Regret-Optimal Control

Thesis by
Gautam Goel

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended May 26, 2022

Gautam Goel
ORCID: 0000-0002-7054-7218

To my father, Ashok Goel, who inspired me to become a scientist

ACKNOWLEDGEMENTS

I would like to thank my advisor, Babak Hassibi, for all of his guidance and support during my journey through graduate school. In my mind, Babak embodies the very best qualities of a scientist: he is generous with students, uncompromising in rigor, and allergic to hype and self-promotion. I am proud to call him my advisor, and I can only hope to emulate his example in my own career.

I would also like to thank the other members of my committee: Elad Hazan, Yisong Yue, and Adam Wierman. I have learned something from each of them over the years, and I am grateful for their encouragement and advice.

Many people helped orient me when I was a junior student, including Desmond Cai, Niangjun Chen, Hu Fu, Nikolai Matni, Madeleine Udell, and Steven Wu. Their guidance has proven to be invaluable, and I thank them for it. I am also grateful for my colleagues, Ehsan Abbasi, Navid Azizan, Ahmed Douik, Sahin Lale, Taylan Kargin, Yiheng Lin, Fariborz Salehi, Guanya Shi, and Hikmet Yildiz, who provided a welcoming and stimulating environment during my graduate studies.

My time at Caltech has been immeasurably brightened by my excellent friends: Navid Azizan, Dylan Bannon, Spencer Gordon and Rie Ohta, Dimitar and Nina Ho, Mason McGill, Jenish Mehta, Riley Murray, Florian Schäfer, Jialin Song, Lucien Werner, Eric Zhan, and Juba Ziani. Most of all, I would like to thank Anish Sarma, who is among the most thoughtful and decent of men. I have enjoyed our conversations immensely, and I will miss our long walks across Pasadena.

The affection of my girlfriend, Megan Schill, made the travails of graduate school infinitely more bearable. I have never met a woman who is more generous or more loyal, and I'm grateful to have her in my life.

The unconditional love of my family is a blessing which has sustained me over my long journey through graduate school, and I thank them for it. I would especially like to thank my father, Ashok Goel, to whom this thesis is dedicated. My father has always supported my aspiration to become a scientist; I will never forget how we used to wake early in the morning so that he could drive me to school for my AP calculus practice tests. My values, such as they are, I inherited from him.

ABSTRACT

Optimal controllers are usually designed to minimize cost under the assumption that the disturbance they encounter is drawn from some specific class. For example, in H_2 control the disturbance is assumed to be generated by a stochastic process and the controller is designed to minimize its expected cost, while in H_∞ control the disturbance is assumed to be generated adversarially and the controller is designed to minimize its worst-case cost. This approach suffers from an obvious drawback: a controller which encounters a disturbance which falls outside of the class it was designed to handle may perform poorly. This observation naturally motivates the design of adaptive controllers which dynamically adjust their control strategy as they causally observe the disturbance instead of blindly following a prescribed strategy.

Inspired by online learning, we propose *data-dependent regret* as a criterion for controller design. In the regret-optimal control paradigm, causal controllers are designed to minimize regret against a hypothetical *optimal noncausal controller*, which selects the cost-minimizing sequence of control actions given noncausal access to the disturbance sequence. Controllers with low regret retain a performance guarantee irrespective of how the disturbance is generated; it is this universality which makes our approach an attractive alternative to traditional H_2 and H_∞ control. The regret of the causal controller is bounded by some measure of the complexity of the disturbance sequence; we consider several different complexity measures, including the energy of the disturbance sequence, which measures the size of the disturbance, and the pathlength of the disturbance, which measures its variation over time. We also consider the alternative metric of *competitive ratio*, which is the worst-case ratio between the cost incurred by the causal controller and the cost incurred by the optimal noncausal controller. This metric can also be viewed as a special case of data-dependent regret, where the complexity measure is simply the offline optimal cost. For each of these complexity measures, we derive a corresponding control algorithm with optimal data-dependent regret. The key technique we introduce is an operator-theoretic reduction from regret-optimal control to H_∞ control; each of the regret-optimal controllers we obtain can be interpreted as an H_∞ controller in a synthetic system of larger dimension. We also extend regret-optimal control to the more challenging *measurement-feedback* setting, where the online controller must choose control actions without directly observing the disturbance sequence, using only noisy linear measurements of the state.

We show that the competitive controller can be arbitrarily well-approximated by the class of disturbance-action-controller (DAC) policies. The convexity of this class of policies makes it amenable to online optimization via a reduction to online convex optimization with memory, and this class has hence attracted much recent attention in online learning. Using our approximation result, we show how to obtain algorithms which achieve the "best-of-both-worlds": sublinear policy regret against DAC policies and approximate competitive ratio. These performance guarantees can even be extended to the "adaptive control" setting, where the controller does not know the system dynamics ahead of time and must perform online system identification.

We present numerical experiments in a linear dynamical system which demonstrate how the performance of regret-optimal controllers varies as a function of the complexity of the disturbance. We extend regret-optimal control to nonlinear dynamical systems using model-predictive control (MPC) and present experiments which suggest that regret-optimal control is a promising approach to adapting to model error in nonlinear control.

PUBLISHED CONTENT AND CONTRIBUTIONS

[GH22a] Gautam Goel and Babak Hassibi. "Competitive Control". In: *IEEE Transactions on Automatic Control* (2022). URL: https://arxiv.org/abs/2107.13657.
Accepted for publication. G.G proved the results of the paper and wrote the manuscript.

[GH22b] Gautam Goel and Babak Hassibi. "Online Estimation and Control with Optimal Pathlength Regret". In: *Learning for Dynamics and Control* (2022). URL: https://arxiv.org/abs/2110.12544.
Accepted for publication. G.G proved the results of the paper and wrote the manuscript.

[GH22c] Gautam Goel and Babak Hassibi. "Regret-Optimal Estimation and Control". In: *IEEE Transactions on Automatic Control, Special Issue on Learning and Control* (2022). URL: https://arxiv.org/abs/2106.12097.
Accepted for publication. G.G proved the results of the paper and wrote the manuscript.

[GH22d] Gautam Goel and Babak Hassibi. "The Power of Linear Controllers in LQR Control". In: *Conference on Decision and Control* (2022). URL: https://arxiv.org/abs/2002.02574.
Under consideration. G.G proved the results of the paper and wrote the manuscript.

[Goe+22] Gautam Goel et al. "Best of Both Worlds: Competitive Ratio and Policy Regret in Online Control". In: *Neural Information Processing Systems* (2022). URL: https://arxiv.org/abs/2106.12097.
Under consideration. G.G helped proved the main result of the paper and contributed substantially to the writing and simulations.

TABLE OF CONTENTS

Acknowledgements iv
Abstract v
Published Content and Contributions vii
Table of Contents vii
List of Illustrations ix
Chapter I: Introduction 1
1.1 Optimal data-dependent regret through H_∞ control 3
1.2 Preliminaries 5
Chapter II: The Optimal Noncausal Controller 14
2.1 An operator-theoretic model of the optimal noncausal controller . . . 14
2.2 A factorization of the offline optimal cost 15
2.3 A state-space model of the optimal noncausal controller 16
2.4 An $\Omega(T)$ lower bound on dynamic regret 18
Chapter III: Regret-Optimal Full-Information Control 22
3.1 Competitive Control 22
3.2 Energy-Optimal Control 31
3.3 Pathlength-Optimal Control 35
Chapter IV: Regret-Optimal Measurement-Feedback Control 41
4.1 Non-existence results 42
4.2 Regret bounded by the joint energy of w and v 45
4.3 Regret bounded by the pathlength of w and the energy of v 49
Chapter V: Connections to Online Learning 55
5.1 Approximation of the competitive controller by DAC policies 56
5.2 Best of both worlds: sublinear policy regret implies approximate competitive ratio 60
Chapter VI: Numerical Experiments 65
6.1 Double Integrator 65
6.2 Inverted Pendulum 71
Chapter VII: Conclusion and Future Work 78
Bibliography 80
Appendix A: Some Useful Lemmas 83
Appendix B: H_∞ control 86
B.1 Full-Information H_∞ control 86
B.2 Measurement-Feedback H_∞ control 88

LIST OF ILLUSTRATIONS

Number *Page*

6.1 Frequency responses of causal controllers in the double integrator system. 66
6.2 Relative performance of causal controllers in the double integrator system driven by an i.i.d. Gaussian disturbance. 68
6.3 Relative performance of the competitive and H_2 controllers in the double integrator system driven by an i.i.d. Gaussian disturbance. . . 69
6.4 Relative performance of causal controllers in the double integrator system driven by a constant disturbance. 70
6.5 Relative performance of controllers in the double integrator system driven by a sinusoidal disturbance with frequency π (log scale). . . . 70
6.6 Relative performance of the competitive and H_2 controllers in the double integrator system driven by a Gaussian random walk. 71
6.7 Relative performance of causal controllers in an inverted pendulum system driven by an i.i.d. Gaussian disturbance. 73
6.8 Relative performance of causal controllers in an inverted pendulum system driven by a constant disturbance. 74
6.9 Relative performance of causal controllers in an inverted pendulum system driven by sinusoidal noise with frequency 0.01π. 75
6.10 Relative performance of causal controllers in an inverted pendulum system driven by sinusoidal noise with frequency 0.1π (log scale). . . 75
6.11 Relative performance of causal controllers in an inverted pendulum system driven by sinusoidal noise with frequency π (log scale). . . . 76
6.12 Relative performance of causal controllers in an inverted pendulum system driven by a Gaussian random walk (log scale). 76

Chapter 1

INTRODUCTION

Optimal control models the regulation of a dynamical system as an optimization problem, where the goal of the controller is to steer the evolution of the system to minimize cost. Suppose the dynamical system has state x which evolves in discrete-time according to the evolution equation

$$x_{t+1} = \mathcal{G}(x_t, u_t, w_t),$$

where the time index t ranges over a (possibly infinite) time horizon and u_t and w_t are a control input and an exogenous disturbance, respectively. The function $\mathcal{G}$ determines how a new state x_{t+1} is produced from the previous state x_t and the inputs u_t and w_t; it is generally assumed that $\mathcal{G}$ is continuous and differentiable. In each timestep, the controller incurs a loss which varies according to the desirability of the current state and control action. The goal of the controller is to minimize its cumulative loss over the entire time horizon by dynamically selecting the control inputs in response to the disturbances so as to steer the system onto a loss-minimizing trajectory. The controller is constrained to be *causal*, i.e., it must select control actions sequentially using only its observations of previous disturbances and without knowing which disturbances it will encounter in the future. The problem of designing the controller to minimize cost is ill-posed without imposing some assumption about how the disturbance sequence is generated; it is precisely this generative assumption which differentiates different branches of optimal control.

The H_2 control paradigm, first proposed by Kalman in [Kal+60], poses control as a stochastic optimization problem. The controller posits that the disturbances are stochastic and drawn i.i.d. from a fixed distribution. In each timestep, the controller picks the control action which minimizes the expected future losses under this assumption.

The H_∞ control paradigm, first proposed by George Zames in [Zam81], poses control as a minimax game played between the controller and an adversary which selects the disturbance sequence. The controller posits that the goal of the adversary is to maximize the ratio between the cumulative loss incurred by the controller across all timesteps and the energy of the disturbance sequence. In each timestep,

the controller picks the control action which minimizes the future cumulative loss under this assumption. This formulation of control has the natural interpretation of ensuring robustness of the resulting control policy; intuitively, this policy ensures that only large disturbances can force the controller to incur large losses. The H_∞ approach to control has rich connections to risk-sensitive control [Whi81] and dynamic games [BB08].

Both H_2 and H_∞ control suffer from an obvious drawback: both paradigms presuppose a specific model of how the disturbance is generated. If an H_2 or H_∞ controller encounters a disturbance which is different from the type of of disturbance it was designed to handle, then neither paradigm can ensure the optimality of the resulting trajectory. In fact, the performance of the closed-loop system can be extremely poor, as was shown by Doyle in [Doy78].

Motivated by this observation, this thesis proposes a radically different approach to control, which we call 'regret-optimal control.' The regret-optimal control paradigm is inspired by online learning, which models sequential decision-making through the lens of *regret minimization*. In this framework, an online decision-maker seeks to make decisions which are almost as good as the optimal decisions chosen with the benefit of hindsight. The difference between the losses incurred by the online decision-maker and the optimal losses in hindsight is called the *regret*; by minimizing regret, the online decision-maker hopes to achieve near-optimal cost no matter which sequence of losses it encounters.

In the regret-optimal control framework, we design causal controllers so as to approximate a hypothetical "optimal noncausal" controller which selects the globally cost-minimizing sequence of control actions in hindsight, given perfect knowledge of the disturbances. We thus shift our focus from minimizing the costs incurred by the causal controller to instead minimizing the gap in performance between the causal controller and this hypothetical offline controller. Since the cost incurred by the offline controller is a lower bound on the cost achievable by any controller, causal or noncausal, any controller with low regret is thus guaranteed to perform almost as well as any other controller, including an H_2 or H_∞ controller. It is this universality which makes our regret-optimal control paradigm an attractive alternative to standard approaches to optimal control.

The regret-optimal control problem we study is harder than those usually studied in online learning in two ways. First, we seek controllers which minimize regret in the infinite-dimensional, non-parametric space of *all possible causal controllers*.

This space includes controllers which select control actions as a nonlinear function of the disturbance sequence; we are not content to merely optimize regret over parametric families of linear controllers, such as linear state-feedback policies. This is in stark contrast to most work in online learning, where both the online policy and the comparator policy usually belong to a finite-dimensional, parametric class. Second, the system dynamics serve to couple the losses incurred by the controller across rounds; in each round, the action selected by the controller affects the state of the system and hence affects all future losses experienced by the controller. The controller must thus anticipate how its decisions will affect the future evolution of the system. The controller must also take into account how the disturbances selected by the adversary will propagate through the dynamics in future rounds. This is very different from classical online learning problems like multi-armed bandits (MAB) and online convex optimization (OCO), where a suboptimal decision in a single round might result in a low reward in that round but does not affect the ability of the online algorithm to collect future rewards. We refer to [Haz19] and [LS20] for more background on online learning and bandits.

1.1 Optimal data-dependent regret through H_∞ control

The online learning community has traditionally focused on obtaining online algorithms whose regret is bounded *uniformly* over bounded loss sequences and has sublinear dependence on the time horizon T. For example, the well-known Exp3 algorithm for MAB attains $O(\sqrt{KT}\ln K)$ regret for any bounded sequence of losses over K arms [Aue+02; Haz19], while the Online Gradient Descent Algorithm (OGD) for OCO attains $O(\sqrt{T})$ regret for general convex costs [Zin03] and $O(\ln T)$ regret when the costs are strongly convex [HAK07]. Such uniform bounds imply that the time-averaged losses of the online algorithm converge asymptotically to the time-averaged losses of the comparator, irrespective of how the loss sequence is generated. Online algorithms which such sublinear uniform regret bounds are sometimes call "no-regret" algorithms, because their time-averaged regret is guaranteed to converge to zero as T tends to infinity. Recent work [Aga+19; Sim20; FS20; CH21; AGL22] has described online control algorithms with this no-regret property; however, in all of these works the comparator policy is chosen to be a member of some finite-dimensional, parametric class, which is a much weaker comparator than the optimal noncausal controller we consider. We show in Chapter 2 that no online control algorithm can guarantee sublinear regret against this policy for all bounded disturbance sequences.

We thus seek online control algorithms with optimal *data-dependent* regret against the optimal noncausal controller. A data-dependent regret bound is a regret bound which depends on actual instance encountered by the algorithm, and is usually stated in terms of some measure of the "complexity" of the instance. One example of such a complexity measure is *pathlength*, which measures how much the losses vary over time. Intuitively, learning should be easier when the losses vary only slowly over time, because the losses observed by the learner in the past are predictive of the losses the learner can expect to encounter in the future.

The key insight of this thesis is that controllers with optimal data-dependent regret can be derived using a reduction to H_∞ control. It somewhat surprising that we are able to establish a connection between regret minimization and H_∞ control; these two optimization frameworks have been developed independently over several decades by two different communities with two very different sets of goals. We briefly sketch the main idea of this connection here, and present our reductions in detail in Chapters 3 and 4.

Suppose we would like to obtain a causal policy π whose regret against the optimal noncausal policy π_0 is bounded by some complexity measure C of the disturbance. In other words, we would like the inequality

$$J(\pi, w) - J(\pi_0, w) \leq \gamma^2 \cdot C(w) \tag{1.1}$$

to hold for all disturbances $w \in \ell_2$, where γ is a scaling parameter which we would like to be as small as possible. Rearranging, we see that this condition is equivalent to

$$\frac{J(\pi, w)}{\gamma^2 \cdot C(w) + J(\pi_0, w)} \leq 1.$$

Since this inequality is supposed to hold for all $w \in \ell_2$, it suffices to establish the inequality for worst-case disturbances:

$$\sup_{w \in \ell_2} \frac{J(\pi, w)}{\gamma^2 \cdot C(w) + J(\pi_0, w)} \leq 1.$$

Suppose we could construct a synthetic disturbance $\hat{w}$ and a synthetic dynamical system driven by $\hat{w}$ with the following two properties. First, suppose the cost incurred by π in the original system in response to w is equal to the cost incurred by π in the synthetic system in response to $\hat{w}$, so that $J(\pi, w) = J(\pi, \hat{w})$. Second, suppose that

$$\|\hat{w}\|^2 = \gamma^2 \cdot C(w) + J(\pi_0, w).$$

The inequality (1.1) becomes

$$\sup_{\hat{w}\in\ell_2} \frac{J(\pi, \hat{w})}{\|\hat{w}\|^2} \leq 1.$$

This is an H_∞ condition; its has the interpretation of ensuring that the cost incurred by the controller π in the synthetic system is smaller than the energy of the synthetic disturbance $\hat{w}$. A central contribution of this thesis is to demonstrate how to construct the synthetic disturbance and the synthetic system, for several different complexity measures C. One particularly interesting choice of complexity measure we consider is $C(w) = J(\pi_0, w)$; in this case the bound (1.1) can be rearranged to give a bound on the *competitive ratio*, which is the worst-case ratio of the cost incurred by the causal policy π to the cost incurred by the optimal noncausal policy π_0:

$$\sup_{w\in\ell_2} \frac{J(\pi, w)}{J(\pi_0, w)}.$$

1.2 Preliminaries

Linear-Quadratic Control

We restrict our attention to discrete-time linear-quadratic (LQ) control over a doubly-infinite horizon. In this setting, the dynamics are given by the linear evolution equation

$$x_{t+1} = Ax_t + B_u u_t + B_w w_t. \tag{1.2}$$

Here $x_t \in \mathbb{R}^n$ is a state variable we seek to regulate, $u_t \in \mathbb{R}^m$ is a control variable which we can dynamically adjust to influence the evolution of the system, and $w_t \in \mathbb{R}^p$ is an exogenous *driving disturbance*. We formulate control as an online optimization problem, where the goal of the controller is to select the control actions so as to minimize the quadratic cost

$$\sum_{t=-\infty}^{\infty} \left(x_t^\top Q x_t + u_t^\top R u_t\right), \tag{1.3}$$

where $Q \succeq 0, R \succ 0$. We assume that the dynamics $A \in \mathbb{R}^{n\times n}, B_u \in \mathbb{R}^{n\times m}, B_w \in \mathbb{R}^{n\times p}$ and costs $Q \in \mathbb{R}^{n\times n}, R \in \mathbb{R}^{m\times m}$ are known to the controller, so the only uncertainty in the evolution of the system comes from the external disturbance w. For notational convenience, we assume throughout this thesis that the system is parameterized such that $R = I_m$; we emphasize that this imposes no real restriction, since we can always rescale the control input u to ensure that $R = I_m$.

Observation models: Full-Information and Measurement-Feedback

Control can be studied in several distinct settings, each of which posits a different model for the process by which the controller makes observations and decisions. In full-information control, the controller is able to directly observe the state x_t in each timestep. In addition, we say that a controller is *causal* if in each timestep it is able to observe all previous disturbances up to and including the disturbance at the current timestep, e.g., $u = \pi(\ldots, w_{t-1}, w_t)$ for some function π. Similarly, a controller is *strictly causal* if in each timestep it is able to observe all previous disturbances up to but not including the current timestep, e.g., $u = \pi(\ldots, w_{t-1})$. A controller which is not causal is called noncausal; in particular, the control action it selects at time t might depend on some w_s where $s > t$. In computer science, it is common to refer to algorithms which make decisions sequentially as new data arrives as *online algorithms*; we use the words 'causal' and 'online' interchangeably.

Given a control policy π and a disturbance w, we define $J(\pi, w)$ to be the cost that π incurs on the instance w. More formally, $J(\pi, w)$ is given by

$$\begin{aligned} J(\pi, w) = \quad & \sum_{t=-\infty}^{\infty} \left(x_t^\top Q x_t + u_t^\top u_t\right) \\ & \text{where} \quad x_{t+1} = Ax_t + Bu_t + w_t, \\ & \qquad\quad u = \pi(w). \end{aligned}$$

While most of our results are for infinite-horizon control, we also present a few results which describe control over a finite-horizon $\{0, \ldots, T\}$. In this setting a causal control policy is a causal function π which maps the disturbance sequence $w = (w_0, \ldots, w_T)$ to a control signal $u = (u_0, \ldots, w_T)$; a strictly causal control policy is defined analogously. In our finite-horizon results we always assume the initialization $x_0 = 0$ for simplicity. We define the cost incurred by the policy π to be

$$\begin{aligned} J_T(\pi, w) = \quad & \sum_{t=0}^{T} \left(x_t^\top Q x_t + u_t^\top u_t\right) \\ & \text{where} \quad x_{t+1} = Ax_t + Bu_t + w_t, \\ & \qquad\quad x_0 = 0, \\ & \qquad\quad u = \pi(w). \end{aligned}$$

Notice that this definition only includes the costs incurred by π up to time T. In some situations it is useful to measure the additional costs that π would incur were it to keep driving the state to zero after time T, when the disturbances have ceased

to perturb the system. Through a slight abuse of notation, we define the infinite-horizon cost incurred by a policy π_0 in response to a finite disturbance w to be $J(\pi, w)$, where w is understood to have been padded with leading and trailing zeros to obtain a doubly-infinite disturbance sequence.

In measurement-feedback control, the controller is able to directly observe neither the state nor the disturbance. Instead, in each timestep the controller has access to the noisy observation

$$y_t = Cx_t + v_t,$$

where $C \in \mathbb{R}^{r \times n}$ and $v_t \in \mathbb{R}^r$ is a *measurement disturbance*. We emphasize that measurement-feedback control is generally much more challenging than full-information control; for example, the observation y_t will not contain much information about the state x_t if $r \ll n$.

A causal measurement-feedback control policy π is a causal function which maps the observations $y = (\ldots, y_{-1}, y_0, y_1, \ldots)$ to a control signal $u = (\ldots, u_{-1}, u_0, u_1, \ldots)$. We emphasize that the observations depend on π, because they are a function of the state, which itself is generated by the disturbance w and the control u. Given a measurement-feedback policy π, a driving disturbance w, and a measurement disturbance v, we define $J(\pi, w, v)$ to be the cost that π incurs on the instance (w, v). More formally, $J(\pi, w, v)$ is given by

$$\begin{aligned} J(\pi, w, v) = \quad & \sum_{t=-\infty}^{\infty} \left(x_t^\top Q x_t + u_t^\top u_t\right) \\ & \text{where} \quad x_{t+1} = Ax_t + Bu_t + w_t, \\ & \qquad\quad y_t = Cx_t + v_t, \\ & \qquad\quad u = \pi(y). \end{aligned}$$

Policy classes

A central goal of this thesis is to find controllers which minimize regret in the infinite-dimensional class of *all possible causal controllers*, including controllers which select control actions as a nonlinear function of the disturbance sequence. This is very different from the approach which is usually taken in online learning, where the goal is to minimize regret with respect to some parametric class of policies Π; this narrower notion of regret is usually called *policy regret*. In Chapter 5 we derive "best-of-both-worlds" controllers which are simultaneously near-optimal relative to the infinite-dimensional class of all causal controllers, and are near-optimal (in a sharper sense) to specific classes of parametric control policies. We consider two

parametric classes, both of which have attracted much recent attention in the online learning community:

1. The first class we consider is the class of stabilizing linear state-feedback policies, which we denote by $\mathcal{K}$; both the strictly causal H_2 controller and the strictly causal H_∞ controller belong to this class. Each policy in this class is parameterized by a matrix $K \in \mathbb{R}^{m \times n}$. In each timestep, π selects the control action $u_t = Kx_t$; we say that K is *stabilizing* if $\rho(A + BK) < 1$. This condition implies that a unit impulse disturbance applied to the closed-loop system at time $t = 0$ will dissipate as $t \to \infty$. A non-asymptotic form of stability called *strong stability* was introduced in [Coh+18] and is defined as follows. Suppose K is stabilizing; then there exist matrices S, L such that $A + BK = SLS^{-1}$ and $\|L\| < 1$. We say that K is (κ, δ)-*strongly stable* if

$$\|S\|\|S^{-1}\| \leq \kappa$$

and

$$\|L\| \leq 1 - \delta.$$

Notice that with this definition, powers of $A + BK$ obey a simple bound:

$$\|(A + BK)^i\| \leq \kappa(1 - \delta)^i.$$

2. The second class we consider is the class of *disturbance-action controller* (DAC) policies, which we denote by $\mathcal{M}$. Each policy π in this class is parameterized by a stabilizing controller K, a horizon H and weights $M = \left(M^{[0]}, \ldots M^{[H-1]}\right)$. In each timestep, π chooses the control action

$$u_t = Kx_t + \sum_{t=1}^{H} M^{[i-1]} w_{t-i}.$$

It is immediately clear that every linear state-feedback policy is a DAC policy with all H weights set to zero. We define a (H, θ, δ)-DAC policy class to be the set of all H-horizon DAC policies where the weights satisfy the geometric decay condition $\|M^{[i]}\| \leq \theta(1-\delta)^i$. Intuitively, this condition implies that the control action in each timestep is not greatly affected by the disturbances which were observed by the controller many timesteps in the past. We additionally require that K is (κ, δ)-stabilizing, for some $\kappa > 0$.

Input-output approach to control

As is standard in the input-output approach to control, we encode controllers as linear *transfer operators* mapping the disturbances to the quadratic cost we wish to minimize. Recall that an operator F is *bounded* if it maps ℓ_2-bounded sequences to ℓ_2-bounded sequences. Let x be some signal and let $y = Fx$. We say that F is *causal* if y_t depends only on $(\dots, x_{t-1}, x_t)$ and *strictly causal* if y_t depends only on $(\dots, x_{t-1})$. It is easy to check that the product of two causal operators is causal, and the product of a causal and a strictly causal operator is strictly causal.

Let $L = Q^{1/2}$ and let $s_t = Lx_t$. With this notation, the quadratic costs (1.3) can be written in a very simple form:

$$\|s\|_2^2 + \|u\|_2^2.$$

The dynamics (1.2) can be written as

$$s = Fu + Gw,$$

where F and G are strictly causal operators depending on A, B_u, B_w, L. Over a finite-horizon $t = 0, \dots, T$, the operators F and G can be explicitly written as block Toeplitz matrices:

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & \\ LB_u & 0 & 0 & 0 & & \\ LAB_u & LB_u & 0 & 0 & & \\ LA^2B_u & LAB_u & LB_u & 0 & & \\ \vdots & & & & & \ddots \end{bmatrix},$$

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & \\ LB_w & 0 & 0 & 0 & & \\ LAB_w & LB_w & 0 & 0 & & \\ LA^2B_w & LAB_w & LB_w & 0 & & \\ \vdots & & & & & \ddots \end{bmatrix}.$$

In the infinite-horizon setting, we can think of F and G as doubly-infinite block Toeplitz matrices which map the infinite sequences u and w to s. In the z-domain, F and G can be compactly expressed as

$$F(z) = L(zI - A)^{-1}B_u, \quad G(z) = L(zI - A)^{-1}B_w.$$

Full-information controllers which are linear in the disturbance (e.g., $u = Kw$ for some causal operator K) are associated with a linear transfer operator T_K which maps the driving disturbance w to s and u:

$$T_K = \begin{bmatrix} FK + G \\ K \end{bmatrix}.$$

The cost incurred by the controller K is simply

$$w^* T_K^* T_K w.$$

In the measurement-feedback setting, the controller does not directly observe x_t and w_t, but instead only observes $y_t = Cx_t + v_t$. This observation model can be captured by the relation $y = Hu + Jw + v$, where H and J are strictly causal operators depending on A, B_u, B_w, C. Over a finite-horizon $t = 0, \ldots, T$, the operators H and J can be explicitly written as block Toeplitz matrices:

$$H = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ CB_u & 0 & 0 & 0 & \\ CAB_u & CB_u & 0 & 0 & \\ CA^2B_u & CAB_u & CB_u & 0 & \\ \vdots & & & & \ddots \end{bmatrix},$$

$$J = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ CB_w & 0 & 0 & 0 & \\ CAB_w & CB_w & 0 & 0 & \\ CA^2B_w & CAB_w & CB_w & 0 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

In the infinite-horizon setting, we can think of H and J as doubly-infinite block Toeplitz matrices which map the infinite sequences u and w to y. In the z-domain, H and J can be compactly expressed as

$$H(z) = C(zI - A)^{-1}B_u, \quad J(z) = C(zI - A)^{-1}B_w.$$

Measurement-feedback controllers which are linear in the observations (e.g., $u = Ky$ for some causal operator K) are associated with a linear transfer operator T_K which maps the driving disturbance w and the measurement disturbance v to s and u:

$$T_K = \begin{bmatrix} FQJ + G & FQ \\ QJ & Q \end{bmatrix},$$

where we define the Youla parameter $Q = K(I - HK)^{-1}$. Note that we can easily recover K from Q by setting $K = (I + QH)^{-1}Q$. The cost incurred by the controller K is

$$\begin{bmatrix} w \\ v \end{bmatrix}^* T_K^* T_K \begin{bmatrix} w \\ v \end{bmatrix}.$$

We refer the reader to [HSK99] for more background on the input-output approach to control.

Performance metrics

The goal of this thesis is to derive causal controllers which approximate the performance of an optimal noncausal controller. The *optimal noncausal controller* (sometimes called the offline optimal controller) selects the control actions in each timestep with access to the full disturbance sequence w so as to minimize the cost (1.3). The cost incurred by the optimal noncausal controller is called the *offline optimal cost*; it is a lower bound on the cost achievable by any controller, causal or noncausal. We describe the optimal noncausal controller in great detail in Chapter 2. Throughout this thesis, we denote the optimal noncausal controller by π_0.

There are two standard performance metrics which compare the performance of an online algorithm relative to the performance of an offline optimal algorithm: *regret* and *competitive ratio*. The regret of a policy π on the disturbance w is simply the difference in the cost incurred by online policy π and the cost which could have been achieved by an optimal noncausal policy π_0:

$$\textsc{Regret}(w) = J(\pi, w) - J(\pi_0, w).$$

This notion of regret is sometimes called *dynamic regret* to emphasize that the comparator policy π_0 is unconstrained and may vary over time; it is also sometimes called the *competitive difference*. The competitive ratio is the worst-case ratio in costs:

$$\textsc{Competitive Ratio} = \sup_{w \in \ell_2} \frac{J(\pi, w)}{J(\pi_0, w)}.$$

We emphasize that regret is a function of the input w, whereas the competitive ratio bounds the ratio in costs over all inputs w. Our goal when studying regret is usually to bound the regret of a policy π in terms of some *data-dependent* quantity associated with the disturbance w. Some natural choices are the energy of w, which is a measure of how "large" w is, and the pathlength of w, which measures how much w varies over time.

A somewhat weaker notion of regret is *policy regret*, which is the difference in the cost incurred by online policy π and the cost which could have been counterfactually achieved by the best policy selected in hindsight from some class of polices Π:

$$\textsc{Policy Regret}(w) = J(\pi, w) - \min_{\pi \in \Pi} J(\pi, w).$$

The key distinction between dynamic regret and policy regret is that policy regret restricts the comparator policy to lie in some specific class Π, whereas in dynamic regret the comparator is the optimal noncausal policy. Since the cost incurred by the optimal noncausal policy is a lower bound on the cost incurred by any policy, including all of the policies contained in Π, the following inequality holds for all disturbances w:

$$\textsc{Policy Regret}(w) \leq \textsc{Regret}(w).$$

The definitions of regret and competitive ratio can be easily extended to the measurement-feedback setting by including the measurement disturbance v:

$$\textsc{Regret}(w, v) = J(\pi, w, v) - J(\pi_0, w)$$

and

$$\textsc{Competitive Ratio} = \sup_{w, v \in \ell_2} \frac{J(\pi, w, v)}{J(\pi_0, w)}.$$

We refer to [BE05] for background on competitive analysis.

Notation and Terminology

We define the *energy* of a p-dimensional signal w to be squared ℓ_2 norm of w:

$$\|w\|_2^2 = \sum_{t=-\infty}^{\infty} \|w_t\|_2^2.$$

We define the *pathlength* of w to be

$$\sum_{t=-\infty}^{\infty} \|w_t - w_{t-1}\|_2^2.$$

We define the derivative operator D_p to be the linear operator which maps a sequence $w = (\ldots, w_{-1}, w_0, w_1, \ldots)$ to its discrete derivative $(\ldots, w_{-1} - w_{-2}, w_0 - w_{-1}, w_1 - w_0, \ldots)$. The pathlength of w can hence be compactly represented as $\|D_p w\|_2^2$. We note that D_p has the z-domain representation

$$D_p = (1 - z^{-1}) I_p.$$

We use I_n to denote the $n \times n$ identity matrix. We let $\bar{\sigma}(M)$ denote the largest singular value of a matrix M and let $\bar{\lambda}(A)$ denote the largest eigenvalue of a square matrix A. We denote the spectral radius of a square matrix A by $\rho(A)$; we say that A is *stable* if $\rho(A) < 1$; otherwise A is *unstable*. We define $\|v\|$ to be the standard ℓ_2 norm of a finite-dimensional vector v; similarly, we define $\|M\|$ to be the ℓ_2-induced operator norm (the spectral norm) of a finite-dimensional matrix M. Given a disturbance sequence w, either finite or infinite, we let $\|w\|_\infty = \sup_t \|w_t\|$. We let $\|F\|$ denote the H_∞ norm of a transfer operator F; the H_∞ norm can be viewed as an analog of the ℓ_2-induced operator norm for infinite-dimensional operators.

Suppose $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$. The pair (A, B) is *controllable* if

$$\operatorname{rank}\left(\begin{bmatrix} \lambda I - A & B \end{bmatrix}\right) = n$$

for every eigenvalue λ of A. The pair (A, B) is *stabilizable* if

$$\operatorname{rank}\left(\begin{bmatrix} \lambda I - A & B \end{bmatrix}\right) = n$$

for every unstable eigenvalue λ of A. The pair (A, B) is *unit-circle controllable* if

$$\operatorname{rank}\left(\begin{bmatrix} \lambda I - A & B \end{bmatrix}\right) = n$$

for every unit-circle eigenvalue λ of A. Let $C \in \mathbb{R}^{r \times n}$. The pair (A, C) is *observable* if and only if (A^*, C^*) is controllable. The pair (A, C) is *detectable* if and only if (A^*, C^*) is stabilizable. The pair (A, C) is *unit-circle observable* if and only if (A^*, C^*) is unit-circle controllable.

Chapter 2

THE OPTIMAL NONCAUSAL CONTROLLER

The goal of this thesis is to derive online controllers which approximate the performance of the optimal noncausal controller as closely as possible. We give two equivalent descriptions of the optimal noncausal controller, one in terms of transfer operators and one in state-space form. We also show that no causal controller can attain sublinear regret relative to the optimal noncausal controller.

2.1 An operator-theoretic model of the optimal noncausal controller

Recall that $F(z)$ and $G(z)$ are the transfer operators mapping $u(z)$ and $w(z)$ to $s(z)$, respectively:

$$F(z) = L(zI - A)^{-1}B_u, \quad G(z) = L(zI - A)^{-1}B_w.$$

Both the optimal noncausal controller and the offline optimal cost have well-known expressions in terms of F and G:

Theorem 1 (Theorem 11.2.1 in [HSK99])**.** *The optimal noncausal controller is*

$$\pi_0(w) = -(I + F^*F)^{-1}F^*Gw$$

and the offline optimal cost is

$$J(\pi_0, w) = w^*G^*(I + FF^*)^{-1}Gw.$$

Proof. Recall that the cost incurred by a controller which selects the control sequence u in response to the disturbance sequence w is

$$\|Fu + Gw\|^2 + \|u\|^2.$$

Completing the square, this cost can be rewritten as

$$w^*G^*(I+FF^*)^{-1}Gw + (u+(I+F^*F)^{-1}F^*Gw)^*(I+F^*F)(u+(I+F^*F)^{-1}F^*Gw).$$

Notice that both of these terms are non-negative. The first term clearly does not depend on u, whereas the second term can be set to zero by setting

$$u = -(I + F^*F)^{-1}F^*Gw.$$

It is hence clear that this is the cost-minimizing choice of u, and

$$w^* G^* (I + FF^*)^{-1} Gw$$

is the minimal achievable cost. □

2.2 A factorization of the offline optimal cost

Recall that the offline optimal cost is

$$J(\pi_0, w) = w^* G^* (I + FF^*)^{-1} Gw.$$

Let $\Delta(z)$ be the unique causal and causally invertible operator such that

$$I + FF^* = \Delta\Delta^*.$$

We note that the existence and uniqueness of Δ follows from the positive-definiteness of $I + FF^*$. The offline cost can be rewritten in terms of Δ as

$$J(\pi_0, w) = \|\Delta^{-1} Gw\|^2.$$

This description of the offline optimal cost plays a central role in our derivation of regret-optimal controllers. We can easily derive $\Delta(z)$ in closed form:

Theorem 2. *The following canonical factorization holds:*

$$I + F(z)F(z^{-*})^* = \Delta(z)\Delta^*(z^{-*}),$$

where we define

$$\Delta(z) = (I + L(zI - A)^{-1}K)\Sigma^{1/2}, \tag{2.1}$$
$$K = APL\Sigma^{-1}, \quad \Sigma = I + LPL^*,$$

and P is the unique Hermitian solution to the Riccati equation

$$P = B_u B_u^* + APA^* - APL(I + LPL^*)^{-1} LPA^*.$$

Proof. We expand $I + F(z)F(z^{-*})^*$ as

$$\begin{bmatrix} L(zI - A)^{-1} & I \end{bmatrix} \begin{bmatrix} B_u B_u^* & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} (z^{-*} I - A)^{-*} L^* \\ I \end{bmatrix}.$$

Applying Lemma 2, we see that this equals

$$\begin{bmatrix} L(zI - A)^{-1} & I \end{bmatrix} \Lambda(P) \begin{bmatrix} (z^{-*} I - A)^{-*} L^* \\ I \end{bmatrix},$$

where P is an arbitrary Hermitian matrix and we define

$$\Lambda(P) = \begin{bmatrix} B_u B_u^* - P + APA^* & APL^* \\ LPA^* & I + LPL^* \end{bmatrix}.$$

Notice that the $\Lambda(P)$ can be factored as

$$\begin{bmatrix} I & K(P) \\ 0 & I \end{bmatrix} \begin{bmatrix} \Gamma(P) & 0 \\ 0 & \Sigma(P) \end{bmatrix} \begin{bmatrix} I & 0 \\ K^*(P) & I \end{bmatrix},$$

where we define

$$\Gamma(P) = B_u B_u^* - P + APA^* - K(P)\Sigma(P)K^*(P),$$

$$K(P) = APL\Sigma(P)^{-1},$$

$$\Sigma(P) = I + LPL^*.$$

By assumption, (A, B_u) is stabilizable and (A, L) is detectable, therefore the Riccati equation $\Gamma(P) = 0$ has a unique stabilizing solution (Theorem E.6.2 in [KSH00]). Suppose P is chosen to be this solution, and define $K = K(P)$, $\Sigma = \Sigma(P)$. We immediately obtain the canonical factorization

$$I + F(z)F(z^{-*})^* = \Delta(z)\Delta^*(z^{-*}),$$

where we define

$$\Delta(z) = (I + L(zI - A)^{-1}K)\Sigma^{1/2}. \tag{2.2}$$

□

2.3 A state-space model of the optimal noncausal controller

In this section we derive a state-space model of the optimal noncausal controller over a finite horizon $1, \ldots, T$; this gives a computationally efficient way to compute the offline optimal control actions, which we use extensively in the numerical experiments presented in Chapter 6. The technical machinery we develop here also plays a crucial role in our proof that no online algorithm can achieve sublinear regret against the optimal noncausal controller (Theorem 4).

Given a noise sequence $w = (w_0, \ldots w_T)$, the optimal noncausal controller selects the control actions which minimizes the cumulative cost from $t = 0, \ldots T$:

$$\begin{aligned} &\min_{u_0,\ldots,u_T} \quad \sum_{t=0}^{T} x_t^\top Q x_t + u_t^\top u_t \\ &\text{where} \quad x_{t+1} = Ax_t + B_u u_t + B_w w_t. \end{aligned} \tag{2.3}$$

We emphasize that the optimal offline control actions are defined with respect to the actual realizations $w_0, \ldots w_T$; the optimal offline control actions are the optimal actions in hindsight, with full knowledge of the realization w.

We use dynamic programming to recursively compute the optimal control actions, starting from the last time step and moving backwards in time; this approach mirrors the well-known derivation of the Linear Quadratic Regulator. For any fixed disturbance $w = (w_0, \ldots w_T)$, define the "offline cost-to-go" function

$$J_t^w(x) = \min_u [x^\top Q x + u^\top u + J_{t+1}^w(x)(Ax + B_u u + B_w w_t)]$$

for $t = 1.. \ldots T$, where we set $J_{T+1}(x) = 0$. This function measures the aggregate cost over the future time horizon starting at the state x at time t, under the assumption that in each time step, the offline controller picks the control action which minimizes the future cost given the current state and the realizations $w_t \ldots w_T$.

We will show that $J_t^w(x)$ can be written as $x^\top P_t x + v_t^\top x_t + q_t$ for all $t \in \{1 \ldots T+1\}$, where P_t is the solution to the Riccati recurrence which appears in the derivation of the H_2-optimal policy. The claim is clearly true for $t = T + 1$, since we can take $(P_{T+1}, v_{T+1}, q_{T+1}) = (0, 0, 0)$. Proceeding by backwards induction, suppose

$$J_{t+1}^w(x) = x^\top P_{t+1} x + v_{t+1}^\top x + q_{t+1}$$

for some v_{t+1}, q_{t+1}. Then $J_t^w(x)$ is

$$\begin{aligned} J_t^w(x) &= \min_u \left[x^\top Q x + u^\top u + J_{t+1}^w(Ax_t + B_u u_t + B_w w_t)\right] \\ &= \min_u \left[x^\top Q x + u^\top u + (Ax + B_u u + B_w w_t)^\top P_{t+1}(Ax + B_u u + B_w w_t)\right. \\ &\quad \left. + \; v_{t+1}^\top (Ax + B_u u + B_w w_t) + q_{t+1}\right]. \end{aligned}$$

Solving for the minimizing u, we see that the offline optimal control action is

$$u_t = -(I + B_u^\top P B_u)^{-1} B_u^\top \left(P_{t+1} A x_t + P_{t+1} B_w w_t + \frac{1}{2} v_{t+1}\right).$$

We note that

$$-(I + B^\top P B)^{-1} B^\top P_{t+1}(Ax_t + B_w w_t)$$

is precisely the H_2-optimal control action. In other words, the optimal offline control action in timestep t is the sum of the H_2-optimal online control action and a correction term which depends only on the future disturbances $w_t \ldots w_T$; it is this correction term which gives the offline controller its advantage relative over every causal controller.

Plugging this choice of u_t into our expression for $J_t^w(x)$ and collecting terms, we see that $J_t^w(x) = x^\top P_t x + v_t^\top x + q_t$ where P_t, v_t, and q_t satisfy the backwards recurrences

$$\begin{aligned} P_t &= Q + A^\top P_{t+1} A - A^\top P_{t+1} B (I + B_u^\top P_{t+1} B_u)^{-1} B_u^\top P A && (2.4) \\ v_t &= 2A^\top S_t B_w w_t + A^\top S_t P_{t+1}^{-1} v_{t+1}, && (2.5) \\ q_t &= w_t^\top B_w^\top S_{t+1} B_w w_t + v_{t+1}^\top P_{t+1}^{-1} S_t B_w w_t + q_{t+1} - \frac{1}{4} v_{t+1}^\top B_u H_t^{-1} B_u^\top v_{t+1}, && (2.6) \end{aligned}$$

where we define

$$\begin{aligned} S_t &= P_{t+1} - P_{t+1} B_u H_t^{-1} B_u^\top P_{t+1}, && (2.7) \\ H_t &= I + B_u^\top P_{t+1} B_u. && (2.8) \end{aligned}$$

We have proven:

Theorem 3. *The control actions selected by the optimal noncausal optimal controller are given by*

$$u_t = -(I + B_u^\top P_{t+1} B_u)^{-1} B_u^\top \left(P_{t+1} A x_t + P_{t+1} B_w w_t + \frac{1}{2} v_{t+1} \right),$$

where P_t satisfies the backwards recurrence (2.4) *and v_t satisfies the backwards recurrence (2.5), and we initialize $P_{T+1} = 0, v_{T+1} = 0$.*

We note that this theorem parallels various results from the filtering literature, which express the solutions to smoothing problems in terms of the corresponding filtering problems and future observations, e.g., [RTS65].

2.4 An $\Omega(T)$ lower bound on dynamic regret

We now show that no online control algorithm can achieve sublinear regret against the optimal noncausal controller. We prove:

Theorem 4. *Suppose (A, B) is stabilizable and $(A, Q^{1/2})$ is observable on the unit circle, and suppose the disturbances are generated i.i.d from a distribution $\mathcal{D}$ with mean zero and bounded covariance $\Sigma \succ 0$. There exists a constant $c_0 > 0$ depending on (A, B, Q, Σ) such that the regret of any causal controller π satisfies*

$$\lim_{t \to T} \mathop{\mathbb{E}}_{w \sim \mathcal{D}} \left[\frac{J_T(\pi, w) - J_T(\pi_0, w)}{T} \right] \geq c_0. \tag{2.9}$$

Informally, this theorem says that the dynamic regret of any causal controller must grow at rate $\approx c_0 T$, up to lower-order terms.

Proof. The key idea is to focus on the setting where the disturbances are picked i.i.d. from a fixed distribution $\mathcal{D}$ in each round; in this setting the regret-minimizing policy is simply the H_2-optimal policy. To see this, fix any causal controller π and time horizon T. The regret of π is simply

$$\mathop{\mathbb{E}}_{w\sim\mathcal{D}}[J_T(\pi, w) - J_T(\pi_0, w)].$$

By linearity of expectation, this is

$$\mathop{\mathbb{E}}_{w\sim\mathcal{D}}[J_T(\pi, w)] - \mathop{\mathbb{E}}_{w\sim\mathcal{D}}[J_T(\pi_0, w)].$$

It is now clear that the online algorithm which minimizes regret is exactly the one which minimizes the expected cost

$$\mathop{\mathbb{E}}_{w\sim\mathcal{D}}[J_T(\pi, w)].$$

The H_2-optimal controller is the unique causal controller which minimizes this expected cost. It follows that if we can show that the H_2-optimal controller satisfies the lower bound in (2.9) then this will establish the lower bound for all online controllers. It is well-known that the expected time-averaged cost of the H_2-optimal controller converges to $\mathrm{Tr}(B_w \Sigma B_w^\top P)$ as $T \to \infty$, where P is the solution of the Riccati equation

$$P = Q + A^\top P A - A^\top P B_u (I + B_u^\top P B_u)^{-1} B_u^\top P A. \tag{2.10}$$

In Theorem 5 we show that the expected time-averaged cost of the optimal noncausal controller converges to

$$\mathrm{Tr}(B_w \Sigma B_w^\top S) - \frac{1}{4}\mathrm{Tr}(B_u H^{-1} B_u^\top V),$$

as $T \to \infty$, where we define

$$S = P - P B_u H^{-1} B_u^\top P,$$

$$H = I + B_u^\top P B_u,$$

and V is the unique solution to the Lyapunov equation

$$V = 4A^\top S B w \Sigma B_w^\top S A + A^\top S P^{-1} V P^{-1} S A.$$

It is now clear that the expected time-averaged regret of the H_2-optimal controller converges to

$$\mathrm{Tr}(B_w \Sigma B_w^\top P B_u H^{-1} B_u^\top P) + \frac{1}{4}\mathrm{Tr}(B_u H^{-1} B_u^\top V).$$

This quantity is positive since H, Σ, and V are positive-semidefinite. □

The expected cost of the optimal noncausal controller

We now turn to the problem of computing the expected infinite-horizon cost of the optimal noncausal controller under the assumption that the disturbances are generated i.i.d from $\mathcal{D}$. We prove:

Theorem 5. *Suppose (A, B) is stabilizable and $(A, Q^{1/2})$ is observable on the unit circle, and suppose the disturbances are generated i.i.d from a distribution $\mathcal{D}$ with mean zero and bounded covariance $\Sigma \succ 0$. The expected time-averaged cost of the optimal noncausal controller converges to*

$$\operatorname{Tr}(B_w \Sigma B_w^\top S) - \frac{1}{4} \operatorname{Tr}(B_u H^{-1} B_u^\top V) \tag{2.11}$$

as $T \to \infty$, where P is the solution to the algebraic Riccati Equation (2.10),

$$S = P - P B_u H^{-1} B_u^\top P,$$

$$H = I + B_u^\top P B_u,$$

and V is the unique solution to the Lyapunov equation

$$V = 4A^\top S B_w \Sigma B_w^\top S A + A^\top S P^{-1} V P^{-1} S A.$$

Proof. Using the notation we introduced in the proof of Theorem 3, the infinite-horizon cost of the optimal offline policy is

$$\begin{aligned} &\lim_{T\to\infty} \mathop{\mathbb{E}}_{w\sim\mathcal{D}} \left[\frac{1}{T} J_0^w(x_0) \right] \\ &= \lim_{T\to\infty} \frac{1}{T} \mathop{\mathbb{E}}_{w\sim\mathcal{D}} \left[x_0^\top P_0 x_0 + v_0^\top x_0 + q_0 \right]. \end{aligned}$$

Recall that we assumed $x_0 = 0$. Using the recursion for v_t given by (2.5) and the fact that $v_{T+1} = 0$ and $\mathbb{E}[w_t] = 0$ for all $t \in \{0 \dots T\}$, we easily see that $\mathbb{E}[v_t] = 0$ for all $t \in \{0 \dots T\}$. In particular, $\mathbb{E}_w[v_0] = 0$, so all that remains is to calculate $\mathbb{E}[q_0]$. Using the recurrence (2.6) we derived for q_t, we see that

$$\mathbb{E}[q_t] = \operatorname{Tr}(B_w^\top \Sigma S_t B_w) - \frac{1}{4} \operatorname{Tr}(B_u H_t^{-1} B_u^\top V_{t+1}) + \mathop{\mathbb{E}}_{w}[q_{t+1}],$$

where we defined $V_t = \mathbb{E}[v_t v_t^\top]$. Here we used the fact that $\mathbb{E}[v_{t+1}^\top (P_{t+1}^{-1} S_t) w_t] = 0$, since v_{t+1} and w_t are independent and $\mathbb{E}[w_t] = 0$. We see that V_t is given by

$$\begin{aligned} V_t &= \mathbb{E}[v_t v_t^\top] \\ &= 4A^\top S_t B_w \Sigma B_w^\top S_t A + A^\top S_t P_{t+1}^{-1} V_{t+1} P_{t+1}^{-1} S_t A, \end{aligned}$$

where we applied the recurrence (2.5) and observed that the cross-terms vanish by independence of v_{t+1} and w_t and the fact that $\mathbb{E}[w_t] = 0$.

Let us now consider the limiting behavior of V_t as $t \to \infty$. The assumption that (A, B) is stabilizable and $(A, Q^{1/2})$ is observable on the unit circle together imply that P_t converges to P, the solution of the algebraic Riccati equation (2.10), as $t \to \infty$ (Theorem 14.5.1 in [HSK99]). Applying the definition of S_t given in (2.7), we see that S_t converges to

$$S = P - PB_u(I + B_u^\top PB_u)^{-1}B_u^\top P.$$

To determine the convergence of V_t, it suffices to show that $\rho(A^\top SP^{-1}) < 1$ (Lemma D.1.2 in [KSH00]), in which case V_t will converge to the solution of the equation

$$V = 4A^\top SB_w\Sigma B_w^\top SA + A^\top SP^{-1}VP^{-1}SA. \tag{2.12}$$

Notice that

$$\begin{aligned} A^\top SP^{-1} &= A^\top - A^\top PB_u(I + B_u^\top PB_u)^{-1}B_u^\top \\ &= (A + B_uK)^\top, \end{aligned}$$

where K is the strictly causal H_2-optimal controller. The Kalman gain $A + B_uK$ always has spectral radius strictly less than one, establishing the convergence of V_t to the solution of equation (2.12). We see that the infinite-horizon optimal offline cost is

$$\begin{aligned} & \lim_{T\to\infty} \frac{1}{T} \mathop{\mathbb{E}}_{w\sim\mathcal{D}}[J_0^w(x_0)] \\ = & \lim_{T\to\infty} \frac{1}{T} \mathop{\mathbb{E}}_{w\sim\mathcal{D}}[q_0] \\ = & \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \left(\operatorname{Tr}(B_w\Sigma B_w^\top S_t) - \frac{1}{4}\operatorname{Tr}(B_uH_t^{-1}B_u^\top V_{t+1}) \right) \\ = & \operatorname{Tr}(B_w\Sigma B_w^\top S) - \frac{1}{4}\operatorname{Tr}(B_uH^{-1}B_u^\top V). \end{aligned}$$

□

Chapter 3

REGRET-OPTIMAL FULL-INFORMATION CONTROL

The goal of this thesis is to derive causal controllers which approximate the performance of the optimal noncausal controller as closely as possible. There are, however, several different notions of approximation. In *competitive control*, our goal is to approximate the optimal noncausal controller in a multiplicative sense, i.e., to find a causal controller whose cost is always at most a constant more than the cost incurred by the optimal noncausal controller. We can instead choose to approximate the optimal noncausal controller in an additive sense, by bounding the difference between the costs incurred by the causal and noncausal controllers (i.e. the regret). The regret is a function of the disturbance w, so it is natural that our bound on the regret be some function of w as well; we call such a regret bound a "data-dependent" regret bound. In *energy-optimal* control we bound the regret of the online controller by the energy of w, and in *pathlength-optimal* control we bound the regret of the online controller by the pathlength of w. In this chapter, we derive state-space models of the competitive, energy-optimal, and pathlength-optimal controllers in the full-information setting, where the online controller is able to directly observe the state and disturbance sequence when selecting control actions.

3.1 Competitive Control

The competitive control problem is to find an online controller with optimal competitive ratio:

Problem 1 (Competitive control)**.** *Find a causal controller which minimizes the competitive ratio*

$$\sup_{w} \frac{J(\pi, w)}{J(\pi_0, w)}.$$

We call the controller with the smallest possible competitive ratio the *competitive controller*. The competitive approach to control was first proposed in [GW19], who described a competitive algorithm in a narrow class of linear systems using the Online Balanced Descent (OBD) algorithm introduced in [CGW18]; this approach was further explored in [Goe+19; Shi+20]. In this section, we derive the controller with optimal competitive ratio in both LTI systems over an infinite horizon and

in time-varying systems over a finite horizon, without making any non-standard structural assumptions about the system dynamics.

Instead of minimizing the competitive ratio directly, we instead solve the following relaxation:

Problem 2 (Suboptimal competitive control)**.** *Given $\gamma > 0$, find a causal controller such that*

$$\sup_{w} \frac{J(\pi, w)}{J(\pi_0, w)} < \gamma^2$$

for all disturbances w, or determine whether no such controller exists.

We call such a controller the *competitive controller at level γ*. It is clear that if we can solve this suboptimal problem then we can easily recover the competitive controller via bisection on γ.

We derive necessary and sufficient conditions for the existence of a competitive controller at level γ, along with a state-space model of the controller:

Theorem 6. *Suppose (A, B_u) is stabilizable and (A, L) is detectable. Fix $\gamma > 0$ and define $\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{L}$ as in (3.5). A causal controller π such that*

$$\sup_{w \in \ell_2} \frac{J(\pi, w)}{J(\pi_0, w)} < \gamma^2 \tag{3.1}$$

exists if and only if the DARE

$$\widehat{P} = \widehat{L}^* \widehat{L} + \widehat{A}^* \widehat{P} \widehat{A} - \widehat{A}^* \widehat{P} \widetilde{B} \widetilde{H}^{-1} \widetilde{B}^* \widehat{P} \widehat{A},$$

where we define

$$\begin{cases} \widetilde{B} = \begin{bmatrix} \widehat{B}_u & \widehat{B}_w \end{bmatrix}, \\ \widetilde{R} = \begin{bmatrix} I_m & 0 \\ 0 & -\gamma^2 I_n \end{bmatrix}, \\ \widetilde{H} = \widetilde{R} + \widetilde{B}^* P \widetilde{B}, \end{cases}$$

has a solution $\widehat{P}$ such that

1. *$\widehat{A} - \widetilde{B} \widetilde{H}^{-1} \widetilde{B}^* \widehat{P} \widehat{A}$ is stable;*

2. *$\widetilde{R}$ and $\widetilde{H}$ have the same inertia;*

3. *$\widehat{P} \succeq 0$.*

In this case, one possible causal controller satisfying (3.1) is

$$u_t = -(I_{2n} + \widehat{B}_u^* \widehat{P} \widehat{B}_u)^{-1} \widehat{B}_u^* \widehat{P} (\widehat{A}\xi_t + \widehat{B}_w \widehat{w}_{t+1}),$$

where the synthetic state $\xi \in \mathbb{R}^{2n}$ evolves according to the linear dynamics equation

$$\xi_{t+1} = \widehat{A}\xi_t + \widehat{B}_u u_t + \widehat{B}_w \widehat{w}_{t+1}$$

and the synthetic disturbance $\widehat{w}$ is given in (3.9). A strictly causal controller satisfying (3.1) exists if and only if conditions (1) and (3) hold, and additionally

$$\widehat{B}_u^* \widehat{P} \widehat{B}_u \prec \gamma^2 I_m$$

and

$$I_{2n} + \widehat{B}_u^* \widehat{P} (I_{2n} - \gamma^{-2} \widehat{B}_w \widehat{B}_w^* \widehat{P})^{-1} \widehat{B}_u \succ 0.$$

In this case, one possible strictly causal controller satisfying (3.1) is

$$u_t = -(I_m + \widehat{B}_u^* \widetilde{P} \widehat{B}_u)^{-1} \widehat{B}_u^* \widetilde{P} \widehat{A} \xi_t,$$

where we define

$$\widetilde{P} = \widehat{P} - \widehat{P}\widehat{B}_w(-\gamma^2 I_p + \widehat{B}_w^* \widehat{P} \widehat{B}_w)^{-1} \widehat{B}_w^* \widehat{P}.$$

Proof. Recall that the offline optimal cost is

$$J(\pi_0, w) = w^* G^* (I + FF^*)^{-1} G w.$$

We see that condition (3.1) is equivalent to

$$J(\pi, w) < \gamma^2 w^* G^* \left(I + FF^*\right)^{-1} G w.$$

Let $\Delta(z)$ be the unique causal and causally invertible operator such that

$$I + FF^* = \Delta\Delta^*.$$

With this factorization, condition 3.1 becomes the H_∞ condition

$$J(\pi, \widehat{w}) < \gamma^2 \|\widehat{w}\|^2,$$

where the synthetic disturbance $\widehat{w}$ is given by

$$\widehat{w}(z) = \Delta_2^{-1}(z)G(z)w(z)$$

and the system dynamics in the frequency domain are

$$s(z) = F(z)u(z) + \Delta_2(z)\widehat{w}(z). \tag{3.2}$$

In Theorem 2 we found $\Delta(z)$ is given by

$$\Delta(z) = (I + L(zI - A)^{-1}K)\Sigma^{1/2},$$

where we define

$$K = APL\Sigma^{-1}, \quad \Sigma = I + LPL^*,$$

and P is the unique Hermitian solution to the Riccati equation

$$B_u B_u^* - P + APA^* - APL(I + LPL^*)^{-1}LPA^* = 0.$$

With this factorization, we can easily recover the optimal infinite-horizon competitive controller; it is simply the H_∞-optimal infinite-horizon controller in the system whose dynamics in the frequency domain are

$$s(z) = F(z)u(z) + \Delta(z)\widehat{w}(z), \tag{3.3}$$

where the synthetic disturbance $\widehat{w}$ is

$$\widehat{w}(z) = \Delta^{-1}(z)G(z)w(z). \tag{3.4}$$

Define

$$\begin{cases} \widehat{A} = \begin{bmatrix} A & K\Sigma^{1/2} \\ 0 & 0 \end{bmatrix}, \\ \widehat{B}_u = \begin{bmatrix} B_u \\ 0 \end{bmatrix}, \\ \widehat{B}_w = \begin{bmatrix} 0 \\ I_n \end{bmatrix}, \\ \widehat{L} = \begin{bmatrix} L & \Sigma^{1/2} \end{bmatrix}. \end{cases} \tag{3.5}$$

Notice that $\widehat{\Delta}(z) = z^{-1}\Delta(z)$ can be cleanly expressed as

$$\widehat{\Delta}(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_w. \tag{3.6}$$

Similarly, $F(z)$ can be written as

$$F(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_u. \tag{3.7}$$

We can rewrite the frequency domain dynamics (3.2) in terms of $\widehat{\Delta}(z)$:

$$s(z) = F(z)u(z) + \widehat{\Delta}(z)(z\widehat{w}(z)). \tag{3.8}$$

It is easy to check that the stabilizability of (A, B_u) implies the stabilizability of $(\widehat{A}, \widehat{B}_u)$, and similarly the detectability of (A, L) implies the unit-circle observability of $(\widehat{A}, \widehat{L})$. We have shown that the competitive-suboptimal controller at level γ in the system $\{A, B_u, B_w, L\}$ is the H_∞-suboptimal controller at level γ in the system $\{\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{L}\}$. Plugging these parameters into Theorem 17 immediately yields necessary and sufficient conditions for the existence of a causal competitive-suboptimal controller at level γ, along with a state-space model for the controller, if it exists. It is easy to recover the strictly causal competitive-suboptimal controller in an analogous fashion.

We now construct the synthetic disturbance $\widehat{w}$. Recall that $\widehat{w}(z) = \Delta^{-1}(z)G(z)w(z)$. We have

$$\Delta^{-1}(z) = \Sigma^{-1/2}\left(I - L(zI - (A - KL))^{-1}K\right),$$

$$G(z) = L(zI - A)^{-1}B_w.$$

We note that $A - KL$ is stable and hence $\Delta^{-1}(z)$ is causal and bounded since its poles are strictly contained in the unit circle. The operator $\Delta^{-1}(z)G(z)$ is given by

$$\Delta^{-1}(z)G(z) = \Sigma^{-1/2}L(zI - (A - KL))^{-1}B_w,$$

therefore a state-space model for $\widehat{w}$ is

$$\nu_{t+1} = (A - KL)\nu_t + B_w w_t, \quad \widehat{w}_t = \Sigma^{-1/2}L\nu_t. \tag{3.9}$$

We emphasize that the system whose frequency domain dynamics are given by (3.8) is driven by $\widehat{w}_{t+1}$, not $\widehat{w}_t$, since the driving disturbance in (3.8) is $z\widehat{w}(z)$, not $\widehat{w}(z)$. We reiterate that $\widehat{w}$ is a strictly causal function of w; in particular, $\widehat{w}_{t+1}$ depends only on $(\ldots, w_{t-1}, w_t)$. □

Competitive Control in Time-Varying Systems

We show that a controller with optimal competitive ratio can also be systems which vary over a finite horizon from $t = 1$ to T; similar techniques can be used to obtain pathlength-optimal and energy-optimal controllers in time-varying systems. The dynamics are

$$x_{t+1} = A_t x_t + B_{u,t} w_t + B_{w,t} w_t,$$

where $A_t \in \mathbb{R}^{n\times n}$, $B_{u,t} \in \mathbb{R}^{n\times m}$, $B_{w,t} \in \mathbb{R}^{n\times p}$ for $t = 1$ to T. The state cost in each round is $\|s_t\|^2$, where $s_t = Q_t^{1/2} x_t$ and $Q_t \succeq 0$.

Theorem 7 (Finite-horizon competitive control)**.** *A causal controller π such that*

$$\sup_w \frac{J_T(\pi, w)}{J_T(\pi_0, w)} < \gamma^2 \tag{3.10}$$

exists if and only if

$$\widehat{B}_{w,t}^\top \left[\widehat{P}_{t+1} - \widehat{P}_{t+1} \widehat{B}_{u,t} \left(I_m + \widehat{B}_{u,t}^\top \widehat{P}_{t+1} \widehat{B}_{u,t} \right)^{-1} \widehat{B}_{u,t}^\top \widehat{P}_{t+1} \right] \widehat{B}_{w,t} \prec \gamma^2 I_n$$

for $t = 0, \ldots, T$, where we define

$$\widehat{A}_t = \begin{bmatrix} A_t & K_t \Sigma_t^{1/2} \\ 0 & 0 \end{bmatrix}, \quad \widehat{B}_{u,t} = \begin{bmatrix} B_{u,t} \\ 0 \end{bmatrix}, \quad \widehat{B}_{w,t} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad \widehat{L}_t = \begin{bmatrix} Q_t^{1/2} & \Sigma_t^{1/2} \end{bmatrix},$$

we define $\widehat{P}_t$ to be the solution of the backwards-time Riccati recursion

$$\widehat{P}_t = \widehat{L}_t^\top \widehat{L}_t + \widehat{A}_t^\top \widehat{P}_{t+1} \widehat{A}_t - \widehat{A}_t^\top \widehat{P}_{t+1} \tilde{B}_t \tilde{H}_t^{-1} \tilde{B}_t^\top \widehat{P}_{t+1} \widehat{A}_t$$

where we initialize $\widehat{P}_{T+1} = 0$, and we define

$$\tilde{B}_t = \begin{bmatrix} \widehat{B}_{u,t} & \widehat{B}_{w,t} \end{bmatrix}, \quad \tilde{H}_t = \begin{bmatrix} I & 0 \\ 0 & -\gamma^2 I \end{bmatrix} + \tilde{B}_t^\top \widehat{P}_{t+1} \tilde{B}_t,$$

and K_t, Σ_t are defined as in (3.14). In this case, a causal controller with competitive ratio bounded above by γ^2 is given by

$$u_t = -\left(I_m + \widehat{B}_{u,t}^\top \widehat{P}_{t+1} \widehat{B}_{u,t} \right)^{-1} \widehat{B}_{u,t}^\top \widehat{P}_{t+1} \left(\widehat{A}_t \xi_t + \widehat{B}_{w,t} \widehat{w}_{t+1} \right),$$

where the dynamics of $\xi \in \mathbb{R}^{2n}$ are

$$\xi_{t+1} = \widehat{A}_t \xi_t + \widehat{B}_{u,t} u_t + \widehat{B}_{w,t} \widehat{w}_{t+1} \tag{3.11}$$

and we initialize $\xi_0 = 0$. The synthetic disturbance $\widehat{w}$ can be computed using the recursion

$$\nu_{t+1} = (A_t - K_t Q_t^{1/2}) \nu_t + B_{w,t} w_t, \quad \widehat{w}_t = \Sigma_t^{-1/2} Q_t^{1/2} \nu_t,$$

where we initialize $v_0 = 0$*. A strictly causal finite-horizon controller with competitive ratio bounded above by* γ^2 *exists if and only if*

$$\widehat{B}_{w,t}^\top \widehat{P}_{t+1} \widehat{B}_{w,t} \prec \gamma^2 I_n$$

for $t = 0, \ldots, T$*. In this case, a strictly causal controller with competitive ratio bounded above by* γ^2 *is given by*

$$u_t = -\left(I_m + \widehat{B}_{u,t}^\top \widehat{P}_{t+1} \widehat{B}_{u,t}\right)^{-1} \widehat{B}_{u,t}^\top \widetilde{P}_{t+1} \widehat{A}_t \xi_t,$$

where we define

$$\widetilde{P}_{t+1} = \widehat{P}_{t+1} - \widehat{P}_{t+1} \widehat{B}_{w,t} (-\gamma^2 I_p + \widehat{B}_{w,t}^\top \widehat{P}t + 1 \widehat{B}_{w,t})^{-1} \widehat{B}_{w,t}^\top \widehat{P}_{t+1}.$$

Proof. Let $L_t = Q_t^{1/2}$ for $t = 0, \ldots, T$. The offline optimal cost is

$$J_T(\pi_0, w) = w^* G^* (I + FF^*)^{-1} G w,$$

where F and G are the transfer matrices mapping $u = (u_0, \ldots, u_T)$ and $w = (w_0, \ldots, w_T)$ to $s = (s_0, \ldots, s_T)$:

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ L_1 B_{u,0} & 0 & 0 & 0 & \\ L_2 A_1 B_{u,0} & L_2 B_{u,1} & 0 & 0 & \\ L_3 A_2 A_1 B_{u,0} & L_3 A_1 B_{u,1} & L_3 B_{u,2} & 0 & \\ \vdots & & & & \ddots \end{bmatrix},$$

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ L_1 B_{w,0} & 0 & 0 & 0 & \\ L_2 A_1 B_{w,0} & L_2 B_{w,1} & 0 & 0 & \\ L_3 A_2 A_1 B_{w,0} & L_3 A_1 B_{w,1} & L_3 B_{w,2} & 0 & \\ \vdots & & & & \ddots \end{bmatrix}.$$

We note that due to the time-varying dynamics, F and G lack the block-Toeplitz structure of the transfer operators associated to LTI systems.

We see that condition (3.10) is equivalent to

$$J_T(\pi, w) \leq \gamma^2 w^* G^* \left(I + FF^*\right)^{-1} G w.$$

Let Δ be the unique causal and causally invertible matrix such that

$$I + FF^* = \Delta\Delta^*.$$

With this factorization, condition 3.10 becomes the H_∞ condition

$$J_T(\pi, \widehat{w}) \leq \gamma^2 \|\widehat{w}\|^2,$$

where the synthetic disturbance $\widehat{w}$ is given by

$$\widehat{w} = \Delta_2^{-1} G w$$

and the synthetic system dynamics are

$$s = Fu + \Delta_2 \widehat{w}. \tag{3.12}$$

A state-space model for F is given by

$$\epsilon_{t+1} = A_t \epsilon_t + B_{u,t} u_t, \quad s_t = L_t \epsilon_t.$$

Given this state-space model, we wish to obtain the factorization $\Delta\Delta^\top = I + FF^\top$ where Δ is causal. We interpret $I + FF^\top$ as the covariance matrix of an appropriately defined random variable and use the Kalman filter to obtain a state-space model for Δ. Suppose that u and v are zero-mean random variables such that $\mathbb{E}[uu^\top] = I$, $\mathbb{E}[vv^\top] = I$ and $\mathbb{E}[uv^\top] = 0$. Define $y = Fu + v$; notice that $\mathbb{E}[yy^\top] = I + FF^\top$. Given a state-space model for F, the Kalman filter can be used to construct a state-space model for a causal matrix Δ such that $y = \Delta e$, where e is a zero-mean random variable such that $\mathbb{E}[ee^\top] = I$; this is the so-called “whitening" property of the Kalman filter. Notice that since $y = Fu + v$, $\mathbb{E}[yy^\top] = I + FF^\top$; on the other hand, $y = \Delta e$, so $\mathbb{E}[yy^\top] = \Delta\Delta^\top$. Therefore $I + FF^\top = \Delta\Delta^\top$, as desired.

Using the Kalman filter as described in Theorem 9.2.1 in [KSH00], we obtain a state-space model for Δ:

$$\eta_{t+1} = A_t \eta_t + K_t \Sigma_t^{1/2} e_t, \quad y_t = L_t \eta_t + \Sigma_t^{1/2} e_t, \tag{3.13}$$

where we define

$$K_t = A_t P_t L_t \Sigma_t^{-1}, \quad \Sigma_t = I + L_t P_t L_t^\top, \tag{3.14}$$

and P_t is defined recursively as

$$P_{t+1} = A_t P_t A_t^\top + B_{u,t} B_{u,t}^\top - K_t \Sigma_t K_t^\top$$

where we initialize $P_0 = 0$.

Now that we have state-space models for F and Δ, we can form a state-space model for the overall system (3.12). Letting $\alpha_t = \epsilon_t + \eta_t$, we see that a state-space model for this system is

$$\alpha_{t+1} = A_t\alpha_t + B_{u,t}u_t + K_t\Sigma_t^{1/2}\widehat{w}_t, \quad s_t = L_t\alpha_t + \Sigma_t^{\frac{1}{2}}\widehat{w}_t.$$

This system can be rewritten as

$$\xi_{t+1} = \widehat{A}_t\xi_t + \widehat{B}_{u,t}u_t + \widehat{B}_{w,t}\widehat{w}_{t+1}, \quad s_t = \widehat{L}_t\xi_t, \tag{3.15}$$

where we define

$$\widehat{A}_t = \begin{bmatrix} A_t & K_t\Sigma_t^{1/2} \\ 0 & 0 \end{bmatrix}, \quad \widehat{B}_{u,t} = \begin{bmatrix} B_{u,t} \\ 0 \end{bmatrix}, \quad \widehat{B}_{w,t} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad \widehat{L}_t = \begin{bmatrix} L_t & \Sigma_t^{1/2} \end{bmatrix}$$

and we initialize $\xi_0 = 0$. Recall that our goal is to find a controller π in the synthetic system (3.15) such that $J_T(\pi, \widehat{w}) < \gamma^2\|\widehat{w}\|_2^2$ for all disturbances $\widehat{w}$, or to determine whether no such controller exists. Theorem 18 gives necessary and sufficient conditions for the existence of such a controller, along with an explicit state-space description of the controller, if it exists.

We emphasize that the driving disturbance in the synthetic system (3.15) is not w, but rather the synthetic disturbance $\widehat{w} = \Delta^{-1}Gw$. Notice that $\Delta^{-1}G$ is strictly causal, since Δ^{-1} is causal and G is strictly causal. Exchanging inputs and outputs in (3.13), we see that a state-space model for Δ^{-1} is

$$\eta_{t+1} = (A_t - K_tL_t)\eta_t + K_ty_t, \quad e_t = \Sigma_t^{-1/2}(y_t - L_t\eta_t).$$

A state-space model for G is

$$\delta_{t+1} = A_t\delta_t + B_{w,t}w_t, \quad s_t = L_t\delta_t.$$

Equating s and y, we see that a state-space model for $\Delta^{-1}G$ is

$$\begin{bmatrix} \eta_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} A_t - K_tL_t & K_tL_t \\ 0 & A_t \end{bmatrix} \begin{bmatrix} \eta_t \\ \delta_t \end{bmatrix} + \begin{bmatrix} 0 \\ B_{w,t} \end{bmatrix} w_t,$$

$$e_t = \Sigma_t^{-1/2}L_t(\delta_t - \eta_t).$$

Setting $\nu_t = \delta_t - \eta_t$ and simplifying, we see that a minimal representation for $\widehat{w}$ is

$$\nu_{t+1} = (A_t - K_tL_t)\nu_t + B_{w,t}w_t, \quad \widehat{w}_t = \Sigma_t^{-1/2}L_t\nu_t.$$

We reiterate that $\widehat{w}$ is a strictly causal function of w; in particular, $\widehat{w}_{t+1}$ depends only on $w_0, w_1, \ldots, w_t$. □

3.2 Energy-Optimal Control

The energy-optimal control problem is to find an online controller which minimizes the ratio of the regret to the energy in the disturbance:

Problem 3 (Energy-optimal control)**.** *Find an online controller which minimizes*

$$\sup_{w} \frac{J(\pi, w) - J(\pi_0, w)}{\|w\|_2^2}.$$

We call the controller with the smallest possible ratio the *energy-optimal controller*. Instead of minimizing the ratio of regret to energy directly, we instead solve the following relaxation:

Problem 4 (Energy-suboptimal control)**.** *Given $\gamma > 0$, find an online controller such that*

$$\sup_{w} \frac{J(\pi, w) - J(\pi_0, w)}{\|w\|_2^2} < \gamma^2$$

for all disturbances w, or determine whether no such controller exists.

We call such a controller the *energy-suboptimal controller at level γ*. It is clear that if we can solve this suboptimal problem then we can easily recover the energy-optimal controller via bisection on γ.

We derive necessary and sufficient conditions for the existence of an energy-suboptimal controller at level γ, along with a state-space model of the controller:

Theorem 8. *Suppose (A, B_u) is stabilizable and (A, L) is detectable. Fix $\gamma > 0$ and define $\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{L}$ as in (3.26). A causal controller π such that*

$$\sup_{w \in \ell_2} \frac{J(\pi, w) - J(\pi_0, w)}{\|w\|_2^2} < \gamma^2 \tag{3.16}$$

exists if and only if the DARE

$$\widehat{P} = \widehat{L}^* \widehat{L} + \widehat{A}^* \widehat{P} \widehat{A} - \widehat{A}^* \widehat{P} \widetilde{B} \widetilde{H}^{-1} \widetilde{B}^* \widehat{P} \widehat{A},$$

where we define

$$\begin{cases} \widetilde{B} = \begin{bmatrix} \widehat{B}_u & \widehat{B}_w \end{bmatrix}, \\ \widetilde{R} = \begin{bmatrix} I_m & 0 \\ 0 & -I_p \end{bmatrix}, \\ \widetilde{H} = \widetilde{R} + \widetilde{B}^* P \widetilde{B}, \end{cases}$$

has a solution $\widehat{P}$ *such that*

1. $\widehat{A} - \widetilde{B}\widetilde{H}^{-1}\widetilde{B}^*\widehat{P}\widehat{A}$ *is stable;*
2. $\widetilde{R}$ *and* $\widetilde{H}$ *have the same inertia;*
3. $\widehat{P} \succeq 0$.

In this case, one possible causal controller satisfying (3.16) is

$$u_t = -(I + \widehat{B}_u^*\widehat{P}\widehat{B}_u)^{-1}\widehat{B}_u^*\widehat{P}(\widehat{A}\xi_t + \widehat{B}_w\widehat{w}_t),$$

where the synthetic state $\xi \in \mathbb{R}^{2n}$ *evolves according to the linear dynamics equation*

$$\xi_{t+1} = \widehat{A}\xi_t + \widehat{B}_u u_t + \widehat{B}_w\widehat{w}_t$$

and the synthetic disturbance $\widehat{w}$ *is given in (3.21). A strictly causal controller satisfying (3.16) exists if and only if conditions (1) and (3) hold, and additionally*

$$\widehat{B}_u^*\widehat{P}\widehat{B}_u \prec \gamma^2 I,$$

$$I + \widehat{B}_w^*\widehat{P}(I - \widehat{B}_u(-\gamma^2 I + \widehat{B}_u^*\widehat{P}\widehat{B}_u)^{-1}\widehat{B}_u^*\widehat{P})\widehat{B}_w \succ 0.$$

In this case, one possible strictly causal controller satisfying (3.16) is

$$u_t = -(I_m + \widehat{B}_u^*\widetilde{P}\widehat{B}_u)^{-1}\widehat{B}_u^*\widetilde{P}\widehat{A}\xi_t,$$

where we define

$$\widetilde{P} = \widehat{P} - \widehat{P}\widehat{B}_w(-\gamma^2 I_p + \widehat{B}_w^*\widehat{P}\widehat{B}_w)^{-1}\widehat{B}_w^*\widehat{P}.$$

Proof. Recall that the offline optimal cost is

$$J(\pi_0, w) = w^* G^*(I + FF^*)^{-1}Gw.$$

We see that condition (3.16) is equivalent to

$$J(\pi, w) < w^*\left[\gamma^2 I + G^*\left(I + FF^*\right)^{-1}G\right]w.$$

Our goal is to obtain the factorization

$$\gamma^2 I + G^*\left(I + FF^*\right)^{-1}G = \Delta_2^*\Delta_2, \tag{3.17}$$

where $\Delta_2(z)$ is causal and causally invertible. With this factorization, condition (3.16) becomes the H_∞ condition

$$J(\pi, \widehat{w}) < \|\widehat{w}\|^2,$$

where the synthetic disturbance $\widehat{w}$ is given by

$$\widehat{w}(z) = \Delta_2(z) w(z)$$

and the system dynamics in the frequency domain are

$$s(z) = F(z)u(z) + G(z)\Delta_2^{-1}(z)\widehat{w}(z). \tag{3.18}$$

In Theorem 2 we found that a causal and causally invertible operator Δ satisfying

$$I + F(z)F(z^{-*})^* = \Delta(z)\Delta(z^{-*})^*$$

is given by

$$\Delta(z) = (I + L(zI - A)^{-1}K)\Sigma^{1/2},$$

where we define

$$K = APL\Sigma^{-1}, \quad \Sigma = I + LPL^*,$$

and P is the unique Hermitian solution to the Riccati equation

$$B_u B_u^* - P + APA^* - APL(I + LPL^*)^{-1}LPA^* = 0.$$

The operator Δ^{-1} is given by

$$\Delta^{-1}(z) = \Sigma^{-1/2}\left(I - L(zI - (A - KL))^{-1}K\right),$$

therefore $\Delta^{-1}(z)G(z)$ is

$$\Delta^{-1}(z)G(z) = \Sigma^{-1/2}L(zI - (A - KL))^{-1}B_w.$$

Define $\widetilde{A} = A - KL$. We can now recover the factorization (3.17). Notice that the left-hand side of (3.17) can be written as

$$\begin{bmatrix} B_w^*(z^{-*}I - \widetilde{A})^{-*} & I \end{bmatrix} \begin{bmatrix} L\Sigma^{-1}L & 0 \\ 0 & \gamma^2 I \end{bmatrix} \begin{bmatrix} (zI - \widetilde{A})^{-1}B_w \\ I \end{bmatrix}.$$

Applying Lemma 1, we see that this equals

$$\begin{bmatrix} B_w^*(z^{-*}I - \widetilde{A})^{-*} & I \end{bmatrix} \Lambda_2(P_2) \begin{bmatrix} (zI - \widetilde{A})^{-1}B_w \\ I \end{bmatrix},$$

where P_2 is an arbitrary Hermitian operator and we define

$$\Lambda_2(P_2) = \begin{bmatrix} L\Sigma^{-1}L - P_2 + \widetilde{A}^* P_2 \widetilde{A} & \widetilde{A}^* P_2 B_w \\ B_w^* P_2 \widetilde{A} & \gamma^2 I + B_w^* P_2 B_w \end{bmatrix}.$$

Notice that the $\Lambda_2(P_2)$ can be factored as

$$\begin{bmatrix} I & K_2^*(P_2) \\ 0 & I \end{bmatrix} \begin{bmatrix} \Gamma_2(P_2) & 0 \\ 0 & \Sigma_2(P_2) \end{bmatrix} \begin{bmatrix} I & 0 \\ K_2(P_2) & I \end{bmatrix},$$

where we define

$$\Gamma_2(P_2) = L\Sigma^{-1}L - P_2 + \widetilde{A}^* P_2 \widetilde{A} - K_2^*(P_2)\Sigma_2 K_2(P_2),$$

$$K_2(P_2) = \Sigma_2^{-1}(P_2) B_w^* P_2 \widetilde{A},$$

$$\Sigma_2(P_2) = \gamma^2 I + B_w^* P_2 B_w.$$

Notice that $\widetilde{A}$ is stable, therefore the Riccati equation $\Gamma_2(P_2) = 0$ has a unique stabilizing solution (Theorem E.6.2 in [KSH00]). Suppose P_2 is chosen to be this solution, and define $K_2 = K_2(P_2)$, $\Sigma_2 = \Sigma_2(P_2)$. We immediately obtain the factorization (3.17), where we define

$$\Delta_2(z) = \Sigma_2^{1/2}(I + K_2(zI - \widetilde{A})^{-1} B_w). \tag{3.19}$$

Recall that the energy-suboptimal controller at level γ is the H_∞-suboptimal controller at level 1 in the system (3.18). The operator Δ_2^{-1} is given by

$$\Delta_2^{-1}(z) = (I - K_2(zI - (\widetilde{A} - B_w K_2))^{-1} B_w) \Sigma_2^{-1/2}.$$

We note that $\widetilde{A} - B_w K_2$ is stable and hence $\Delta_2^{-1}(z)$ is causal and bounded since its poles are strictly contained in the unit circle. Define

$$\begin{cases} \widehat{A} = \begin{bmatrix} A & -B_w K_2 \\ 0 & \widetilde{A} - B_w K_2 \end{bmatrix}, \\ \widehat{B}_u = \begin{bmatrix} B_u \\ 0 \end{bmatrix}, \\ \widehat{B}_w = \begin{bmatrix} B_w \Sigma_2^{-1/2} \\ B_w \Sigma_2^{-1/2} \end{bmatrix}, \\ \widehat{L} = \begin{bmatrix} L & 0 \end{bmatrix}. \end{cases} \tag{3.20}$$

It is easy to verify that

$$F(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_u, \quad G(z)\Delta^{-1}(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_w.$$

Furthermore, one can check that the stabilizability of (A, B_u) implies the stabilizability of $(\widehat{A}, \widehat{B}_u)$, and similarly the detectability of (A, L) implies the unit-circle observability of $(\widehat{A}, \widehat{B}_u)$. We have shown that the energy-suboptimal controller at level γ in the system $\{A, B_u, B_w, L\}$ is the H_∞-suboptimal controller at level 1 in the system $\{\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{L}\}$. Plugging these parameters into Theorem 17 immediately yields necessary and sufficient conditions for the existence of a causal energy-suboptimal controller at level γ, along with a state-space model for the controller, if it exists. It is easy to recover the strictly causal energy-suboptimal controller in an analogous fashion. Recall that the synthetic disturbance $\widehat{w}$ is given by $\widehat{w}(z) = \Delta_2(z)w(z)$; it immediately follows from (3.19) that a state-space model for $\widehat{w}$ is

$$\nu_{t+1} = \widetilde{A}\nu_t + B_w w_t, \quad \widehat{w}_t = \Sigma_2^{1/2}(K_2\nu_t + w_t). \tag{3.21}$$

□

3.3 Pathlength-Optimal Control

It is natural to bound the regret of an online learning algorithm by the temporal variation in the data it encounters. Intuitively, it should be easier to achieve low regret when the data changes slowly over time, since the past observations of the algorithm are predictive of the future; conversely, when the data sequence changes frequently or abruptly one should expect a learning algorithm to incur high regret. An alternative idea, first proposed by Zinkevich in [Zin03], is to bound the regret of the online algorithm by the variation of the comparator sequence instead of the variation of the data. Both types of regret bounds are referred to as *pathlength* bounds. A series of works [BGZ15; Bub+19; Zha+20] describe bandit algorithms whose regret is bounded pathlength. We also note [ZWZ22], which describes an online control algorithm with regret which is bounded by the variation in a comparator sequence of DAC policies.

The pathlength-optimal control problem is to find an online controller which minimizes the ratio of its regret (relative to the optimal noncausal controller) to the pathlength the disturbance:

Problem 5 (Pathlength-optimal control)**.** *Find an online controller which minimizes*

$$\sup_w \frac{J(\pi, w) - J(\pi_0, w)}{\|D_p w\|_2^2}.$$

We call the controller with the smallest possible ratio the *pathlength-optimal controller*. Instead of minimizing the ratio of regret to pathlength directly, we instead solve the following relaxation:

Problem 6 (Pathlength-suboptimal control)**.** *Given $\gamma > 0$, find an online controller such that*

$$\sup_w \frac{J(\pi, w) - J(\pi_0, w)}{\|D_p w\|_2^2} < \gamma^2$$

for all disturbances w, or determine whether no such controller exists.

We call such a controller the *pathlength-suboptimal controller at level γ*. It is clear that if we can solve this suboptimal problem then we can easily recover the pathlength-optimal controller via bisection on γ.

We derive necessary and sufficient conditions for the existence of an pathlength-suboptimal controller at level γ, along with a state-space model of the controller:

Theorem 9. *Suppose (A, B_u) is stabilizable and (A, L) is detectable. Fix $\gamma > 0$ and define $\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{L}$ as in (3.26). A causal controller π satisfying*

$$\sup_{w \in \ell_2} \frac{J(\pi, w) - J(\pi_0, w)}{\|D_p w\|_2^2} < \gamma^2 \tag{3.22}$$

exists if and only if the DARE

$$\widehat{P} = \widehat{L}^* \widehat{L} + \widehat{A}^* \widehat{P} \widehat{A} - \widehat{A}^* \widehat{P} \widetilde{B} \widetilde{H}^{-1} \widetilde{B}^* \widehat{P} \widehat{A},$$

where we define

$$\begin{cases} \widetilde{B} = \begin{bmatrix} \widehat{B}_u & \widehat{B}_w \end{bmatrix}, \\ \widetilde{R} = \begin{bmatrix} I_m & 0 \\ 0 & -I_p \end{bmatrix}, \\ \widetilde{H} = \widetilde{R} + \widetilde{B}^* P \widetilde{B}, \end{cases}$$

has a solution $\widehat{P}$ such that

1. *$\widehat{A} - \widetilde{B} \widetilde{H}^{-1} \widetilde{B}^* \widehat{P} \widehat{A}$ is stable;*
2. *$\widetilde{R}$ and $\widetilde{H}$ have the same inertia;*

3. $\widehat{P} \succeq 0$.

In this case, one possible causal controller satisfying (3.22) is

$$u_t = -(I_{2n+p} + \widehat{B}_u^* \widehat{P} \widehat{B}_u)^{-1} \widehat{B}_u^* \widehat{P} (\widehat{A}\xi_t + \widehat{B}_w \widehat{w}_t),$$

where the synthetic state $\xi \in \mathbb{R}^{2n+p}$ evolves according to the linear dynamics equation

$$\xi_{t+1} = \widehat{A}\xi_t + \widehat{B}_u u_t + \widehat{B}_w \widehat{w}_t$$

and the synthetic disturbance $\widehat{w}$ is given in (3.27). A strictly causal controller satisfying (3.22) exists if and only if conditions (1) and (3) hold, and additionally

$$\widehat{B}_u^* \widehat{P} \widehat{B}_u \prec \gamma^2 I,$$

$$I + \widehat{B}_w^* \widehat{P} (I - \widehat{B}_u (-\gamma^2 I + \widehat{B}_u^* \widehat{P} \widehat{B}_u)^{-1} \widehat{B}_u^* \widehat{P}) \widehat{B}_w \succ 0.$$

In this case, one possible strictly causal controller satisfying (3.22) is

$$u_t = -(I_m + \widehat{B}_u^* \widetilde{P} \widehat{B}_u)^{-1} \widehat{B}_u^* \widetilde{P} \widehat{A} \xi_t,$$

where we define

$$\widetilde{P} = \widehat{P} - \widehat{P}\widehat{B}_w(-\gamma^2 I_p + \widehat{B}_w^* \widehat{P} \widehat{B}_w)^{-1} \widehat{B}_w^* \widehat{P}.$$

Proof. Recall that the offline optimal cost is

$$J(\pi_0, w) = w^* G^* (I + FF^*)^{-1} G w$$

and that the pathlength of the disturbance sequence is $\|Dw(z)\|^2$, where $D(z) = 1 - z$. We see that condition (3.22) is equivalent to

$$J(\pi, w) < w^* \left[\gamma^2 D^* D + G^* (I + FF^*)^{-1} G \right] w.$$

Our goal is to obtain a canonical factorization

$$\gamma^2 D^* D + G^* (I + FF^*)^{-1} G = \Delta_2^* \Delta_2, \tag{3.23}$$

where $\Delta_2(z)$ is causal and causally invertible. With this factorization, condition (3.22) becomes the H_∞ condition

$$J(\pi, \widehat{w}) < \|\widehat{w}\|^2,$$

where the synthetic disturbance $\widehat{w}$ is given by

$$\widehat{w}(z) = \Delta_2(z) w(z)$$

and the system dynamics in the frequency domain are

$$s(z) = F(z)u(z) + G(z)\Delta_2^{-1}(z)\widehat{w}(z). \tag{3.24}$$

In Theorem 2 we found that a causal and causally invertible operator Δ satisfying

$$I + F(z)F(z^{-*})^* = \Delta(z)\Delta(z^{-*})^*$$

is given by

$$\Delta(z) = (I + L(zI - A)^{-1}K)\Sigma^{1/2},$$

where we define

$$K = APL\Sigma^{-1}, \quad \Sigma = I + LPL^*,$$

and P is the unique Hermitian solution to the Riccati equation

$$B_u B_u^* - P + APA^* - APL(I + LPL^*)^{-1}LPA^* = 0.$$

The operator $\Delta^{-1}(z)$ is given by

$$\Delta^{-1}(z) = \Sigma^{-1/2}\left(I - L(zI - (A - KL))^{-1}K\right),$$

therefore a minimal representation of $\Delta^{-1}(z)G(z)$ is

$$\Delta^{-1}(z)G(z) = \Sigma^{-1/2}L(zI - (A - KL))^{-1}B_w.$$

We can now recover the factorization (3.23). Notice that the left-hand side of (3.23) can be written as

$$\begin{bmatrix} \widetilde{B}_w^*(z^{-*}I - \widetilde{A})^{-*} & I \end{bmatrix} \begin{bmatrix} \widetilde{L}^*\widetilde{L} & \widetilde{S} \\ \widetilde{S}^* & \gamma^2 I \end{bmatrix} \begin{bmatrix} (zI - \widetilde{A})^{-1}\widetilde{B}_w \\ I \end{bmatrix},$$

where we define

$$\widetilde{L} = \begin{bmatrix} \Sigma^{-1/2}L & 0 \\ 0 & \gamma I_p \end{bmatrix}, \quad \widetilde{S} = \begin{bmatrix} 0 \\ \gamma^2 I_p \end{bmatrix}, \quad \widetilde{A} = \begin{bmatrix} A - KL & 0 \\ 0 & 0 \end{bmatrix}, \quad \widetilde{B}_w = \begin{bmatrix} B_w \\ -I_p \end{bmatrix}.$$

Applying Lemma 1, we see that this equals

$$\begin{bmatrix} \widetilde{B}_w^*(z^{-*}I - \widetilde{A})^{-*} & I \end{bmatrix} \Lambda_2(P_2) \begin{bmatrix} (zI - \widetilde{A})^{-1}\widetilde{B}_w \\ I \end{bmatrix},$$

where P_2 is an arbitrary Hermitian operator and we define

$$\Lambda_2(P_2) = \begin{bmatrix} \widetilde{L}^*\widetilde{L} - P_2 + \widetilde{A}^* P_2 \widetilde{A} & \widetilde{S} + \widetilde{A}^* P_2 \widetilde{B}_w \\ \widetilde{S}^* + \widetilde{B}_w^* P_2 \widetilde{A} & \gamma^2 I + \widetilde{B}_w^* P_2 \widetilde{B}_w \end{bmatrix}.$$

Notice that the $\Lambda_2(P_2)$ can be factored as

$$\begin{bmatrix} I & K_2^*(P_2) \\ 0 & I \end{bmatrix} \begin{bmatrix} \Gamma_2(P_2) & 0 \\ 0 & \Sigma_2(P_2) \end{bmatrix} \begin{bmatrix} I & 0 \\ K_2(P_2) & I \end{bmatrix},$$

where we define

$$\Gamma_2(P_2) = \widetilde{L}^*\widetilde{L} - P_2 + \widetilde{A}^* P_2 \widetilde{A} - K_2^*(P_2)\Sigma_2 K_2(P_2),$$

$$K_2(P_2) = \Sigma_2^{-1}(P_2)(\widetilde{S}^* + \widetilde{B}_w^* P_2 \widetilde{A}),$$

$$\Sigma_2(P_2) = \gamma^2 I + \widetilde{B}_w^* P_2 \widetilde{B}_w.$$

It is clear that $(\widetilde{A}, \widetilde{B}_w)$ is stabilizable, therefore the Riccati equation $\Gamma_2(P_2) = 0$ has a unique stabilizing solution (Theorem E.6.2 in [KSH00]). Suppose P_2 is chosen to be this solution, and define $K_2 = K_2(P_2)$, $\Sigma_2 = \Sigma_2(P_2)$. We immediately obtain the factorization (3.23), where we define

$$\Delta_2(z) = \Sigma_2^{1/2}(I + K_2(zI - \widetilde{A})^{-1}\widetilde{B}_w). \tag{3.25}$$

Recall that the pathlength-suboptimal controller at level γ is the H_∞-suboptimal controller at level 1 in the system (3.24). A state-space model for Δ_2^{-1} is given by

$$\Delta_2^{-1}(z) = (I - K_2(zI - (\widetilde{A} - \widetilde{B}_w K_2))^{-1}\widetilde{B}_w)\Sigma_2^{-1/2}.$$

We note that $\widetilde{A} - \widetilde{B}_w K_2$ is stable and hence $\Delta_2^{-1}(z)$ is causal and bounded since its poles are strictly contained in the unit circle. Define

$$\begin{cases} \widehat{A} = \begin{bmatrix} A & -B_w K_2 \\ 0 & \widetilde{A} - \widetilde{B}_w K_2 \end{bmatrix} \\ \widehat{B}_u = \begin{bmatrix} B_u \\ 0 \end{bmatrix}, \\ \widehat{B}_w = \begin{bmatrix} B_w \Sigma_2^{-1/2} \\ \widetilde{B}_w \Sigma_2^{-1/2} \end{bmatrix} \\ \widehat{L} = \begin{bmatrix} L & 0 \end{bmatrix}. \end{cases} \tag{3.26}$$

It is easy to verify that

$$F(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_u, \quad G(z)\Delta^{-1}(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_w.$$

Furthermore, one can check that the stabilizability of (A, B_u) implies the stabilizability of $(\widehat{A}, \widehat{B}_u)$, and similarly the detectability of (A, L) implies the unit-circle observability of $(\widehat{A}, \widehat{B}_u)$. We have shown that the pathlength-suboptimal controller at level γ in the system $\{A, B_u, B_w, L\}$ is the H_∞-suboptimal controller at level 1 in the system $\{\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{L}\}$. Plugging these parameters into Theorem 17 immediately yields necessary and sufficient conditions for the existence of a causal pathlength-suboptimal controller at level γ, along with a state-space model for the controller, if it exists. It is easy to recover the strictly causal pathlength-suboptimal controller in an analogous fashion. Recall that the synthetic disturbance $\widehat{w}$ is given by $\widehat{w}(z) = \Delta_2(z)w(z)$; it immediately follows from (3.25) that a state-space model for $\widehat{w}$ is

$$\nu_{t+1} = \widetilde{A}\nu_t + \widetilde{B}_w w_t, \qquad \widehat{w}_t = \Sigma_2^{1/2}(K_2\nu_t + w_t). \tag{3.27}$$

□

Chapter 4

REGRET-OPTIMAL MEASUREMENT-FEEDBACK CONTROL

We now turn to the more challenging problem of measurement-feedback control, where the online controller is unable to directly observe the state or disturbance when selecting the control action, but instead only has access to a noisy linear observation of the state in each timestep:

$$y_t = Cx_t + v_t.$$

We let H and J be the transfer operators mapping u and w to the observations y:

$$y = Hu + Jw + v.$$

We restrict our attention to causal control policies which are a linear function of the observations y, i.e., policies which set $u = Ky$ for some causal matrix K. Solving for y, we see that

$$y = (I - HK)^{-1}(Jw + v),$$

implying that

$$u = K(I - HK)^{-1}(Jw + v).$$

We introduce the Youla parameterization $Q = K(I - HK)^{-1}$; we can easily recover K from Q by setting $K = (I+QH)^{-1}Q$. To each K, we associate the transfer operator

$$T_K : \begin{bmatrix} w \\ v \end{bmatrix} \to \begin{bmatrix} s \\ u \end{bmatrix}$$

given by

$$T_K = \begin{bmatrix} G & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} F \\ I \end{bmatrix} Q \begin{bmatrix} J & I \end{bmatrix}. \tag{4.1}$$

Recall that the optimal noncausal controller selects the control $u = K_0 w$, where

$$K_0 = -(I + F^*F)^{-1}F^*G.$$

The transfer operator associated to the optimal noncausal controller is therefore

$$T_{K_0} = \begin{bmatrix} -F(I + F^*F)^{-1}F^*G + G & 0 \\ -(I + F^*F)^{-1}F^*G & 0 \end{bmatrix}.$$

The zeros in the second column represent the fact that the optimal noncausal controller observes the actual disturbance w and the control signal it selects is not at all affected by the measurement noise v; it is this disparity between the information available to the online and offline controllers which makes regret-optimal measurement-feedback control considerably more challenging than regret-optimal full-information control. In fact, we show in this chapter that it is often impossible to attain any regret bound at all, depending on the stability of the system dynamics and which data-dependent regret bound we consider.

4.1 Non-existence results

Non-existence of a controller with bounded competitive ratio

In this section we establish that in unstable systems there is no controller with bounded competitive ratio. The key observation is that the competitive ratio bound implies that the online controller must incur zero cost on every disturbance sequences on which the optimal noncausal controller incurs zero cost. In particular, in order to have a bounded competitive ratio, a controller must always set the control u to be zero when $w = 0$. In the measurement-feedback setting the controller is unable to directly observe w, and hence must set $u = 0$ at all times. The only setting under which this "zero controller" can be competitive is when the system is stable.

Theorem 10. *Fix $\gamma > 0$ and suppose A is unstable. There does not exist a causal measurement-feedback controller π such that*

$$\frac{J(\pi, w, v)}{J(\pi_0, w)} < \gamma^2 \tag{4.2}$$

for all driving disturbances w and and all measurement disturbances v. If A is stable, then the only competitive controller is the "zero controller" which always sets $u = 0$. This controller has competitive ratio $1 + \|F\|^2$.

Proof. We restrict our attention to controllers π which set $\pi(y) = Ky$ for some causal operator K. Condition (4.2) can be rewritten as

$$T_K^* T_K \prec \gamma^2 T_{K_0}^* T_{K_0}. \tag{4.3}$$

The (2, 2) block of $T_K^* T_K$ is $Q^*(I + FF^*)Q$. This operator is clearly positive-semidefinite, and zero only when $K = 0$. If $K \neq 0$ then condition (4.3) clearly cannot hold because the (2, 2) block of $T_{K_0}^* T_{K_0}$ is zero. If $K = 0$, then the competitive ratio is simply

$$\sup_{w \in \ell_2} \frac{\|Gw\|^2}{\|\Delta^{-1} Gw\|^2},$$

where Δ is defined in Theorem 2. Let $w_2 = \Delta^{-1} G w$. The competitive ratio becomes

$$\begin{aligned}
\sup_{w_2 \in \ell_2} \frac{\|\Delta w_2\|^2}{\|w_2\|^2} &= \sup_{w_2 \in \ell_2} \frac{\|\Delta^* w_2\|^2}{\|w_2\|^2} \\
&= \sup_{w_2 \in \ell_2} \frac{w_2^*(I + FF^*)w_2}{\|w_2\|^2} \\
&= 1 + \sup_{w_2 \in \ell_2} \frac{w_2^* F F^* w_2}{\|w_2\|^2} \\
&= 1 + \|F\|^2.
\end{aligned}$$

It is easy to check that F is bounded if and only if A is stable. □

Non-existence of a controller with regret bounded by the pathlength of the driving disturbance and the measurement disturbance

In this section we establish that there is no controller whose regret bounded by the joint pathlength of the driving disturbance and the measurement disturbance; this non-existence result holds for all linear systems, stable or unstable. The key observation is that the regret bound implies that the online controller must incur zero cost whenever w and v are constant. In the measurement-feedback setting the controller is unable to directly observe w and v, and hence must set $u = 0$ at all times. However, there exist choices of w and v with zero pathlength such that this "zero controller" incurs positive regret, contradicting the pathlength bound.

Theorem 11. *Fix $\gamma > 0$. There does not exist a causal controller π such that*

$$J(\pi, w, v) - J(\pi_0, w) < \gamma^2 \left(\|D_p w\|_2^2 + \|D_r v\|_2^2 \right) \tag{4.4}$$

for all driving disturbances w and and all measurement disturbances v.

Proof. We restrict our attention to controllers π which set $\pi(y) = Ky$ for some causal operator K. We argue by way of contradiction; suppose there is some causal K satisfying (4.4). Let us first assume $K \neq 0$. Let $w = 0$ and let v be nonzero constant sequence such that $(I - HK)^{-1} v$ is orthogonal to the nullspace of K. The left-hand side of (4.4) is $v^* Q^* (I + FF^*) Q v$, which is strictly positive due to our assumption on v. The right-hand side of (4.4) is clearly zero for all choices of γ, therefore condition (4.4) cannot hold. Now assume that $K = 0$. The left-hand side of (4.4) is

$$\begin{aligned}
\|Gw\|^2 - \|\Delta^{-1} G w\|^2 &= w^* G^* (I - (I + FF^*)^{-1}) G w \\
&= w^* G^* F (I + F^* F)^{-1} F^* G w,
\end{aligned}$$

where Δ is defined in Theorem 2. Set w to be a nonzero constant sequence which is orthogonal to the nullspace of F^*G; such a choice always exists since F and G are strictly causal. Set $v = 0$. Then the left-hand side of (4.4) is strictly positive, but the right-hand side is zero for all γ. □

Non-existence of a controller with regret bounded by the energy of the driving disturbance and the pathlength of the measurement disturbance

In this section we establish that in unstable systems there is no controller whose regret is bounded by the energy of the driving disturbance and the pathlength of the measurement disturbance. The key observation is that the regret bound implies that the online controller must incur zero cost whenever $w = 0$ and v is constant. In the measurement-feedback setting the controller is unable to directly observe w and v, and hence must set $u = 0$ at all times. The only setting under which this "zero controller" can have regret bounded by the energy of the driving disturbance and the pathlength of the measurement disturbance is when the system is stable.

Theorem 12. *Fix $\gamma > 0$ and suppose A is unstable. There does not exist a causal controller π such that*

$$J(\pi, w, v) - J(\pi_0, w) < \gamma^2 \left(\|w\|_2^2 + \|D_r v\|_2^2 \right) \tag{4.5}$$

for all driving disturbances w and and all measurement disturbances v. If A is stable, then the only controller which satisfies (4.5) for any value of γ is the "zero controller" which always sets $u = 0$. This controller satisfies (4.5) with $\gamma = \|\Delta_2^{-*} F^* G\|$, where Δ_2 is the unique causal and causally invertible operator such that $\Delta_2^* \Delta_2 = I + F^* F$.

Proof. We restrict our attention to controllers π which set $\pi(y) = Ky$ for some causal operator K. We argue by way of contradiction; suppose there is some causal K satisfying (4.5). Let us first assume $K \neq 0$. Let $w = 0$ and let v be nonzero constant sequence such that $(I - HK)^{-1} v$ is orthogonal to the nullspace of K. The left-hand side of (4.5) is $v^* Q^* (I + FF^*) Q v$, which is strictly positive due to our assumption on v. The right-hand side of (4.5) is clearly zero for all choices of γ, therefore condition (4.5) cannot hold. Now assume that $K = 0$. The left-hand side of (4.5) is

$$\begin{aligned} \|Gw\|^2 - \|\Delta^{-1} G w\|^2 &= w^* G^* (I - (I + FF^*)^{-1}) G w \\ &= w^* G^* F (I + F^* F)^{-1} F^* G w, \end{aligned}$$

where Δ is defined in Theorem 2. The smallest possible value of γ^2 such that (4.5) holds is given by

$$\sup_{w \in \ell_2} \frac{w^* G^* F (I + F^* F)^{-1} F^* G w}{\|w\|^2} = \|\Delta_2^{-*} F^* G\|^2,$$

where Δ_2 is the unique causal and causally invertible operator such that $\Delta_2^* \Delta_2 = I + F^* F$. We note that the operator $\Delta_2^{-*} F^* G$ is bounded if and only if A is stable. □

4.2 Regret bounded by the joint energy of w and v

In this section we describe a causal controller π whose regret is bounded by the joint energy of the driving disturbance and the measurement disturbance. This presents a sharp contrast with the negative results of the previous sections; the key difference is that the only pair of disturbances (w, v) whose joint energy is zero is simply $w = 0, v = 0$. It is easy to guarantee that π sets $u = 0$ on this specific instance without also requiring that π sets $u = 0$ on all other instances.

Analogously with the full-information setting, we derive the measurement-feedback controller whose regret has optimal dependence on the joint energy of the driving disturbance and the measurement disturbance via a reduction to H_∞ measurement-feedback control.

Theorem 13. *Fix $\gamma > 0$ and define $\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{C}, \widehat{L}$ as in (4.11). There exists a causal measurement-feedback controller π such that*

$$J(\pi, w, v) - J(\pi_0, w) < \gamma^2 (\|w\|_2^2 + \|v\|_2^2) \tag{4.6}$$

if and only if the control DARE

$$P_c = \widehat{A}^* P_c \widehat{A} + \widehat{L}^* \widehat{L} - K_c^* R_c K_c,$$

and the estimation DARE

$$P_e = \widehat{A} P_e \widehat{A}^* + \widehat{B}_w \widehat{B}_w^* - K_e R_e K_e^*,$$

where we define

$$\begin{cases} K_c = R_c^{-1} \begin{bmatrix} \widehat{B}_u^* \\ \widehat{B}_w^* \end{bmatrix} P_c \widehat{A} \\ R_c = \begin{bmatrix} I_m & 0 \\ 0 & -I_p \end{bmatrix} + \begin{bmatrix} \widehat{B}_u^* \\ \widehat{B}_w^* \end{bmatrix} P_c \begin{bmatrix} \widehat{B}_u & \widehat{B}_w \end{bmatrix}, \\ K_e = \widehat{A} P^d \begin{bmatrix} \widehat{C} & L \end{bmatrix} R_e^{-1}, \\ R_e = \begin{bmatrix} I_r & 0 \\ 0 & -I_n \end{bmatrix} + \begin{bmatrix} \widehat{C} \\ \widehat{L} \end{bmatrix} P_c \begin{bmatrix} \widehat{C}^* & \widehat{L}^* \end{bmatrix}, \end{cases}$$

have solutions $P_c \succeq 0$ and $P_e \succeq 0$ such that

1. *The matrix $\widehat{A} - \begin{bmatrix} \widehat{B}_u & \widehat{B}_w \end{bmatrix} K_c$ is stable.*
2. *The matrix R_c has m positive eigenvalues and p negative eigenvalues.*
3. *The matrix $\widehat{A} - K_e \begin{bmatrix} \widehat{C} \\ \widehat{L} \end{bmatrix}$ is stable.*
4. *The matrix R_e has r positive eigenvalues and n negative eigenvalues.*
5. *$\rho(P_c P_e) < 1$.*

If these conditions are satisfied, then one possible choice of π is given by

$$u_t = -K_u(\xi_t + P\widehat{C}^*(I_r + \widehat{C}P\widehat{C}^*)^{-1}(y_t - \widehat{C}\xi_t)),$$

where the synthetic state $\xi \in \mathbb{R}^{2n}$ evolves according to the linear dynamics equation

$$\xi_{t+1} = (\widehat{A} - \widehat{B}_w K_w)(\xi_t + P\widehat{C}^*(I_r + \widehat{C}P\widehat{C}^*)^{-1}(y_t - \widehat{C}\xi_t)) + \widehat{B}_u u_t.$$

The matrices $K_u \in \mathbb{R}^{m \times 2n}$ and $K_w \in \mathbb{R}^{p \times 2n}$ are defined as

$$\begin{bmatrix} K_u \\ K_w \end{bmatrix} = K_c,$$

and we define $P \in \mathbb{R}^{2n \times 2n}$ as

$$P = P_e(I_n - P_c P_e)^{-1}.$$

We note that the regret-optimal controller can be easily obtained via bisection on γ.

Proof. The regret condition (4.6) can be rewritten as

$$\begin{bmatrix} w \\ v \end{bmatrix}^* T_K^* T_K \begin{bmatrix} w \\ v \end{bmatrix} < \begin{bmatrix} w \\ v \end{bmatrix}^* \left(\gamma^2 \begin{bmatrix} I_p & 0 \\ 0 & I_r \end{bmatrix} + T_{K_0}^* T_{K_0} \right) \begin{bmatrix} w \\ v \end{bmatrix}. \tag{4.7}$$

Let Δ_2 be the unique causal and causally invertible operator such that

$$\gamma^2 I_p + G^* \left(I + FF^*\right)^{-1} G = \Delta_2^* \Delta_2. \tag{4.8}$$

Then

$$\gamma^2 \begin{bmatrix} I_p & 0 \\ 0 & I_r \end{bmatrix} + T_{K_0}^* T_{K_0} = \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix}^* \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix}.$$

Define

$$\begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix} = \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix}.$$

Condition (4.7) can be rewritten as

$$\begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix}^* T_{\widehat{K}}^* T_{\widehat{K}} \begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix} < \left\| \begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix} \right\|_2^2,$$

or equivalently as

$$\|T_{\widehat{K}}\| < 1,$$

where we define

$$T_{\widehat{K}} = T_K \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix}^{-1}.$$

Using the parameterization (4.1) of T_K, we see that

$$T_{\widehat{K}} = \begin{bmatrix} G\Delta_2^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} F \\ I \end{bmatrix} (\gamma^{-1} Q) \begin{bmatrix} \gamma J \Delta_2^{-1} & I \end{bmatrix}.$$

Notice that $T_{\widehat{K}}$ itself has the form of a transfer operator described in (4.1); it is the transfer operator with Youla parameter $\widehat{Q} = \gamma^{-1} Q$ in the system

$$\widehat{s} = \widehat{F}\widehat{u} + \widehat{G}\widehat{w}, \quad \widehat{y} = \widehat{H}\widehat{u} + \widehat{J}\widehat{w} + \widehat{v}, \tag{4.9}$$

where we define $\widehat{F} = F, \widehat{G} = G\Delta_2^{-1}, \widehat{H} = \gamma H, \widehat{J} = \gamma J \Delta_2^{-1}$. It is now clear that a controller K satisfying (4.6) exists if and only if there exists a controller $\widehat{K}$ in the system (4.9) such that $\|T_{\widehat{K}}\| < 1$. If such a controller $\widehat{K}$ exists, then we can easily

recover K from $\widehat{K}$ by setting $K = \gamma\widehat{K}$; notice that this is the unique choice of K which is consistent with the relations $\widehat{H} = \gamma H, \widehat{Q} = \gamma^{-1}Q$.

In order to assign state-space structure to $\widehat{F}, \widehat{G}, \widehat{H}, \widehat{J}$, we must first find $\Delta_2(z)$. Let Δ be the unique causal and causally invertible operator such that

$$I + FF^* = \Delta\Delta^*.$$

Theorem 2 shows that

$$\Delta(z) = (I + L(zI - A)^{-1}K)\Sigma^{1/2}.$$

It follows that $\Delta^{-1}(z)G(z)$ is given by

$$\Delta^{-1}(z)G(z) = \Sigma^{-1/2}L(zI - (A - KL))^{-1}B_w.$$

Define $\widetilde{A} = A - KL$. Notice that

$$\gamma^2 I_p + G^*\Delta^*\Delta^{-1}G = \begin{bmatrix} B_w^*(z^{-*}I_n - \widetilde{A})^{-*} & I_p \end{bmatrix} \begin{bmatrix} L^*\Sigma^{-1}L & 0 \\ 0 & \gamma^2 I_p \end{bmatrix} \begin{bmatrix} (zI_n - \widetilde{A})^{-1}B_w \\ I_p \end{bmatrix}.$$

Applying Lemma 1, we see that this equals

$$\begin{bmatrix} B_w^*(z^{-*}I_n - \widetilde{A})^{-*} & I \end{bmatrix} \Lambda_2(P_2) \begin{bmatrix} (zI_n - \widetilde{A})^{-1}B_w \\ I \end{bmatrix},$$

where P_2 is an arbitrary Hermitian matrix and we define

$$\Lambda_2(P_2) = \begin{bmatrix} L^*\Sigma^{-1}L - P_2 + \widetilde{A}^*P_2\widetilde{A} & \widetilde{A}^*P_2B_w \\ B_w^*P_2\widetilde{A} & \gamma^2 I + B_w^*P_2B_w \end{bmatrix}.$$

Notice that the $\Lambda_2(P_2)$ can be factored as

$$\begin{bmatrix} I & K_2^*(P_2) \\ 0 & I \end{bmatrix} \begin{bmatrix} \Gamma_2(P_2) & 0 \\ 0 & \Sigma_2(P_2) \end{bmatrix} \begin{bmatrix} I & 0 \\ K_2(P_2) & I \end{bmatrix},$$

where we define

$$\Gamma_2(P_2) = L^*\Sigma^{-1}L - P_2 + \widetilde{A}^*P_2\widetilde{A} - K_2^*(P_2)\Sigma_2K_2(P_2),$$

$$K_2(P_2) = \Sigma_2^{-1}(P_2)B_w^*P_2\widetilde{A}, \quad \Sigma_2(P_2) = \gamma^2 I + B_w^*P_2B_w.$$

It is clear that $(\widetilde{A}, B_w)$ is stabilizable (in fact, $\widetilde{A}$ is stable), therefore the Riccati equation $\Gamma_2(P_2) = 0$ has a unique stabilizing solution (Theorem E.6.2 in [KSH00]).

Suppose P_2 is chosen to be this solution, and define $K_2 = K_2(P_2)$, $\Sigma_2 = \Sigma_2(P_2)$. We immediately obtain the factorization (4.8), where we define

$$\Delta_2(z) = \Sigma_2^{1/2}(I + K_2(zI - \widetilde{A})^{-1}B_w). \tag{4.10}$$

We have

$$\Delta_2^{-1}(z) = (I - K_2(zI - (\widetilde{A} - B_w K_2))^{-1}B_w)\Sigma_2^{-1/2}.$$

We note that $\widetilde{A} - B_w K_2$ is stable and hence $\Delta^{-1}(z)$ is causal and bounded since its poles are strictly contained in the unit circle. Define

$$\begin{cases} \widehat{A} = \begin{bmatrix} A & -B_w K_2 \\ 0 & \widetilde{A} - B_w K_2 \end{bmatrix}, \\ \widehat{B}_w = \begin{bmatrix} B_w \Sigma_2^{-1/2} \\ B_w \Sigma_2^{-1/2} \end{bmatrix}, \\ \widehat{B}_u = \begin{bmatrix} B_u \\ 0 \end{bmatrix}, \\ \widehat{C} = \begin{bmatrix} \gamma C & 0 \end{bmatrix}, \\ \widehat{L} = \begin{bmatrix} L & 0 \end{bmatrix}. \end{cases} \tag{4.11}$$

Recall that $\widehat{F} = F, \widehat{G} = G\Delta_2^{-1}, \widehat{H} = \gamma H, \widehat{J} = \gamma J \Delta_2^{-1}$; it follows that $\widehat{F}, \widehat{G}, \widehat{H}, \widehat{J}$ are given by

$$\widehat{F}(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_u, \quad \widehat{G}(z) = \widehat{L}(zI - \widehat{A})\widehat{B}_w,$$
$$\widehat{H}(z) = \widehat{C}(zI - \widehat{A})^{-1}\widehat{B}_u, \quad \widehat{J}(z) = \widehat{C}(zI - \widehat{A})\widehat{B}_w.$$

□

4.3 Regret bounded by the pathlength of w and the energy of v

In this section we describe a causal controller π whose regret is bounded by the pathlength of the driving disturbance and the energy of the measurement disturbance. This presents a sharp contrast with the negative results of the previous sections; the key difference is that the only pairs of disturbances (w, v) such that the pathlength

of the w is zero and the energy of v is zero are pairs where w is constant and v is zero. It is easy to guarantee that π matches the offline controller π_0 on these specific instances, without constraining the behavior of π on all other instances.

Analogously with the full-information setting, we derive the measurement-feedback controller whose regret has optimal dependence on the pathlength of the driving disturbance and the energy of the measurement disturbance via a reduction to H_∞ measurement-feedback control.

Theorem 14. *Fix $\gamma > 0$ and define $\widehat{A}, \widehat{B}_u, \widehat{B}_w, \widehat{C}, \widehat{L}$ as in (4.19). There exists a causal measurement-feedback controller π such that*

$$J(\pi, w, v) - J(\pi_0, w) < \gamma^2(\|D_p w\|_2^2 + \|v\|_2^2) \tag{4.12}$$

if and only if the control DARE

$$P_c = \widehat{A}^* P_c \widehat{A} + \widehat{L}^* \widehat{L} - K_c^* R_c K_c,$$

and the estimation DARE

$$P_e = \widehat{A} P_e \widehat{A}^* + \widehat{B}_w \widehat{B}_w^* - K_e R_e K_e^*,$$

where we define

$$\begin{cases} K_c = R_c^{-1} \begin{bmatrix} \widehat{B}_u^* \\ \widehat{B}_w^* \end{bmatrix} P_c \widehat{A} \\ R_c = \begin{bmatrix} I_m & 0 \\ 0 & -I_p \end{bmatrix} + \begin{bmatrix} \widehat{B}_u^* \\ \widehat{B}_w^* \end{bmatrix} P_c \begin{bmatrix} \widehat{B}_u & \widehat{B}_w \end{bmatrix}, \\ K_e = \widehat{A} P^d \begin{bmatrix} \widehat{C} & L \end{bmatrix} R_e^{-1}, \\ R_e = \begin{bmatrix} I_r & 0 \\ 0 & -I_n \end{bmatrix} + \begin{bmatrix} \widehat{C} \\ \widehat{L} \end{bmatrix} P_c \begin{bmatrix} \widehat{C}^* & \widehat{L}^* \end{bmatrix}, \end{cases}$$

have solutions $P_c \succeq 0$ and $P_e \succeq 0$ such that

1. *The matrix $\widehat{A} - \begin{bmatrix} \widehat{B}_u & \widehat{B}_w \end{bmatrix} K_c$ is stable.*

2. *The matrix R_c has m positive eigenvalues and p negative eigenvalues.*
3. *The matrix $\widehat{A} - K_e \begin{bmatrix} \widehat{C} \\ \widehat{L} \end{bmatrix}$ is stable.*
4. *The matrix R_e has r positive eigenvalues and n negative eigenvalues.*
5. $\rho(P_c P_e) < 1$.

If these conditions are satisfied, then one possible choice of π is given by

$$u_t = -K_u(\xi_t + P\widehat{C}^*(I_r + \widehat{C}P\widehat{C}^*)^{-1}(y_t - \widehat{C}\xi_t)),$$

where the synthetic state $\xi \in \mathbb{R}^{2n+p}$ evolves according to the linear dynamics equation

$$\xi_{t+1} = (\widehat{A} - \widehat{B}_w K_w)(\xi_t + P\widehat{C}^*(I_r + \widehat{C}P\widehat{C}^*)^{-1}(y_t - \widehat{C}\xi_t)) + \widehat{B}_u u_t.$$

The matrices $K_u \in \mathbb{R}^{m\times(2n+p)}$ and $K_w \in \mathbb{R}^{p\times(2n+p)}$ are defined as

$$\begin{bmatrix} K_u \\ K_w \end{bmatrix} = K_c,$$

and we define $P \in \mathbb{R}^{(2n+p)\times(2n+p)}$ as

$$P = P_e(I_n - P_c P_e)^{-1}.$$

We note that the regret-optimal controller can be easily obtained via bisection on γ.

Proof. The regret condition (14) can be rewritten in matrix form as

$$\begin{bmatrix} w \\ v \end{bmatrix}^* T_K^* T_K \begin{bmatrix} w \\ v \end{bmatrix} < \begin{bmatrix} w \\ v \end{bmatrix}^* \left(\gamma^2 \begin{bmatrix} D_p^* D_p & 0 \\ 0 & I \end{bmatrix} + T_{K_0}^* T_{K_0} \right) \begin{bmatrix} w \\ v \end{bmatrix}. \tag{4.13}$$

Let Δ_2 be the unique causal and causally invertible operator such that

$$\gamma^2 D_p^* D_p + G^* (I + FF^*)^{-1} G = \Delta_2^* \Delta_2. \tag{4.14}$$

Then

$$\gamma^2 \begin{bmatrix} D_p^* D_p & 0 \\ 0 & I \end{bmatrix} + T_{K_0}^* T_{K_0} = \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix}^* \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix}.$$

Define

$$\begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix} = \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix} \begin{bmatrix} w \\ v \end{bmatrix}.$$

Condition (4.13) can be rewritten as

$$\begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix}^* T_{\widehat{K}}^* T_{\widehat{K}} \begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix} < \left\| \begin{bmatrix} \widehat{w} \\ \widehat{v} \end{bmatrix} \right\|_2^2,$$

or equivalently as

$$\|T_{\widehat{K}}\| < 1,$$

where we define

$$T_{\widehat{K}} = T_K \begin{bmatrix} \Delta_2 & 0 \\ 0 & \gamma I \end{bmatrix}^{-1}.$$

Using the parameterization (4.1) of T_K, we see that

$$T_{\widehat{K}} = \begin{bmatrix} G\Delta_2^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} F \\ I \end{bmatrix} (\gamma^{-1} Q) \begin{bmatrix} \gamma J \Delta_2^{-1} & I \end{bmatrix}.$$

Notice that $T_{\widehat{K}}$ itself has the form of a transfer operator described in (4.1); it is the transfer operator with Youla parameter $\widehat{Q} = \gamma^{-1} Q$ in the system

$$\widehat{s} = \widehat{F}\widehat{u} + \widehat{G}\widehat{w}, \quad \widehat{y} = \widehat{H}\widehat{u} + \widehat{J}\widehat{w} + \widehat{v}, \tag{4.15}$$

where we define $\widehat{F} = F, \widehat{G} = G\Delta_2^{-1}, \widehat{H} = \gamma H, \widehat{J} = \gamma J \Delta_2^{-1}$. It is now clear that a controller K satisfying (4.13) exists if and only if there exists a controller $\widehat{K}$ in the system (4.15) such that $\|T_{\widehat{K}}\| < 1$. If such a controller $\widehat{K}$ exists, then we can easily recover K from $\widehat{K}$ by setting $K = \gamma \widehat{K}$; notice that this is the unique choice of K which is consistent with the relations $\widehat{H} = \gamma H, \widehat{Q} = \gamma^{-1} Q$.

In order to assign state-space structure to $\widehat{F}, \widehat{G}, \widehat{H}, \widehat{J}$, we must first find $\Delta_2(z)$. Let Δ be the unique causal and causally invertible operator such that

$$I + FF^* = \Delta\Delta^*.$$

Theorem 2 shows that

$$\Delta(z) = (I + L(zI - A)^{-1} K)\Sigma^{1/2}.$$

It follows that $\Delta^{-1}(z)G(z)$ is given by

$$\Delta^{-1}(z)G(z) = \Sigma^{-1/2} L(zI - (A - KL))^{-1} B_w.$$

Define

$$\widetilde{L} = \begin{bmatrix} \Sigma^{-1/2} L & 0 \\ 0 & \gamma I_p \end{bmatrix}, \quad \widetilde{S} = \begin{bmatrix} 0 \\ \gamma^2 I_p \end{bmatrix}, \quad \widetilde{A} = \begin{bmatrix} A - KL & 0 \\ 0 & 0 \end{bmatrix}, \quad \widetilde{B}_w = \begin{bmatrix} B_w \\ -I_p \end{bmatrix}. \tag{4.16}$$

Notice that we can rewrite

$$\begin{bmatrix} \gamma^2 D_p^*(z^{-*}) D_p(z) + G^* \Delta^{-*} \Delta^{-1} G & 0 \\ 0 & \gamma^2 I \end{bmatrix}$$

as

$$\begin{bmatrix} \widetilde{B}_w^*(z^{-*} I - \widetilde{A})^{-*} & I \end{bmatrix} \begin{bmatrix} \widetilde{L}^* \widetilde{L} & \widetilde{S} \\ \widetilde{S}^* & \gamma^2 I \end{bmatrix} \begin{bmatrix} (zI - \widetilde{A})^{-1} \widetilde{B}_w \\ I \end{bmatrix}.$$

Applying Lemma 1, we see that this equals

$$\begin{bmatrix} \widetilde{B}_w^*(z^{-*} I - \widetilde{A})^{-*} & I \end{bmatrix} \Lambda_2(P_2) \begin{bmatrix} (zI - \widetilde{A})^{-1} \widetilde{B}_w \\ I \end{bmatrix},$$

where P_2 is an arbitrary Hermitian matrix and we define

$$\Lambda_2(P_2) = \begin{bmatrix} \widetilde{L}^* \widetilde{L} - P_2 + \widetilde{A}^* P_2 \widetilde{A} & \widetilde{S} + \widetilde{A}^* P_2 \widetilde{B}_w \\ \widetilde{S}^* + \widetilde{B}_w^* P_2 \widetilde{A} & \gamma^2 I + \widetilde{B}_w^* P_2 \widetilde{B}_w \end{bmatrix}.$$

Notice that the $\Lambda_2(P_2)$ can be factored as

$$\begin{bmatrix} I & K_2^*(P_2) \\ 0 & I \end{bmatrix} \begin{bmatrix} \Gamma_2(P_2) & 0 \\ 0 & \Sigma_2(P_2) \end{bmatrix} \begin{bmatrix} I & 0 \\ K_2(P_2) & I \end{bmatrix},$$

where we define

$$\Gamma_2(P_2) = \widetilde{L}^* \widetilde{L} - P_2 + \widetilde{A}^* P_2 \widetilde{A} - K_2^*(P_2) \Sigma_2 K_2(P_2),$$

$$K_2(P_2) = \Sigma_2^{-1}(P_2)(\widetilde{S}^* + \widetilde{B}_w^* P_2 \widetilde{A}), \quad \Sigma_2(P_2) = \gamma^2 I + \widetilde{B}_w^* P_2 \widetilde{B}_w. \tag{4.17}$$

It is clear that $(\widetilde{A}, \widetilde{B}_w)$ is stabilizable (this follows from the stability of $A - KL$), therefore the Riccati equation $\Gamma_2(P_2) = 0$ has a unique stabilizing solution (Theorem E.6.2 in [KSH00]). Suppose P_2 is chosen to be this solution, and define $K_2 = K_2(P_2)$, $\Sigma_2 = \Sigma_2(P_2)$. We immediately obtain the factorization (4.14), where we define

$$\Delta_2(z) = \Sigma_2^{1/2}(I + K_2(zI - \widetilde{A})^{-1} \widetilde{B}_w). \tag{4.18}$$

We have

$$\Delta_2^{-1}(z) = (I - K_2(zI - (\widetilde{A} - \widetilde{B}_w K_2))^{-1} \widetilde{B}_w) \Sigma_2^{-1/2}.$$

We note that $\widetilde{A} - \widetilde{B}_w K_2$ is stable and hence $\Delta_2^{-1}(z)$ is causal and bounded since its poles are strictly contained in the unit circle. Define

$$\begin{cases} \widehat{A} = \begin{bmatrix} A & -B_w K_2 \\ 0 & \widetilde{A} - \widetilde{B}_w K_2 \end{bmatrix} \\ \widehat{B}_u = \begin{bmatrix} B_u \\ 0 \end{bmatrix}, \\ \widehat{B}_w = \begin{bmatrix} B_w \Sigma_2^{-1/2} \\ \widetilde{B}_w \Sigma_2^{-1/2} \end{bmatrix}, \\ \widehat{C} = \begin{bmatrix} \gamma C & 0 \end{bmatrix}, \\ \widehat{L} = \begin{bmatrix} L & 0 \end{bmatrix}. \end{cases} \tag{4.19}$$

Recall that $\widehat{F} = F, \widehat{G} = G\Delta_2^{-1}, \widehat{H} = \gamma H, \widehat{J} = \gamma J \Delta_2^{-1}$; it follows that $\widehat{F}, \widehat{G}, \widehat{H}, \widehat{J}$ are given by

$$\widehat{F}(z) = \widehat{L}(zI - \widehat{A})^{-1}\widehat{B}_u, \quad \widehat{G}(z) = \widehat{L}(zI - \widehat{A})\widehat{B}_w,$$
$$\widehat{H}(z) = \widehat{C}(zI - \widehat{A})^{-1}\widehat{B}_u, \quad \widehat{J}(z) = \widehat{C}(zI - \widehat{A})\widehat{B}_w.$$

□

Chapter 5

CONNECTIONS TO ONLINE LEARNING

In this section we demonstrate a surprising connection between regret-optimal control and online learning. Specifically, we show that the competitive controller is closely approximated by disturbance-action-control (DAC) policies, which generalize linear state-feedback policies. A DAC policy has the form

$$u_t = Kx_t + \sum_{i=1}^{H} M^{[i-1]} w_{t-1},$$

where K is a stabilizing state-feedback controller and $M = (M^{[0]}, \ldots, M^{[H-1]})$ is a series of geometrically decreasing weights. The class of DAC policies has attracted much recent attention in the online learning community [Aga+19; FS20; Sim20; CH21] due to the fact that that for any fixed K, the states generated by a DAC policy have a linear dependence on the weights M; it follows that the problem of selecting the weights to minimize cost can be formulated as a convex program. In particular, [Aga+19] showed that the weights can be optimized online via a gradient-based algorithm using a reduction to online convex optimization with memory. This problem, introduced in [AHM15], is a generalization of OCO where the costs incurred by the learner depend not only on the learner's most recent decision but rather on the learner's past H decisions.

Using this approximation of the competitive controller by DAC policies, we show that online algorithms with sublinear policy regret against DAC policies can be instantiated so as to retain sublinear policy regret while also achieving an approximate competitive ratio guarantee on bounded disturbance sequences. The proof can be summarized as follows. We first construct a DAC policy $\tilde{\pi}_c$ which ε-approximates the strictly causal competitive policy π_c, in the sense that

$$J_T(\tilde{\pi}_c, w) - J_T(\pi_c, w) \leq \varepsilon$$

for all bounded disturbances w. Our construction of $\tilde{\pi}_c$ involves rewriting the competitive controller as the sum of a stabilizing component $\widehat{K}_0$ and a linear combination of all previous disturbances; the weights on the most recent H disturbances are used to instantiate $\tilde{\pi}_c$. We then observe that an algorithm which obtains sublinear policy regret against a DAC class containing $\tilde{\pi}_c$ will automatically attain the optimal

competitive ratio of the strictly causal competitive controller π_c, up to a sublinear correction term. Surprisingly, these results can be extended even to the "adaptive control" setting, where the online algorithm does not know the system dynamics ahead of time and must perform online system identification.

Our results shows that combining regret-optimal controllers, which are derived using H_∞ control, can be fruitfully combined with gradient-based algorithms from online learning to give new performance guarantees in control. While the results we present here are stated in terms of the competitive controller, we expect that analogous results can be derived for other flavors of regret-optimal control using similar techniques.

5.1 Approximation of the competitive controller by DAC policies

We prove that we can find a DAC policy which generates a sequence of states and control actions which closely tracks the sequence of states and control actions generated by the optimal competitive policy by taking the history H of the DAC to be sufficiently large, and choosing the stabilizing component and weights appropriately. Before we state our approximation result, we introduce some convenient notation. We introduce the partition

$$\widehat{K} = \begin{bmatrix} \widehat{K}_0 & \widehat{K}_1 \end{bmatrix}$$

of $\widehat{K}$ into two $n \times n$ block matrices, where $\widehat{K}$ is the state-feedback matrix appearing in the strictly causal controller we obtained in Theorem 6. Recall that $\widehat{A} + \widehat{B}\widehat{K}$ is stable; this matrix is block upper-triangular, and the matrix in the (1, 1) block is $A + B\widehat{K}_0$. It follows that $A + B\widehat{K}_0$ is also stable, and hence there exist matrices S_1, L_1 such

$$A + B\widehat{K}_0 = S_1 L_1 S_1^{-1}$$

and $\|L_1\| < 1$. Similarly, there exist matrices S_2, L_2 such that $A - KQ^{1/2} = S_2 L_2 S_2^{-1}$ and $\|L_2\| < 1$.

We now show that the competitive controller can be arbitrarily well-approximated by DAC policies:

Theorem 15. *Fix a time horizon T and a disturbance bound W. Define*

$$\kappa = \max(1, \widehat{K}_0, \widehat{K}_1, \|B_u\|, \|B_w\|, \|S_1\|\|S_1^{-1}\|, \|S_2\|\|S_2^{-1}\|)$$

and

$$\delta = 1 - \max(1/2, \|L_1\|, \|L_2\|).$$

For any $\varepsilon > 0$, set

$$H = \frac{1}{\log(1-\delta/2)} \log\left(\frac{988W^2\kappa^{12}\max(1, \|Q\|^2)}{\delta^4\varepsilon}T\right)$$

and define $\mathcal{M}$ be the set of $(H, \kappa^2(1+\|Q^{1/2}\|), \delta/2)$-DAC policies with stabilizing component $\widehat{K}_0$. Let π_c be the strictly causal competitive controller described in Theorem 6. There exists a DAC policy $\tilde{\pi}_c \in \mathcal{M}$ such that

$$J_T(\tilde{\pi}_c, w) - J_T(\pi_c, w) \leq \varepsilon$$

for all disturbance sequences $w = (w_0, \ldots, w_T)$ such that $\|w\|_\infty \leq W$.

Proof. Unrolling the recursive definition of ν_t in Theorem 6, we see that

$$\nu_t = \sum_{i=1}^{t} (A - KQ^{1/2})^{i-1} B_w w_{t-i}.$$

Let $\{x_t\}_{t=0}^T$ be the state sequence generated by the strictly causal competitive control policy π_c in the original m-dimensional system and let $\{\xi_t\}_{t=0}^T$ is the sequence of states generated by π_c in the $2m$-dimensional synthetic system. Recall that $\widehat{w}_t = \Sigma^{-1/2}Q^{1/2}\nu_t$ and $u_t = \hat{K}\xi_t$, so Theorem 6 implies that the first n states of ξ_t are precisely $x_t - \nu_t$. We decompose x_t as a linear combination of the disturbance terms:

$$\begin{aligned}
x_{t+1} &= Ax_t + B_u\widehat{K}\xi_t + B_w w_t \\
&= Ax_t + B_u \begin{bmatrix} \widehat{K}_0 & \widehat{K}_1 \end{bmatrix} \begin{bmatrix} x_t - \nu_t \\ \Sigma^{-1/2}Q^{1/2}\nu_t \end{bmatrix} + B_w w_t \\
&= (A + B_u\widehat{K}_0)x_t + B_u(\widehat{K}_1\Sigma^{-1/2}Q^{1/2} - \widehat{K}_0)\nu_t + B_w w_t \\
&= \sum_{i=0}^{t} (A + B_u\widehat{K}_0)^i \left(B_w w_{t-i} + B_u(\widehat{K}_1\Sigma^{-1/2}Q^{1/2} - \widehat{K}_0)\nu_{t-i}\right) \\
&= \sum_{i=0}^{t} (A + B_u\widehat{K}_0)^i \left(B_w w_{t-i} + B_u \sum_{j=1}^{t-i} M^{[j-1]} B_w w_{t-i-j}\right) \\
&= \sum_{i=0}^{t} \left((A + B_u\widehat{K}_0)^i + \sum_{j=1}^{i} (A + B_u\widehat{K}_0)^{i-j} B_u M^{[j-1]}\right) B_w w_{t-i},
\end{aligned}$$

where we define the weights

$$M^{[i-1]} = (\widehat{K}_1\Sigma^{-1/2}Q^{1/2} - \widehat{K}_0)(A - KQ^{1/2})^{i-1}$$

for all $i \geq 1$. We now describe a DAC policy $\tilde{\pi}_c$ which approximates π_c; recall that every DAC policy is parameterized by a stabilizing controller and a set of H weights. We take the stabilizing controller to be $\widehat{K}_0$ and the set of weights to be $M = (M^{[0]}, \ldots, M^{[H-1]})$. The control action selected by $\tilde{\pi}_c$ in each timestep is therefore

$$\tilde{u}_t = \widehat{K}_0 \tilde{x}_t + \sum_{i=1}^{H} M^{[i-1]} B_w w_{t-i}, \tag{5.1}$$

where $\{\tilde{x}_t\}_{t=0}^{T}$ is the state sequence generated by $\tilde{\pi}_c$.

We now show that the weights $\{M^{[i-1]}\}_{i=1}^{\infty}$ decay geometrically in time. Recall that $A + B\widehat{K}_0 = S_1 L_1 S_1^{-1}$, where $\|S_1\|\|S_1^{-1}\| \leq \kappa$ and $\|L_1\| \leq 1 - \delta$. Similarly, $A - KQ^{1/2} = S_2 L_2 S_2^{-1}$, where $\|S_2\|\|S_2^{-1}\| \leq \kappa$ and $\|L_2\| \leq 1 - \delta$. It follows that

$$\max\left(\|(A + B\widehat{K}_0)^{i-1}\|, (A - KQ^{1/2})^{i-1})\|\right) \leq \kappa(1-\delta)^{i-1}. \tag{5.2}$$

Recall that $\Sigma = I + Q^{1/2} P Q^{1/2} \geq I$, therefore $\|\Sigma^{-1/2}\| \leq 1$. It follows that

$$\begin{aligned} \|\widehat{K}_1 \Sigma^{-1/2} Q^{1/2} - \widehat{K}_0\| &\leq \|\widehat{K}_1 \Sigma^{-1/2} Q^{1/2}\| + \|\widehat{K}_0\| \\ &\leq \kappa(1 + \|Q^{1/2}\|) \end{aligned} \tag{5.3}$$

Using (5.2) and (5.3) together, we immediately obtain

$$\|M^{[i-1]}\| \leq \kappa^2 (1 + \|Q^{1/2}\|)(1-\delta)^{i-1}. \tag{5.4}$$

We now bound the distance between x_t and $\tilde{x}_t$. Plugging our choice of $\tilde{u}_t$ given in (5.1) into the dynamics

$$x_{t+1} = A x_t + B_u u_t + B_w w_t,$$

we see that

$$\begin{aligned} \tilde{x}_{t+1} &= A\tilde{x}_t + B_u \left(\widehat{K}_0 \tilde{x}_t + \sum_{i=1}^{H} M^{[i-1]} B_w w_{t-i} \right) + B_w w_t \\ &= \sum_{i=0}^{t} (A + B_u \widehat{K}_0)^i \left(B_w w_{t-i} + B_u \sum_{j=1}^{H} M^{[j-1]} B_w w_{t-i-j} \right) \\ &= \sum_{i=0}^{t} \left((A + B_u \widehat{K}_0)^i + \sum_{j=1}^{\min\{H,i\}} (A + B_u \widehat{K}_0)^{i-j} B_u M^{[j-1]} \right) B_w w_{t-i}. \end{aligned}$$

Applying (5.2) and (5.4) we obtain

$$\begin{aligned}
\|x_{t+1} - \tilde{x}_{t+1}\| &\leq \sum_{i=H+1}^{t} \sum_{j=H+1}^{i} \|(A + B_u \widehat{K}_0)^{i-j}\| \|B_u\| \|M^{[j-1]}\| \|B_w\| W \\
&\leq W\kappa^5 (1 + \|Q^{1/2}\|) \sum_{i=H+1}^{t} \sum_{j=H+1}^{i} (1-\delta)^{i-1} \\
&\leq W\kappa^5 (1 + \|Q^{1/2}\|) \sum_{i \geq H+1} i(1-\delta)^{i-1} \\
&\leq \frac{4W\kappa^5 (1 + \|Q^{1/2}\|)}{\delta} \sum_{i \geq H+1} (1-\delta/2)^i, \\
&\leq \frac{8W\kappa^5 (1 + \|Q^{1/2}\|)}{\delta^2} (1-\delta/2)^H \qquad (5.5)
\end{aligned}$$

where in the penultimate line we applied Lemma 4 in Appendix A .

We now bound the distance between u_t and $\tilde{u}_t$. Unrolling the dynamics, we see that

$$\begin{aligned}
u_t &= \widehat{K} \xi_t \\
&= \widehat{K}_0 x_t + (\widehat{K}_1 \Sigma^{-1/2} Q^{1/2} - \widehat{K}_0) \nu_t \\
&= \widehat{K}_0 x_t + \sum_{i=1}^{t-1} (\widehat{K}_1 \Sigma^{-1/2} Q^{1/2} - \widehat{K}_0)(A - KQ^{1/2})^{i-1} w_{t-i} \\
&= \widehat{K}_0 x_t + \sum_{i=1}^{t-1} M^{[i-1]} w_{t-i}.
\end{aligned}$$

The definition of $\tilde{u}_t$ given in (5.1) and the bounds (5.4) and (5.5) imply that

$$\begin{aligned}
\|u_t - \tilde{u}_t\| &\leq \|\widehat{K}_0\| \|x_t - \tilde{x}_t\| + \sum_{i=H+1}^{t-1} \|M^{[i-1]}\| w_{t-i} \\
&\leq \frac{9W\kappa^6 (1 + \|Q^{1/2}\|)}{\delta^2} (1-\delta/2)^H,
\end{aligned}$$

where we used the fact that $\delta < 1$ and $\kappa \geq \max(1, \|\widehat{K}_0\|)$.

We now show that by taking H to be sufficiently large we can ensure that

$$J_T(\tilde{\pi}_c, w) - J_T(\pi_c, w) \leq \varepsilon.$$

Lemma 5 shows that

$$\max(\|x_t\|, \|u_t\|) \leq \frac{3W\kappa^5 (1 + \|Q^{1/2}\|)}{\delta^2}.$$

We put the pieces together to bound the difference in costs incurred by the competitive policy and our DAC approximation in each timestep. Using the easily-verified inequality $\|x\|^2 - \|y\|^2 \leq (\|x - y\|)(\|2x\| + \|x - y\|)$ and the fact that $\kappa \geq 1$, $1 - \delta/2 < 1$, we see that

$$\begin{aligned}\|Q^{1/2}x_t\|^2 - \|Q^{1/2}\tilde{x}_t\|^2 &\leq \|Q\|\|x_t - \tilde{x}_t\|\,(2\|x_t\| + \|x_t - \tilde{x}_t\|) \\ &\leq \frac{112W^2\kappa^{10}(1 + \|Q^{1/2}\|)^2\|Q\|}{\delta^4}(1 - \delta/2)^H\end{aligned}$$

Similarly,

$$\begin{aligned}\|u_t\|^2 - \|\tilde{u}_t\|^2 &\leq \|u_t - \tilde{u}_t\|\,(2\|u_t\| + \|u_t - \tilde{u}_t\|) \\ &\leq \frac{135W^2\kappa^{12}(1 + \|Q^{1/2}\|)^2}{\delta^4}(1 - \delta/2)^H.\end{aligned}$$

We observe that

$$(1 + \|Q^{1/2}|)^2 \max(1, \|Q\|) \leq 4\max(1, \|Q\|^2).$$

The difference in aggregate cost is therefore

$$\begin{aligned}J_T(\tilde{\pi}_c, w) - J_T(\pi_c, w) &= \sum_{t=1}^{T}(x_t^\top Q x_t + \|u_t\|^2 - (\tilde{x}_t^\top Q\tilde{x}_t + \|\tilde{u}_t\|^2) \\ &\leq \frac{988W^2\kappa^{12}\max(1, \|Q\|^2)}{\delta^4}(1 - \delta/2)^H T.\end{aligned}$$

Taking

$$H \geq \frac{1}{\log(1 - \delta/2)}\log\left(\frac{988W^2\kappa^{12}\max(1, \|Q\|^2)}{\delta^4\varepsilon}T\right)$$

is thus sufficient to guarantee that

$$J_T(\tilde{\pi}_c, w) - J_T(\pi_c, w) \leq \varepsilon.$$

□

We now show that this result implies best-of-both-worlds.

5.2 Best of both worlds: sublinear policy regret implies approximate competitive ratio

We now show that any online learning algorithm $\mathcal{A}$ which minimizes regret relative to the class $\mathcal{M}$ of DAC policies described in Theorem 15 automatically achieves the best of both worlds: sublinear regret against the best DAC policy selected in hindsight from $\mathcal{M}$, and optimal competitive ratio, up to a sublinear regret term:

Theorem 16. *Fix $\varepsilon, W > 0$ and the system parameters (A, B_u, B_w, Q), and let $\mathcal{M}$ be defined as in Theorem 15. Let $\mathcal{A}$ be any algorithm such that*

$$J_T(\mathcal{A}, w) - \min_{\pi \in \mathcal{M}} J_T(\pi, w) \leq R(T)$$

for all disturbances w such that $\|w\|_\infty \leq W$, where $R(T)$ is a bound on the regret incurred by $\mathcal{A}$ over the time horizon T. Then $\mathcal{A}$ also satisfies the "approximate competitive ratio" condition

$$J_T(\mathcal{A}, w) < \gamma^2 \cdot J_T(\pi_0, w) + R(T) + O(1),$$

where γ^2 is the optimal competitive ratio attained by the strictly causal competitive controller.

Before turning to the proof, we note that the approximate competitive ratio guarantee described in Theorem 16 is substantially weaker than that offered by the standard competitive controller π_c. Recall that π_c offers the following guarantee:

$$\sup_{w \in \ell_2} \frac{J(\pi_c, w)}{J(\pi_0, w)} < \gamma^2.$$

Note that this holds for all disturbances w, including those which grow over time. As we showed in Theorem 7, this guarantee can also be obtained in time-varying systems. By contrast, Theorem 16 holds only for bounded disturbances and LTI systems. Furthermore, the competitive ratio bound described in Theorem 16 is not necessarily constant in the time horizon T; it says that

$$\sup_{\|w\|_\infty < W} \frac{J_T(\mathcal{A}, w)}{J_T(\pi_0, w)} < \gamma^2 + \sup_{\|w\|_\infty < W} \frac{R(T) + O(1)}{J_T(\pi_0, w)}.$$

This bound on the competitive ratio can grow rapidly as $T \to \infty$, provided that w is constructed such that $J_T(\pi_0, w) \ll R(T)$. It is an open question whether there exist control algorithms with constant competitive ratio and sublinear policy regret.

We now present the proof of Theorem 16.

Proof. By Theorem 15, we know that there exists a DAC policy $\tilde{\pi}_c \in \mathcal{M}$ such that

$$J_T(\tilde{\pi}_c, w) - J_T(\pi_c, w) \leq \varepsilon,$$

where π_c is the strictly causal competitive controller described in Theorem 6. The cost incurred by $\mathcal{A}$ can be bounded as follows:

$$\begin{aligned} J_T(\mathcal{A}, w) &\leq \min_{\pi^\star \in \mathcal{M}} J_T(\pi^\star, w) + R(W, T) \\ &\leq J_T(\tilde{\pi}_c, w) + R(T) \\ &\leq J_T(\pi_c, w) + R(T) + \varepsilon \\ &\leq J(\pi_c, w) + R(T) + \varepsilon \\ &\leq \gamma^2 \cdot J(\pi_0, w) + R(T) + \varepsilon \\ &\leq \gamma^2 \cdot J_T(\pi_0, w) + R(T) + O(1), \end{aligned}$$

where in the penultimate line we applied the competitive ratio bound of the infinite-horizon strictly causal competitive controller, and in the last line we used the fact that ε is constant with respect to T and used the following bound which relates the finite-horizon cost of the optimal noncausal controller to its infinite-horizon cost:

$$J(\pi_0, w) \leq J_T(\pi_0, w) + O(1).$$

This bound can be proven as follows. The infinite-horizon cost incurred by π_0 on the disturbance sequence w is the sum of the cost incurred by π_0 up to time T and the cost incurred by π_0 from time $T+1$ to infinity. Recall that we define $w_t = 0$ for all $t \geq T+1$. Using the characterization of π_0 we obtained in Theorem 3, we see that π_0 simply selects the H_2-optimal action starting at time $T+1$. The H_2 policy is stabilizing and the state x_{T+1} generated by the optimal noncausal policy has norm which is $O(1)$, so the residual $J(\pi_0, w) - J_T(\pi_0, w)$ is also $O(1)$. □

By leveraging recently proposed online control algorithms with sublinear policy regret against DAC classes, we immediately obtain several concrete best-of-both-worlds results as corollaries of Theorem 16. We begin by considering the case when the learner knows the linear system (A, B_u, B_w); in this case, we can apply the algorithm from [Sim20], which has $O(\text{polylog}(T))$ regret against the best DAC policy selected in hindsight from $\mathcal{M}$. This algorithm hence exhibit the following best-of-both-worlds behavior, where we suppress dependence on ϵ, W, and the system parameters and focus on how the cost incurred by the algorithm scales in the time horizon:

Corollary 1. *Fix $\varepsilon, W > 0$ and the system parameters (A, B_u, B_w), and let $\mathcal{M}$ be defined as in Theorem 15. There exists a computationally efficient online control algorithm $\mathcal{A}$ which, given the system parameters (A, B_u, B_w), simultaneously achieves the following performance guarantees:*

1. *(Approximate competitive ratio) The cost of $\mathcal{A}$ satisfies*

$$J_T(\mathcal{A}, w) < \gamma^2 \cdot J_T(\pi_0, w) + O\left(\text{polylog}(T)\right)$$

for all disturbances w such that $\|w\|_\infty < W$, where γ^2 is the optimal competitive ratio attained by the strictly causal competitive controller.

2. *(Sublinear regret) The cost of $\mathcal{A}$ satisfies*

$$J_T(\mathcal{A}, w) < \min_{\pi \in \mathcal{M}} J_T(\pi, w) + O\left(\text{polylog}(T)\right)$$

for all disturbances w such that $\|w\|_\infty < W$.

We next consider the "adaptive control" setting, where the online control algorithm does not know the system dynamics. The algorithm proposed in [CH21] achieves $O(\sqrt{T})$ regret even when the system dynamics are unknown; we note that this is exponentially worse than the $O(\text{polylog}(T))$ regret which is attainable when the dynamics are known. The algorithm operates in two phases; in the first phase it performs system identification by exciting the system using the control inputs and then estimating the system parameters by observing the system response to the excitation. In the second phase, it uses the estimates obtained in the first phase to construct an H_2 controller in the estimated system; this H_2 controller is stabilizing in the true system provided that the estimate is sufficiently accurate. The $O(\sqrt{T})$ regret bound leads to the following corollary of Theorem 16, where we again suppress dependence on ϵ, W, and the system parameters and focus on how the cost incurred by the algorithm scales in the time horizon:

Corollary 2. *Fix $\varepsilon, W > 0$ and the system parameters (A, B_u, B_w), and let $\mathcal{M}$ be defined as in Theorem 15. There exists a computationally efficient online control algorithm $\mathcal{A}$ which, without knowing the system parameters simultaneously achieves the following performance guarantees:*

1. *(Approximate competitive ratio) The cost of $\mathcal{A}$ satisfies*

$$J_T(\mathcal{A}, w) < \gamma^2 \cdot J_T(\pi_0, w) + O\left(\sqrt{T}\right)$$

for all disturbances w such that $\|w\|_\infty < W$, where γ^2 is the optimal competitive ratio attained by the strictly causal competitive controller.

2. *(Sublinear regret) The cost of $\mathcal{A}$ satisfies*

$$J_T(\mathcal{A}, w) < \min_{\pi \in \mathcal{M}} J_T(\pi, w) + O\left(\sqrt{T}\right)$$

for all disturbances w such that $\|w\|_\infty < W$.

We note that [Aga+19] showed that algorithms with low regret relative to a DAC policy class $\mathcal{M}$ automatically also achieve low regret relative to the class of stabilizing linear state-feedback policies $\mathcal{K}$; this implies that Corollaries 1 and 2 can also be translated into statements about regret relative to $\mathcal{K}$. We refer to their paper for details.

Chapter 6

NUMERICAL EXPERIMENTS

6.1 Double Integrator

The double integrator is a simple dynamical system that models the one-dimensional kinematics of a moving object. The states of the system are the object's position and velocity, which are represented by the variables $x \in \mathbb{R}$ and $\dot{x} \in \mathbb{R}$, respectively. The continuous-time dynamics of the system are represented by the Newtonian equation

$$\frac{d}{dt}\begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} = \begin{bmatrix} \dot{x}(t) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ u(t) \end{bmatrix} + \begin{bmatrix} 0 \\ w(t) \end{bmatrix},$$

where $u(t) \in \mathbb{R}$ and $w(t) \in \mathbb{R}$ are the control input and the exogenous disturbance at time t, respectively. The goal of the controller is to stabilize the object by keeping $(x, \dot{x})$ as close to $(0, 0)$ as possible, while using little energy. The discrete-time dynamics are

$$\begin{bmatrix} x_{t+1} \\ \dot{x}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & \delta_t \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} + \begin{bmatrix} 0 \\ \delta_t \end{bmatrix} u_t + \begin{bmatrix} 0 \\ \delta_t \end{bmatrix} w_t,$$

where δ_t is the discretization parameter; in our experiments we set $\delta_t = 0.1$ seconds. In our experiments we take $Q = I_2$ and initialize x and $\dot{x}$ to zero.

We compare the regret-optimal controllers to standard H_2 and H_∞ controllers in the causal, full-information setting. Each of the regret-optimal and H_∞ controllers is associated with a parameter γ which quantifies its performance guarantee:

1. The competitive controller has $\gamma = 2.14$; in other words, the competitive ratio of the competitive controller is 4.58.

2. The energy-optimal controller has $\gamma = 0.63$; in other words, the regret of the energy-optimal controller against the optimal noncausal controller is bounded by 0.4 times the energy of the disturbance.

3. The pathlength-optimal controller has $\gamma = 934.99$; in other words, the regret of the pathlength-optimal controller against the optimal noncausal controller is bounded by 8.74×10^5 times the pathlength of the disturbance.

4. The H_∞-optimal controller has $\gamma = 1.00$, so the cost incurred by the H_∞ controller is at most the energy of the disturbance.

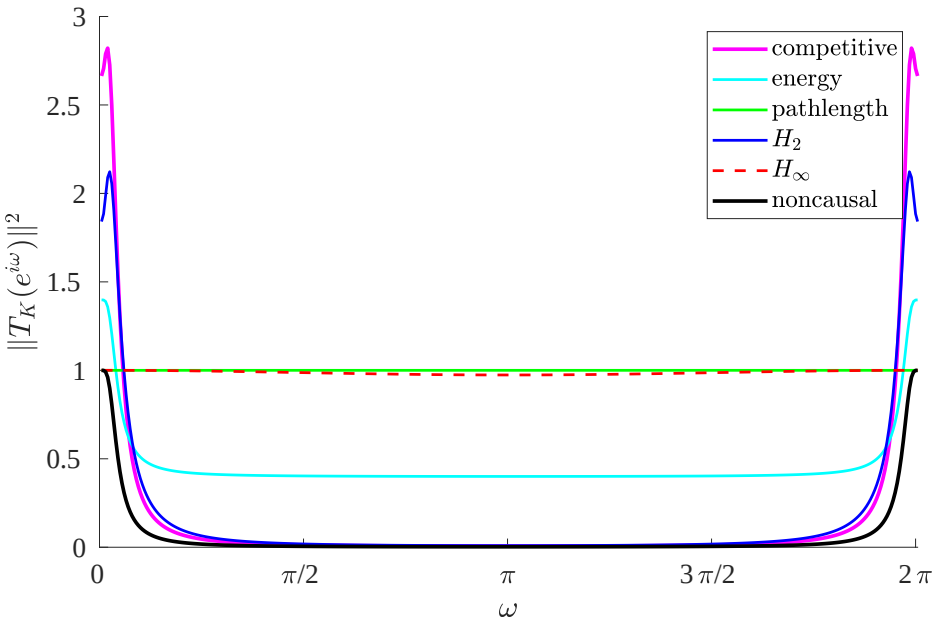

Figure 6.1: Frequency responses of causal controllers in the double integrator system.

Frequency Response

In Figure 6.1, we plot the frequency response $\|T_K(e^{i\omega})\|^2$ at different frequencies ω, where T_K is the transfer operator associated to the controller K and K is alternately taken to be the competitive, energy-optimal, pathlength-optimal, H_2, H_∞, and optimal noncausal controller. The frequency response measures how much energy is transferred to the closed-loop control system from a sinusoidal disturbance with frequency ω. The optimal noncausal controller has the lowest frequency response at all frequencies; its frequency response is a lower bound on the frequency response of any other controller. The H_∞ controller is the causal controller which minimizes

$$\sup_{\omega\in[0,2\pi]} \bar{\sigma}\left(T_K^*(e^{i\omega})T_K(e^{i\omega})\right).$$

We see that the frequency response of the H_∞ controller peaks at $\omega = 0$, where it matches the frequency response of the optimal noncausal controller; all other causal controllers have a higher peak frequency response. The H_2 controller is the causal controller which minimizes

$$\frac{1}{2\pi}\int_{\omega=0}^{2\pi} \mathrm{Tr}\left(T_K^*(e^{i\omega})T_K(e^{i\omega})\right),$$

therefore the H_2 controller is the controller with the smallest area under its frequency response curve; we note that the disturbance in the double integrator system is scalar,

therefore $\bar{\sigma}\left(T_K^*(e^{i\omega})T_K(e^{i\omega})\right) = \mathrm{Tr}\left(T_K^*(e^{i\omega})T_K(e^{i\omega})\right)$. The competitive ratio of a controller K at frequency ω is the spectral radius of the matrix

$$T_K^*(e^{i\omega})T_K(e^{i\omega})\left(T_{K_0}^*(e^{i\omega})T_{K_0}(e^{i\omega})\right)^{-1},$$

where T_K is the transfer operator associated to K and T_{K_0} is the transfer operator associated with the optimal noncausal controller. The disturbance is scalar, so the competitive ratio simplifies to

$$\frac{T_K^*(e^{i\omega})T_K(e^{i\omega})}{T_{K_0}^*(e^{i\omega})T_{K_0}(e^{i\omega})},$$

which is simply the ratio of frequency responses.

We see that the competitive controller has a lower frequency response than all other causal controllers at all frequencies except near $\omega = 0$ and $\omega = 2\pi$, where its frequency response is highest. Intuitively, this is because the competitive controller is constrained to maintain a frequency response within a factor of 4.58 of the frequency response of the optimal noncausal controller, so it has more leeway at frequencies near $\omega = 0$ and $\omega = 2\pi$, where the optimal noncausal controller also has a high frequency response.

The energy-optimal controller minimizes

$$\sup_{\omega\in[0,2\pi]} \bar{\sigma}\left(T_K^*(e^{i\omega})T_K(e^{i\omega}) - T_{K_0}^*(e^{i\omega})T_{K_0}(e^{i\omega})\right).$$

In other words, the energy-optimal controller minimizes the gap between its own frequency response and the frequency response of the optimal noncausal controller. Consulting Figure 6.1, we see that every other causal controller indeed has a larger peak difference in frequency response relative to the optimal noncausal controller.

Time Domain Analysis

We next benchmark our regret-optimal controllers in the double integrator system across a wide variety of disturbances. We first consider an i.i.d. standard Gaussian disturbance. In Figure 6.2 we see that the competitive controller closely tracks the performance of the H_2 controller, while the pathlength-optimal and H_∞ controller incur almost identical cost. Intuitively, an i.i.d. Gaussian disturbance has no correlation across timesteps and hence is expected to have high pathlength, so it is unsurprising that the pathlength-optimal controller performs poorly. We see that the energy-optimal controller incurs cost which is roughly halfway between that of the

H_2 and H_∞ controllers. In Figure 6.3 we plot the costs of only the competitive and H_2 controllers to better illustrate how closely the competitive controller is able to approximate the performance of the H_2 controller.

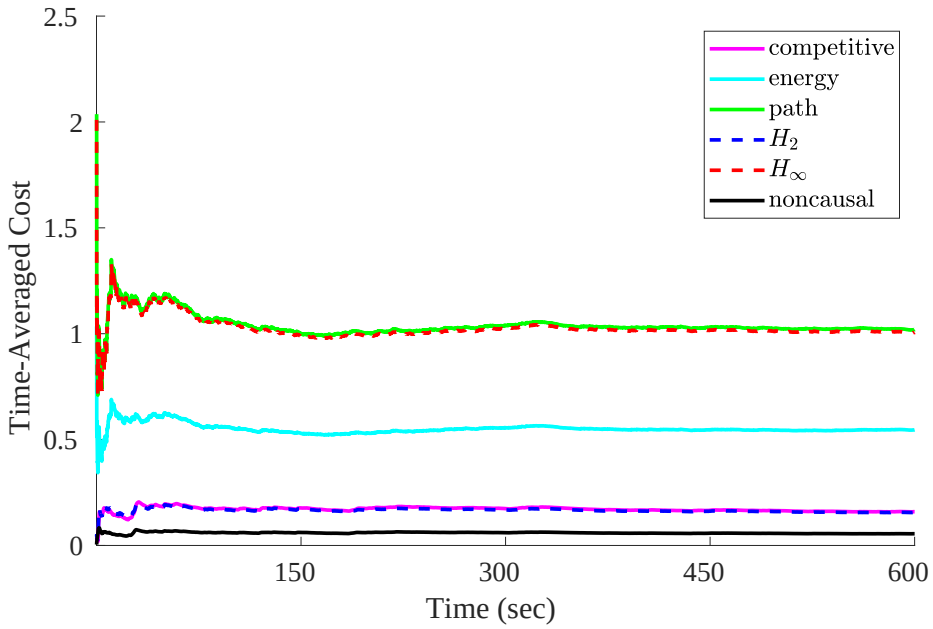

Figure 6.2: Relative performance of causal controllers in the double integrator system driven by an i.i.d. Gaussian disturbance.

We next consider sinusoidal disturbances of frequency $\omega = 0$, i.e., constant disturbances. The frequency responses shown in Figure 6.1 predict that the competitive controller will perform the worst, while the pathlength-optimal and H_∞ controllers will match the performance of the optimal noncausal controller; this prediction is confirmed in Figure 6.4. As expected, the pathlength-optimal controller incurs zero regret, since the pathlength of the disturbance is zero. We note that while the competitive controller incurs the most cost, its cost is still less than 4.58 times the optimal noncausal cost, which is consistent with its competitive ratio bound. We also note that the time-averaged cost of the H_∞ controller converges to 1.00 in steady-state, which is consistent with the fact that its H_∞ gain is 1.00.

We next consider a sinusoidal disturbance of frequency $\omega = \pi$, i.e., the disturbances alternate between +1 and −1 in each timestep. Due to the very large variation in the costs incurred by the causal controllers, we plot the costs on a log scale in Figure 6.5. The frequency responses shown in Figure 6.1 predict that the competitive controller will perform the best out of all causal controllers; this prediction is confirmed

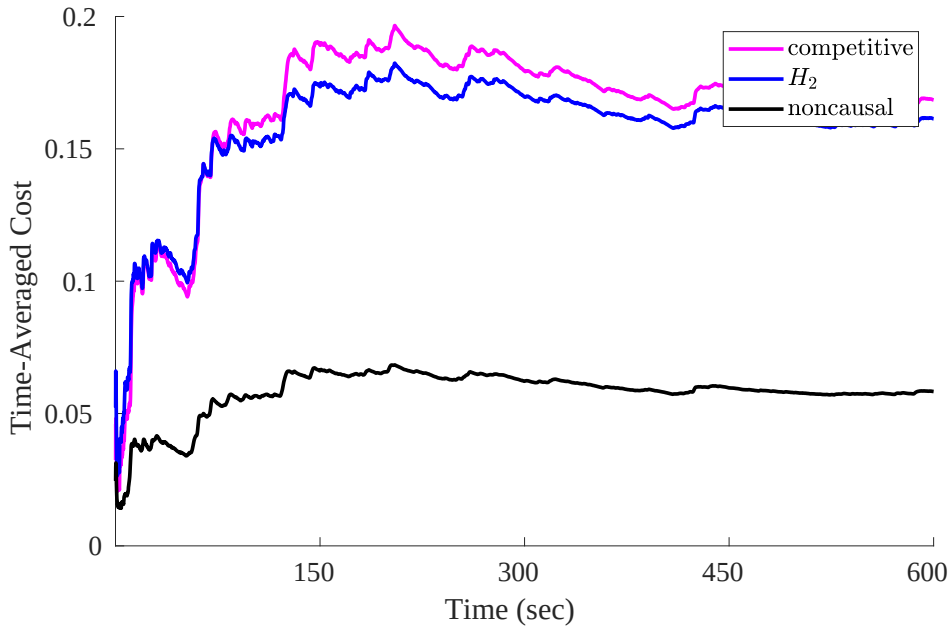

Figure 6.3: Relative performance of the competitive and H_2 controllers in the double integrator system driven by an i.i.d. Gaussian disturbance.

in Figure 6.5. The energy-optimal, pathlength-optimal, and H_∞ controllers incur several hundred times more cost than the competitive controller. This huge variation in costs highlights the value proposition of the competitive controller; while it does not always perform better than the other causal controllers, it is impossible to construct disturbances on which it incurs hundreds of times more cost than the other controllers. The H_∞ controller, however, can be extremely suboptimal for some disturbances.

We note that Figure 6.4 and Figure 6.5 together illustrate the key difference between the competitive and energy-optimal controllers: the gap in performance between the energy-optimal controller and the optimal noncausal controller is roughly constant at all frequencies, whereas the competitive controller closely tracks the noncausal controller when the noncausal controller incurs low cost, but can incur much higher cost when the noncausal controller also incurs high cost. This difference in behavior stems from the fact that the energy-optimal controller seeks to minimize the worst-case *difference* in cost relative to the optimal noncausal controller, whereas the competitive controller instead minimizes the worst-case *ratio* of costs.

We next consider a "Gaussian random walk" disturbance, where a series of standard Gaussian random variables is sampled once, before the experiment begins, and

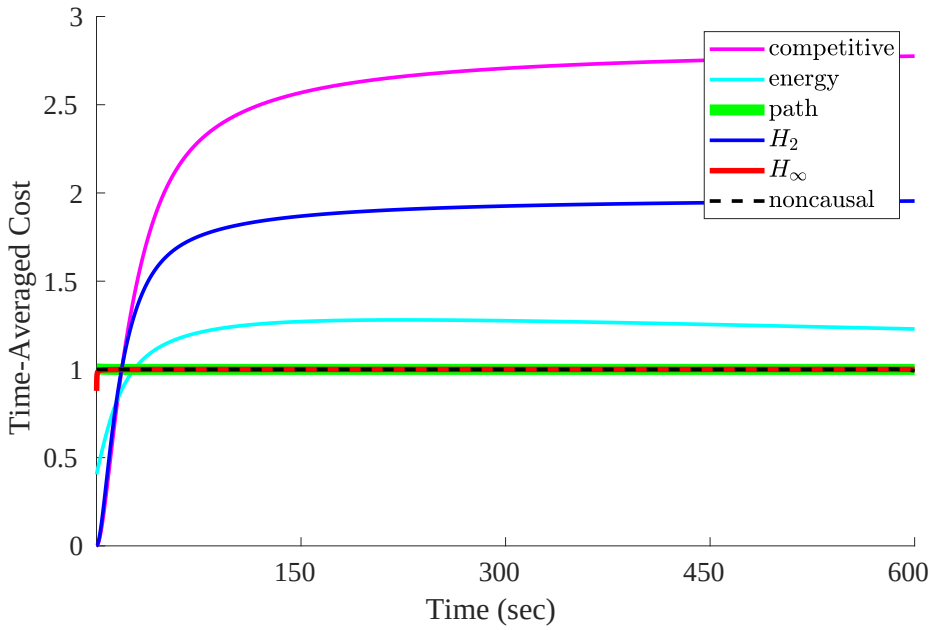

Figure 6.4: Relative performance of causal controllers in the double integrator system driven by a constant disturbance.

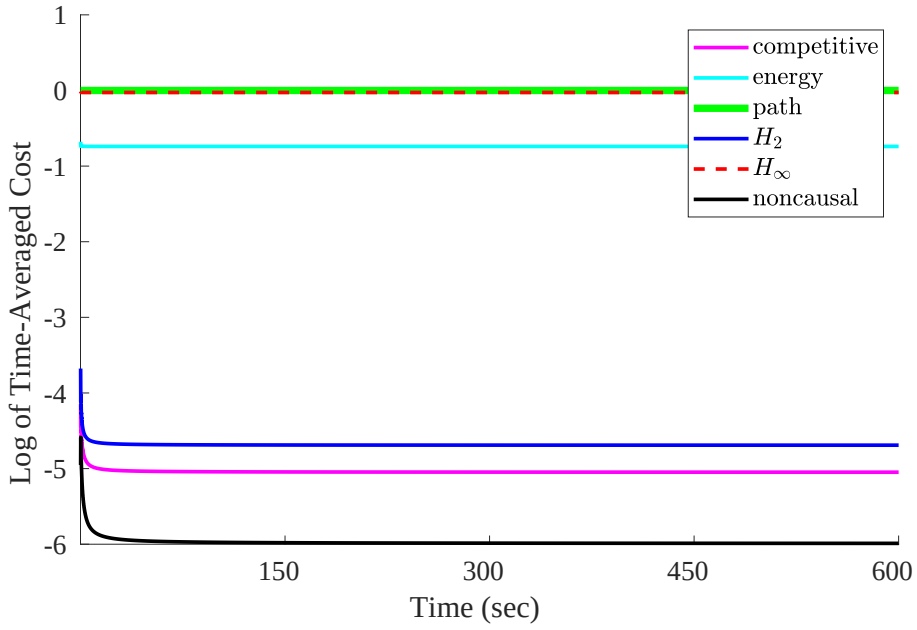

Figure 6.5: Relative performance of controllers in the double integrator system driven by a sinusoidal disturbance with frequency π (log scale).

the disturbance in timestep t is the cumulative sum of the first t variables. In

Figure 6.6, we see that the pathlength-optimal controller almost exactly matches the performance of the optimal noncausal controller; intuitively, this is because the pathlength of a random Gaussian walk is very small compared to its energy. The energy-optimal controller incurs roughly 25% more cost than the optimal noncausal cost, while the competitive controller incurs around 2.7 times more cost, well below the 4.58 factor guaranteed by its competitive ratio bound.

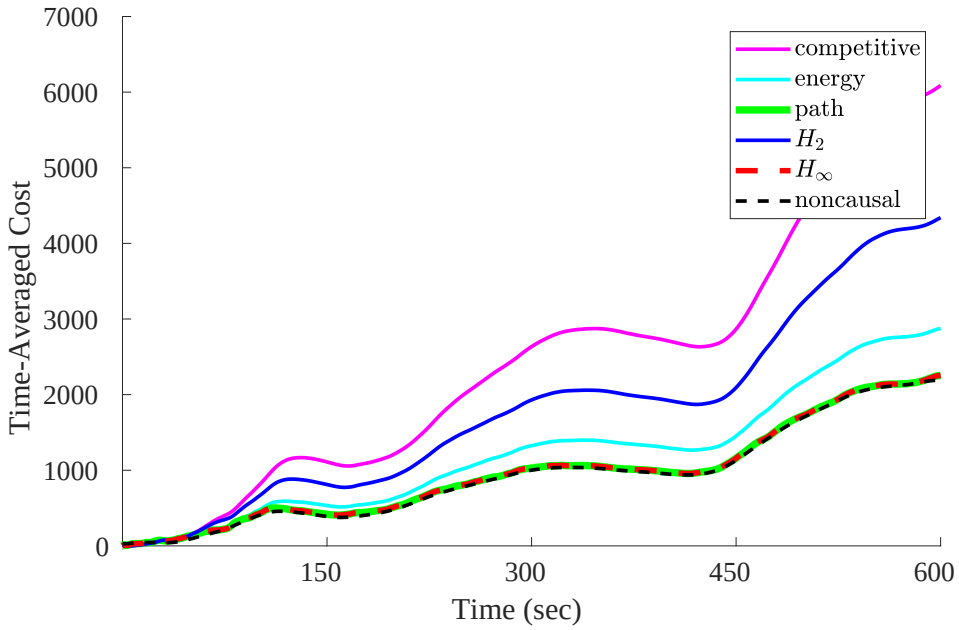

Figure 6.6: Relative performance of the competitive and H_2 controllers in the double integrator system driven by a Gaussian random walk.

6.2 Inverted Pendulum

In this section we study the behavior of the competitive controller in the classic nonlinear inverted pendulum system. This system models the dynamics of a pendulum which attached to one end of a rigid rod; the other end of the rod is attached to a mobile base which the controller can push to the left or the right. The pendulum is initially suspended above the base but is subjected to environmental forces which push it off the vertical axis; the goal of the controller is to dynamically adjust the position of the base to keep the pendulum in the vertical position. This system has two scalar states, θ and $\dot{\theta}$, representing the angle between the rod and the vertical direction the and angular velocity, respectively, and a single scalar control input u, which represents the force applied by the controller to the base. The state $(\theta, \dot{\theta})$

evolves according to the nonlinear evolution equation

$$\frac{d}{dt}\begin{bmatrix}\theta \\ \dot{\theta}\end{bmatrix} = \begin{bmatrix}\dot{\theta} \\ \sin\theta + u\cos\theta + w\cos\theta\end{bmatrix},$$

where we have assumed that the coordinate system is scaled so that all physical parameters of the pendulum system are equal to 1.

Although these dynamics are nonlinear, we can benchmark the competitive controller against the H_2, H_∞, and optimal noncausal optimal controllers using Model Predictive Control (MPC). In the MPC framework, we iteratively linearize the model dynamics around the current state, compute the optimal control signal in the linearized system, and then update the state in the original nonlinear system using this control signal. We emphasize that the noncausal controller is no longer guaranteed to outperform all other controllers in the MPC setting; the decisions it makes in each timestep are globally optimal only in the linearized dynamics, and may be highly suboptimal in the true nonlinear dynamics.

Intuitively, one might expect that our regret-optimal algorithms would outperform H_2 and H_∞ controllers in the MPC setting for the following reason. An MPC controller must account for two kinds of deviation from the reference trajectory. The first is simply the effect of the disturbance, which pushes the state off the reference trajectory in each timestep; this deviation is also present in the linear setting we have studied throughout this thesis. The second is linearization error which arises from the fact that the linearized model the controller uses to select control actions is only a local approximation to the true nonlinear dynamics. This linearization error means that the controller itself may inadvertently add to the deviation from the reference trajectory, since the control signal selected by the controller generally will not be optimal in the actual system, even if it is optimal in the linearized model. Both H_2 and H_∞ controllers are designed with respect to a specific generative model of the disturbance sequence and do not account for linearization error. Even if the disturbance encountered by the controller is generated according to this model, the actions selected by the controller will not be optimal in the nonlinear system. For example, even if the disturbance is generated i.i.d. from a zero-mean Gaussian distribution, the H_2 controller will not select the optimal control signal for the nonlinear system due to linearization error. The regret-optimal controllers, however, do not posit any specific model of the disturbance sequence, and hence may be better able to account for linearization error.

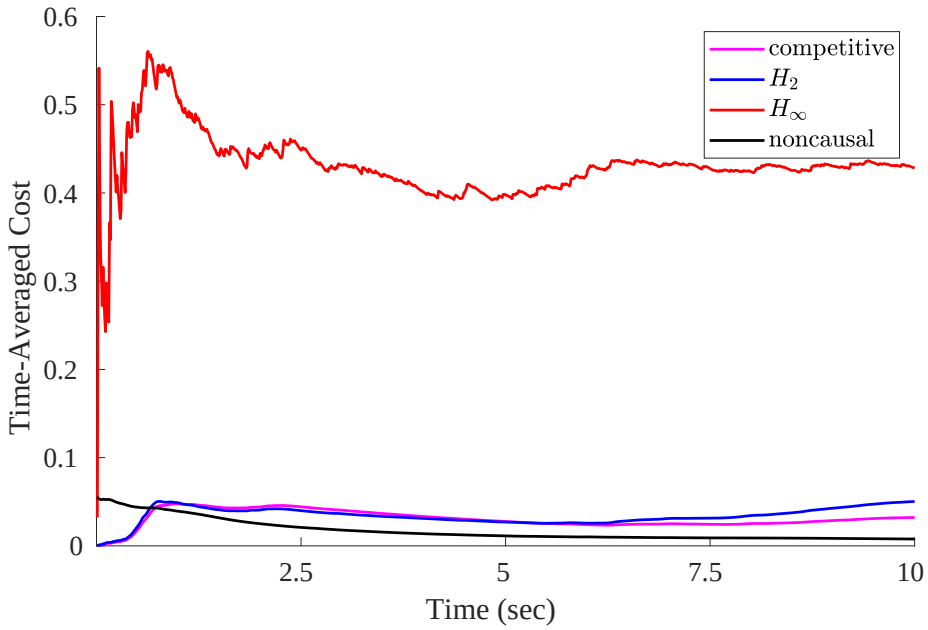

Figure 6.7: Relative performance of causal controllers in an inverted pendulum system driven by an i.i.d. Gaussian disturbance.

We plot the relative costs incurred by the competitive, H_2, H_∞ and optimal noncausal controllers across a wide variety of disturbances. In our experiments we take $Q = I_2$ and initialize θ and $\dot{\theta}$ to zero. We set the discretization parameter $\delta_t = 0.01$ and sample the dynamics at intervals of δ_t over the time interval $[0, 10]$; the total number of timesteps in each experiment is thus 1000.

In Figure 6.7, we plot the costs incurred by the causal controllers when the disturbance is generated i.i.d. from a standard Gaussian distribution in each timestep. We see that the competitive controller performs almost as well as the H_2 controller; this is somewhat surprising, since the H_2 controller is specifically tuned for i.i.d zero-mean noise. The H_∞ controller incurs nearly an order of magnitude more cost than the competitive controller.

We next consider sinusoidal disturbances with amplitude 1 across a range of frequencies. We first consider a sinusoidal disturbance of frequency $\omega = 0$, i.e., a constant disturbance. We see in Figure 6.8 that the H_∞ controller exactly matches the performance of the optimal noncausal controller, while the competitive and H_2 controllers incur an order of magnitude more cost. However, at higher frequencies the competitive controller starts to significantly outperform the H_2 and H_∞ controllers. In Figures 6.9 - 6.11 we plot the relative performance of the causal

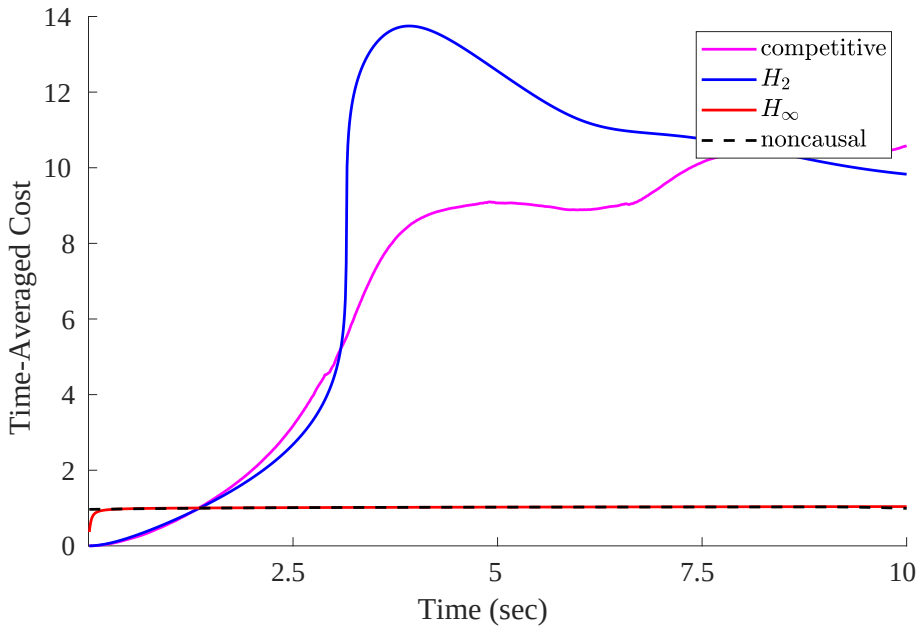

Figure 6.8: Relative performance of causal controllers in an inverted pendulum system driven by a constant disturbance.

controllers at a range of frequencies ranging from $\omega = 0.01\pi$ to $\omega = \pi$. We see that the competitive controller performs the best, and outperforms the H_∞ controller by a three orders of magnitude at $\omega = \pi$.

We next consider a "Gaussian random walk" disturbance, where a series of standard Gaussian random variables is sampled once, before the experiment begins, and the disturbance in timestep t is the cumulative sum of the first t variables. In Figure 6.12, we plot the costs incurred by the causal controllers and the optimal noncausal controller, averaged across ten trials. Perhaps surprisingly, the noncausal controller performs the worst, incurring more than 400 times as much cost as the H_∞ controller. This experiment highlights how the nonlinear dynamics radically influence the relative performance of various controllers; the noncausal controller does not account for linearization error and hence can incur high cost, despite having full knowledge of the disturbance sequence. While the competitive controller does not perform as well as the H_∞ controller, it still manages to perform significantly better than the H_2 controller.

Together, our experiments highlight the value proposition offered by the competitive controller: while it doesn't always perform the best out of all causal controllers, it always incurs at most a small constant factor more cost than the optimal noncausal

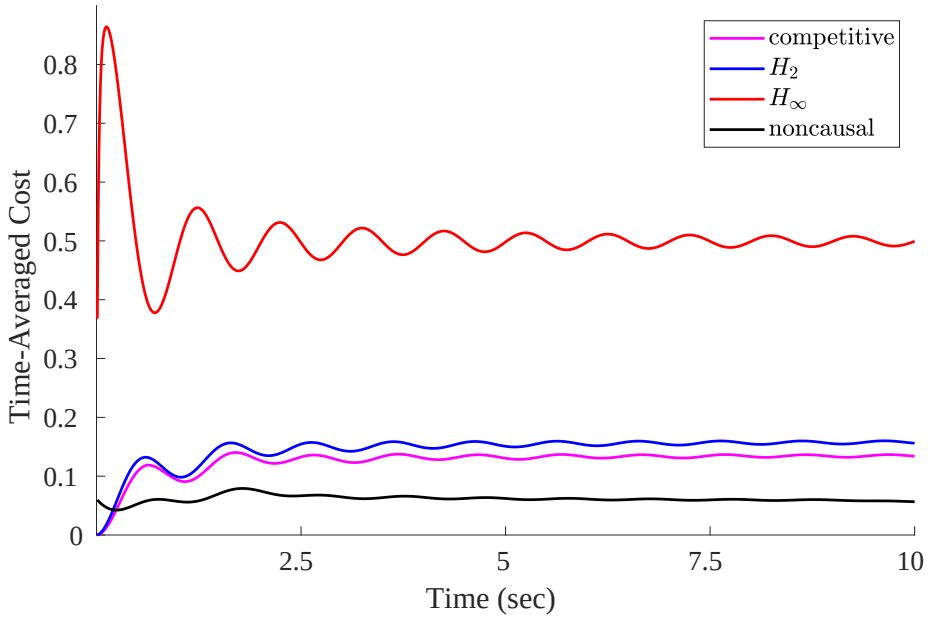

Figure 6.9: Relative performance of causal controllers in an inverted pendulum system driven by sinusoidal noise with frequency 0.01π.

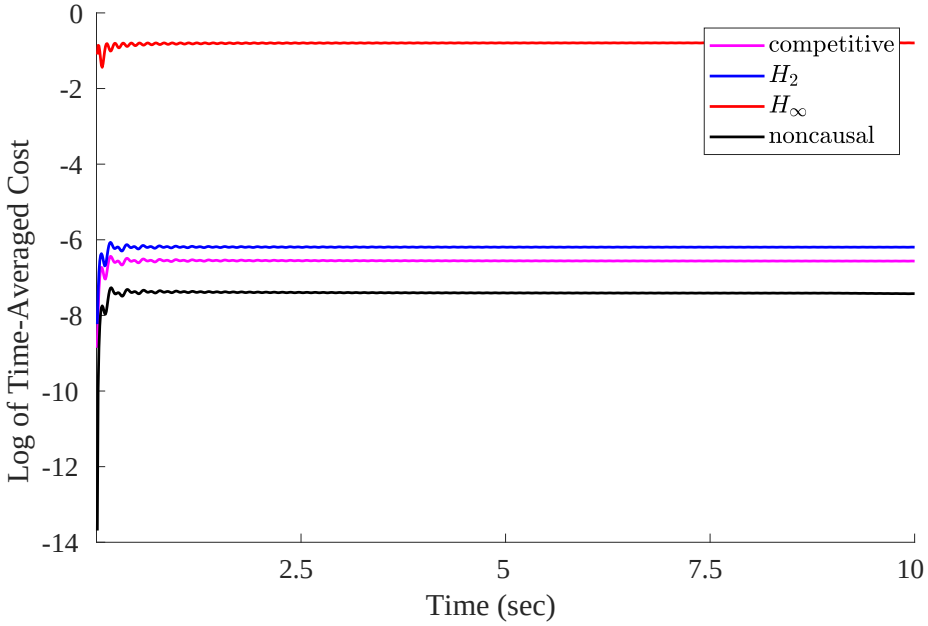

Figure 6.10: Relative performance of causal controllers in an inverted pendulum system driven by sinusoidal noise with frequency 0.1π (log scale).

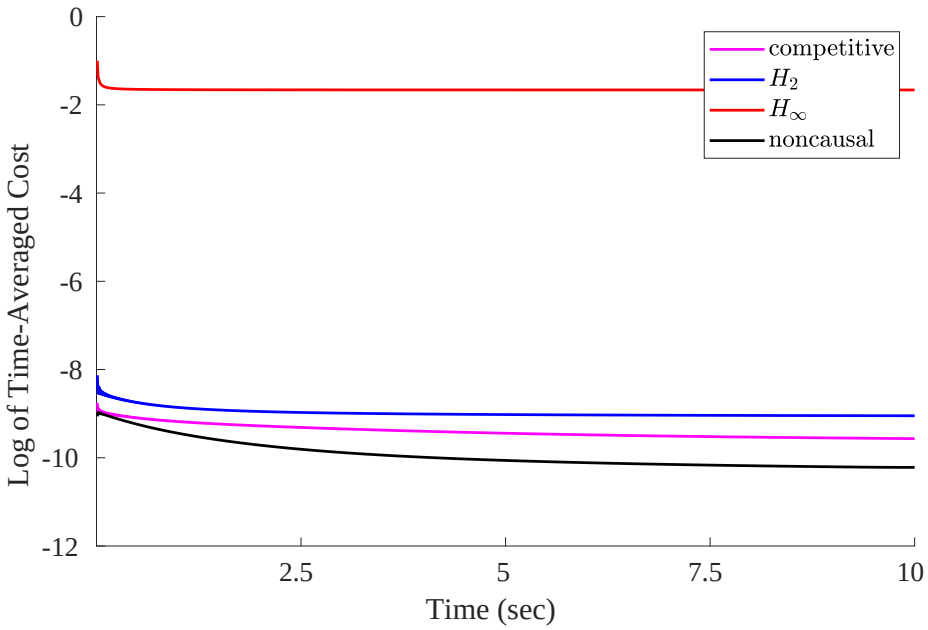

Figure 6.11: Relative performance of causal controllers in an inverted pendulum system driven by sinusoidal noise with frequency π (log scale).

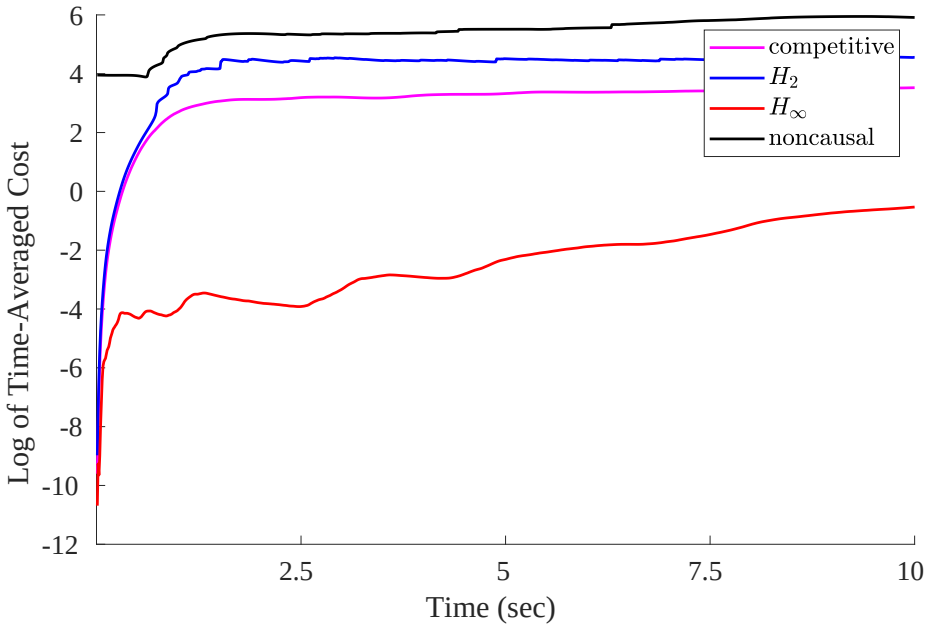

Figure 6.12: Relative performance of causal controllers in an inverted pendulum system driven by a Gaussian random walk (log scale).

controller. This is in stark contrast to the H_∞ controller, which incurs hundreds of times more cost than the optimal noncausal controller on some adversarially constructed disturbances. Furthermore, when the disturbance is stochastic, the competitive controller is able to nearly match the performance of the H_2 controller, which is specifically tuned for stochastic noise.

Chapter 7

CONCLUSION AND FUTURE WORK

In this thesis we proposed data-dependent regret as a criterion for controller design and showed that causal controllers with optimal data-dependent regret can be derived via a reduction to H_∞ control in both the full-information and measurement-feedback settings. We also described a surprising connection between regret-optimal control and online learning, and used that connection to show that it is possible to achieve "best-of-both-worlds" performance guarantees, at least in LTI systems which are perturbed by norm-bounded disturbances. We also presented numerical simulations which suggest that our regret-optimal control paradigm is a promising approach to adapting to model error in nonlinear control.

There are a few open problems which we think would be interesting to explore in future work. First, our negative results in Chapter 4 suggest that the optimal noncausal controller is too strong to use as a benchmark in some regret minimization problems. The fact that the optimal noncausal controller is able to observe the actual disturbances when selecting the control actions represents a huge advantage relative to the online controller, which only has access to noisy linear measurements of the state; it is this disparity which precludes the existence of measurement-feedback controllers with bounded competitive ratio, or regret which is bounded by the pathlength of the measurement disturbance. In recent work [GH21] we described an alternative choice of noncausal benchmark controller, which is able to evaluate the costs that would result from choosing a specific sequence of control actions but does not have access to the actual disturbance sequence. We believe such noncausal benchmarks are an interesting choice to study in future work on regret-optimal measurement-feedback control.

Another topic to explore in future work is regret-optimal control in settings where it is more natural to measure the size of signals via the the ℓ_∞ norm rather than the ℓ_2 norm. In the late 1980s and early 1990s a theory of robust control was developed [Vid86; DP87; DK93] to bound the worst-case amplification of ℓ_∞ bounded signals; this approach to control is called L_1 control. It naturally parallels the H_∞ approach to robust control, which focuses on bounding the worst-case amplification of ℓ_2 bounded signals. It would be interesting to develop an analogous theory of regret-

optimal control using a reduction to L_1 control.

Finally, while we have studied the problem of obtaining controllers with sublinear policy regret and approximate competitive ratio, it is natural to consider the more challenging metric of *adaptive policy regret*, which was recently studied in [GHM20]. This metric generalizes adaptive regret, which was introduced by Hazan and Seshadhri in [HS09], and measures the performance of an online control algorithm by its regret relative to a fixed comparator in any contiguous subinterval. It is hence a more challenging notion than policy regret, since the controller must achieve low regret in every subinterval, not merely over the full time horizon. One advantage of this metric is that is more meaningful when the dynamics are time-varying and it makes little sense to compare to a fixed controller.

BIBLIOGRAPHY

[Aga+19] Naman Agarwal et al. "Online Control with Adversarial Disturbances". In: *International Conference on Machine Learning*. 2019, pp. 111–119.

[AGL22] Mohammad Akbari, Bahman Gharesifard, and Tamas Linder. "Logarithmic Regret in Online Linear Quadratic Control using Riccati Updates". In: *Mathematics of Control, Signals, and Systems* (2022), pp. 1–32.

[AHM15] Oren Anava, Elad Hazan, and Shie Mannor. "Online Learning for Adversaries with Memory: Price of Past Mistakes". In: *Advances in Neural Information Processing Systems* 28 (2015).

[Aue+02] Peter Auer et al. "The Nonstochastic Multiarmed Bandit Problem". In: *SIAM journal on computing* 32.1 (2002), pp. 48–77.

[BB08] Tamer Başar and Pierre Bernhard. H_∞ *Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Springer Science & Business Media, 2008.

[BE05] Allan Borodin and Ran El-Yaniv. *Online computation and Competitive Analysis*. Cambridge University Press, 2005.

[BGZ15] Omar Besbes, Yonatan Gur, and Assaf Zeevi. "Non-Stationary Stochastic Optimization". In: *Operations research* 63.5 (2015), pp. 1227–1244.

[Bub+19] Sébastien Bubeck et al. "Improved Path-length Regret Bounds for Bandits". In: *Conference On Learning Theory*. PMLR. 2019, pp. 508–528.

[CGW18] Niangjun Chen, Gautam Goel, and Adam Wierman. "Smoothed Online Convex Optimization in High Dimensions via Online Balanced Descent". In: *Conference On Learning Theory*. PMLR. 2018, pp. 1574–1594.

[CH21] Xinyi Chen and Elad Hazan. "Black-Box Control for Linear Dynamical Dystems". In: *Conference on Learning Theory*. PMLR. 2021, pp. 1114–1143.

[Coh+18] Alon Cohen et al. "Online Linear Quadratic Control". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 1029–1038.

[DK93] Munther A Dahleh and Mustafa H Khammash. "Controller Design for Plants with Structured Uncertainty". In: *Automatica* 29.1 (1993), pp. 37–56.

[Doy78] John C. Doyle. "Guaranteed Margins for LQG Regulators". In: *IEEE Transactions on Automatic Control* 23.4 (1978), pp. 756–757.

[DP87] Munther A. Dahleh and J. Boyd Pearson. "ℓ^1-Optimal Feedback Controllers for MIMO Discrete-Time Systems". In: *IEEE Transactions on Automatic Control* 32.4 (1987), pp. 314–322.

[FS20] Dylan Foster and Max Simchowitz. "Logarithmic Regret for Adversarial Online Control". In: *International Conference on Machine Learning*. PMLR. 2020, pp. 3211–3221.

[GH21] Gautam Goel and Babak Hassibi. "Regret-Optimal Measurement-Feedback Control". In: *Learning for Dynamics and Control*. PMLR. 2021, pp. 1270–1280.

[GHM20] Paula Gradu, Elad Hazan, and Edgar Minasyan. "Adaptive Regret for Control of Time-Varying Dynamics". In: *arXiv preprint arXiv:2007.04393* (2020).

[Goe+19] Gautam Goel et al. "Beyond Online Balanced Descent: An Optimal Algorithm for Smoothed Online Optimization". In: *Advances in Neural Information Processing Systems*. 2019, pp. 1875–1885.

[GW19] Gautam Goel and Adam Wierman. "An Online Algorithm for Smoothed Regression and LQR Control". In: *Proceedings of Machine Learning Research* 89 (2019), pp. 2504–2513.

[HAK07] Elad Hazan, Amit Agarwal, and Satyen Kale. "Logarithmic Regret Algorithms for Online Convex Optimization". In: *Machine Learning* 69.2 (2007), pp. 169–192.

[Haz19] Elad Hazan. "Introduction to Online Convex Optimization". In: *arXiv preprint arXiv:1909.05207* (2019).

[HS09] Elad Hazan and Comandur Seshadhri. "Efficient Learning Algorithms for Changing Environments". In: *Proceedings of the 26th International Conference on Machine Learning*. 2009, pp. 393–400.

[HSK99] Babak Hassibi, Ali H. Sayed, and Thomas Kailath. *Indefinite-Quadratic Estimation and Control: A Unified Approach to H_2 and H_∞ Theories*. SIAM, 1999.

[Kal+60] Rudolf Emil Kalman et al. "Contributions to the Theory of Optimal Control". In: *Bol. soc. mat. mexicana* 5.2 (1960), pp. 102–119.

[KSH00] Thomas Kailath, Ali H. Sayed, and Babak Hassibi. *Linear Estimation*. Prentice Hall, 2000.

[LS20] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[RTS65] Herbert E. Rauch, F. Tung, and Charlotte T. Striebel. "Maximum Likelihood Estimates of Linear Dynamic Systems". In: *AIAA journal* 3.8 (1965), pp. 1445–1450.

[Shi+20] Guanya Shi et al. “Online Optimization with Memory and Competitive Control”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 20636–20647.

[Sim20] Max Simchowitz. “Making Non-Stochastic Control (Almost) as Easy as Stochastic”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 18318–18329.

[Vid86] Mathukumalli Vidyasagar. “Optimal Rejection of Persistent Bounded Disturbances”. In: *IEEE Transactions on Automatic Control* 31.6 (1986), pp. 527–534.

[Whi81] Peter Whittle. “Risk-Sensitive Linear/Quadratic/Gaussian Control”. In: *Advances in Applied Probability* 13.4 (1981), pp. 764–777.

[Zam81] George Zames. “Feedback and Optimal Sensitivity: Model Reference Transformations, Multiplicative Seminorms, and Approximate Inverses”. In: *IEEE Transactions on Automatic Control* 26.2 (1981), pp. 301–320.

[Zha+20] Peng Zhao et al. “A Simple Approach for Non-Stationary Linear Bandits”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2020, pp. 746–755.

[Zin03] Martin Zinkevich. “Online Convex Programming and Generalized Infinitesimal Gradient Ascent”. In: *Proceedings of the 20th International Conference on Machine Learning*. 2003, pp. 928–936.

[ZWZ22] Peng Zhao, Yu-Xiang Wang, and Zhi-Hua Zhou. “Non-Stationary Online Learning with Memory and Non-Stochastic Control”. In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2022, pp. 2101–2133.

Appendix A

SOME USEFUL LEMMAS

Lemma 1. *For all H, F and all Hermitian matrices P, we have*

$$\begin{bmatrix} H^*(z^{-1}I - F^*)^{-1} & I \end{bmatrix} \Omega(P) \begin{bmatrix} (zI - F)^{-1}H \\ I \end{bmatrix} = 0,$$

where we define

$$\Omega(P) = \begin{bmatrix} -P + F^*PF & F^*PH \\ H^*PF & H^*PH \end{bmatrix}.$$

Proof. This identity is essentially the "transpose" of Lemma 2 and is easily verified via direct calculation. □

Lemma 2. *For all H, F and all Hermitian matrices P, we have*

$$\begin{bmatrix} H(zI - F)^{-1} & I \end{bmatrix} \Omega(P) \begin{bmatrix} (z^{-1}I - F^*)^{-1}H^* \\ I \end{bmatrix} = 0,$$

where we define

$$\Omega(P) = \begin{bmatrix} -P + FPF^* & FPH^* \\ HPF^* & HPH^* \end{bmatrix}.$$

Proof. This identity is a special case of Lemma 3; it also appears as Lemma 12.3.3 in [HSK99]. □

Lemma 3. *For all H_1, H_2, F_1, F_2 and all matrices W, we have*

$$\begin{bmatrix} H_1(zI - F_1)^{-1} & I \end{bmatrix} \Omega(W) \begin{bmatrix} (z^{-1}I - F_2^*)^{-1}H_2^* \\ I \end{bmatrix} = 0,$$

where we define

$$\Omega(W) = \begin{bmatrix} -W + F_1WF_2^* & F_1WH_2^* \\ H_1WF_2^* & H_1WH_2^* \end{bmatrix}.$$

Proof. Notice that $\Omega(W)$ can be rewritten as

$$\Omega(W) = \begin{bmatrix} F_1 \\ H_1 \end{bmatrix} W \begin{bmatrix} F_2 & H_2 \end{bmatrix} - \begin{bmatrix} I \\ 0 \end{bmatrix} W \begin{bmatrix} I & 0 \end{bmatrix}.$$

The proof is immediate after observing that

$$\begin{bmatrix} H_1(zI - F_1)^{-1} & I \end{bmatrix} \begin{bmatrix} F_1 \\ H_1 \end{bmatrix} = H_1(zI - F_1)^{-1}z,$$

$$\begin{bmatrix} F_2 & H_2 \end{bmatrix} \begin{bmatrix} (z^{-1}I - F_2^*)^{-1}H_2^* \\ I \end{bmatrix} = z^{-1}(z^{-1}I - F_2^*)^{-1}H_2^*.$$

□

Lemma 4. *For all $\delta \in [0, 1/2]$ and all $i \geq 0$, the following inequality holds:*

$$i(1-\delta)^{i-1} \leq \frac{4}{\delta}(1-\delta/2)^i.$$

Proof. Recall that $xe^{-x} \leq 1$ for all x, therefore $ie^{-\delta i/2} \leq \frac{2}{\delta}$. Also for all $\delta \in [0, 1/2]$ we have the inequality $\sqrt{1-\delta} \leq 1 - \delta/2$. We see that

$$\begin{aligned} i(1-\delta)^{i-1} &= \frac{i}{1-\delta}(1-\delta)^{i/2}(1-\delta)^{i/2} \\ &\leq \frac{i}{1-\delta}e^{-\delta i/2}(1-\delta)^{i/2} \\ &\leq \frac{2}{\delta(1-\delta)}(1-\delta)^{i/2} \\ &\leq \frac{2}{\delta(1-\delta)}(1-\delta/2)^i \\ &\leq \frac{4}{\delta}(1-\delta/2)^i. \end{aligned}$$

□

Lemma 5. *Let π be a policy from an (H, θ, δ)-DAC policy class whose stabilizing component $\mathbb{K}$ is (κ, δ)-strongly stable. The states and actions generated by π are pointwise bounded; in each timestep $t \geq 1$ the states and control actions satisfy*

$$\max(\|x_t\|, \|u_t\|) \leq \frac{3\kappa^3\theta W}{\delta}.$$

Proof. Since $\mathbb{K}$ is (κ, δ)-strongly stable, there exists matrices S, L such that $A+B\mathbb{K} = SLS^{-1}$ with $\max(\|\mathbb{K}\|, \|B\|, \|S\|\|S^{-1}, 1)\| \leq \kappa$ and $\|L\| \leq 1 - \delta$. Recall that for every member of an (H, R, δ)-DAC class satisfies $\|M^{[j-1]}\| \leq \theta(1-\delta)^{j-1}$, where $\theta > 1$ and $\delta < 1$. It follows that

$$\sum_{j=1}^{\infty} \|M^{[j-1]}\| \leq \frac{\theta}{\delta}.$$

Observe that for any $t \geq 0$ the states x_{t+1} satisfies

$$\begin{aligned}
\|x_{t+1}\| &= \left\| \sum_{i=0}^{t-1} (A + B\mathbb{K})^i \left(w_{t-i} + B \sum_{j=1}^{H} M^{[j-1]} w_{t-i-j} \right) \right\| \\
&\leq \left(\sum_{i=0}^{t-1} \|(A + B\mathbb{K})^i\| \right) \cdot \max_{0 \leq i \leq t-1} \left\| w_{t-i} + B \sum_{j=1}^{H} M^{[j-1]} w_{t-i-j} \right\| \\
&\leq \left(\sum_{i=0}^{t-1} \|(A + B\mathbb{K})^i\| \right) \cdot \max_{0 \leq i \leq t-1} \left(\|w_{t-i}\| + \|B\| \max_j \|w_{t-i-j}\| \sum_{j=1}^{H} \|M^{[j-1]}\| \right) \\
&\leq \frac{\kappa}{\delta} W \left(1 + \frac{\kappa\theta}{\delta} \right) \\
&\leq \frac{2\kappa^2 \theta W}{\delta^2},
\end{aligned}$$

Turning to u_t, we see that

$$\begin{aligned}
\|u_t\| &= \left\| \mathbb{K} x_t + \sum_{i=1}^{H} M^{[i-1]} w_{t-i} \right\| \\
&\leq \frac{2\kappa^3 \theta W}{\delta^2} + \frac{\theta W}{\delta}.
\end{aligned}$$

Since $\kappa \geq 1$ and $\delta < 1$, both $\|u_t\|$ and $\|x_t\|$ are bounded above by $\frac{3\kappa^3\theta W}{\delta^2}$ for all $t \geq 1$. □

Appendix B

H_∞ CONTROL

In this chapter we review the H_∞ control paradigm and present state-space models of H_∞ controllers in both the full-information and measurement-feedback settings.

B.1 Full-Information H_∞ control

The H_∞-optimal control problem is to find the controller which minimizes the worst-case gain from the energy in the disturbance w to the cost incurred by the controller. More formally, the H_∞-optimal control problem is:

Problem 7 (H_∞-optimal control)**.** *Find a causal controller π that minimizes*

$$\sup_w \frac{J(\pi, w)}{\|w\|_2^2}.$$

The finite-horizon H_∞ problem is identical, except that the infinite-horizon cost $J(\pi, w)$ is replaced by the finite-horizon cost $J_T(\pi, w)$. In general, it is not known how to derive a closed-form for the H_∞-optimal controller, so it is common to consider a relaxation:

Problem 8 (Suboptimal H_∞ control at level γ)**.** *Given $\gamma > 0$, find an online controller such that*

$$\frac{J(\pi, w)}{\|w\|_2^2} < \gamma^2$$

for all disturbances w, or determine whether no such controller exists.

We call such a controller an H_∞ *controller at level* γ. It is clear that if we can solve this suboptimal problem then we can easily recover the H_∞-optimal controller via bisection on γ. The suboptimal H_∞ control problem has a well-known state-space solution:

Theorem 17 (Theorem 13.3.3 in [HSK99])**.** *Suppose (A, B_u) is stabilizable and (A, L) is observable on the unit circle. A causal controller K such that*

$$\|T_K\| < \gamma$$

exists if and only if the DARE

$$P = Q + A^\top P A - A^\top P \widetilde{B} \widetilde{H}^{-1} \widetilde{B}^\top P A$$

where we define

$$\begin{cases} \widetilde{B} = \begin{bmatrix} B_u & B_w \end{bmatrix}, \\ \\ \widetilde{R} = \begin{bmatrix} I_m & 0 \\ 0 & -\gamma^2 I_p \end{bmatrix}, \\ \\ \widetilde{H} = \widetilde{R} + \widetilde{B}^\top P \widetilde{B}, \end{cases}$$

has a solution P such that

1. *$A - \widetilde{B}\widetilde{H}^{-1}\widetilde{B}^\top PA$ is stable;*
2. *$\widetilde{R}$ and $\widetilde{H}$ have the same inertia;*
3. *$P \geq 0$.*

In this case, one possible infinite-horizon H_∞ controller at level γ is

$$u_t = -(I_m + B_u^\top P B_u)^{-1} B_u^\top P(Ax_t + B_w w_t).$$

A strictly causal H_∞ controller at level γ exists if and only if conditions (1) and (3) hold, and additionally

$$B_u^\top P B_u \prec \gamma^2 I_m$$

and

$$I_m + B_u^\top P(I_n - \gamma^{-2} B_w B_w^\top P)^{-1} B_u \succ 0.$$

In this case, one possible strictly causal H_∞ controller at level γ is

$$u_t = -(I_m + B_u^\top \widetilde{P} B_u)^{-1} B_u^\top \widetilde{P} A x_t,$$

where we define

$$\widetilde{P} = P - P B_w(-\gamma^2 I_p + B_w^\top P B_w)^{-1} B_w^\top P.$$

H_∞ Control in Time-Varying Systems

We can also describe a state-space model for the H_∞ controller at level γ in a time-varying system over a finite horizon:

Theorem 18 (Theorems 9.5.1 and 9.5.2 in [HSK99])**.** *Given $\gamma > 0$, a causal finite-horizon H_∞ controller at level γ exists if and only if*

$$B_{w,t}^\top \left[P_{t+1} - P_{t+1} B_{u,t} H_t^{-1} B_{u,t}^\top P_{t+1} \right] B_{w,t} \prec \gamma^2 I_p$$

for all $t = 0, \ldots T$, where we define

$$H_t = I_m + B_{u,t}^\top P_{t+1} B_{u,t}$$

and P_t is the solution of the backwards-time Riccati recurrence

$$P_t = Q_t + A_t^\top P_{t+1} A_t - A_t^\top P_{t+1} \tilde{B}_t \tilde{H}_t^{-1} \tilde{B}_t^\top P_{t+1} A_t,$$

where we initialize $P_{T+1} = 0$ and define

$$\tilde{B}_t = \begin{bmatrix} B_{u,t} & B_{w,t} \end{bmatrix}, \quad \tilde{R} = \begin{bmatrix} I_m & 0 \\ 0 & -\gamma^2 I_p \end{bmatrix}, \quad \tilde{H}_t = \tilde{R} + \tilde{B}_t^\top P_{t+1} \tilde{B}_t.$$

In this case, one possible causal finite-horizon H_∞ controller at level γ is given by

$$u_t = -H_t^{-1} B_{u,t}^\top P_{t+1} (A_t x_t + B_{w,t} w_t).$$

A strictly causal finite-horizon controller at level γ exists if and only if

$$B_{u,t}^\top P_{t+1} B_{u,t} \prec \gamma^2 I_m$$

for $t = 0 \ldots T$. In this case, one possible strictly causal finite-horizon controller at level γ is given by

$$u_t = -H_t^{-1} B_{u,t}^\top \widetilde{P}_{t+1} A_t x_t,$$

where we define

$$\widetilde{P}_{t+1} = P_{t+1} - P_{t+1} B_{w,t} (-\gamma^2 I_p + B_{w,t}^\top P_{t+1} B_{w,t})^{-1} B_{w,t}^\top P_{t+1}.$$

If the dynamics are time-invariant then this controller converges to the infinite-horizon controller described in Theorem 17 as $T \to \infty$.

B.2 Measurement-Feedback H_∞ control

The H_∞-optimal measurement-feedback control problem is to find the controller which minimizes the worst-case gain from the energy in the disturbances w and v to the cost incurred by the controller. More formally, the H_∞-optimal measurement-feedback control problem is:

Problem 9 (H_∞-optimal measurement-feedback control)**.** *Find an online controller that minimizes*

$$\sup_{w,v} \frac{J(\pi, w, v)}{\|w\|_2^2 + \|v\|_2^2}.$$

In general, it is not known how to derive a closed-form for the H_∞-optimal measurement-feedback controller, so it is common to consider a relaxation:

Problem 10 (Suboptimal H_∞ measurement-feedback control at level γ)**.** *Given $\gamma > 0$, find an online controller such that*

$$\frac{J(\pi, w)}{\|w\|_2^2 + \|v\|_2^2} < \gamma^2$$

for all disturbances w, or determine whether no such controller exists.

We call such a controller an H_∞ *measurement-feedback controller at level* γ. It is clear that if we can solve this suboptimal problem then we can easily recover the H_∞-optimal controller via bisection on γ. The suboptimal H_∞ measurement-feedback control problem has a well-known state-space solution:

Theorem 19 (Theorem 13.3.5 in [HSK99])**.** *A causal measurement-feedback controller K such that*

$$\|T_K\| < 1$$

exists if and only if the control DARE

$$P_c = A^* P_c A + L^* L - K_c^* R_c K_c$$

and the estimation DARE

$$P_e = A P_e A^* + B_w B_w^* - K_e R_e K_e^*,$$

where we define

$$\begin{cases} K_c = R_c^{-1} \begin{bmatrix} B_u^* \\ B_w^* \end{bmatrix} P_c A \\ \\ R_c = \begin{bmatrix} I_m & 0 \\ 0 & -I_p \end{bmatrix} + \begin{bmatrix} B_u^* \\ \\ B_w^* \end{bmatrix} P_c \begin{bmatrix} B_u & B_w \end{bmatrix}, \\ \\ K_e = A P^d \begin{bmatrix} C^* & L^* \end{bmatrix} R_e^{-1}, \\ \\ R_e = \begin{bmatrix} I_r & 0 \\ 0 & -I_n \end{bmatrix} + \begin{bmatrix} C \\ L \end{bmatrix} P_c \begin{bmatrix} C^* & L^* \end{bmatrix}, \end{cases}$$

have solutions $P_c \succeq 0$ *and* $P_e \succeq 0$ *such that*

1. *The matrix* $A - \begin{bmatrix} B_u & B_w \end{bmatrix} K_c$ *is stable.*
2. *The matrix* R_c *has* m *positive eigenvalues and* p *negative eigenvalues.*
3. *The matrix* $A - K_e \begin{bmatrix} C \\ L \end{bmatrix}$ *is stable.*
4. *The matrix* R_e *has* r *positive eigenvalues and* n *negative eigenvalues.*
5. $\rho(P_c P_e) < 1$.

If these conditions are satisfied, then one possible choice of K *is given by*

$$u_t = -K_u(\widehat{x}_t + PC^*(I_r + CPC^*)^{-1}(y_t - C\widehat{x}_t)),$$

where the state-estimate $\widehat{x}_t$ *is given by the recursion*

$$\widehat{x}_{t+1} = (A - B_w K_w)(\widehat{x}_t + PC^*(I_r + CPC^*)^{-1}(y_t - C\widehat{x}_t)) + B_u u_t,$$

and $K_u \in \mathbb{R}^{m \times n}$ *and* $K_w \in \mathbb{R}^{p \times n}$ *are defined as*

$$\begin{bmatrix} K_u \\ K_w \end{bmatrix} = K_c,$$

and we define $P \in \mathbb{R}^{n \times n}$ *as*

$$P = P_e(I_n - P_c P_e)^{-1}.$$

While Theorem 19 tells us how to determine whether there exists a controller K such that $\|T_K\| < 1$, it does not directly answer the more general question if there exists a controller K such that $\|T_K\| < \gamma$, for any fixed $\gamma > 0$. This condition is equivalent to $\|T_{\widehat{K}}\| < 1$, where we define $T_{\widehat{K}} = \gamma^{-1} T_K$. Notice that

$$T_{\widehat{K}} = \begin{bmatrix} \gamma^{-1}G & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} F \\ I \end{bmatrix} (\gamma^{-1}Q) \begin{bmatrix} J & I \end{bmatrix}.$$

Recall that $Q = K(I - HK)^{-1}$, implying that $\gamma^{-1}Q = (\gamma^{-1}K)[I - (\gamma H)(\gamma^{-1}K)]^{-1}$. It follows that the controller K satisfying $\|T_K\| < \gamma$ in the system F, G, H, J is precisely $\gamma\widehat{K}$, where $\widehat{K}$ is the controller satisfying $\|T_{\widehat{K}}\| < 1$ in the system $\widehat{F}, \widehat{G}, \widehat{J}, \widehat{H}$

and $\widehat{F} = F, \widehat{G} = \gamma^{-1}G, \widehat{H} = \gamma H, \widehat{J} = J$. We can assign state-space structure to $\widehat{F}, \widehat{G}, \widehat{H}, \widehat{J}$ as follows. Recall that

$$F(z) = L(zI - A)^{-1}B_u, \quad G(z) = L(zI - A)^{-1}B_w,$$

$$H(z) = C(zI - A)^{-1}B_u, \quad J(z) = C(zI - A)^{-1}B_w.$$

Define

$$\widehat{C} = \gamma C, \quad \widehat{B}_w = \gamma^{-1}B_w.$$

We have

$$\widehat{F}(z) = L(zI - A)^{-1}\widehat{B}_u, \quad \widehat{G}(z) = L(zI - A)\widehat{B}_w,$$
$$\widehat{H}(z) = \widehat{C}(zI - A)^{-1}B_u, \quad \widehat{J}(z) = \widehat{C}(zI - A)\widehat{B}_w.$$

We can hence use Theorem 19 to check the existence of a controller K such that $\|T_K\| < \gamma$, for any fixed $\gamma > 0$.