ is our control, $W(t)$ is Brownian motion, and we incur a cost

$$\int_0^T \{(q(t))^2 + (u(t))^2\}\, dt.$$

We would like to understand optimal agnostic control for this system, i.e., we'd like to find strategies that minimize worst-case regret. In analogy with our work on the system (1), we first attempt to understand optimal Bayesian control.

In the simplest case of Bayesian control, suppose we know *a priori* that $a = 1$ or $a = -1$, each with probability $1/2$.

We write $\text{ECost}(\sigma)$ to denote the expected cost incurred by executing a strategy σ, and we set

$$\text{ECost}^* = \inf\{\text{ECost}(\sigma) : \text{ All strategies } \sigma\}. \tag{26}$$

For this simple problem, we make the following conjectures.

- The infimum in (26) is not achieved by any strategy σ, because there is a regime in which we would like to set $u(t) = \pm\infty$, in order to gain instant information about a.
- A nearly optimal strategy will determine $u(t)$ as a function of position $q(t)$, time t, and $p(t)$ = posterior probability that $a = +1$, given history up to time t. Thus, $u(t) = \tilde{u}(q(t), t, p(t))$ for a function $\tilde{u}(q, t, p)$ on the "state space" $\Omega = \mathbb{R} \times [0, T] \times [0, 1]$.
- The state space Ω is partitioned into two regimes Ω_0 and Ω_1. In Ω_0, we would like to set $\tilde{u} = \pm\infty$, so we set $\tilde{u} = \mathcal{U}$, a large positive number.† In Ω_1, we take $\tilde{u}$ to be a solution of a relevant Bellman equation. A free boundary condition determines how we partition Ω into Ω_0 and Ω_1.
- As $\mathcal{U} \to \infty$, such strategies approach optimality. Perhaps one should define strategies in a way that allows $u = \pm\infty$. If so, this had better be done carefully.

We emphasize that the above are speculations—we have no rigorous results on optimal agnostic control for the system (25). We remark, however, that in [4] the first-named author has found a strategy that achieves bounded multiplicative regret for a more general system than (25).‡

Clearly, there is much to be done before we can claim to understand agnostic control theory.

References

[1] Karl Aström. *Introduction to Stochastic Control Theory*. Academic Press, 1970.

[2] Dimitri Bertsekas. *Dynamic Programming and Optimal Control: Volume I*, volume 1. Athena scientific, 2012.

[3] D.P. Brazy. Group chairman's factual report of investigation. *National Transportation Safety Board Docket No. SA-532, Exhibit No. 12*, 2009.

[4] Jacob Carruth. A bounded regret strategy for linear dynamics with unknown control. *arXiv:2311.13365 (preprint)*, 2023.

†We could just as well set $\tilde{u} = -\mathcal{U}$.

‡In fact, the results of [4] assume that the starting position q_0 satisfies $|q_0| \geq 1$. After an easy modification, however, the strategy defined in that paper for the system (25) achieves bounded regret for arbitrary $q_0 \in \mathbb{R}$.

[5] Jacob Carruth, Maximilian F. Eggl, Charles Fefferman, and Clarence W. Rowley. Controlling unknown linear dynamics with almost optimal regret. *arXiv:2309.10142 (preprint)*, 2023.

[6] Jacob Carruth, Maximilian F. Eggl, Charles Fefferman, and Clarence W. Rowley. Optimal agnostic control of unknown linear dynamics in a bounded parameter range. *arXiv:2309.10138 (preprint)*, 2023.

[7] Jacob Carruth, Maximilian F. Eggl, Charles Fefferman, Clarence W. Rowley, and Melanie Weber. Controlling unknown linear dynamics with bounded multiplicative regret. *Revista Matemática Iberoamericana*, 38(7):2185–2216, 2022.

[8] Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, Learning, and Games.* Cambridge university press, 2006.

[9] Charles L. Fefferman, Bernat Guillen Pegueroles, Clarence W. Rowley, and Melanie Weber. Optimal control with learning on the fly: a toy problem. *Revista Matematica Iberoamericana*, 37(1), 2021.

[10] Daniel Gurevich, Debdipta Goswami, Charles L Fefferman, and Clarence W Rowley. Optimal control with learning on the fly: System with unknown drift. In *Learning for Dynamics and Control Conference*, pages 870–880. PMLR, 2022.

[11] Elad Hazan and Karan Singh. Introduction to online nonstochastic control. *arXiv:2211.09619 (preprint)*, 2022.

[12] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.