Title

Reliable Machine Learning in Feedback Systems

Author

Dean, Sarah

Publication Date

2021

Peer reviewed|Thesis/dissertation

Reliable Machine Learning in Feedback Systems

by

Sarah Ankaret Anderson Dean

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering—Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Benjamin Recht, Chair
Professor Francesco Borrelli
Associate Professor Moritz Hardt

Summer 2021

Reliable Machine Learning in Feedback Systems

Abstract

Reliable Machine Learning in Feedback Systems

by

Sarah Ankaret Anderson Dean

Doctor of Philosophy in Engineering—Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Benjamin Recht, Chair

Machine learning is a promising tool for processing complex information, but it remains an unreliable tool for control and decision making. Applying techniques developed for static datasets to real world problems requires grappling with the effects of feedback and systems that change over time. In these settings, classic statistical and algorithmic guarantees do not always hold. How do we anticipate the dynamical behavior of machine learning systems before we deploy them? Towards the goal of ensuring reliable behavior, this thesis takes steps towards developing an understanding of the trade-offs and limitations that arise in feedback settings.

In Part I, we focus on the application of machine learning to automatic feedback control. Inspired by physical autonomous systems, we attempt to build a theoretical foundation for the data-driven design of optimal controllers. We focus on systems governed by linear dynamics with unknown components that must be characterized from data. We study unknown dynamics in the setting of the Linear Quadratic Regulator (LQR), a classical optimal control problem, and show that a procedure of least-squares estimation followed by robust control design guarantees safety and bounded sub-optimality. Inspired by the use of cameras in robotics, we also study a setting in which the controller must act on the basis of complex observations, where a subset of the state is encoded by an unknown nonlinear and potentially high dimensional sensor. We propose using a perception map, which acts as an approximate inverse, and show that the resulting perception-control loop has favorable properties, so long as either a) the controller is robustly designed to account for perception errors or b) the perception map is learned from sufficiently dense data.

In Part II, we shift our attention to algorithmic decision making systems, where machine learning models are used in feedback with people. Due to the difficulties of measurement, limited predictability, and the indeterminacy of translating human values into mathematical objectives, we eschew the framework of optimal control. Instead, our goal is to articulate the impacts of simple decision rules under one-step feedback models. We

first consider consequential decisions, inspired by the example of lending in the presence of credit score. Under a simple model of impact, we show that several group fairness constraints, proposed to mitigate inequality, may harm the groups they aim to protect. In fact, fairness criteria can be viewed as a special case of a broader framework for designing decision policies that trade off between private and public objectives, in which notions of impact and wellbeing can be encoded directly. Finally, we turn to the setting of recommendation systems, which make selections from a wide array of choices based on personalized relevance predictions. We develop a novel perspective based on reachability that quantifies agency and access. While empirical audits show that models optimized for accuracy may limit reachability, theoretical results show that this is not due to an inherent trade-off, suggesting a path forward. Broadly, this work attempts to re-imagine the goals of predictive models ubiquitous in machine learning, moving towards new design principles that prioritize human values.

To my family.

Contents

Acknowledgments

I received a huge amount of support from my collaborators, mentors, family, friends, and the broader UC Berkeley community during my PhD work. It has been wonderful to be surrounded by so many great people, and I have thoroughly enjoyed the time I have spent in this beautiful place.

It's hard to say what my PhD research would have looked like without the guidance of my advisor Benjamin Recht. When I began my graduate career, I wasn't sure exactly where I wanted to focus, and Ben showed me that I didn't have to. Ben modeled for me a research agenda with a wide breadth and encouraged me to follow my curiousity while maintaining a unified rigorous perspective. Whether the topic is robust control, recommendation systems, or computational microscopy, Ben offers wisdom, relevant references, and abundant hot takes.

There are many others whose mentorship has been invaluable to my success. Moritz Hardt introduced me to the world of "fair" machine learning, and taught me how to think critically and with technical clarity about the deployment of machine learning in social contexts. During our many joint group meetings with the MPC lab, Francesco Borrelli challenged me to articulate the goals of my work relative to traditional controls perspectives. When Nikolai Matni joined Ben's group as a post-doc, he patiently taught us about something called System Level Synthesis, a framework which has proven to be literally fundamental to much of the work in this thesis. My education in control theory was rounded out by the perspectives of Murat Arcak, Andrew Packard, and Claire Tomlin. Special thanks to Moritz, Francesco, and Claire for serving on my quals and dissertation committee, and to Shirley, Ria, Jon, and Kosta for smoothing out bureaucratic bumps and making sure my compute needs were met.

Perhaps the most wonderful thing about my time as a PhD student has been all my truly excellent collaborators, who helped to shape the work presented in this dissertation. It has been a pleasure to formulate and work through problems of data-driven control with Horia Mania, Nik, Stephen Tu, and Vickie Ye. I deeply enjoyed thinking about fairness and wellbeing with Lydia Liu, Esther Rolf, and Max Simchowitz, and I'm thankful for the wisdom of Joshua Blumenstock and Daniel Björkegren, who helped to ground us in the perspective of welfare economics. I am indebted to everyone at Canopy, especially Brian Whitman and Sarah Rich, for introducing me to the rich problem space of recommender systems in an inspiring real-world setting. Working through reachability with Mihaela Curmei helped immensely with clarifying my thoughts on these ideas.

I have been fortunate to work with many others. My collaborations with Andrew Taylor, Ryan Cosner, Victor Doronbantu, Aaron Ames, and Yisong Yue brought me into the world of nonlinear control. Ugo Rosolia taught me about receding horizon control, including how to race a car autonomously. I would never have gotten my hands dirty with hardware without Aurelia Guy or Rohan Sinha, who, along with Vickie, turned a dream of racing-from-pixels into a reality. I'm grateful to Deirdre Quillen for entertaining my fascination with ice skating robots. The best research code I've written was in collaboration

with Mihaela, Wenshuo Guo, Karl Krauth, and Alex Zhao. Zack Phillips and Laura Waller taught me how to think about optimization within the context of a larger system, a perspective which carries over into the rest of my work. Co-organizing GEESE with McKane Andrus, Roel Dobbe, Nitin Kohli, and Tom Gilbert, Nate Lambert, and Tom Zick expanded my horizons, and our many conversations changed the way I think about technology. The entire Modest Yachts slack has been a source of entertainment and information and it would be remiss not to mention Yasaman Bahri, Ross Boczar, Orianna DeMasi, Sara Fridovich-Keil, Eric Jonas, John Miller, Juanky Perdomo, Becca Roelofs, Ludwig Schmidt, Vaishaal Shankar, and Shivaram Venkataraman.

Finally, crucial to my happiness and success are my friends, whose support has been a constant since our days at Scotia-Glenville schools and on Penn's campus; my partner Pavlo Manovi, who I am lucky to share a life with; and my family, who made me who I am today.

Chapter 1

Introduction

Many modern digital systems—from automotive vehicles to social media platforms—have unprecedented abilities to measure, store, and process data. Optimism about the potential to benefit from this data is driven by parallel progress in machine learning, where huge datasets and vast computational power have led to advances in complex tasks like image recognition and machine translation. However, many applications go beyond processing complex information to acting on the basis of it—moving from classifying, categorizing, and translating to making decisions and taking actions. Applying techniques developed for static datasets to real world problems requires grappling with the effects of feedback and systems that change over time. In these settings, classic statistical and algorithmic guarantees do not always hold. Even rigorously evaluating performance can be difficult. How do we anticipate the behavior of machine learning systems before we deploy them? Can we design them to ensure good outcomes? What are the fundamental limitations and trade-offs?

In this thesis, we develop principled techniques for a variety of dynamical settings, towards a vision of reliable machine learning. This work draws on tools and concepts from control theory, which has a long history of formulating guarantees about the behavior of dynamical systems, optimization, which provides a language to articulate goals and tradeoffs, and of course machine learning, which uses data to understand and act on the world. Machine learning models are designed to make accurate predictions, whether about the trajectory of an autonomous vehicle, the likelihood of a loan repayment, or the level of engagement with a news article. Traditionally concieved of in the framework of static supervised learning, these models become part of a dynamical system as soon as they are used to take actions that affect the environment (Figure 1.1). Whether the context is steering an autonomous vehicle, approving a loan, or recommending a piece of content, incorporating the learned model into *policy* gives rise to a feedback loop.

There are problems associated with the use of static models in dynamic environments. Whether due to distribution shift, partial observability, or error accumulation, their predictive abilities may fail in feedback settings. Supervised learning is usually designed to guaranteed good average case performance, but a lane detector that works well on

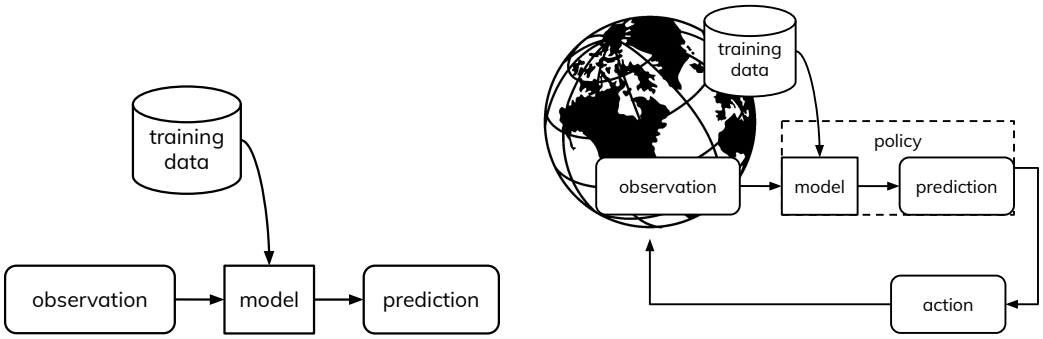

Figure 1.1: Though machine learning models are often trained with a static supervised learning framing in mind (left), when deployed, they become part of a feedback loop (right).

average may still misclassify a particular image and cause a crash. Furthermore, the statistical correlations exploited to make accurate predictions may in fact contain biases or other harmful patterns that we wish to avoid propagating. Considering an applicant's zip code in a lending decision may be statistically optimal, but lead to practices of redlining. Recommending videos with objectionable content might increase engagement, but at a detriment to the viewer's mental health. Contending with these challenges requires thinking carefully about how machine learning models are used, and designing policies that ensure desirable outcomes and are robust to errors.

In the following chapters, we consider several settings, broadly categorized into two parts: data-driven optimal control and feedback in social-digital systems. In Part I, we show how to combine machine learning and robust control to design data-driven policies with non-asymptotic performance and safety guarantees. Chapter 2 reviews a framework that enables policy analysis and synthesis for systems with uncertain dynamics and measurement errors. In Chapter 3, we consider the setting of a linear system with unknown dynamics and study the sample complexity of a classic optimal control problem with safety constraints. In Chapter 4, we look instead to challenges presented by complex sensing modalities and develop guarantees for perception-based control. Turning from the dynamics of physical systems to impact in social ones, in Part II, we consider learning algorithms that interact with people. In Chapter 5, we characterize the relationship between fairness and wellbeing in consequential decision making. We focus on the setting of content recommendation in Chapter 6, and develop a method for characterizing user agency in interactive systems. In the remainder of this chapter, we introduce and motivate the settings for the chapters to follow.

1.1 Data-Driven Optimal Control

Having surpassed human performance in video games [MKS+15] and Go [SHM+16], there has been a renewed interest in applying machine learning techniques to planning and control. In particular, there has been a considerable amount of effort in developing new techniques for *continuous control* where an autonomous system interacts with a physical environment [DCHSA16; LFDA16]. Despite some impressive results in domains like manipulation [ABC+20], recent years have seen both driver and pedestrian fatalities due to malfunctions in automated vehicle control systems [Boa17; Boa20; Nat17]. Contending with errors arising from learned models is different from traditional notions of process and measurement noise. How can we ensure that our new data-driven automated systems are safe and robust?

In Part I of this thesis, we attempt to build a foundation for the theoretical understanding of how machine learning interfaces with control by analyzing simple optimal control problems. We develop baselines delineating the possible control performance achievable given a fixed amount of data collected from a system with an unknown component. The standard optimal control problem aims to find a control sequence that minimizes a given cost. We assume a dynamical system with *state* $x_t \in \mathbb{R}^n$ can be acted on by a *control* $u_t \in \mathbb{R}^m$ and obeys the dynamics

$$x_{t+1} = f_t(x_t, u_t, w_t) \tag{1.1.1}$$

where w_t is the process noise. The control action is allowed to depend on observations $y_t \in \mathbb{R}^d$ of the system state, which may be partial and imperfect: $y_t = g_t(x_t, v_t)$ where v_t is the measurement noise. Optimal control then seeks to minimize

$$\begin{array}{ll} \text{minimize} & c(x_0, u_0, x_1, \ldots, x_{T-1}, u_{T-1}, x_T) \\ \text{subject to} & x_{t+1} = f_t(x_t, u_t, w_t) \\ & y_t = g_t(x_t, v_t) \end{array} \quad . \tag{1.1.2}$$

Here, $c(\cdot)$ denotes a cost function which depends on the trajectory, and the input u_t is allowed to depend on all previous measurements and actions. In this generality, problem (1.1.2) encapsulates many of the problems considered in the reinforcement learning literature. It is also a difficult problem to solve generally, but for restricted settings, classic approaches in control theory offer tractable solutions when the dynamics and measurement models are known.

We study this problem when components of it are unknown and must be estimated from data. Even in the case of linear dynamics, it is challenging to reason about the effects of machine learning errors on the evolution of an uncertain system. Chapter 2 covers background on linear systems and controllers that is crucial to our investigations. It presents an overview of *System Level Synthesis*, a recently developed framework for optimal control that allows us to handle uncertainty in a transparent and analytically tractable manner.

In Chapter 3, we study how machine learning interfaces with control when the dynamics of the system are unknown and the state can be observed exactly. We analyze one of the most well-studied problems in classical optimal control, the *Linear Quadratic Regulator* (LQR). In this setting, the system to be controlled obeys *linear* dynamics, and we wish to minimize some *quadratic* function of the system state and control action. We further investigate tradeoffs with safety by considering the additional requirement that both the state and input satisfy *linear constraints*. This problem has been studied for decades in control. The unconstrained version has a simple, closed form solution on the infinite time horizon and an efficient, dynamic programming solution on finite time horizons. The constrained version has received much attention within the Model Predictive Control (MPC) community. By combining linear regression with robust control, we bound the number of samples necessary to guarantee safety and performance.

In Chapter 4, we turn to a setting inspired by the fact that incorporating rich, perceptual sensing modalities such as cameras remains a major challenge in controlling complex autonomous systems. We focus on the practical scenario where the underlying dynamics of a system are well understood, and it is the interaction with a complex sensor that is the limiting factor. Specifically, we consider controlling a known linear dynamical system for which partial state information can only be extracted from nonlinear and potentially high dimensional observations. Our approach is to design a *virtual sensor* by learning a perception map, i.e., a map from complex observations to a subset of the state. Showing that errors in the perception map do not accumulate and lead to instability requires generalization guarantees stronger than are typical in machine learning. We show that either robust control or sufficiently dense data can guarantee the closed-loop stability and performance of such a vision based control system.

1.2 Feedback in Social-Digital Systems

From credit scores to video recommendations, many machine learning systems that interact with people have a temporal feedback component, reshaping a population over time. Lending practices, for example, can shift the distribution of debt and wealth in the population. Job advertisements allocate opportunity. Video recommendations shape interests. Machine learning algorithms used in these contexts are mostly trained to optimize a single metric of performance. The decisions made by such algorithms can have unintended adverse side effects: profit-maximizing loans can have detrimental effects on borrowers [ST19] and fake news can undermine democratic institutions [Per17].

However, it is difficult to explicitly model or plan around the dynamical interactions between populations and algorithms. Unlike in physical systems, there are difficulties of measurement, limited predictability [SLK+20], and the indeterminacy of translating human values into mathematical objectives. Actions are usually discrete: an acceptance or a rejection, choosing a particular piece of content to recommend. Rather than attempt to design a policy that optimizes a questionable objective subject to incorrect dynamical

models, our goal is to develop a framework for articulating the impacts of simple decision rules. We therefore investigate methods for quantifying and incorporating considerations of impact without using the full framing of optimal control. This work attempts to reimagine the goals of predictive models ubiquitous in machine learning, moving towards new design principles that prioritize human values.

Chapter 5 focuses on consequential decision making. From medical diagnosis and criminal justice to financial loans and humanitarian aid, consequential decisions increasingly rely on data-driven algorithms. Existing scholarship on fairness in automated decision making criticizes unconstrained machine learning for its potential to *harm* historically underrepresented or disadvantaged groups in the population [Exe16; BS16]. Consequently, a variety of *fairness criteria* have been proposed as constraints on standard learning objectives. Even though these constraints are clearly intended to *protect* the disadvantaged group by an appeal to intuition, a rigorous argument to that effect is often lacking. In Chapter 5, we contextualize group fairness criteria by characterizing their delayed impact. By framing the problem in terms of a temporal measure of wellbeing, we see that static criteria alone cannot ensure favorable outcomes. We then consider an alternate framework: dual optimization of institutional (e.g. profit) and individual (e.g. welfare) objectives directly. Decisions constrained to obey fairness criteria can be equivalently viewed through the dual objective lens by defining welfare in a particular group-dependent way. This insight, arising from the equivalence between constrained and regularized optimization, shows that fairness constraints can be viewed as a special case of balancing multiple objectives.

Chapter 6 focuses on recommendation systems, which offer a distinct set of challenges. Through recommendation systems, personalized preference models mediate access to many types of information on the internet. Aiming to surface content that will be consumed, enjoyed, and highly rated, these models are primarily designed to accurately predict individuals' preferences. The focus on improving model accuracy favors systems in which human behavior becomes as predictable as possible—effects which have been implicated in unintended consequences like polarization or radicalization. In Chapter 6, we attempt to formalize some of the values at stake by considering notions of user control and access. We investigate *reachability* as a way to characterize user agency in interactive systems. We develop a metric that is computationally tractable and can be used to audit dynamical properties of a recommender system prior to deployment. Our experimental results show that accurate predictive models, when used to sort information, can unintentionally make portions of the content library inaccessible. Our theoretical results show that there is no inherent trade-off, suggesting that it is possible to design learning algorithms which provide agency while maintaining accuracy.

Ultimately, the integration of data-driven automation into important domains requires us to understand and guarantee properties like safety, equity, agency, and wellbeing. This is a challenge in dynamic and uncertain systems. The work presented in Part I takes a step towards building a theoretical foundation for what it takes to guarantee safety in data-driven optimal control. There is a further challenge in formally defining important

properties as tractable technical specifications. This is especially true for qualitative and contextual concepts like agency and wellbeing. The work presented in Part II takes a step towards evaluating proposed technical formalisms and articulating new ones. Progress along both of these thrusts is necessary to enable reliable machine learning in feedback systems.

Part I

Data-Driven Optimal Control

Chapter 2

System Level Analysis and Synthesis

2.1 Introduction

It is difficult to reason about the trajectory of uncertain dynamical systems. One source of uncertainty is noise: process noise perturbs the evolution of the system, measurement noise corrupts measurements of the system state. Another source of uncertainty arises when the dynamics or measurement models are not fully known. The behavior of the system depends on not only on the values of the unknown components, be they noise processes or models, but on the control inputs. Designing control laws that are safe and stable requires accounting for all possible trajectories of an uncertain system.

Of course, control theory has long dealt with these challenges. The problem of feedback control is to mitigate the perturbation of process noise based on measurements of the system state, noisy though they may be. The field of robust control has developed an abundance of design methods for systems with uncertain dynamics and measurement models. Despite this rich history, classical methods do not readily offer an accounting for how the magnitude of the uncertainty degrades the optimality of the controller, or how much uncertainty is tolerable when needing to keep a system safe.

In this section, we review a recently developed framework for linear control synthesis that allows us to grapple with uncertainty in a transparent and analytically tractable manner. The machinery that we review plays an important role in Chapters 3 and 4. This chapter uses material first presented in papers coauthored with Horia Mania, Nikolai Matni, Benjamin Recht, Stephen Tu, and Vickie Ye [DMMRT20; DTMR19; DMRY20; DR21].

2.2 From Linear Controllers to System Responses

The evolution of a linear system is determined by its initial condition x_0 and the dynamics equation

$$x_{t+1} = Ax_t + Bu_t + w_t \tag{2.2.1}$$

where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$ are the *state transition matrices*, $u_t \in \mathbb{R}^m$ is the control input, and $w_t \in \mathbb{R}^n$ is the process noise, also called the disturbance. The process noise is often assumed to be stochastic, zero mean, and independent over time. It can alternatively be modeled as bounded and adversarial; both can be handled in a straightforward manner.

Measurements, or system outputs, take the form

$$y_t = Cx_t + v_t \tag{2.2.2}$$

where $C \in \mathbb{R}^{d\times n}$ is the *measurement matrix* and $v_t \in \mathbb{R}^d$ is the measurement noise. As before, this noise process is often assumed to be stochastic, but bounded and adversarial models can be handled as well.

We now show that when linear systems are in feedback with linear controllers, the trajectory can be written as a linear function of the process and measurement noise.

State Feedback

A *state feedback* controller relies on perfect measurements of the system state. As a motivating example, a static state feedback control policy K, i.e., let $u_t = Kx_t$. Then, the closed loop map from the disturbance process $(w_0, w_1, \dots)$ to the state x_t and control input u_t at time t is given by

$$\begin{array}{rcl} x_t &=& (A+BK)^t x_0 + \sum_{k=0}^{t-1}(A+BK)^{k-1} w_{t-k}\ , \\ u_t &=& K(A+BK)^t x_0 + \sum_{k=0}^{t-1} K(A+BK)^{k-1} w_{t-k}\ . \end{array} \tag{2.2.3}$$

This follows by substituting the static control policy and unrolling the recursion in (2.2.1). Letting $w_{-1} = x_0$, $\Phi_x(k) := (A+BK)^{k-1}$, and $\Phi_u(k) := K(A+BK)^{k-1}$, we can rewrite (2.2.3) as

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t+1} \begin{bmatrix} \Phi_x(k) \\ \Phi_u(k) \end{bmatrix} w_{t-k}\ , \tag{2.2.4}$$

where $(\Phi_x(k), \Phi_u(k))$ are called the *closed-loop system response elements* induced by the static controller K. For any stable closed-loop system, i.e. the spectral radius $\rho(A+BK) < 1$, the matrix powers decay,

$$\lim_{t\to\infty} (A+BK)^t = 0\ .$$

Therefore, as long as K *stabilizes* the system (A, B), the system response elements are well defined on infinite horizons.

Though more difficult to write in closed-form, a similar argument can be made for any controller which is a linear function of the state and its history. Therefore, the expression (2.2.4) is valid for any linear dynamic controller. Though we conventionally think of the control policy as a function mapping states to input, whenever such a mapping is linear, both the control input and the state can be written as linear functions of the

disturbance signal $(w_t)_{t\geq 0}$. This immediately makes transparent the effect of the process noise on the system trajectories.

With such an identification, the linear dynamics require that the system response variables $(\Phi_x(k), \Phi_u(k))$ must obey the constraints

$$\Phi_x(1) = I \ , \ \ \Phi_x(k+1) = A\Phi_x(k) + B\Phi_u(k) \ , \ \ \forall k \geq 1 \ . \tag{2.2.5}$$

As we describe in detail in Section 2.4, these constraints are in fact both necessary and sufficient. Therefore, designing linear controllers is equivalent to designing system responses to noise. This is useful because the state feedback parameterization generally leads to non-convex expressions. Inspecting (2.2.3), it is clear that convex expressions of x_t and u_t will be non-convex in K. On the other hand, since the relation in (2.2.4) is linear, convex constraints on state and input translate to convex constraints on the system response elements.

Output Feedback

An *output feedback* controller must compute control inputs on the basis of measurements. Whenever this controller is a linear function of the measurements, it is also a linear function of the states and measurement noise variables. Therefore, like in the state feedback case, it will be possible to view the closed-loop system in terms of linear system response variables. The main difference is that there are two sources of noise driving the system, and therefore four system response variables of interest.

As an illustrative example, consider the classic Luenberger observer combined with a static feedback policy. The observer uses measurements to update a state estimate

$$\widehat{x}_{t+1} = A\widehat{x}_t + Bu_t + L(y_t - C\widehat{x}_t) \ , \tag{2.2.6}$$

where L is the static *gain matrix* and $\widehat{x}_0$ is the initial estimate. The static controller is $u_t = K\widehat{x}_t$. By defining the auxiliary state variable $e_t = \widehat{x}_t - x_t$ the closed-loop dynamics can be written as

$$\begin{bmatrix} x_{t+1} \\ e_{t+1} \end{bmatrix} = \begin{bmatrix} A+BK & BK \\ 0 & A-LC \end{bmatrix} \begin{bmatrix} x_t \\ e_t \end{bmatrix} + \begin{bmatrix} I & 0 \\ 0 & L \end{bmatrix} \begin{bmatrix} w_k \\ v_k \end{bmatrix} .$$

Therefore, we can unroll the recursion in a similar manner to the state feedback case and write the trajectory in terms of a convolution with the system response variables

$$\begin{bmatrix} x_t \\ u_t \end{bmatrix} = \sum_{k=1}^{t+1} \begin{bmatrix} \Phi_{xw}(k) & \Phi_{xv}(k) \\ \Phi_{uw}(k) & \Phi_{uv}(k) \end{bmatrix} \begin{bmatrix} w_{t-k} \\ v_{t-k} \end{bmatrix} \ . \tag{2.2.7}$$

As long as the controller and observer are both stable, i.e. $\rho(A+BK) < 1$ and $\rho(A-LC) < 1$, the system response elements are well-defined on infinite horizons.

The expression (2.2.7) holds for any controller that is a linear function of the history of system outputs. Therefore, for linear output feedback control, the trajectory can be written as a linear function of the process and measurement noise. This makes transparent the effect of these two noise processes on the system trajectories.

The dynamics and measurement model require that the system responses obey the constraints

$$\Phi_{xw}(1) = I, \quad \begin{bmatrix}\Phi_{xw}(k+1) & \Phi_{xv}(k+1)\end{bmatrix} = A\begin{bmatrix}\Phi_{xw}(k) & \Phi_{xv}(k)\end{bmatrix} + B\begin{bmatrix}\Phi_{uw}(k) & \Phi_{uv}(k)\end{bmatrix},$$
$$\begin{bmatrix}\Phi_{xw}(k+1)\\ \Phi_{uw}(k+1)\end{bmatrix} = \begin{bmatrix}\Phi_{xw}(k)\\ \Phi_{uw}(k)\end{bmatrix} A + \begin{bmatrix}\Phi_{xv}(k+1)\\ \Phi_{uv}(k+1)\end{bmatrix} C\,. \tag{2.2.8}$$

As in the state feedback case, these constraints are in fact both necessary and sufficient. Therefore, designing linear controllers is equivalent to designing system responses, and because the relation in (2.2.7) is linear, convex constraints on state and input translate to convex constraints on the system response elements.

Signals and Transfer Functions

To make guarantees about the behavior of systems on arbitrarily long time horizons, it is necessary to reason about their infinitely long trajectories $(x_0, x_1, \dots)$. As the previous subsections illustrate, these trajectories are the result of a convolution between system response variables and noise signals. It is therefore pertinent to develop notation and machinery for dealing with these objects.

First, for notational convenience, we turn to the more compact representation of *transfer functions*. The signal $(x_0, x_1, \dots)$ that results from the convolution between the operator $(\Phi(1), \Phi(2), \dots)$ and the signal $(w_{-1}, w_0, w_1, \dots)$ is written as

$$\mathbf{x} = \mathbf{\Phi}\mathbf{w}\,.$$

One way to obtain this representation is by definition. We can directly define the notation of signals $\mathbf{x} = (x_0, x_1, \dots)$ and linear operators $\mathbf{\Phi} = (\Phi(1), \Phi(2), \dots)$. Then, define the multiplication operation between these objects to correspond to a linear convolution, so that

$$\mathbf{x} = \mathbf{\Phi}\mathbf{w} \iff x_t = \sum_{k=1}^{t+1} \Phi(k) w_{t-k} \;\; \forall t \geq 0\,.$$

An alternative is to take a z-transform and work in the frequency domain. The frequency domain variable z can informally be thought of as a time-shift operator acting on signals, i.e., $z(x_t, x_{t+1}, \dots) = (x_{t+1}, x_{t+2}, \dots)$. We define $\mathbf{\Phi}(z) = \sum_{t=1}^{\infty} z^{-t}\Phi(t)$, $\mathbf{x}(z) = \sum_{t=0}^{\infty} z^{-t} x_t$ and similarly for $\mathbf{w}(z)$. By manipulating summations and polynomials, we have

$$\mathbf{x}(z) = \left(\sum_{k=1}^{\infty} z^{-k}\Phi(k)\right)\left(\sum_{k=-1}^{\infty} z^{-k} w_k\right) = \sum_{t=0}^{\infty}\sum_{k=1}^{t+1} z^{-t}\Phi(k) w_{t-k}\,,$$

as desired. The argument z of the transfer function and signals is often dropped. This representation is common in the controls literature.

Finally, it can be convenient to alternatively represent these objects as semi-infinite vectors and block Toeplitz matrices. Here, we define

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} \Phi(0) & 0 & 0 & \dots \\ \Phi(1) & \Phi(0) & 0 & \dots \\ \Phi(2) & \Phi(1) & \Phi(0) & \\ \vdots & \vdots & & \ddots \end{bmatrix}.$$

Then the convolution follows by matrix-vector multiplication. This representation makes some properties evident by analogy to matrices. It can also be useful for implementation and computation when considering finite horizons or truncations of signals and system responses.

The three representations above are in some sense equivalent (see, e.g. the introductory chapters of the textbook by Dahleh and Diaz-Bobillo [DD94]). In the remainder of this chapter, Chapter 3, and Chapter 4, we use letters such as x and A to denote vectors and matrices, and boldface letters such as $\mathbf{x}$ and $\mathbf{G}$ to denote infinite horizon signals and linear convolution operators. We write $x_{0:t} = (x_0, x_1, \dots, x_t)$ for the history of signal $\mathbf{x}$ up to time t. For a function $x_t \mapsto f_t(x_t)$, we write $\mathbf{f}(\mathbf{x})$ to denote the signal $\left(f_t(x_t)\right)_{t=0}^{\infty}$. We will denote the tth element as $\mathbf{G}[t] = G(t)$ and $\mathbf{x}[t] = x_t$. We will also denote $G[t:1]$ as the block row vector of system response elements of $\mathbf{G}$

$$G[t:1] = \begin{bmatrix} G(t) & \dots & G(1) \end{bmatrix},$$

and similarly for $G[1:t]$ with indices reversed. Linear dynamic controllers can also be written in this notation, i.e. $\mathbf{u} = \mathbf{K}\mathbf{y}$. We use the shorthand $u_t = \mathbf{K}(y_{0:t})$ to indicate the dependence between inputs and measurements at the tth time step.

Under this notation, we have for state and output feedback respectively,

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \mathbf{w} \quad \text{or} \quad \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix}. \tag{2.2.9}$$

The affine *realizability constraints* can be rewritten for state feedback as

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I,$$

and for output feedback as

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} \begin{bmatrix} zI - A \\ -C \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix}.$$

This follows from (2.2.5) and (2.2.7). I can also been derived from writing the linear dynamics in signal notation and then rearranging the expressions,

$$z\mathbf{x} = A\mathbf{x} + B\mathbf{u} + \mathbf{w}, \quad \mathbf{y} = C\mathbf{x} + \mathbf{v}\,.$$

We finish this discussion by introducing norms and normed spaces on signals and transfer functions. For a thorough introduction to the functional analysis commonly used in control theory, see the text by Zhou, Doyle, and Glover [ZDG96]. As is standard, we let $\|x\|_p$ denote the ℓ_p-norm of a vector x. For a matrix M, we let $\|M\|_p$ denote its $\ell_p \to \ell_p$ operator norm. We will consider the $\mathcal{H}_2$, $\mathcal{H}_\infty$, and $\mathcal{L}_1$ norms, which are infinite horizon analogs of the Frobenius, spectral, and $\ell_\infty \to \ell_\infty$ operator norms of a matrix, respectively:

$$\|\mathbf{G}\|_{\mathcal{H}_2} = \sqrt{\sum_{k=0}^{\infty} \|G(k)\|_F^2}, \;\; \|\mathbf{G}\|_{\mathcal{H}_\infty} = \sup_{\|\mathbf{w}\|_2=1} \|\mathbf{G}\mathbf{w}\|_2, \;\; \|\mathbf{G}\|_{\mathcal{L}_1} = \sup_{\|\mathbf{w}\|_\infty=1} \|\mathbf{G}\mathbf{w}\|_\infty\,.$$

As the $\mathcal{H}_\infty$ and $\mathcal{L}_1$ norms are induced norms, they satisfy the sub-multiplicative property $\|\mathbf{GH}\| \le \|\mathbf{G}\|\|\mathbf{H}\|$. The $\mathcal{H}_2$ norm satisfies

$$\|\mathbf{GH}\|_{\mathcal{H}_2} \le \|\mathbf{G}\|_{\mathcal{H}_\infty}\|\mathbf{H}\|_{\mathcal{H}_2}\,.$$

We remark that it is also possible to define the $\mathcal{H}_\infty$ system norm in terms of the power norm [YG14], defined as $\|\mathbf{x}\|_{pow} := (\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T} \|x_k\|_2^2)^{1/2}$.

In this thesis, we restrict our attention to the function space $\mathcal{RH}_\infty$, consisting of discrete-time stable matrix-valued transfer functions. We use $\frac{1}{z}\mathcal{RH}_\infty$ to denote the set of transfer functions $\mathbf{G}$ such that $z\mathbf{G} \in \mathcal{RH}_\infty$. A linear time-invariant transfer function is stable if and only if it is exponentially stable. Therefore, we further define for positive values C and $\rho \in [0, 1)$

$$\mathcal{RH}_\infty(C, \rho) := \left\{ \mathbf{G} = \sum_{k=0}^{\infty} G(k) z^{-k} \mid \|G(k)\|_2 \le C\rho^k,\ k = 0, 1, 2, \ldots \right\}. \tag{2.2.10}$$

This set contains transfer functions that satisfy a specified decay rate in the spectral norm of their impulse response elements.

2.3 Optimal Linear Control

We now turn to optimal control. While the previous section developed machinery useful for understanding the general behavior of linear systems in feedback with linear controllers, this section focuses on the design of linear controllers. Putting aside our discussion of system responses for now, we introduce optimal control problems written

in terms of state and control signals. In particular we consider

$$\begin{array}{ll} \text{minimize}_{\mathbf{K}} & c(\mathbf{x}, \mathbf{u}) \\ \text{subject to} & x_{t+1} = Ax_t + Bu_t + w_t \\ & y_t = Cx_t + v_t \\ & u_t = \mathbf{K}(y_{0:t}), \end{array} \tag{2.3.1}$$

for x_t the state, u_t the control input, w_t the process noise, v_t the measurement noise, $\mathbf{K}$ a linear time-invariant operator, and $c(\mathbf{x}, \mathbf{u})$ a suitable cost function.

Control design depends on how the disturbance $\mathbf{w}$ and measurement error $\mathbf{v}$ are modeled, as well as performance objectives. Table 2.1 summarizes several common cost functions that arise from different system desiderata and different classes of disturbances and measurement errors $\boldsymbol{\nu} := (\mathbf{w}, \mathbf{v})$. By modeling the disturbance and sensor noise as being drawn from different signal spaces, and by choosing correspondingly suitable cost functions, we can incorporate practical performance, safety, and robustness considerations into the design process. We now review prominent examples of optimal control problems that we revisit in Chapters 3 and 4.

Example 2.1 (Linear Quadratic Regulator)**.** Suppose that the cost function is given by

$$c(\mathbf{x}, \mathbf{u}) = \mathbb{E}_{\mathbf{w}} \left[\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} x_t^\top Q x_t + u_t^\top R u_t \right],$$

for some specified positive definite matrices Q and R and $w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I)$. Further suppose that the controller is given full information about the system, i.e. $C = I$ and $v_t = 0$ such that the measurement model collapses to $y_t = x_t$. Then the optimal control problem reduces to the familiar Linear Quadratic Regulator (LQR) problem

$$\begin{array}{ll} \text{minimize} & \mathbb{E}_w \left[\lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} x_t^\top Q x_t + u_t^\top R u_t \right] \\ \text{subject to} & x_{t+1} = Ax_t + Bu_t + w_t \,. \end{array} \tag{2.3.2}$$

For stabilizable (A, B), and detectable (A, Q), this problem has a closed-form stabilizing controller based on the solution of the discrete algebraic Riccati equation (DARE) [ZDG96]. This optimal control policy is linear, and given by

$$u_t^{\text{LQR}} = -(B^\top P B + R)^{-1} B^\top P A x_t =: K_{\text{LQR}} x_t, \tag{2.3.3}$$

where P is the positive-definite solution to the DARE defined by (A, B, Q, R).

Example 2.2 (Linear Quadratic Gaussian Control)**.** Suppose that we have the same setup as the previous example, but that now the measurement is instead given by $y_t = Cx_t + v_t$

for some C such that the pair (A, C) is detectable, and that $v_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I)$. Then the optimal control problem reduces to the Linear Quadratic Gaussian (LQG) control problem, the solution to which is:

$$u_t^{\text{LQG}} = K_{\text{LQR}} \hat{x}_t, \tag{2.3.4}$$

where $\hat{x}_t$ is the Kalman filter estimate of the state at time t. The steady state update rule for this estimate also linear, and is given by

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + L_{\text{LQG}}(y_{t+1} - C(A\widehat{x}_t + Bu_t)) \,,$$

for filter gain $L_{\text{LQG}} = -PC^\top(CPC^\top + I)^{-1}$ where P is the solution to the DARE defined by $(A^\top, C^\top, I, I)$. This optimal output feedback controller satisfies the *separation principle*, meaning that the optimal controller K_{LQR} is computed independently of the optimal estimator gain L_{LQG}.

The LQR and LQG problems best model sensor noise, aggregate behavior, and natural processes arising from statical-mechanical systems. The robust version of these problems considers instead the worst-case quadratic cost for ℓ_2 or power-norm bounded noise. This is the $\mathcal{H}_\infty$ optimal control problem, which has a rich history [ZDG96].

In our final example, we consider robustness in the sense of worst-case deviations and ℓ_∞ bounded noise. In particular, the $\mathcal{L}_1$ control problem [DP87] minimizes the cost function

$$c(\mathbf{x}, \mathbf{u}) = \sup_{\substack{\mathbf{w},\mathbf{v} \\ t \geq 0}} \left\| \begin{matrix} Q^{1/2} x_t \\ R^{1/2} u_t \end{matrix} \right\|_\infty$$

for w_t and v_t such that $\|w_t\|_\infty \leq 1$, $\|v_t\|_\infty \leq 1$ for all k. The optimal controller does not obey the separation principle, and as such, there is no clear notion of an estimated state. This formulation best accommodates real-time safety constraints and actuator saturation, which we motivate in the following example.

Example 2.3 (Reference Tracking)**.** Consider a reference tracking problem where it is known that both the distances between waypoints and sensor errors are instantaneously ℓ_∞ bounded, and we want to ensure that the system remains within a bounded distance of the waypoints. Denoting the system state as $\boldsymbol{\xi}$ and the waypoint sequence as $\mathbf{r}$, the cost function is

$$c(\boldsymbol{\xi}, \mathbf{u}) = \sup_{\substack{\|r_{t+1} - r_t\|_\infty \leq 1, \\ \|v_t\|_\infty \leq 1, t \geq 0}} \left\| \begin{matrix} Q^{1/2}(\xi_t - r_t) \\ R^{1/2} u_t \end{matrix} \right\|_\infty .$$

If we specify costs by $Q = \text{diag}(1/b_{x,i}^2)$ and $R = \text{diag}(1/b_{u,i}^2)$, then as long as the optimal cost is less than 1, we can guarantee bounded tracking error $|\xi_{i,t} - r_{i,k}| \leq b_{x,i}$ and actuation $|u_{i,t}| \leq b_{u,i}$ for all possible realizations of the waypoint and sensor error processes. Considering the one-step lookahead case, we can define an augmented state, i.e. $x_t := [\xi_t; r_t]$,

Name	Disturbance class	Cost function	Use cases
LQR/$\mathcal{H}_2$	$\mathbb{E}[\boldsymbol{\nu}] = 0$, $\mathbb{E}[\boldsymbol{\nu}^4] < \infty$, ν_t i.i.d.	$\mathbb{E}_\nu \left[\lim_{T\to\infty} \sum_{t=0}^{T} \frac{1}{T} x_t^\top Q x_t + u_t^\top R u_t \right]$	Sensor noise, aggregate behavior, natural processes
$\mathcal{H}_\infty$	$\|\boldsymbol{\nu}\|_{pow} \le 1$, or $\|\boldsymbol{\nu}\|_2 \le 1$	$\sup_{\|\boldsymbol{\nu}\|_{pow}\le 1} \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} x_t^\top Q x_t + u_t^\top R u_t$, or $\sup_{\|\boldsymbol{\nu}\|_2 \le 1} \sum_{t=0}^{\infty} x_t^\top Q x_t + u_t^\top R u_t$	Modeling error, energy/power constraints
$\mathcal{L}_1$	$\|\boldsymbol{\nu}\|_\infty \le 1$	$\sup_{\|\boldsymbol{\nu}\|_\infty \le 1, t\ge 0} \left\| \begin{matrix} Q^{1/2} x_t \\ R^{1/2} u_t \end{matrix} \right\|_\infty$	Real-time safety constraints, actuator saturation/limits

Table 2.1: Different noise model classes induce different cost functions, and can be used to model different phenomenon, or combinations thereof. See texts by Zhou, Doyle, and Glover [ZDG96] and Dahleh and Pearson [DP87] for more details.

and model waypoints as bounded disturbances, i.e. $w_t := r_{t+1} - r_t$. A similar formulation exists for any T-step lookahead of the reference trajectory.

We can then formulate the following $\mathcal{L}_1$ optimal control problem,

$$\begin{aligned} \text{minimize} \quad & \sup_{\|\nu\|_\infty \le 1, t\ge 0} \left\| \begin{matrix} \bar{Q}^{1/2} x_t \\ R^{1/2} u_t \end{matrix} \right\|_\infty \\ \text{subject to} \quad & x_{t+1} = \bar{A} x_t + \bar{B} u_t + \bar{H} w_t \,, \\ & y_t = \bar{C} x_t + v_t, \end{aligned} \tag{2.3.5}$$

where

$$\bar{A} = \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \quad \bar{C} = \begin{bmatrix} C & 0 \end{bmatrix}, \quad \bar{H} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \quad \bar{Q}^{1/2} = \begin{bmatrix} Q^{1/2} & -Q^{1/2} \end{bmatrix} .$$

2.4 System Level Synthesis

We now formally discuss the *System Level Synthesis* (SLS) framework, which shows that optimal control problems can be equivalently cast in terms of system response variables.

To motivate the state feedback case, consider an arbitrary transfer function $\mathbf{K}$ denoting the map from state to control action, $\mathbf{u} = \mathbf{K}\mathbf{x}$. Then the closed-loop transfer matrices from the process noise $\mathbf{w}$ to the state $\mathbf{x}$ and control action $\mathbf{u}$ satisfy

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} (zI - A - B\mathbf{K})^{-1} \\ \mathbf{K}(zI - A - B\mathbf{K})^{-1} \end{bmatrix} \mathbf{w}. \tag{2.4.1}$$

This expression is non-convex in $\mathbf{K}$, posing a problem for efficiently solving optimal control problems. Therefore, we turn to an alternate parametrization. The following

theorem parameterizes the set of stable closed-loop transfer matrices that are achievable by some stabilizing controller $\mathbf{K}$.

Theorem 2.4.1 (state feedback Parameterization [WMD19])**.** *The following are true:*

- *The affine subspace defined by*

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I, \ \mathbf{\Phi_x}, \mathbf{\Phi_u} \in \frac{1}{z}\mathcal{RH}_\infty \tag{2.4.2}$$

  parameterizes all system responses (2.4.1) *from* $\mathbf{w}$ *to* $(\mathbf{x}, \mathbf{u})$*, achievable by an internally stabilizing state feedback controller* $\mathbf{K}$*.*

- *For any transfer matrices* $(\mathbf{\Phi_x}, \mathbf{\Phi_u})$ *satisfying* (2.4.2)*, the controller* $\mathbf{K} = \mathbf{\Phi_u}\mathbf{\Phi_x}^{-1}$ *is internally stabilizing and achieves the desired system response* (2.4.1)*.*

In the output feedback case, for $\mathbf{u} = \mathbf{K}\mathbf{y}$, closed-loop transfer matrices from the process $\mathbf{w}$ and measurement noise $\mathbf{v}$ to the state $\mathbf{x}$ and control action $\mathbf{u}$ satisfy

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} (zI - A - B\mathbf{K}C)^{-1} & (zI - A - B\mathbf{K}C)^{-1}B\mathbf{K} \\ \mathbf{K}C(zI - A - B\mathbf{K})^{-1} & \mathbf{K}C(zI - A - B\mathbf{K}C)^{-1}B\mathbf{K} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix}. \tag{2.4.3}$$

Again, this expression is non-convex in $\mathbf{K}$, so we turn to the system level parametrization.

Theorem 2.4.2 (output feedback Parameterization [WMD19])**.** *The following are true:*

- *The affine subspace defined by* $\mathbf{\Phi_{xw}}, \mathbf{\Phi_{xv}}, \mathbf{\Phi_{uw}} \in \frac{1}{z}\mathcal{RH}_\infty$, $\mathbf{\Phi_{uv}} \in \mathcal{RH}_\infty$,

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} = \begin{bmatrix} I & 0 \end{bmatrix}, \ \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} \begin{bmatrix} zI - A \\ -C \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix} \tag{2.4.4}$$

  parameterizes all system responses (2.4.3) *from* $(\mathbf{w}, \mathbf{v})$ *to* $(\mathbf{x}, \mathbf{u})$*, achievable by an internally stabilizing state feedback controller* $\mathbf{K}$*.*

- *For any transfer matrices* $(\mathbf{\Phi_{xw}}, \mathbf{\Phi_{xv}}, \mathbf{\Phi_{uw}}, \mathbf{\Phi_{uv}})$ *satisfying* (2.4.2)*, the controller* $\mathbf{K} = \mathbf{\Phi_{uv}} - \mathbf{\Phi_{uw}}\mathbf{\Phi_{xw}}^{-1}\mathbf{\Phi_{xv}}$ *is internally stabilizing and achieves the desired system response* (2.4.3)*.*

Theorem 2.4.1 and 2.4.2 make formal the intuition that linear controllers are equivalent to system responses constrained to lie in an affine space. As a result, optimization problems over stabilizing linear controllers can be reparameterized into optimization problems over system response variables. We will sometimes denote that a system response $\mathbf{\Phi}$ satisfies the *realizability constraints* in (2.4.2) or (2.4.4) as $\mathbf{\Phi} \in \mathcal{A}$, where $\mathcal{A}$ denotes affine space defined by (A, B) or (A, B, C) depending on the context. Theorem 2.4.1 and 2.4.2 also describe how to recover the control law from the system responses. This controller can be implemented via a state-space realization [AM17] or as an interconnection of the system response elements [WMD19].

Control Costs as System Norms

In this SLS framework, many control costs (including those in Table 2.1) can be written as system norms. We first consider the $\mathcal{H}_\infty$ and $\mathcal{L}_1$ costs, which can be written in signal notation as

$$c(\mathbf{x}, \mathbf{u}) = \sup_{\substack{\|\mathbf{w}\| \le \varepsilon_w \\ \|\mathbf{v}\| \le \varepsilon_v}} \left\| \begin{bmatrix} Q^{1/2}\mathbf{x} \\ R^{1/2}\mathbf{u} \end{bmatrix} \right\| ,$$

where ε_w and ε_v respectively bound the norms of $\mathbf{w}$ and $\mathbf{v}$ and we allow $\|\cdot\|$ to represent the ℓ_2/power norm or ℓ_∞ norm and associated induced norm. Then by substituting the identity (2.2.9), the cost can be written as

$$\sup_{\substack{\|\mathbf{w}\| \le \varepsilon_w \\ \|\mathbf{v}\| \le \varepsilon_e}} \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{xw}} & \boldsymbol{\Phi}_{\mathbf{xv}} \\ \boldsymbol{\Phi}_{\mathbf{uw}} & \boldsymbol{\Phi}_{\mathbf{uv}} \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} \right\| = \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{xw}} & \boldsymbol{\Phi}_{\mathbf{xv}} \\ \boldsymbol{\Phi}_{\mathbf{uw}} & \boldsymbol{\Phi}_{\mathbf{uv}} \end{bmatrix} \begin{bmatrix} \varepsilon_w I & \\ & \varepsilon_v I \end{bmatrix} \right\| .$$

For LQR and LQG control, the control objective is equivalent to a system $\mathcal{H}_2$ norm, a fact that we now derive. From the expression (2.2.7), we have that

$$x_t^\top Q x_t = \sum_{k=1}^{t+1} \sum_{\ell=1}^{t+1} \begin{bmatrix} w_{t-k} \\ v_{t-k} \end{bmatrix}^\top \begin{bmatrix} \Phi_{xw}(k)^\top \\ \Phi_{xv}(k)^\top \end{bmatrix} Q \begin{bmatrix} \Phi_{xw}(\ell) & \Phi_{xv}(\ell) \end{bmatrix} \begin{bmatrix} w_{t-\ell} \\ v_{t-\ell} \end{bmatrix} . \tag{2.4.5}$$

Then as long as the process and measurement noise are zero mean, independent from each other and across time, and have variances σ_w^2 and σ_v^2 respectively,

$$\mathbb{E}\left[x_t^\top Q x_t\right] = \sum_{k=1}^{t+1} \mathbf{Tr}\left(\begin{bmatrix} \sigma_w \Phi_{xw}(k)^\top \\ \sigma_v \Phi_{xv}(k)^\top \end{bmatrix} Q \begin{bmatrix} \sigma_w \Phi_{xw}(k) & \sigma_v \Phi_{xv}(k) \end{bmatrix} \right) .$$

A similar expression holds for the input term. We can then write

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[x_t^\top Q x_t + u_t^\top R u_t\right] &= \sum_{t=1}^{\infty} \mathbf{Tr}\left(\begin{bmatrix} \sigma_w \Phi_{xw}(t)^\top \\ \sigma_v \Phi_{xv}(t)^\top \end{bmatrix} Q \begin{bmatrix} \sigma_w \Phi_{xw}(t) & \sigma_v \Phi_{xv}(t) \end{bmatrix} \right) \\ &\quad + \mathbf{Tr}\left(\begin{bmatrix} \sigma_w \Phi_{uw}(t)^\top \\ \sigma_v \Phi_{uv}(t)^\top \end{bmatrix} Q \begin{bmatrix} \sigma_w \Phi_{uw}(t) & \sigma_v \Phi_{uv}(t) \end{bmatrix} \right) \\ &= \sum_{t=1}^{\infty} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Phi_{xw}(t) & \Phi_{xv}(t) \\ \Phi_{uw}(t) & \Phi_{uv}(t) \end{bmatrix} \begin{bmatrix} \sigma_w I & 0 \\ 0 & \sigma_v I \end{bmatrix} \right\|_F^2 \\ &= \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{xw}} & \boldsymbol{\Phi}_{\mathbf{xv}} \\ \boldsymbol{\Phi}_{\mathbf{uw}} & \boldsymbol{\Phi}_{\mathbf{uv}} \end{bmatrix} \begin{bmatrix} \sigma_w I & 0 \\ 0 & \sigma_v I \end{bmatrix} \right\|_{\mathcal{H}_2}^2 . \end{aligned}$$

This derivation requires only fairly general assumptions about the noise processes. However, using system response variables implicitly assumes a linear (possibly dynamic) controller. Thus, the equivalence between the $\mathcal{H}_2$ norm and the LQR/LQG cost only holds under this restricted policy class. The optimal controller is in fact a linear dynamic controller a) in the state feedback case and b) when the noise processes are Gaussian (as in Example 2.1 and 2.2). However, the optimal filtering procedure for more general noise processes will not be a Kalman filter, and it may not be linear [Ber95].

Robustness to Unknown Dynamics

So far we have seen that System Level Synthesis provides a convenient parametrization that makes transparent the effects of noise signals and allows for convex optimization. Neither of these properties is entirely unique to SLS, and other approaches to control synthesis are possible. However, a major benefit is in how SLS handles misspecification in the dynamics model. In this section, we show how uncertainty in the dynamics can be handled in a transparent manner. This allows for the computation of robust controllers that stabilize all systems within an uncertainty set.

Although classic methods exist for computing robust controllers [Fer97; Pag95; SAPT02; WP95], they typically require solving non-convex optimization problems, and it is not readily obvious how to extract interpretable measures of controller performance as a function of the size of the uncertainty. By lifting the system description into a higher dimensional space, SLS makes it tractable to reason analytically about uncertain dynamics.

The robust variant of Theorem 2.4.1 traces the effects of misspecification.

Theorem 2.4.3 (Robust Stability [MWA17])**.** *Let $\mathbf{\Phi_x}$ and $\mathbf{\Phi_u}$ be two transfer matrices in $\frac{1}{z}\mathcal{RH}_\infty$ such that*

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I + \mathbf{\Delta}. \tag{2.4.6}$$

Then the controller $\mathbf{K} = \mathbf{\Phi_u}\mathbf{\Phi_x}^{-1}$ stabilizes the system described by (A, B) if and only if $(I + \mathbf{\Delta})^{-1} \in \mathcal{RH}_\infty$. Furthermore, the resulting system response is given by

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} (I + \mathbf{\Delta})^{-1} \mathbf{w}. \tag{2.4.7}$$

Corollary 2.4.4. *Under the assumptions of Theorem 2.4.3, if $\|\mathbf{\Delta}\| < 1$ for any induced norm $\|\cdot\|$, then the controller $\mathbf{K} = \mathbf{\Phi_u}\mathbf{\Phi_x}^{-1}$ stabilizes the system described by (A, B).*

The proof of the corollary follows immediately from the small gain theorem.

To see why these results are useful, suppose that some estimate $(\widehat{A}, \widehat{B})$ is used for synthesis. Then,

$$\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I \iff \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I + \begin{bmatrix} \widehat{A} - A & \widehat{B} - B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix}.$$

Therefore Theorem 2.4.3 allows us to write a closed-form expression for the trajectory of the uncertain system in terms of the designed response ($\mathbf{\Phi_x}, \mathbf{\Phi_u}$), and uncertainties $A - \widehat{A}$ and $B - \widehat{B}$.

A similar result holds for output feedback controllers which makes transparent the effect of uncertainty in the measurement matrix C (see e.g. work by Boczar, Matni, and Recht [BMR18]).

2.5 Finite Dimensional Approximations

The System Level Synthesis framework introduced in Section 2.4 shows that optimal control problems like (2.3.1) can be cast in terms of system response variables. However, system responses are semi-infinite, so it is not clear how to solve the resulting optimization problem efficiently. An elementary approach to reducing the semi-infinite program to a finite dimensional one is to only optimize over the first L elements of the transfer functions, effectively taking a finite impulse response (FIR) approximation. Since these are stable maps, we expect the effects of such an approximation to be negligible as long as the optimization horizon L is chosen to be sufficiently large. Later in this section, we show that this is indeed the case.

We first outline how the optimization variables and constraints admit finite-dimensional representations. To derive the finite expressions, it is useful to consider the (truncated) Toeplitz matrix representation of the transfer functions. The FIR approximation optimizes over $((\Phi_x(t), \Phi_u(t))_{t=1}^L$ in the state feedback case and $((\Phi_{xw}(t), \Phi_{uw}(t), \Phi_{xv}(t), \Phi_{uv}(t))_{t=1}^L$ in the output feedback case. The affine realizability constraints reduce to a finite number of linear equality constraints in the form of (2.2.5) or (2.2.8). The $\mathcal{H}_2$ norm can be cast as a second order cone constraint. The $\mathcal{H}_\infty$ norm can be reduced to a compact SDP as in Theorem 5.8 of Dumitrescu [Dum07], described explicitly for SLS in Appendix G.3 of Dean et al. [DMMRT18]. The $\mathcal{L}_1$ norm becomes an $\ell_\infty \to \ell_\infty$ operator norm on the horizontal concatenation of system response elements, $\Phi[L : 1]$. Finally, the controller given by $\mathbf{K} = \mathbf{\Phi_u}\mathbf{\Phi_x}^{-1}$ or $\mathbf{K} = \mathbf{\Phi_{uv}} - \mathbf{\Phi_{uw}}\mathbf{\Phi_{xw}}^{-1}\mathbf{\Phi_{xv}}$ can be written in an equivalent state-space realization (A_K, B_K, C_K, D_K) via Theorems 2 and 3 of Anderson and Matni [AM17].

In the interest of clarity, for the remainder of this thesis we will present the infinite horizon version of the optimization problems, with the understanding that finite horizon approximations are necessary in practice. The sub-optimality results presented in Chapters 3 and 4 hold up to constant factors for these approximations. We now make precise that for sufficiently long horizons L, the effects of the approximation are negligible.

Sub-optimality of Finite Approximation

Consider a general state feedback optimal control problem cast in system response variables:

$$\begin{aligned} \text{minimize} \quad & \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \right\| \\ \text{subject to} \quad & \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I, \\ & \mathbf{\Phi_x}, \mathbf{\Phi_u} \in \tfrac{1}{z}\mathcal{RH}_\infty \end{aligned} \tag{2.5.1}$$

where $\|\cdot\|$ is any induced norm. We define cost($\mathbf{K}$) to be the norm of the system response generated by applying the controller $\mathbf{K} = \mathbf{\Phi_u}^{-1}\mathbf{\Phi_x}$ to the system with linear dynamics (A, B).

The finite approximation to this problem optimizes over only the first L impulse response elements:

$$\begin{aligned} \min_{0\le\gamma<1} \quad & \min_{\mathbf{\Phi_x},\mathbf{\Phi_u},V} \frac{1}{1-\gamma} \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \right\| \\ \text{subject to} \quad & \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I + z^{-L}V, \\ & \mathbf{\Phi_x} = \textstyle\sum_{t=1}^{L} z^{-t}\Phi_x(t), \;\; \mathbf{\Phi_u} = \sum_{t=1}^{L} z^{-t}\Phi_u(t)\,, \\ & \|V\| \le \gamma\,. \end{aligned} \tag{2.5.2}$$

The slack term V accounts for the error introduced by truncating the infinite response transfer functions. This allows us to optimize over controllers that don't necessarily force the system to have a finite impulse response. We remark that it is possible to enforce that the system is FIR by setting $V = 0$ and $\gamma = 0$; the problem remains feasible whenever (A, B) is controllable and L is large enough. While that makes for a cruder approximation, it avoids the additional complexity of searching over the scalar variables γ.

Intuitively, if the truncated tail captured by V is sufficiently small, then the finite approximation has near optimal performance. The next result formalizes this intuition. It shows that the cost penalty incurred decays exponentially in the horizon L over which the approximation is taken.

Theorem 2.5.1. *Let* $\mathbf{K}_\star = \mathbf{\Phi}_\mathbf{u}^\star(\mathbf{\Phi}_\mathbf{x}^\star)^{-1}$ *be the controller resulting from the optimal control problem* (2.5.1) *and suppose that* $\mathbf{\Phi}_\mathbf{x}^\star \in \mathcal{RH}_\infty(C_\star, \rho_\star)$. *Let* $\mathbf{K}_L = \mathbf{\Phi}_\mathbf{u}^L(\mathbf{\Phi}_\mathbf{x}^L)^{-1}$ *be controller resulting from the finite approximation* (2.5.2). *Then for any* $0 < \varepsilon < 1$, *as long as*

$$L \ge \log(2C_\star/\varepsilon)/\log(1/\rho_\star) + 1\,,$$

the relative sub-optimality of the finite approximation is bounded by ε

$$\frac{\text{cost}(\mathbf{K}^L) - \text{cost}(\mathbf{K}^\star)}{\text{cost}(\mathbf{K}^\star)} \le \varepsilon\,.$$

Proof. First, notice that since the affine realizability constraint is not exactly met, by the robustness result in Theorem 2.4.3

$$\operatorname{cost}(\mathbf{K}_L) = \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}}^L \\ \mathbf{\Phi}_{\mathbf{u}}^L \end{bmatrix} (I + z^{-L} V_L)^{-1} \right\| .$$

The expression can be further simplified using the properties of induced norms,

$$\operatorname{cost}(\mathbf{K}_L) \le \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}}^L \\ \mathbf{\Phi}_{\mathbf{u}}^L \end{bmatrix} \right\| \frac{1}{1 - \|V_L\|} \le \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}}^L \\ \mathbf{\Phi}_{\mathbf{u}}^L \end{bmatrix} \right\| \frac{1}{1 - \gamma_L} ,$$

where the final inequality holds due to the inequality constraint $\|V\| \le \gamma$.

Next, we construct a feasible solution to (2.5.2):

$$\tilde{\mathbf{\Phi}}_{\mathbf{x}} = \sum_{t=1}^{L} z^{-t} \Phi_x^\star(t), \quad \tilde{\mathbf{\Phi}}_{\mathbf{u}} = \sum_{t=1}^{L} z^{-t} \Phi_u^\star(t), \quad \tilde{V} = -\Phi_x^\star(L+1), \quad \tilde{\gamma} = C_\star \rho_\star^{L+1} . \tag{2.5.3}$$

Because $(\mathbf{\Phi}_{\mathbf{x}}^\star, \mathbf{\Phi}_{\mathbf{u}}^\star)$ is a solution to the original optimal control problem, the infinite system response satisfies the affine realizability constraint for (A, B). It is straightforward to check that the under the definition of $\tilde{V}$, the truncated affine constraints are satisfied as well. We also have that $\|\tilde{V}\| = \|\Phi_x^\star(L+1)\| \le C_\star \rho_\star^{L+1} = \tilde{\gamma}$ by the definition of $C_\star$ and $\rho_\star$. By the assumption on L and ε, $C_\star \rho_\star^{L+1} < 1$.

Then by the optimality of $(\mathbf{\Phi}_{\mathbf{x}}^L, \mathbf{\Phi}_{\mathbf{u}}^L, \gamma_L)$,

$$\frac{1}{1 - \gamma_L} \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}}^L \\ \mathbf{\Phi}_{\mathbf{u}}^L \end{bmatrix} \right\| \le \frac{1}{1 - \tilde{\gamma}} \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Phi}}_{\mathbf{x}} \\ \tilde{\mathbf{\Phi}}_{\mathbf{u}} \end{bmatrix} \right\| .$$

Notice that $\tilde{\mathbf{\Phi}}_{\mathbf{x}} = \mathbf{\Phi}_{\mathbf{x}}^\star + \sum_{t=L+1}^{\infty} z^{-t} \Phi_x^\star(t)$ and $\tilde{\mathbf{\Phi}}_{\mathbf{u}} = \mathbf{\Phi}_{\mathbf{u}}^\star + \sum_{t=L+1}^{\infty} z^{-t} \Phi_u^\star(t)$. Therefore, by triangle inequality,

$$\left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{\Phi}}_{\mathbf{x}} \\ \tilde{\mathbf{\Phi}}_{\mathbf{u}} \end{bmatrix} \right\| \le \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}}^\star \\ \mathbf{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\| .$$

Combining this chain of inequalities,

$$\operatorname{cost}(\mathbf{K}^L) \le \frac{1}{1 - C \star \rho_\star^{L+1}} \operatorname{cost}\left(\mathbf{K}^\star\right) \le \left(1 + 2C \star \rho_\star^{L+1}\right) \operatorname{cost}\left(\mathbf{K}^\star\right) .$$

where the final inequality holds by the observation that $\frac{1}{1-x} \le 1 + 2x$ whenever $x \le 1/2$, which is implied by the assumption on ε. □

A nearly identical result also holds for the $\mathcal{H}_2$ norm, where in the finite problem, the constraint on V is in terms of the $\ell_2 \to \ell_2$ operator norm. Similar results hold when the optimal control problem has additional constraints so long as care is taken to incorporate the variable V into them analogously. Finally, it is possible to generalize this to the output feedback case, though the analogous truncated synthesis problem will search over two auxiliary variables.

Chapter 3

Learning to Control the Linear Quadratic Regulator

3.1 Introduction

In this chapter, we attempt to build a foundation for the theoretical understanding of how machine learning interfaces with control by analyzing one of the most well-studied problems in classical optimal control, the *Linear Quadratic Regulator* (LQR). This chapter uses material first presented in papers coauthored with Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu [DTMR19; DMMRT20; DMMRT18].

We assume that the system to be controlled obeys *linear* dynamics, and we wish to minimize some *quadratic* function of the system state and control action. The optimal control problem can be written as

$$\begin{array}{ll} \text{minimize} & \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} x_t^\top Q x_t + u_{t-1}^\top R u_{t-1}\right] \\ \text{subject to} & x_{t+1} = Ax_t + Bu_t + w_t \end{array} . \tag{3.1.1}$$

In what follows, we will be concerned with the *infinite time horizon* variant of the LQR problem where we let the time horizon T go to infinity. When the dynamics are known, this problem has a celebrated closed form solution based on the solution of matrix Riccati equations [ZDG96].

We will further consider the *constrained* version of the LQR problem, in which the system state and control actions are *linearly* constrained. We define the polytopic constraint sets:

$$\mathcal{X} := \{x \ : \ F_x x \le b_x\}, \ \ \mathcal{U} := \{u \ : \ F_u u \le b_u\} . \tag{3.1.2}$$

The incorporation of such constraints can ensure system safety (by avoiding unsafe regions of the state space) and reliability (by preventing controller saturation). This richer modeling comes at a cost: the constrained LQR problem does not have a simple closed form solution. Nevertheless, by restricting the policy class to linear controllers, we will develop a framework in which the problem becomes tractable.

We analyze the LQR problem when the dynamics of the system are unknown, and we can measure the system's response to varied inputs. We will assume that we can conduct experiments of the following form: given some initial state x_0, we can evolve the dynamics for T time steps using any control sequence $(u_0, \dots, u_{T-1})$, measuring the resulting output $(x_1, \dots, x_T)$. Is it possible to keep the system safe using only the data collected?

We propose a method that couples our uncertainty in estimation with the control design. Our main approach uses the following framework of *Coarse-ID control* to solve the problem of LQR with unknown dynamics:

1. Use supervised learning to learn a coarse model of the dynamical system to be controlled. We refer to the system estimate as the *nominal system*.

2. Build probabilistic guarantees about the distance between the nominal system and the true, unknown dynamics.

3. Solve a robust optimization problem over controllers that optimizes performance of the nominal system while penalizing signals with respect to the estimated uncertainty, ensuring safe and robust execution.

For a sufficient amount of data, this approach is guaranteed to return a control policy with small relative cost which guarantees the safety and asymptotic stability of the closed-loop system. Though simple to state, the analysis of this procedure will take the remainder of the chapter: Section 3.2 focus on the estimation in the second step, Section 3.3 develops a method for synthesizing robust controllers in the third step, and Section 3.4 provides end-to-end performance guarantees. We present numerical experiments demonstrating the capability of this procedure in Section 3.5, and offer concluding remarks and an accounting of open problems in Section 3.6.

Problem Setting

We fix an underlying linear dynamical system with full state observation,

$$x_{t+1} = Ax_t + Bu_t + w_t \,, \tag{3.1.3}$$

where we have initial condition $x_0 \in \mathbb{R}^n$, sequence of inputs $(u_0, u_1, \dots) \subseteq \mathbb{R}^m$, and disturbance process $(w_0, w_1, \dots) \subseteq \mathbb{R}^n$. We consider the expected infinite horizon quadratic cost for the system (A, B) in feedback with a linear controller $\mathbf{K}$:

$$J(A, B, \mathbf{K})^2 := \frac{1}{\sigma_w^2} \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_w [x_t^\top Q x_t + u_t^\top R u_t] \,.$$

We assume that the disturbance process is generated by any distribution which satisfies $\mathbb{E}[w_t] = 0, \mathbb{E}[w_t w_t^\top] = \sigma_w^2 I$, and independence across time, i.e., $w_t \perp w_l$ for $t \neq k$. Note that

any distribution satisfying these constraints induces the same expected quadratic cost, so it is unnecessary to specify a specific distribution. We further assume that the disturbance is norm bounded so that it satisfies $\|w_t\|_\infty \leq r_w$ for all $k \geq 0$. This assumption makes it possible to reason about constraints on the system state and control input on infinite time horizons. We require that the constraints (3.1.2) are satisfied for any possible disturbance sequence.

Putting the cost and the constraints together, the optimal control problem that acts as our baseline is:

$$\begin{array}{ll} \text{minimize}_{\mathbf{K}} & J(A, B, \mathbf{K}) \\ \text{subject to} & x_0 \text{ fixed}, \;\; u_t = \mathbf{K}(x_{0:t}) \;, \\ & x_{t+1} = Ax_t + Bu_t + w_t \;, \\ & F_x x_t \leq b_x, \;\; F_u u_t \leq b_u \;\; \forall t \;, \forall \{w_t : \|w_t\|_\infty \leq r_w\} \;. \end{array} \tag{3.1.4}$$

Above, we search over linear dynamic stabilizing feedback controllers for (A, B) of the form $\mathbf{u} = \mathbf{K}\mathbf{x}$. This is made possible by the system level synthesis framework described in the previous chapter.

As will we show, the optimal control problem given in (3.1.4) is a convex, but infinite-dimensional problem. It is an idealized baseline to compare our actual solutions to; our sub-optimality guarantees will be with respect to the optimal cost achieved by this problem. It is a relevant baseline, since it optimizes for average case performance but ensures safety for the worst-case behavior, consistent with literature on Model Predictive Control (MPC) [MSR05; OJM08]. We remark that an alternative to (3.1.4) is to replace the worst case constraints with probabilistic chance constraints [FGS16]. We do not work with chance constraints because they are generally difficult to directly enforce on an infinite horizon; arguments around recursive feasibility using robust invariant sets are common in the literature to deal with this issue. When the system is unconstrained, i.e. when $\mathcal{X} = \mathbb{R}^n$ and $\mathcal{U} = \mathbb{R}^m$, this baseline is the widely-studied classical LQR problem.

Related Work

We first describe related work in the *estimation of unknown linear systems* and then turn to connections in the literature on *robust control with uncertain models* and the *satisfaction of safety constraints*. We then end this review with a discussion of recent works that directly address the LQR problem and related variants.

Estimation of unknown linear systems. Estimation of unknown linear dynamical systems has a long history in the system identification subfield of control theory. Classical results focus on asymptotic guarantees and/or frequency domain methods [Lju99; CG00; HJN91; Gol98]. Our goal is to analyze the optimality of controllers constructed from a finite amount of data, so we focus on non-asymptotic guarantees for state space identification. Early results on non-asymptotic rates for parameter identification featured conservative

bounds exponential in the system degree [CW02; VK08]. In the past decade, the first polynomial time guarantees were presented in terms of predictive output performance of the model [PIM10; HMR18]. Recently, Hazan, Singh, and Zhang [HSZ17] and Hazan et al. [HLSZZ18] proposed a novel spectral filtering algorithm and showed that one can compete in a regret setting in terms of prediction error. It is not clear how such prediction error bounds can be used in a downstream robust synthesis procedure or converted into a sub-optimality result.

In parallel to the system identification community, identification of auto-regressive time series models is a widely studied topic in the statistics literature [BJRL15; GZ01; KM17; MSS17; MR10]. Many of these works studied generalization for data which is not independent over time, extending the standard learning theory guarantees. At the crux of these arguments lie various mixing assumptions [Yu94], which limits the analysis to stable dynamical systems. Furthermore, results from this line of research suggest that systems with smaller mixing time (i.e. systems that are more stable) are easier to identify (i.e. take less samples). This does not align with our empirical testing, which suggests that identification benefits from more easily excitable systems.

Our results rely on recent work by Simchowitz et al. [SMTJR18] who take a step towards reconciling this issue for stable systems. Since the work presented in this chapter was published, there has been a renewed interest in providing finite data guarantees for classic system identification procedures. For partially observed linear systems, Oymak and Ozay [OO19] and Sarkar, Rakhlin, and Dahleh [SRD21] present non-asymptotic results for a Ho-Kalman-like procedure and Simchowitz, Boczar, and Recht [SBR19] propose a semi-parametric least squares method that is consistent for marginally stable systems. In the state observation setting, Sarkar and Rakhlin [SR19] show that least-squares estimation is consistent for a restricted class of unstable systems. Developing consistent estimators for arbitrary unstable systems remains an open problem.

Robust controller design. For end-to-end guarantees, parameter estimation is only half the picture. It is necessary to ensure that the computed controller guarantees stability and performance for the entire family of system models described by a nominal estimate and a set of unknown but bounded model errors. This problem has a rich history in the controls community. When modeling errors are arbitrary norm bounded linear time-invariant (LTI) operators in feedback with the nominal plant, traditional small gain theorems and robust synthesis techniques exactly solve the problem [DD94; ZDG96]. For more structured errors, there are sophisticated techniques based on structured singular values like μ-synthesis [Doy82; FTD91; PD93; YND91] or integral quadratic constraints (IQCs) [MR97]. While theoretically appealing and much less conservative, the resulting synthesis methods are both computationally intractable (although effective heuristics do exist) and difficult to interpret analytically.

In order to bound the degradation in performance of controlling an uncertain system in terms of the size of the perturbations affecting it, we leverage a novel parameterization

of robustly stabilizing controllers based on the SLS framework [WMD19] that is reviewed in Chapter 2. Originally developed to allow for scaling optimal and robust controller synthesis techniques to large-scale systems, the SLS framework can be viewed as a generalization of the celebrated Youla parameterization [YJB76]. We use SLS because it allows us to account for model uncertainty in a transparent and analytically tractable way.

Since the work presented in this chapter was published, there has been a renewed interest in robust synthesis procedures for the types of uncertainty sets that result from estimation [US18; UFSH19; US20].

Safety constraints. The design of controllers that guarantee robust constraint satisfaction has long been considered in the context of model predictive control [BM99], including methods that model uncertainty in the dynamics directly [KBM96], or model it as an enlarge disturbance process [MSR05; GKM06]. These traditional works do not usually consider identifying the unknown dynamics. Strategies for incorporating safety with estimation of the dynamics include experiment-design inspired costs [LRH11], decoupling learning from constraint satisfaction [AGST13], and set-membership methods rather than parameter estimation [TFSM14; LAC17]. Due to the receding horizon nature of model predictive controllers, this literature relies on set invariance theory for infinite horizon guarantees [Bla99]. In this chapter, we consider the infinite horizon problem directly, and therefore we do not require computation of invariant sets.

The machine-learning community has begun to consider safety in reinforcement learning, where much work positions itself as ensuring safety for arbitrary dynamical systems in lieu of providing statistical guarantees [BS15; BTSK17; DDV+18; CNDG18]. One line of work proposes methods for modifying potentially unsafe inputs generated by arbitrary reinforcement learning algorithms, building on ideas from robust MPC [WZ18; KBTK18]. Despite good empirical results, it remains to show whether the modified inputs successfully excite the system, allowing for a statistical learning rate. Most similar to our work is that of Lu et al. [LZBRS17], who propose a method to allow excitation on top of a safe controller, but consider only finite-time safety and require non-convex optimization to obtain formal guarantees. Later follow-up work by Ahmadi et al. [ACST21] guarantees safety and excitation on the basis of planning over one and two step horizons.

PAC learning and reinforcement learning. There is a body of work that focuses on the unconstrained LQR problem from the perspective of probably-approximately-correct (PAC) and reinforcement learning. In terms of end-to-end guarantee, our work is most comparable to that of Fiechter [Fie97], who analyzes an identify-then-control scheme similar to the one we propose. There are several key differences. First, our probabilistic bounds on identification are much sharper, since we leverage more modern tools from high dimensional statistics. Second, Fiechter implicitly assumes that the estimated controller induces a closed-loop that is not only stable but also contractive. While this is a very strong assumption, contractive closed-loop assumptions are actually pervasive throughout

the early literature, as we describe below. Finally, Fiechter proposes to directly solve the discounted LQR problem with the identified model, and takes a *certainty equivalent* approach to controller synthesis, which does not take into account any uncertainty. This is problematic for two reasons: the *optimal* solution to the discounted LQR problem does not guarantee stabilization (see e.g. [PBND17]) and even if the optimal solution does stabilize the underlying system, the certainty equivalent controller may not. Our experiments demonstrate this behavior.

There is a body of related efforts in RL and online learning. In a seminal work, Abbasi-Yadkori and Szepesvári [AS11] propose to use the optimism in the face of uncertainty (OFU) principle for the LQR problem, by maintaining confidence ellipsoids on the true parameter, and using the controller which, in feedback, minimizes the cost objective the most among all systems in the confidence ellipsoid. They show that the regret of this approach scales as $\widetilde{O}(\sqrt{T})$, where $\widetilde{O}(\cdot)$ hides problem dependent constants and poly-logarithmic factors. However, the method is computationally intractable, and the analysis makes the very strong assumption that the optimal closed-loop systems are contractive for every A, B in the confidence ellipsoid. The regret bound is improved by Ibrahimi, Javanmard, and Roy [IJR12] under additional sparsity constraints on the dynamics. More recently, Cohen, Koren, and Mansour [CKM19] and Abeille and Lazaric [AL20] propose computationally tractable approximations to the OFU principle based on semidefinite programming and Lagrangian relaxation, respectively. They improve the analysis, relaxing some of the more restrictive assumptions, and show that the approximate methods achieve the same regret scaling.

An alternative online approach is to use Thompson sampling [RVKOW18] for exploration. This method has been shown to achieve $\widetilde{O}(\sqrt{T})$ regret for one dimensional systems [AL17; AL18] or within a Bayesian framework of expected regret [OGJ17]. These works also make the same restrictive assumption that the optimal closed-loop systems are uniformly contractive over some known set. Recent work shows that a naive ϵ-greedy exploration scheme in combination with the certainty equivalent controller achieves optimal regret [MTR19; SF20]. In fact, the regret of this scheme does not depend directly on the state dimension; Perdomo et al. [PSAB21] show that rather than requiring consistent parameter recovery, the performance of certainty equivalent control scales with the prediction error of the estimates. The success of the certainty equivalent controller in online settings hinges on large enough horizons T to guarantee small enough estimation errors. In our framework, we construct controllers with stability and performance guarantees even in moderate error regimes.

Many of the aforementioned works assume enough knowledge to construct an initial stabilization controller. In recent work, Treven et al. [TCMK21] propose a data-driven method for constructing a controller for potentially unstable systems under a Bayesian prior.

Since the work presented in this chapter was published, additional variants of the LQR problem have been investigated, including the partially observed setting [TMP20;

LAHA20] and non-stochastic noise models [ABHKS19; HD19]. There has also been a resurgence of interest in model-free methods, which do not rely on estimating the dynamics model directly. Fazel et al. [FGKM18] show that randomized search algorithms similar to policy gradient can learn the optimal controller with a polynomial number samples. These ideas have also been applied to continuous time settings [MZSJ21]. Approaches based on characterizing the value function and Q function are also popular: Yang et al. [YCHW19] study the actor critic algorithm and Krauth, Tu, and Recht [KTR19] investigate least square temporal differences and policy improvement. Tu and Recht [TR19] present a class of dynamics models for which there is a gap in the sample complexity of model free and model based methods, suggesting that model based methods, such as the one investigated here, are more efficient, at least when it is known that the dynamics are linear.

3.2 System Identification through Least-Squares

To estimate a coarse model of the unknown system dynamics, we turn to the simple method of linear least squares. By running experiments in which the system starts at x_0 and the dynamics evolve with a given input, we can record the resulting state observations. The set of inputs and outputs from each such experiment will be called a *rollout*, and the resulting dataset is $\{(x_t, u_t) : 0 \le t \le T\}$. Therefore, we can estimate the system dynamics by

$$(\widehat{A}, \widehat{B}) \in \arg\min_{(A,B)} \sum_{t=0}^{T-1} \frac{1}{2} \|Ax_t + Bu_t - x_{t+1}\|_2^2. \tag{3.2.1}$$

For the Coarse-ID control setting, a good estimate of error is just as important as the estimate of the dynamics. Statistical theory allows us to quantify the error of the least squares estimator. First, we present an illustrative theoretical analysis of the error in a simplified setting. Then, we present results that apply in full generality.

Least Squares Estimation as a Random Matrix Problem

We begin by explicitly writing the form of the least squares estimator. First, fixing notation to simplify the presentation, let $\Theta := \begin{bmatrix} A & B \end{bmatrix}^\top \in \mathbb{R}^{(n+m)\times n}$ and let $z_t := \begin{bmatrix} x_t \\ u_t \end{bmatrix} \in \mathbb{R}^{n+m}$. Then the system dynamics can be rewritten, for all $t \ge 0$,

$$x_{t+1}^\top = z_t^\top \Theta + w_t^\top .$$

Then in a rollout, we will collect

$$X_T := \begin{bmatrix} x_1^\top \\ x_2^\top \\ \vdots \\ x_T^\top \end{bmatrix}, \quad Z_T := \begin{bmatrix} z_0^\top \\ z_1^\top \\ \vdots \\ z_{T-1}^\top \end{bmatrix}, \quad W_T := \begin{bmatrix} w_0^\top \\ w_1^\top \\ \vdots \\ w_{T-1}^\top \end{bmatrix}. \tag{3.2.2}$$

The system dynamics give the identity

$$X_T = Z_T\Theta + W_T .$$

The least squares estimator for Θ is (assuming for now invertibility of $Z_T^\top Z_T$),

$$\widehat{\Theta} = (Z_T^\top Z_T)^{-1} Z_T^\top X_T = \Theta + (Z_T^\top Z_T)^{-1} Z_T^\top W_T . \tag{3.2.3}$$

Then the estimation error is given by

$$E := \widehat{\Theta} - \Theta = (Z_T^\top Z_T)^{-1} Z_T^\top W_T . \tag{3.2.4}$$

The magnitude of this error is the quantity of interest in determining confidence sets around estimates $(\widehat{A}, \widehat{B})$. However, since W_T and Z_T are not independent, this estimator is difficult to analyze using standard methods. We therefore begin with a simplified procedure using Gaussian noise injection and system resets to recover an independent data estimator in the following subsection. After that, we consider estimation from dependent data collected during closed-loop system operation, and present results that rely on recently developed statistical results for linear system estimation.

Simplified Gaussian Setting

To sidestep issues of data dependence, we suppose the ability to reset the system to $x_0 = 0$ and perform multiple system rollouts. Having the ability to reset the system to a state independent of past observations is important for this analysis, and it could also be

Algorithm 1 Estimation of linear dynamics with independent data

1: **for** ℓ from 1 to N **do**
2: $\quad x_0^{(\ell)} = 0$
3: $\quad$ **for** t from 0 to $T-1$ **do**
4: $\quad\quad x_{t+1}^{(\ell)} = Ax_t^{(\ell)} + Bu_t^{(\ell)} + w_t^{(\ell)}$ with $w_t^{(\ell)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2 I_n)$ and $u_t^{(\ell)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2 I_m)$.
5: $\quad$ **end for**
6: **end for**
7: $(\widehat{A}, \widehat{B}) \in \arg\min_{(A,B)} \sum_{\ell=1}^{N} \frac{1}{2} \|Ax_{T-1}^{(\ell)} + Bu_{T-1}^{(\ell)} - x_T^{(\ell)}\|_2^2$

practically important for potentially unstable systems. We excite the system with Gaussian noise for N rollouts, each of length T. The resulting dataset is $\{(x_t^{(\ell)}, u_t^{(\ell)}) : 1 \le \ell \le N, 0 \le t \le T\}$, where t indexes the time in one rollout and ℓ indexes independent rollouts.

We now show the statistical rate for the least squares estimator which uses just the last sample of each trajectory: $(x_T^{(\ell)}, x_{T-1}^{(\ell)}, u_{T-1}^{(\ell)})$. This estimation procedure is made precise in Algorithm 1. Our analysis ideas are analogous to those used to prove statistical rates for standard linear regression, and they leverage tools in non-asymptotic analysis of random matrices.

The fact that this strategy results in independent data can be seen by defining the estimator matrix directly. The previous estimator (3.2.3) is amended to rely on $X_N = \begin{bmatrix} x_T^{(1)} & x_T^{(2)} & \dots & x_T^{(N)} \end{bmatrix}^\top$, Z_N, and W_N, which are similarly defined. The matrices in the estimator thus contain independent rows. To see this, define the matrices

$$G_T = \begin{bmatrix} A^{T-2}B & A^{T-3}B & \dots & B \end{bmatrix} \quad \text{and} \quad F_T = \begin{bmatrix} A^{T-2} & A^{T-3} & \dots & I_n \end{bmatrix}. \tag{3.2.5}$$

We can unroll the system dynamics and see that

$$x_{T-1} = G_T \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{T-2} \end{bmatrix} + F_T \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{T-2} \end{bmatrix}. \tag{3.2.6}$$

Using Gaussian excitation, $u_t \sim \mathcal{N}(0, \sigma_u^2 I_m)$ and assuming Gaussian process noise $w_t \sim \mathcal{N}(0, \sigma_w^2 I_n)$ gives

$$\begin{bmatrix} x_{T-1} \\ u_{T-1} \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_u^2 G_T G_T^\top + \sigma_w^2 F_T F_T^\top & 0 \\ 0 & \sigma_u^2 I_m \end{bmatrix}\right). \tag{3.2.7}$$

Since $F_T F_T^\top > 0$, as long as both σ_u, σ_w are positive, this is a non-degenerate distribution.

Therefore, bounding the estimation error can be achieved via proving a result on the error in random design linear regression with vector valued observations. First, we present a standard lemma which bounds the spectral norm of the product of two independent Gaussian matrices.

Lemma 3.2.1. *Fix a $\delta \in (0,1)$ and $N \ge 2\log(1/\delta)$. Let $f_k \in \mathbb{R}^m$, $g_k \in \mathbb{R}^n$ be independent random vectors $f_k \sim \mathcal{N}(0, \Sigma_f)$ and $g_k \sim \mathcal{N}(0, \Sigma_g)$ for $1 \le k \le N$. With probability at least $1 - \delta$,*

$$\left\| \sum_{k=1}^N f_k g_k^\top \right\|_2 \le 4 \|\Sigma_f\|_2^{1/2} \|\Sigma_g\|_2^{1/2} \sqrt{N(m+n)\log(9/\delta)}.$$

Lemma 3.2.1 shows that if X is $n_1 \times N$ with i.i.d. $\mathcal{N}(0,1)$ entries and Y is $N \times n_2$ with i.i.d. $\mathcal{N}(0,1)$ entries, and X and Y are independent, then with probability at least $1 - \delta$ we

have

$$\|XY\|_2 \le 4\sqrt{N(n_1+n_2)\log(9/\delta)}\,.$$

Next, we state a standard non-asymptotic bound on the minimum singular value of a standard Wishart matrix (see e.g. Corollary 5.35 of [Ver10]).

Lemma 3.2.2. *Let $X \in \mathbb{R}^{N\times n}$ have i.i.d. $\mathcal{N}(0,1)$ entries. With probability at least $1-\delta$,*

$$\sqrt{\lambda_{\min}(X^\top X)} \ge \sqrt{N} - \sqrt{n} - \sqrt{2\log(1/\delta)}\,.$$

We combine the previous lemmas into a statement on the error of random design regression.

Lemma 3.2.3. *Let $z_1, \ldots, z_N \in \mathbb{R}^n$ be i.i.d. from $\mathcal{N}(0,\Sigma)$ with Σ invertible. Let $Z^\top := \begin{bmatrix} z_1 & \ldots & z_N \end{bmatrix}$. Let $W \in \mathbb{R}^{N\times p}$ with each entry i.i.d. $\mathcal{N}(0,\sigma_w^2)$ and independent of Z. Let $E := (Z^\top Z)^\dagger Z^\top W$, and suppose that*

$$N \ge 8n + 16\log(2/\delta)\,. \tag{3.2.8}$$

For any fixed matrix Q, we have with probability at least $1-\delta$,

$$\|QE\|_2 \le 16\sigma_w\|Q\Sigma^{-1/2}\|_2\sqrt{\frac{(n+p)\log(18/\delta)}{N}}\,.$$

Proof. First, observe that Z is equal in distribution to $X\Sigma^{1/2}$, where $X \in \mathbb{R}^{N\times n}$ has i.i.d. $\mathcal{N}(0,1)$ entries. By Lemma 3.2.2, with probability at least $1-\delta/2$,

$$\sqrt{\lambda_{\min}(X^\top X)} \ge \sqrt{N} - \sqrt{n} - \sqrt{2\log(2/\delta)} \ge \sqrt{N}/2\,.$$

The last inequality uses (3.2.8) combined with the inequality $(a+b)^2 \le 2(a^2+b^2)$. Furthermore, by Lemma 3.2.1 and (3.2.8), with probability at least $1-\delta/2$,

$$\|X^\top W\|_2 \le 4\sigma_w\sqrt{N(n+p)\log(18/\delta)}\,.$$

Let $\mathcal{E}$ denote the event which is the intersection of the two previous events. By a union bound, $\mathbb{P}(\mathcal{E}) \ge 1-\delta$. We continue the rest of the proof assuming the event $\mathcal{E}$ holds. Since $X^\top X$ is invertible,

$$QE = Q(Z^\top Z)^\dagger Z^\top W = Q(\Sigma^{1/2}X^\top X\Sigma^{1/2})^\dagger\Sigma^{1/2}X^\top W = Q\Sigma^{-1/2}(X^\top X)^{-1}X^\top W\,.$$

Taking operator norms on both sides,

$$\|QE\|_2 \le \|Q\Sigma^{-1/2}\|_2\|(X^\top X)^{-1}\|_2\|X^\top W\|_2 = \|Q\Sigma^{-1/2}\|_2\frac{\|X^\top W\|_2}{\lambda_{\min}(X^\top X)}\,.$$

Combining the inequalities above,

$$\frac{\|X^\top W\|_2}{\lambda_{\min}(X^\top X)} \leq 16\sigma_w \sqrt{\frac{(n+p)\log(18/\delta)}{N}} .$$

The result now follows. □

This lemma, in combination with the error decomposition in 3.2.4, allows us to show a bound on the estimation errors.

Proposition 3.2.4. *Assume we collect data from the linear, time-invariant system initialized at $x_0 = 0$, using inputs $u_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_u^2 I_m)$ for $t = 0, \ldots, T-1$, with $T \geq 2$. Suppose that the process noise is $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2 I_n)$ and that*

$$N \geq 8(n+m) + 16\log(4/\delta) .$$

Then, with probability at least $1-\delta$, the least squares estimator using only the final sample of each trajectory satisfies both the inequality

$$\|\widehat{A} - A\|_2 \leq \frac{16\sigma_w}{\sqrt{\lambda_{\min}(\sigma_u^2 G_T G_T^\top + \sigma_w^2 F_T F_T^\top)}} \sqrt{\frac{(n+2m)\log(36/\delta)}{N}} , \tag{3.2.9}$$

and the inequality

$$\|\widehat{B} - B\|_2 \leq \frac{16\sigma_w}{\sigma_u} \sqrt{\frac{(n+2m)\log(36/\delta)}{N}} . \tag{3.2.10}$$

Proof. Consider the least squares estimation error (3.2.4) with modified single-sample-per-rollout matrices. Recall that rows of the design matrix Z_N are distributed as independent normals, as in (3.2.7). Then applying Lemma 3.2.3 with $Q_A = \begin{bmatrix} I_n & 0 \end{bmatrix}$ so that $Q_A E$ extracts only the estimate for A, we conclude that with probability at least $1 - \delta/2$, the expression (3.2.9) holds as long as $N \geq 8(n+m) + 16\log(4/\delta)$. Now applying Lemma 3.2.3 under the same condition on N with $Q_B = \begin{bmatrix} 0 & I_m \end{bmatrix}$, we have with probability at least $1-\delta/2$, that the expression in (3.2.10) holds. The result follows by application of the union bound. □

There are several interesting points to make about the guarantees offered by Proposition 3.2.4. First, there are $n(n+m)$ parameters to learn and our bound states that we need $O(n+m)$ measurements, each measurement providing n values. Hence, this appears to be an optimal dependence with respect to the parameters n and m. Second, this proposition illustrates that not all linear systems are equally easy to estimate. The matrices $G_T G_T^\top$ and $F_T F_T^\top$ are finite time *controllability Gramians* for the control and noise inputs, respectively. These are standard objects in control: each eigenvalue/vector pair of such

a Gramian characterizes how much input energy is required to move the system in that particular direction of the state-space. Therefore $\lambda_{\min}(\sigma_u^2 G_T G_T^\top + \sigma_w^2 F_T F_T^\top)$ quantifies the least controllable, and hence most difficult to excite and estimate, mode of the system. This property is captured nicely in our bound, which indicates that for systems for which all modes are easily excitable (i.e., all modes of the system amplify the applied inputs and disturbances), the identification task becomes easier.

Estimation in Closed-Loop

While the bound in Proposition 3.2.4 provides interpretable insights, the procedure in Algorithm 1 may not be practically possible. Injecting Gaussian noise into a system will likely cause violations of state and input constraints. Instead, is necessary to simultaneously excite and regulate the system. We propose to learn by additively injecting bounded noise to the control inputs computed by a safe controller. In what follows, we describe how a controller that guarantees that the closed-loop system remains within the specified constraint set can also ensure that enough noise is injected into the system to obtain a statistical guarantee on learning.

We bound the errors of the least squares estimator when the system is controlled by

$$\mathbf{u} = \mathbf{K}\mathbf{x} + \boldsymbol{\eta}\,, \tag{3.2.11}$$

where each $\boldsymbol{\eta} = (\eta_0, \eta_1, \dots)$ is stochastic, independent across time, and ℓ_∞-bounded with $\|\eta_t\|_\infty \le r_\eta$.

The bulk of the proof for the statistical rate comes from a general theorem regarding linear-response time series data from Simchowitz et al. [SMTJR18]. We assume that $\mathbf{w}$ and $\boldsymbol{\eta}$ are both zero-mean sequences with independent coordinates and finite fourth moments, assumptions which are quickly verified for common distributions such as uniform on a compact interval or over a discrete set of points. Before stating the main estimation result, we introduce the notation: $a \lesssim b$ (resp. $a \gtrsim b$) denotes that there exists an absolute constant $C > 0$ such that $a \le Cb$ (resp. $a \ge Cb$).

Theorem 3.2.5. *Fix a failure probability $\delta \in (0,1)$. Suppose the disturbance $(w_t)_{t\ge 0}$ and the excitation $(\eta_t)_{t\ge 0}$ are zero-mean sequences with independent coordinates and*

$$\mathbb{E}_{w_t}[w_t(i)^2] = \sigma_w^2,\ \ \mathbb{E}_{w_t}[w_k(i)^4] \lesssim \sigma_w^4,\ \ \mathbb{E}_{\eta_t}[\eta_t(i)^2] = \sigma_\eta^2,\ \ \mathbb{E}_{\eta_t}[\eta_t(i)^4] \lesssim \sigma_\eta^4\,.$$

Assume for simplicity that $\sigma_\eta \le \sigma_w$, and that the stabilizing controller $\mathbf{K}$ in (3.2.11) *achieves a SLS response $\boldsymbol{\Phi}_{\mathbf{x}} \in \frac{1}{z}\mathcal{RH}_\infty(C_x, \rho)$, $\boldsymbol{\Phi}_{\mathbf{u}} \in \frac{1}{z}\mathcal{RH}_\infty(C_u, \rho)$. Let $C_K^2 := nC_x^2 + dC_u^2$. Then as long as the trajectory length T satisfies the condition:*

$$T \gtrsim (n+m)\log\left(\frac{mC_u^2}{\delta} + \frac{\sigma_w^2}{\sigma_\eta^2}\frac{\rho^2 C_u^2 C_K^2}{\delta(1-\rho^2)}\left(1 + \|B\|_2^2 + \frac{\|x_0\|_2^2}{\sigma_w^2 T}\right)\right), \tag{3.2.12}$$

we have the following bound on the least-squares estimation errors that holds with probability at least $1 - \delta$,

$$\max\{\|\Delta_A\|_2, \|\Delta_B\|_2\} \lesssim \frac{\sigma_w C_u}{\sigma_\eta}\sqrt{\frac{n+m}{T}}\sqrt{\log\left(\frac{mC_u}{\delta} + \frac{\sigma_w}{\sigma_\eta}\frac{\rho C_u C_K}{\delta(1-\rho^2)}\left(1 + \|B\|_2 + \frac{\|x_0\|_2}{\sigma_w\sqrt{T}}\right)\right)}.$$

We make several observations about the guarantee offered by Theorem 3.2.5. First, we emphasize that this result relies on $\mathbf{K}$ being a stabilizing controller, unlike in the independent data case. This closed-loop stability assumption is pervasive in the literature on least squares estimation from linear time series. Like Proposition 3.2.4, there appears to be an optimal dependence with respect to the parameters n and m. Though the dependence on problem parameters is somewhat less transparent than in the independent data case, there are two main insights to be had. First, the bound on estimation errors decreases with the size of the excitation, and increases with the size of the process noise, illustrating an inverse dependence on the signal-to-noise ratio. Second, the bound increases proportionally to C_u, a constant which bounds the gain from disturbance to control inputs under the stabilizing controller $\mathbf{K}$. This indicates that when the closed-loop system has larger transients, the dynamics matrices are easier to estimate.

We remark on the interpretation of statistical learning bounds. A priori guarantees, like the one presented here, depend on quantities related to the underlying true system. Statistical bounds in terms of data-dependent quantities can also be worked out; however, modern methods like bootstrapping generally provide tighter statistical guarantees [Efr92]. Though a priori guarantees like this one are not directly useful when the system is unknown, they yield insights about qualities of systems that make them easier or harder to estimate.

We now turn to the proof of Theorem 3.2.5, which hinges on a result on the estimation of linear response time-series by Simchowitz et al. [SMTJR18]. We thus present that result in the context of our problem. Recall that $\Theta^\top = [A, B]$, $\widehat{\Theta}^\top = [\widehat{A}, \widehat{B}]$, and $z_t = [x_t^\top, y_t^\top]^\top$. We denote the filtration $\mathcal{F}_t = \sigma(x_0, \eta_0, w_0 \ldots, \eta_{t-1}, w_{t-1}, \eta_t)$. It is clear that the process $(z_t)_{t\geq 0}$ is $(\mathcal{F}_t)_{t\geq 0}$-adapted. The process $(z_t)_{t\geq 0}$ is said to satisfy the (k, ν, β)-*block martingale small-ball* (BMSB) condition if for any $j \geq 0$ and $v \in \mathbb{R}^{n+m}$, one has that

$$\frac{1}{k}\sum_{i=1}^{k} \mathbb{P}\left(|\langle v, z_{j+i}\rangle| \geq \nu\right) \geq \beta \text{ almost surely.}$$

This condition is used for characterizing the size of the minimum eigenvalue of the covariance matrix $\sum_{t=0}^{T-1} z_t z_t^\top$. A larger ν guarantees a larger lower bound of the minimum eigenvalue. In the context of our problem the result by Simchowitz et al. [SMTJR18] translates as follows.

Theorem 3.2.6 (Simchowitz et al. [SMTJR18])**.** *Fix $\delta \in (0,1)$. For every T, k, ν, and β such that $\{z_t\}_{t\geq 0}$ satisfies the (k,ν,β)-BMSB and*

$$\left\lfloor \frac{T}{k} \right\rfloor \gtrsim \frac{n+m}{\beta^2} \log\left(1 + \frac{\sum_{t=0}^{T-1} \mathbf{Tr}(\mathbb{E} z_t z_t^\top)}{k\lfloor T/k \rfloor \beta^2 \nu^2 \delta}\right),$$

the estimate $\widehat{\Theta}$ defined in (3.2.1) *satisfies the following statistical rate*

$$\mathbb{P}\left(\|\widehat{\Theta} - \Theta\|_2 > \frac{O(1)\sigma_w}{\beta\nu} \sqrt{\frac{n+m}{k\lfloor T/k \rfloor} \log\left(1 + \frac{\sum_{t=0}^{T-1} \mathbf{Tr}(\mathbb{E} z_t z_t^\top)}{k\lfloor T/k \rfloor \beta^2 \nu^2 \delta}\right)}\right) \leq \delta.$$

Therefore, in order to apply this result, there are two necessary ingredients. The first is k, ν, and β such that $(z_t)_{t\geq 0}$ satisfies the (k,ν,β)-BMSB condition. The second is an upper bound on the trace of the covariance of z_t. Before addressing these two issues directly, we begin with a simple small-ball result for random variables with finite fourth moments.

Lemma 3.2.7. *Let $X \in \mathbb{R}$ be a zero-mean random variable with finite fourth moment, which satisfies the conditions*

$$\mathbb{E}[X^4] \leq C(\mathbb{E}[X^2])^2 .$$

Let $a \in \mathbb{R}$ be a fixed scalar and $\theta \in (0,1)$. We have that

$$\mathbb{P}\{|a + X| \geq \sqrt{\theta \mathbb{E}[X^2]}\} \geq (1-\theta)^2 / \max\{4, 3C\} .$$

We defer the proof, and proofs of the following lemmas, to Section 3.7. The next lemma addresses the BMSB condition.

Lemma 3.2.8. *Let x_0 be any initial state in $\mathbb{R}^n$ and let $(u_t)_{t\geq 0}$ be the sequence of inputs generated according to* (3.2.11)*, and assume $\sigma_\eta \leq \sigma_w$. Then, the process $z_t = [x_t^\top, u_t^\top]^\top$ satisfies the*

$$\left(1, \frac{\sigma_\eta}{2C_u}, \frac{1}{C \cdot C_w}\right) \text{ BMSB condition,}$$

where C is an absolute constant, and $C_w = \max_{1\leq i\leq d} \mathbb{E}[w_i^4]/(\mathbb{E}[w_i^2])^2$. In particular, we can take $C = 192$.

The next lemma provides an upper bound on the trace of the covariance of z_t.

Lemma 3.2.9. *Let $\sigma_\eta \leq \sigma_w$. Then, the process $(z_t)_{t\geq 0}$ satisfies*

$$\sum_{t=0}^{T-1} \mathbf{Tr}\left(\mathbb{E} z_t z_t^\top\right) \leq \sigma_\eta^2 mT + \sigma_w^2 \frac{\rho^2 C_K^2 T}{(1-\rho^2)} \left(1 + \|B\|^2 + \frac{\|x_0\|_2^2}{\sigma_w^2 T}\right).$$

The estimation result in Theorem 3.2.5 follows from Theorem 3.2.6, Lemma 3.2.8, Lemma 3.2.9, and simple algebra.

3.3 Robust Synthesis

With estimates of the system $(\widehat{A}, \widehat{B})$ and operator norm error bounds $(\varepsilon_A, \varepsilon_B)$ in hand, we now turn to control design. We will make use of tools from System Level Synthesis introduced in Chapter 2. Using the SLS framework, as opposed to traditional techniques from robust control, allows us to (a) compute robust and constraint-satisfying controllers using semidefinite programming, and (b) provide sub-optimality guarantees in terms of the size of the uncertainties on our system estimates.

In this section, we first show how to represent the LQR problem in the SLS framework. Then, we return to the problem setting of model estimates with error bounds to develop a robust synthesis procedure.

System Level Approach to LQR

Before proceeding, we must formulate the LQR problem in terms of the system responses. As developed in the previous chapter, the LQR cost can be equivalently written as a system $\mathcal{H}_2$ norm weighted by the cost matrices Q and R.

$$J(A, B, \mathbf{K})^2 = \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_2}^2 \tag{3.3.1}$$

where $\mathbf{\Phi_x}$ and $\mathbf{\Phi_u}$ satisfy the constraint $\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I$.

It remains to reformulate the inequality constraints. First, making use of the identity $x_t = \Phi_x(t+1)x_0 + \sum_{k=1}^{t} \Phi_x(k) w_{t-k}$ we can write the robust constraint as

$$F_x \Phi_x(t+1)x_0 + \max_{(w_t)_{t\geq 0}} F_x \sum_{k=1}^{t} \Phi_x(k) w_{t-k} \leq b_x \ .$$

Then considering elements in the second term with each j indexing the rows of F_x,

$$\begin{aligned} \max_{(w_t)_{t\geq 0}} F_{x,j}^\top \sum_{k=1}^{t} \Phi_x(k) w_{t-k} &= \sum_{k=1}^{t} \max_{\|w\|_\infty \leq r_w} F_{x,j}^\top \Phi_x(k) w \\ &= \sum_{k=1}^{t} r_w \|F_{x,j}^\top \Phi_x(k)\|_1 = r_w \|F_{x,j}^\top \Phi_x[t:1]\|_1 \ . \end{aligned}$$

We can perform similar manipulations with the input constraints. This inspires the definition of the vector valued constraint functions

$$\begin{aligned} G_x(\mathbf{\Phi_x}; t)_j &:= F_{x,j}^\top \Phi_x(t+1)x_0 + r_w \|F_{x,j}^\top \Phi_x[t:1]\|_1 \ , \\ G_u(\mathbf{\Phi_u}; t)_j &:= F_{u,j}^\top \Phi_u(t+1)x_0 + r_w \|F_{u,j}^\top \Phi_u[t:1]\|_1 \ , \end{aligned}$$

with j indexing the rows of F_x and F_u and entries of G_x and G_u.

Thus, the worst-case polytopic constraints on state and input become closed-form polytopic constraints on the system response: $G_x(\mathbf{\Phi_x}; k) \leq b_x$ and $G_u(\mathbf{\Phi_u}; k) \leq b_u$. The appearance of the row-wise ℓ_1 norm over the multiplication of the the system response elements with the constraint matrix can be understood as an analog to the $\ell_\infty \to \ell_\infty$ operator norm, mediated by the shape of the polytope. In the next section when we introduce robustness to unknown dynamics, the $\mathcal{L}_1$ system norm will come into play for similar reasons. It is also possible to consider worst-case constraints for disturbances from arbitrary bounded convex sets. Chen and Anderson [CA19] show that in the SLS framework, the representation of such worst-case constraints remains convex.

Combining the re-parametrized cost and the constraints, we have the following convex optimization problem over system response variables:

$$\begin{aligned} \min_{\mathbf{\Phi_x}, \mathbf{\Phi_u} \in \frac{1}{z}\mathcal{RH}_\infty} \quad & \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_2} \\ \text{s.t.} \quad & \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I, \\ & G_x(\mathbf{\Phi_x}; t) \leq b_x, \;\; G_u(\mathbf{\Phi_u}; t) \leq b_u \;\; \forall\, t \geq 1\,. \end{aligned} \tag{3.3.2}$$

By the state feedback parameterization result Theorem 2.4.1, the SLS parametrization encompasses all internally stabilizing state feedback controllers acting on the true system (A, B) [WMD19]. Therefore, the equivalence between (3.3.2) and (3.1.4) follows by the consistency between the parameterizations.

Robust LQR Synthesis

We return to the problem setting where estimates $(\widehat{A}, \widehat{B})$ of a true system (A, B) satisfy

$$\|\Delta_A\|_2 \leq \varepsilon_A, \;\; \|\Delta_B\|_2 \leq \varepsilon_B$$

where $\Delta_A := \widehat{A} - A$ and $\Delta_B := \widehat{B} - B$. In light of this, it is natural to pose the following robust variant of the LQR optimal control problem (3.1.4), which computes a robustly stabilizing and constraint-satisfying controller. The controller seeks to minimize performance and guarantee safety of the system in the worst case, given the (high-probability) norm bounds on the perturbations Δ_A and Δ_B:

$$\begin{aligned} \text{minimize} \quad \sup_{\substack{\|\Delta_A\|_2 \leq \varepsilon_A \\ \|\Delta_B\|_2 \leq \varepsilon_B}} \quad & \lim_{T\to\infty} \tfrac{1}{T} \textstyle\sum_{t=1}^{T} \mathbb{E}\left[x_t^\top Q x_t + u_{t-1}^\top R u_{t-1} \right] \\ \text{subject to} \quad & x_{t+1} = (\hat{A} + \Delta_A) x_t + (\hat{B} + \Delta_B) u_t + w_t, \\ & F_x x_t \leq b_x, \;\; F_u u_t \leq b_u \;\; \forall t \;, \forall \{w_t : \|w_t\|_\infty \leq r_w\}\;, \\ & \forall \{(\Delta_A, \Delta_B) : \|\Delta_A\|_2 \leq \varepsilon_A, \; \|\Delta_B\|_2 \leq \varepsilon_B\}. \end{aligned} \tag{3.3.3}$$

Towards developing a tractable formulation, we make use of the ability for SLS to trace the effects of unknown dynamics on closed-loop behavior. The following lemma describes the behavior of the true system in terms of the estimated system. It also presents a condition under which which any controller $\mathbf{K}$ that stabilizes $(\widehat{A}, \widehat{B})$ also stabilizes the true system (A, B).

Lemma 3.3.1. *Let the controller* $\mathbf{K}$ *stabilize* $(\widehat{A}, \widehat{B})$ *and* $(\mathbf{\Phi_x}, \mathbf{\Phi_u})$ *be its corresponding system response on system* $(\widehat{A}, \widehat{B})$*. Let*

$$\hat{\mathbf{\Delta}} := \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} . \tag{3.3.4}$$

Then if $\mathbf{K}$ *stabilizes* (A, B)*, it results in the following system trajectory*

$$\mathbf{x} = \mathbf{\Phi_x}(I + \hat{\mathbf{\Delta}})^{-1}\mathbf{w}, \quad \mathbf{u} = \mathbf{\Phi_u}(I + \hat{\mathbf{\Delta}})^{-1}\mathbf{w} . \tag{3.3.5}$$

Therefore, it achieves the following LQR cost

$$J(A, B, \mathbf{K}) = \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \left(I + \hat{\mathbf{\Delta}} \right)^{-1} \right\|_{\mathcal{H}_2} . \tag{3.3.6}$$

A sufficient condition for $\mathbf{K}$ *to stabilize* (A, B) *is that* $\|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty} < 1$*.*

Proof. Follows immediately from Theorems 2.4.1, 2.4.3 and Corollary 2.4.4 by noting that for system responses $(\mathbf{\Phi_x}, \mathbf{\Phi_u})$ satisfying

$$\begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I,$$

it holds that

$$\begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I + \hat{\mathbf{\Delta}}$$

for $\hat{\mathbf{\Delta}}$ as defined in equation (3.3.4). □

We therefore recast the robust LQR problem (3.3.3) in the following equivalent form

$$\begin{aligned} \text{minimize} \quad & \sup_{\substack{\|\Delta_A\|_2 \le \varepsilon_A \\ \|\Delta_B\|_2 \le \varepsilon_B}} \left\| \begin{bmatrix} Q^{\frac{1}{2}} & 0 \\ 0 & R^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \left(I + \hat{\mathbf{\Delta}} \right)^{-1} \right\|_{\mathcal{H}_2} \\ \text{subject to} \quad & \begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I, \ \ \mathbf{\Phi_x}, \mathbf{\Phi_u} \in \tfrac{1}{z}\mathcal{RH}_\infty, \\ & G_x(\mathbf{\Phi_x}(I + \hat{\mathbf{\Delta}})^{-1}; t) \le b_x, \ \ G_u(\mathbf{\Phi_u}(I + \hat{\mathbf{\Delta}})^{-1}; t) \le b_u \ \ \forall\, t \ge 1, \\ & \hat{\mathbf{\Delta}} = \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} . \end{aligned} \tag{3.3.7}$$

This robust control problem is one subject to real parametric uncertainty, a class of problems known to be computationally intractable even in the absence of state and input constraints [BYDM94]. Although effective computational heuristics (e.g., DK iteration [ZDG96]) exist for unconstrained LQR, the performance of the resulting controller on the true system is difficult to characterize analytically in terms of the size of the perturbations.

To circumvent this issue and to deal with safety constraints, we take a slightly conservative approach. We find an upper-bound to the cost $J(A, B, \mathbf{K})$ and develop sufficient conditions for constraint satisfaction that depend only on the norm of the uncertainties Δ_A and Δ_B. First, note that if $\|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty} < 1$, we can write

$$J(A, B, \mathbf{K}) \leq \|(I + \hat{\mathbf{\Delta}})^{-1}\|_{\mathcal{H}_\infty} J(\widehat{A}, \widehat{B}, \mathbf{K}) \leq \frac{1}{1 - \|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty}} J(\widehat{A}, \widehat{B}, \mathbf{K}). \tag{3.3.8}$$

Because $J(\widehat{A}, \widehat{B}, \mathbf{K})$ captures the performance of the controller $\mathbf{K}$ on the nominal system $(\widehat{A}, \widehat{B})$, it is not subject to any uncertainty.

Next, we derive a sufficient condition for the state and input constraints. Notice that as long as the inverse exists,

$$\mathbf{\Phi_x}(I + \hat{\mathbf{\Delta}})^{-1} = \mathbf{\Phi_x}\left(I - \hat{\mathbf{\Delta}}(I + \hat{\mathbf{\Delta}})^{-1}\right) = \mathbf{\Phi_x} - \mathbf{\Phi_x}\hat{\mathbf{\Delta}}(I + \hat{\mathbf{\Delta}})^{-1}\,.$$

Therefore, each state constraint (indexed by j) is satisfied at time t as long as

$$b_{x,j} \geq \max_{\mathbf{w}} F_{x,j}^\top(\mathbf{\Phi_x}\mathbf{w})[t] - F_{x,j}^\top(\mathbf{\Phi_x}\hat{\mathbf{\Delta}}(I + \hat{\mathbf{\Delta}})^{-1}\mathbf{w})[t]\,.$$

The first term reduces to $G_x(\mathbf{\Phi_x}; t)$ as above. We resort to a sufficient condition to bound the second term,

$$\begin{aligned}
|F_{x,j}^\top(\mathbf{\Phi_x}\hat{\mathbf{\Delta}}(I + \hat{\mathbf{\Delta}})^{-1}\mathbf{w})[t]| &\leq \|F_{x,j}^\top\Phi_x[t+1:1]\|_1 \|\hat{\mathbf{\Delta}}(I + \hat{\mathbf{\Delta}})^{-1}\mathbf{w}\|_\infty \\
&\leq \|F_{x,j}^\top\Phi_x[t+1:1]\|_1 \|\hat{\mathbf{\Delta}}\|_{\mathcal{L}_1} \|(I + \hat{\mathbf{\Delta}})^{-1}\|_{\mathcal{L}_1} \|\mathbf{w}\|_\infty \\
&\leq \|F_{x,j}^\top\Phi_x[t+1:1]\|_1 \frac{\|\hat{\mathbf{\Delta}}\|_{\mathcal{L}_1}}{1 - \|\hat{\mathbf{\Delta}}\|_{\mathcal{L}_1}} \max(r_w, \|x_0\|_\infty)
\end{aligned}$$

The first inequality is generalized Cauchy-Schwarz and the second holds by the submultiplicative property of the operator norm. A similar computation can be performed for input constraints. Recently work by Chen et al. [CWMPM20] shows that the conservatism of these robust constraint sets with respect to the initial condition x_0 can be ameliorated at the expense of increase complexity in the resulting synthesis procedure. Their approach results in an optimization problem requiring a search over an additional scalar parameter.

It therefore remains to compute a tractable bound for $\|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty}$ and $\|\hat{\mathbf{\Delta}}\|_{\mathcal{L}_1}$. Define $\varepsilon_{A,\infty}$ and $\varepsilon_{B,\infty}$ to be ℓ_∞ bounds on the estimates,

$$\|\Delta_A\|_\infty \le \varepsilon_{A,\infty}, \;\; \|\Delta_B\|_2 \le \varepsilon_{B,\infty} .$$

Then we have the following fact.

Proposition 3.3.2. *For $\hat{\mathbf{\Delta}}$ as defined in* (3.3.4)

$$\|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty} \le \sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi_x} \\ \varepsilon_B \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_\infty}, \quad \|\hat{\mathbf{\Delta}}\|_{\mathcal{L}_1} \le 2 \left\| \begin{bmatrix} \varepsilon_{A,\infty} \mathbf{\Phi_x} \\ \varepsilon_{B,\infty} \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{L}_1} . \tag{3.3.9}$$

Proof. Note that for any block matrix of the form $\begin{bmatrix} M_1 & M_2 \end{bmatrix}$, we have

$$\left\| \begin{bmatrix} M_1 & M_2 \end{bmatrix} \right\|_2 \le \left(\|M_1\|_2^2 + \|M_2\|_2^2 \right)^{1/2} . \tag{3.3.10}$$

To verify this assertion, note that

$$\left\| \begin{bmatrix} M_1 & M_2 \end{bmatrix} \right\|_2^2 = \lambda_{\max}(M_1 M_1^* + M_2 M_2^*) \le \lambda_{\max}(M_1 M_1^*) + \lambda_{\max}(M_2 M_2^*) = \|M_1\|_2^2 + \|M_2\|_2^2 .$$

With (3.3.10) in hand, we have

$$\begin{aligned} \left\| \begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_\infty} &= \left\| \begin{bmatrix} \frac{1}{\varepsilon_A}\Delta_A & \frac{1}{\varepsilon_B}\Delta_B \end{bmatrix} \begin{bmatrix} \varepsilon_A \mathbf{\Phi_x} \\ \varepsilon_B \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_\infty} \\ &\le \left\| \begin{bmatrix} \frac{1}{\varepsilon_A}\Delta_A & \frac{1}{\varepsilon_B}\Delta_B \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi_x} \\ \varepsilon_B \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_\infty} \le \sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi_x} \\ \varepsilon_B \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_\infty} . \end{aligned}$$

Noting that the matrix $\ell_\infty \to \ell_\infty$ operator norm is the maximum ℓ_1 norm of a row,

$$\left\| \begin{bmatrix} M_1 & M_2 \end{bmatrix} \right\|_\infty \le \|M_1\|_\infty + \|M_2\|_\infty . \tag{3.3.11}$$

The $\mathcal{L}_1$ bound follows similarly. □

Applying Proposition 3.3.2 in conjunction with the bound (3.3.8), we arrive at the following upper bound to the cost function of the robust LQR problem (3.3.3), which is independent of the perturbations (Δ_A, Δ_B):

$$\sup_{\substack{\|\Delta_A\|_2 \le \varepsilon_A \\ \|\Delta_B\|_2 \le \varepsilon_B}} J(A, B, \mathbf{K}) \le \frac{J(\widehat{A}, \widehat{B}, \mathbf{K})}{1 - \sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi_x} \\ \varepsilon_B \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_\infty}} . \tag{3.3.12}$$

We remark that the bound (3.3.12) is of interest independent of the synthesis procedure for $\mathbf{K}$. For example, it can be applied to static controllers, including the optimal unconstrained LQR controller $\widehat{\mathbf{K}}$ computed using the nominal system $(\widehat{A}, \widehat{B})$.

Motivated by this bound and the sufficient conditions for constraint satisfaction, we introduce the auxiliary variables $0 \leq \gamma, \tau < 1$ and define the following robust optimization problem:

$$\begin{aligned} \widehat{J}(\gamma, \tau) := \min_{\mathbf{\Phi_x}, \mathbf{\Phi_u} \in \frac{1}{z}\mathcal{RH}_\infty} \ & \frac{1}{1-\gamma} J(\widehat{A}, \widehat{B}, \mathbf{K}) \\ \text{s.t. } & \begin{bmatrix} zI - \widehat{A} & -\widehat{B} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_x} \\ \mathbf{\Phi_u} \end{bmatrix} = I, \\ & \sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi_x} \\ \varepsilon_B \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{H}_\infty} \leq \gamma, \ 2 \left\| \begin{bmatrix} \varepsilon_{A,\infty} \mathbf{\Phi_x} \\ \varepsilon_{B,\infty} \mathbf{\Phi_u} \end{bmatrix} \right\|_{\mathcal{L}_1} \leq \tau, \\ & G_x^\tau(\mathbf{\Phi_x}; t) \leq b_x, \ \ G_u^\tau(\mathbf{\Phi_u}; t) \leq b_u \ \ \forall \, t \geq 1 \, . \end{aligned} \tag{3.3.13}$$

The vector valued robust constraint functions are defined at each index j as

$$G_x^\tau(\mathbf{\Phi_x}; t)_j := G_x(\mathbf{\Phi_x}; t)_j + \frac{\tau}{1-\tau} \max(r_w, \|x_0\|_\infty) \|F_{x,j}^\top \Phi_x[t+1:1]\|_1 \, ,$$

and similarly for $G_u^\tau(\mathbf{\Phi_u}; t)_j$.

Theorem 3.3.3. *For a system with true dynamics (A, B) and estimates $(\widehat{A}, \widehat{B})$ satisfying $\|A - \widehat{A}\|_2 \leq \varepsilon_A$ and $\|B - \widehat{B}\|_2 \leq \varepsilon_B$, then any controller designed from a feasible solution to the robust control problem* (3.3.13) *for any $0 \leq \gamma, \tau < 1$ will stabilize the true system. Furthermore, the state and input constraints will be satisfied.*

In this problem, γ determines the increase in the $\mathcal{H}_2$ cost due to the dynamics uncertainty, while τ determines the increase in the state and input values with respect to the constraints. Both values can be viewed as bounding an enlarge noise process $\tilde{\mathbf{w}} = (I + \hat{\mathbf{\Delta}})^{-1}\mathbf{w}$ driving the system. Viewing the effects of the errors in the dynamics model as an enlarged noise signal allows us to make connections to the MPC literature, where the additive disturbance approximation is common. This literature often considers the enlarged noise process $\tilde{w}_t = w_t + \Delta_A x_t + \Delta_B u_t$, or equivalently $\tilde{\mathbf{w}} = \mathbf{w} + \Delta_A \mathbf{x} + \Delta_B \mathbf{u}$. Then the norm $\|\tilde{\mathbf{w}}\|_\infty$ can be bounded using the state and input constraint sets. This strategy is overly conservative for large constraint sets because this bound degrades as the constraint sets increase in size. On the other hand, our approach does not depend on the constraint set. However, it is affected by the size of the initial condition. Further comparison between the two approaches is presented in the Appendix of [DTMR19].

The robust synthesis problem (3.3.13) is convex, albeit infinite dimensional. As discussed in Section 2.5, a simple finite impulse response truncation yields a finite dimensional problem with similar guarantees. Even if the solution is approximated, as long as it

is feasible, the resulting controller will be stabilizing and guarantee constraint satisfaction. As we show in the next section, for sufficiently small estimation error bounds ε_A and ε_B, we can further bound the sub-optimality of the performance achieved by our robustly stabilizing controller relative to that achieved by the optimal LQR controller.

3.4 Sub-Optimality Guarantees

We now return to analyzing the Coarse-ID control framework. How much is control performance degraded by uncertainties about the dynamics? In this section, we derive a sub-optimality bound which answers this question for the LQR problem. For this purpose, we consider the addition of an outer minimization over the auxiliary variables:

$$\min_{\substack{0\le\gamma<1\\0\le\tau<1}} \widehat{J}(\gamma,\tau)\,. \tag{3.4.1}$$

The existence of state and input constraints results in increased complexity, both in the sub-optimality argument and in the optimization problem above. We therefore begin by outlining the argument in the unconstrained case. After, we present the full result and remark on the adjustments necessary when state and input constraints are present.

Unconstrained LQR

When the problem is unconstrained, i.e. when $\mathcal{X}=\mathbb{R}^n$ and $\mathcal{U}=\mathbb{R}^m$, the variable τ does not play a role. Therefore (3.4.1) can be efficiently optimized considering just one auxiliary variable:

$$\begin{aligned}\min_{\gamma\in[0,1)} \min_{\mathbf{\Phi_x},\mathbf{\Phi_u}\in\frac{1}{z}\mathcal{RH}_\infty} \frac{1}{1-\gamma}J(\widehat{A},\widehat{B},\mathbf{K}) \\ \text{s.t. } \begin{bmatrix} zI-\widehat{A} & -\widehat{B}\end{bmatrix}\begin{bmatrix}\mathbf{\Phi_x}\\ \mathbf{\Phi_u}\end{bmatrix}=I,\ \sqrt{2}\left\|\begin{bmatrix}\varepsilon_A\mathbf{\Phi_x}\\ \varepsilon_B\mathbf{\Phi_u}\end{bmatrix}\right\|_{\mathcal{H}_\infty}\le\gamma\end{aligned} \tag{3.4.2}$$

The objective is jointly quasi-convex in $(\gamma,\mathbf{\Phi_x},\mathbf{\Phi_u})$. Hence, as a function of γ alone the objective is quasi-convex, and furthermore is smooth in the feasible domain. Therefore, the outer optimization with respect to γ can effectively be solved with methods like golden section search. The inner optimization is a convex problem, though an infinite dimensional one.

For sufficiently small estimation error bounds ε_A and ε_B, we can bound the sub-optimality of the performance achieved by our robustly stabilizing controller relative to that achieved by the optimal LQR controller. Denote the solution to the true optimal control problem as $(\mathbf{\Phi_x^\star},\mathbf{\Phi_u^\star})$, then define $\mathbf{K}_\star=\mathbf{\Phi_u^\star}\mathbf{\Phi_x^\star}^{-1}$ and $J_\star=J(A_\star,B_\star,\mathbf{K}_\star)$. In the unconstrained setting, the optimal controller is static, and can thus also be written as $K_\star$.

We use the notation $\mathfrak{R}_M = (zI - M)^{-1}$ for the transfer function induced by the closed-loop matrix M.

Proposition 3.4.1. *Let $J_\star$ denote the minimal LQR cost achievable by any controller for the dynamical system with transition matrices (A, B), and let $K_\star$ denote the optimal contoller. Let $(\widehat{A}, \widehat{B})$ be estimates of the transition matrices such that $\|\Delta_A\|_2 \le \varepsilon_A$, $\|\Delta_B\|_2 \le \varepsilon_B$. Then, if $\mathbf{K}$ is synthesized via* (3.4.2)*, the relative error in the LQR cost is*

$$\frac{J(A, B, \mathbf{K}) - J_\star}{J_\star} \le 5(\varepsilon_A + \varepsilon_B \|K_\star\|_2)\|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}\,, \tag{3.4.3}$$

as long as $(\varepsilon_A + \varepsilon_B\|K_\star\|_2)\|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty} \le 1/5$.

This result offers a guarantee on the performance of the robust controller regardless of the estimation procedure used to estimate the transition matrices. Together with our result on system identification, Proposition 3.4.1 yields a sample complexity upper bound on the performance of the robust SLS controller $\mathbf{K}$ when (A, B) are not known; we later make this guarantee precise. The rest of the section is dedicated to proving Proposition 3.4.1.

Recall that $K_\star$ is the optimal LQR static state feedback matrix for the true dynamics (A, B), and let $\mathbf{\Delta} := -\left[\Delta_A + \Delta_B K_\star\right]\mathfrak{R}_{A+BK_\star}$. We begin with a technical result.

Lemma 3.4.2. *Define $\zeta := (\varepsilon_A + \varepsilon_B\|K_\star\|_2)\|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}$, and suppose that $\zeta < (1+\sqrt{2})^{-1}$. Then $(\gamma_0, \tilde{\mathbf{\Phi}}_x, \tilde{\mathbf{\Phi}}_u)$ is a feasible solution of* (3.4.2)*, where*

$$\gamma_0 = \frac{\sqrt{2}\zeta}{1-\zeta}, \quad \tilde{\mathbf{\Phi}}_x = \mathfrak{R}_{A+BK_\star}(I + \mathbf{\Delta})^{-1}, \quad \tilde{\mathbf{\Phi}}_u = K_\star\mathfrak{R}_{A+BK_\star}(I + \mathbf{\Delta})^{-1}. \tag{3.4.4}$$

Proof. Note that $\|\mathbf{\Delta}\|_{\mathcal{H}_\infty} \le (\varepsilon_A + \varepsilon_B\|K_\star\|_2)\|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty} = \zeta < 1$. Then by construction $\tilde{\mathbf{\Phi}}_x, \tilde{\mathbf{\Phi}}_u \in \frac{1}{z}\mathcal{RH}_\infty$. Therefore, we are left to check three conditions:

$$\gamma_0 < 1, \quad \begin{bmatrix} zI - \widehat{A} & -\widehat{B}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{\Phi}}_x \\ \tilde{\mathbf{\Phi}}_u\end{bmatrix} = I\,, \text{ and } \quad \left\|\begin{bmatrix}\varepsilon_A\tilde{\mathbf{\Phi}}_x \\ \varepsilon_B\tilde{\mathbf{\Phi}}_u\end{bmatrix}\right\|_{\mathcal{H}_\infty} \le \frac{\sqrt{2}\zeta}{1-\zeta}. \tag{3.4.5}$$

The first two conditions follow by simple algebraic computations. For the final condition,

$$\begin{aligned}
\left\|\begin{bmatrix}\varepsilon_A\tilde{\mathbf{\Phi}}_x \\ \varepsilon_B\tilde{\mathbf{\Phi}}_u\end{bmatrix}\right\|_{\mathcal{H}_\infty} &= \sqrt{2}\left\|\begin{bmatrix}\varepsilon_A\mathfrak{R}_{A+BK_\star} \\ \varepsilon_B K_\star\mathfrak{R}_{A+BK_\star}\end{bmatrix}(I + \mathbf{\Delta})^{-1}\right\|_{\mathcal{H}_\infty} \\
&\le \sqrt{2}\|(I + \mathbf{\Delta})^{-1}\|_{\mathcal{H}_\infty}\left\|\begin{bmatrix}\varepsilon_A\mathfrak{R}_{A+BK_\star} \\ \varepsilon_B K_\star\mathfrak{R}_{A+BK_\star}\end{bmatrix}\right\|_{\mathcal{H}_\infty} \\
&\le \frac{\sqrt{2}}{1 - \|\mathbf{\Delta}\|_{\mathcal{H}_\infty}}\left\|\begin{bmatrix}\varepsilon_A I \\ \varepsilon_B K_\star\end{bmatrix}\mathfrak{R}_{A+BK_\star}\right\|_{\mathcal{H}_\infty} \\
&\le \frac{\sqrt{2}(\varepsilon_A + \varepsilon_B\|K_\star\|_2)\|\mathfrak{R}_{A+BK_\star}\|_{\mathcal{H}_\infty}}{1 - \|\mathbf{\Delta}\|_{\mathcal{H}_\infty}} \le \frac{\sqrt{2}\zeta}{1-\zeta}\,.
\end{aligned}$$

□

With the feasible solution in hand, we turn to the proof of the sub-optimality result.

Proof of Proposition 3.4.1. Let $(\gamma_\star, \mathbf{\Phi}_x^\star, \mathbf{\Phi}_u^\star)$ be an optimal solution to problem (3.4.2) and let $\mathbf{K} = \mathbf{\Phi}_\mathbf{u}^\star(\mathbf{\Phi}_x^\star)^{-1}$. We can then write

$$J(A, B, \mathbf{K}) \leq \frac{1}{1 - \|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty}} J(\widehat{A}, \widehat{B}, \mathbf{K}) \leq \frac{1}{1 - \gamma_\star} J(\widehat{A}, \widehat{B}, \mathbf{K}),$$

where the first inequality follows from the bound (3.3.8), and the second follows from the fact that $\|\hat{\mathbf{\Delta}}\|_{\mathcal{H}_\infty} \leq \gamma_\star$ due to Proposition 3.3.2 and the constraint in optimization problem (3.4.2).

From Lemma 3.4.2 we know that $(\gamma_0, \tilde{\mathbf{\Phi}}_x, \tilde{\mathbf{\Phi}}_u)$ defined in equation (3.4.4) is also a feasible solution. Therefore, because $K_\star = \tilde{\mathbf{\Phi}}_u \tilde{\mathbf{\Phi}}_x^{-1}$, we have by optimality,

$$\frac{1}{1 - \gamma_\star} J(\widehat{A}, \widehat{B}, \mathbf{K}) \leq \frac{1}{1 - \gamma_0} J(\widehat{A}, \widehat{B}, K_\star) \leq \frac{J(A, B, K_\star)}{(1 - \gamma_0)(1 - \|\mathbf{\Delta}\|_{\mathcal{H}_\infty})} = \frac{J_\star}{(1 - \gamma_0)(1 - \|\mathbf{\Delta}\|_{\mathcal{H}_\infty})},$$

where the second inequality follows by the argument used to derive (3.3.8) with the true and estimated transition matrices switched. Recall that $\|\mathbf{\Delta}\|_{\mathcal{H}_\infty} \leq \zeta$ and that $\gamma_0 = \sqrt{2}\zeta/(1 + \zeta)$. Therefore

$$\frac{J(A, B, \mathbf{K}) - J_\star}{J_\star} \leq \frac{1}{1 - (1 + \sqrt{2})\zeta} - 1 = \frac{(1 + \sqrt{2})\zeta}{1 - (1 + \sqrt{2})\zeta} \leq 5\zeta,$$

where the last inequality follows because $\zeta < 1/5 < 1/(2 + 2\sqrt{2})$. The conclusion follows. □

Constrained LQR

We now return to the LQR problem with state and input constraints. In this section, we present a general result and outline the necessary changes to the sub-optimality argument. Proofs are deferred to Section 3.7.

Recall the robust control synthesis problem which minimizes $\widehat{J}(\gamma, \tau)$ given in (3.3.13) over $0 \leq \gamma, \tau < 1$. This outer minimization can be achieved by searching over the box $[0, 1) \times [0, 1)$. It is possible to reduce the computational complexity by minimizing over only a single variable: $\max(\gamma, \tau)$. The sub-optimality bound would retain the same flavor, but the norm distinctions between cost and constraints would be less clear.

Before we can present a bound on the sub-optimality of robust controllers synthesized using estimated dynamics, there is some care to be taken relating to the constraints. In the previous subsection, our sub-optimality proof hinged on the construction of a feasible solution to the optimization problem. However, because the robust problem tightens the state and input constraints, constructing this feasible solution is more delicate.

Therefore, we define a constraint-tightened version of the optimal controller. Define

$$\zeta = \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi}_{\mathbf{x}}^\star \\ \varepsilon_B \mathbf{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{H}_\infty}, \quad \zeta_\infty = \left\| \begin{bmatrix} \varepsilon_{A,\infty} \mathbf{\Phi}_{\mathbf{x}}^\star \\ \varepsilon_{B,\infty} \mathbf{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{L}_1} .$$

Then the doubly robust constraint functions are defined at each index j as

$$\begin{aligned} \bar{G}_x^\zeta(\mathbf{\Phi}_{\mathbf{x}}; t)_j = F_{x,j}^\top \Phi(t+1) x_0 + \frac{r_w}{1 - 2\zeta_\infty} \| F_{x,j}^\top \Phi_x[t:1] \|_1 \\ + \frac{4\zeta_\infty \max(r_w, \|x_0\|_\infty)}{1 - 4\zeta_\infty} \| F_{x,j}^\top \Phi_x[t+1:1] \|_1 , \end{aligned}$$

and similarly for $\bar{G}_u^\zeta(\mathbf{\Phi}_{\mathbf{u}}; t)$. These doubly robust constraints take the form of the robust constraints with $\tau = 4\zeta_\infty$ and an enlarged noise process where r_w is replaced by $\frac{r_w}{1-2\zeta_\infty}$. Then the optimal robustly constrained controller is defined as

$$\begin{aligned} (\mathbf{\Phi}_{\mathbf{x}}^c, \mathbf{\Phi}_{\mathbf{u}}^c) \in \underset{\mathbf{\Phi}_{\mathbf{x}}, \mathbf{\Phi}_{\mathbf{u}} \in \frac{1}{z}\mathcal{RH}_\infty}{\operatorname{argmin}} & \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}} \\ \mathbf{\Phi}_{\mathbf{u}} \end{bmatrix} \right\|_{\mathcal{H}_2} \\ \text{s.t. } & \begin{bmatrix} zI - A & -B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}} \\ \mathbf{\Phi}_{\mathbf{u}} \end{bmatrix} = I, \\ & \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi}_{\mathbf{x}} \\ \varepsilon_B \mathbf{\Phi}_{\mathbf{u}} \end{bmatrix} \right\|_{\mathcal{H}_\infty} \le \zeta, \ \left\| \begin{bmatrix} \varepsilon_{A,\infty} \mathbf{\Phi}_{\mathbf{x}} \\ \varepsilon_{B,\infty} \mathbf{\Phi}_{\mathbf{u}} \end{bmatrix} \right\|_{\mathcal{L}_1} \le \zeta_\infty, \\ & \bar{G}_x^\zeta(\mathbf{\Phi}_{\mathbf{x}}; t) \le b_x, \ \bar{G}_u^\zeta(\mathbf{\Phi}_{\mathbf{u}}; t) \le b_u \ \ \forall\, t \ge 1 . \end{aligned} \tag{3.4.6}$$

with $\mathbf{K}_c = \mathbf{\Phi}_{\mathbf{u}}^c (\mathbf{\Phi}_{\mathbf{x}}^c)^{-1}$. This optimization problem designs a controller which satisfies more stringent state and input constraints than the optimal controller, without increasing the norms of the system response (as controlled by ζ and ζ_∞). The relative *robustness cost gap* is defined as

$$M_\varepsilon = \frac{J(A, B, \mathbf{K}_c) - J(A, B, \mathbf{K}_\star)}{J(A, B, \mathbf{K}_\star)} . \tag{3.4.7}$$

Now we are ready to state a bound on the sub-optimality of robust controllers synthesized on estimated dynamics.

Theorem 3.4.3. *Suppose that the robust optimal constrained controller problem* (3.4.6) *is feasible. As long as* $\zeta \le \frac{1}{4\sqrt{2}}$ *and* $\zeta_\infty \le \frac{1}{4}$*, the cost achieved by* $\widehat{\mathbf{K}} = \widehat{\mathbf{\Phi}}_{\mathbf{u}} \widehat{\mathbf{\Phi}}_{\mathbf{x}}^{-1}$ *synthesized from the minimizers of* (3.4.1) *satisfies*

$$\frac{J(A, B, \widehat{\mathbf{K}}) - J_\star}{J_\star} \le 4\sqrt{2}(1 + M_\varepsilon) \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi}_{\mathbf{x}}^\star \\ \varepsilon_B \mathbf{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{H}_\infty} + M_\varepsilon .$$

This result is stated in terms of quantities depending only on the true unknown system It therefore highlights properties that make systems easier or harder to robustly control. We see that the bound grows with the $\mathcal{H}_\infty$ norm of the closed-loop responses $\mathbf{\Phi}_{\mathbf{x}}^\star$ and $\mathbf{\Phi}_{\mathbf{u}}^\star$. It also grows with the robustness cost gap M_ε. This cost gap is be difficult to characterize analytically; in general, it requires checking the boundaries of the robust polytopic constraints. Once errors ε_A and ε_B are small enough that the optimal system response $(\mathbf{\Phi}_{\mathbf{x}}^\star, \mathbf{\Phi}_{\mathbf{u}}^\star)$ satisfies the doubly robust constraints, we will have zero cost gap, $M_\varepsilon = 0$. If the optimal controller saturates its constraints, then the cost gap will be nonzero for any nonzero estimation errors. In Section 3.5, we numerically characterize this optimality gap for a double integrator example (Figure 3.5).

As in the unconstrained case, proving Theorem 3.4.3 relies on the construction of a feasible solution to (3.3.13), which we present in the following lemma.

Lemma 3.4.4. *Under the conditions of Theorem 3.4.3, we have that the following is a feasible solution to* (3.4.1)

$$\mathbf{\Phi}_{\mathbf{x}} = \mathbf{\Phi}_{\mathbf{x}}^c(I - \mathbf{\Delta})^{-1},\ \mathbf{\Phi}_{\mathbf{u}} = \mathbf{\Phi}_{\mathbf{u}}^c(I - \mathbf{\Delta})^{-1},\ \tilde{\gamma} = \frac{\sqrt{2}\zeta}{1 - \sqrt{2}\zeta},\ \ \tilde{\tau} = \frac{2\zeta_\infty}{1 - 2\zeta_\infty}\,.$$

where the $(\mathbf{\Phi}_{\mathbf{x}}^c, \mathbf{\Phi}_{\mathbf{u}}^c)$ *is defined as in* (3.4.6) *and we define* $\mathbf{\Delta} := -\begin{bmatrix}\Delta_A & \Delta_B\end{bmatrix}\begin{bmatrix}\mathbf{\Phi}_{\mathbf{x}}^c \\ \mathbf{\Phi}_{\mathbf{u}}^c\end{bmatrix}$.

The proof of Theorem 3.4.3 follows a similar argument to the proof of Proposition 3.4.1 above, except that there is an extra step to account for the difference between the optimal solution $(\mathbf{\Phi}_{\mathbf{x}}^\star, \mathbf{\Phi}_{\mathbf{u}}^\star)$ and the robustly constrained one $(\mathbf{\Phi}_{\mathbf{x}}^c, \mathbf{\Phi}_{\mathbf{u}}^c)$. This step gives rise to the term M_ε that appears in the bound.

End-to-End Guarantees

We now combine our estimation and synthesis results to provide end-to-end guarantees on the Coarse-ID control framework. We begin with the first step: data collection. By designing the controller used during data collection (3.2.11) with the robust synthesis procedure, it is possible to guarantee safety *during* the estimation process. We therefore have a computationally tractable algorithm which returns a controller that (a) guarantees the closed-loop system remains within the specified constraint set and (b) ensures that enough noise can be injected into the system to obtain a statistical guarantee on learning.

To proceed, define an expanded noise process $\tilde{w}_t = B\eta_t + w_t$. Then synthesize $\mathbf{K}_0$ according to (3.3.13) with initial system estimates (A_0, B_0), initial dynamics uncertainties $(\varepsilon_A^0, \varepsilon_B^0)$ and $(\varepsilon_{A,\infty}^0, \varepsilon_{B,\infty}^0)$, noise bound r_w replaced with

$$r_\eta\|B\|_\infty + r_w \le r_\eta(\|B_0\|_\infty + \varepsilon_{B,\infty}^0) + r_w,$$

and input constraints $b_{u,j}$ replaced with $b_{u,j} - r_\eta \|F_{u,j}\|_1$. As long as the synthesis problem is feasible, the control law $\mathbf{u} = \mathbf{K}_0 \mathbf{x} + \boldsymbol{\eta}$ stabilizes the true system, satisfies state and input constraints, and allows for learning at the rate given in Theorem 3.2.5.

Finally, we connect the sub-optimality result to the statistical learning bound for an end-to-end sample complexity bound on the constrained LQR problem. Define

$$\mathsf{num} = mC_u + \frac{\sigma_w}{\sigma_\eta} \frac{\rho C_u C_K}{(1-\rho^2)} \left(1 + \|B\|_2 + \frac{\|x_0\|_2}{\sigma_w \sqrt{T}}\right) .$$

Further define M_T to be the value of M_ε when the definition determined by (3.4.6) has values set as

$$\zeta = \varepsilon \left\| \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{x}}^\star \\ \boldsymbol{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{H}_\infty}, \quad \zeta_\infty = \varepsilon \sqrt{n+m} \left\| \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{x}}^\star \\ \boldsymbol{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{L}_1}, \quad \varepsilon \gtrsim \frac{\sigma_w C_u}{\sigma_\eta} \sqrt{\frac{n+m}{T}} \sqrt{\log(\mathsf{num}/\delta)} .$$

Corollary 3.4.5. *Under the assumptions of Theorem 3.2.5, for a trajectory of length*

$$T \gtrsim (n+m) \log(\mathsf{num}/\delta) \frac{\sigma_w^2 C_u^2}{\sigma_\eta^2} \max \left\{ (n+m) \left\| \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{x}}^\star \\ \boldsymbol{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{L}_1}^2, \left\| \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{x}}^\star \\ \boldsymbol{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{H}_\infty}^2 \right\}, \tag{3.4.8}$$

the cost achieved by $\widehat{\mathbf{K}} = \widehat{\boldsymbol{\Phi}}_{\mathbf{u}} (\widehat{\boldsymbol{\Phi}}_{\mathbf{x}})^{-1}$ *synthesized from* (3.3.13) *on the least-squares estimates* $\widehat{A}, \widehat{B}$ *satisfies with probability at least* $1 - \delta$*,*

$$\frac{J(A, B, \widehat{\mathbf{K}}) - J_\star}{J_\star} \lesssim \frac{\sigma_w C_u}{\sigma_\eta} \sqrt{\frac{n+m}{T}} (1 + M_T) \left\| \begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{x}}^\star \\ \boldsymbol{\Phi}_{\mathbf{u}}^\star \end{bmatrix} \right\|_{\mathcal{H}_\infty} \sqrt{\log(\mathsf{num}/\delta)} + M_T .$$

Proof Sketch. This result follows by combining the statistical guarantee in Theorem 3.2.5 with the sub-optimality bound in Theorem 3.4.3. Note that we use the naïve bound $\varepsilon_{A,\infty} \le \sqrt{n}\varepsilon_A$ and similarly $\varepsilon_{B,\infty} \le \sqrt{m}\varepsilon_B$; this results in an extra factor of $(n+m)$ appearing in (3.4.8) and in the definition of M_T. □

Our final result depends both on the true system and the initial system estimates by way of the data collecting controller, which affects constants C_x, C_u, and ρ. The largest possible excitation covariance σ_η also depends on the initial system estimates through the feasibility of the robust control problem. There is a trade-off here: for fast estimation, larger excitation σ_η and transients C_u are beneficial. However, the safety constraints will preclude these values from being too large. The constraints also have an effect on sub-optimality through M_T.

3.5 Numerical Experiments

We demonstrate the utility of our framework and illustrate our results on estimation, controller synthesis, and LQR performance with numerical experiments of the end-to-end Coarse-ID control scheme.

Unconstrained LQR

The least squares estimation procedure (3.2.1) is carried out on a simulated system in Python. The synthesis and performance experiments are run in MATLAB. We make use of the YALMIP package for prototyping convex optimization [Löf04] and use the MOSEK solver under an academic license [ApS15]. In particular, when using the FIR approximation described in Section 2.5, we find it effective to make use of YALMIP's `dualize` function, which considerably reduces the computation time.

We focus experiments on a particular example system. Consider the LQR problem instance specified by

$$A = \begin{bmatrix} 1.01 & 0.01 & 0 \\ 0.01 & 1.01 & 0.01 \\ 0 & 0.01 & 1.01 \end{bmatrix}, \quad B = I, \quad Q = 10^{-3} I, \quad R = I \,. \tag{3.5.1}$$

The dynamics correspond to a marginally unstable graph Laplacian system where adjacent nodes are weakly connected, each node receives direct input, and input size is penalized relatively more than state. Dynamics described by graph Laplacians arise naturally in consensus and distributed averaging problems. For this system, we collect data using inputs with variance $\sigma_u^2 = 1$ and noise with variance $\sigma_w^2 = 1$. The rollout length is fixed to $T = 6$, and the number of rollouts used in the estimation is varied. The system estimates are constructed with all collected data rather than just the final time step.

We synthesize robust controllers with two different approaches: the FIR approximation with filters of length $L = 32$ and slack variable V set to 0, and a common Lyapunov (CL) relaxation of the static synthesis problem described in [DMMRT20]. Once the FIR responses $(\Phi_x(t))_{t=1}^L$ and $(\Phi_u(t))_{t=1}^L$ are found, we represent the dynamic controller $\mathbf{K} = \mathbf{\Phi_u}\mathbf{\Phi_x}^{-1}$ by finding an equivalent state-space realization (A_K, B_K, C_K, D_K) via Theorem 2 of [AM17]. In what follows, we compare the performance of these controllers with the nominal LQR controller (the solution to (3.1.1) with $\widehat{A}$ and $\widehat{B}$ as model parameters), and explore the trade-off between robustness, complexity, and performance.

The relative performance of the nominal controller is compared with robustly synthesized controllers in Figure 3.1. For both robust synthesis procedures, two controllers are compared: one using the true errors on A and B, and the other using the bootstrap estimates of the errors (as described in [DMMRT20]). The robust static controller generated via the common Lyapunov approximation performs slightly worse than the more complex FIR controller, but it still achieves reasonable control performance. Moreover, the conservative bootstrap estimates also result in worse control performance, but the degradation of performance is again modest.

Furthermore, the experiments show that the nominal controller often outperforms the robust controllers *when it is stabilizing*. On the other hand, the nominal controller is not guaranteed to stabilize the true system, and as shown in Figure 3.1, it only does so in roughly 80 of the 100 instances after $N = 60$ rollouts. It is also important to note a

distinction between stabilization for nominal and robust controllers. When the nominal controller is not stabilizing, there is no indication or warning (though sufficient conditions for stability can be checked using our result in Corollary 3.3.1 or structured singular value methods [QBR+95]). On the other hand, the robust synthesis procedure will return as infeasible, giving a warning by default that the uncertainties are too high.

Figure 3.2 explores the trade-off between performance and complexity for the computational approximations, both for FIR truncation and the common Lyapunov relaxation. We examine the trade-off both in terms of the bound on the LQR cost (given by the value of the objective) as well as the actual achieved value. It is interesting that for smaller numbers of rollouts (and therefore larger uncertainties), the benefit of using more complex FIR models is negligible, both in terms of the actual costs and the upper bound. This trend makes sense: as uncertainties decrease to zero, the best robust controller should approach the nominal controller, which is associated with infinite impulse response (IIR) transfer functions. Furthermore, for the experiments presented here, FIR length of $L = 32$ seems to be sufficient to characterize the performance of the robust synthesis procedure in (3.4.2). Additionally, we note that static controllers are able to achieve costs of a similar magnitude.

The SLS framework guarantees a stabilizing controller for the true system provided that the computational approximations are feasible for *any* value of γ between 0 and 1, as long as the system errors $(\varepsilon_A, \varepsilon_B)$ are upper bounds on the true errors. Figure 3.3 displays the controller performance for robust synthesis when γ is set to 0.999. Simply ensuring a stable model and neglecting to optimize the nominal cost yields controllers that perform nearly an order of magnitude better than those where we search for the optimal value of γ. This observation aligns with common practice in robust control: constraints ensuring stability are only active when the cost tries to drive the system up against a safety limit. We cannot provide end-to-end sample complexity guarantees for this method and leave such bounds as an enticing challenge for future work.

Constrained LQR

We now turn to a constrained LQR problem. In this case, we consider true dynamics are given by a double integrator

$$x_{t+1} = \begin{bmatrix} 1 & 0.1 \\ 0 & 1 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_t + w_t \ .$$

Because the double integrator is under-actuated (having fewer control inputs than states), it is a challenging control task when safety constraints must be satisfied. We consider constraints that require states to be bounded between -8 and 8, and inputs to be bounded in between -4 and 4. The process noise is uniformly distributed on $[-0.1, 0.1] \times [-0.1, 0.1]$. Our initial estimate comes from a randomly generated initial perturbation of the true system with $\varepsilon_{A,\infty} = \varepsilon_{B,\infty} = 0.1$. Safe controllers are generated with finite truncation

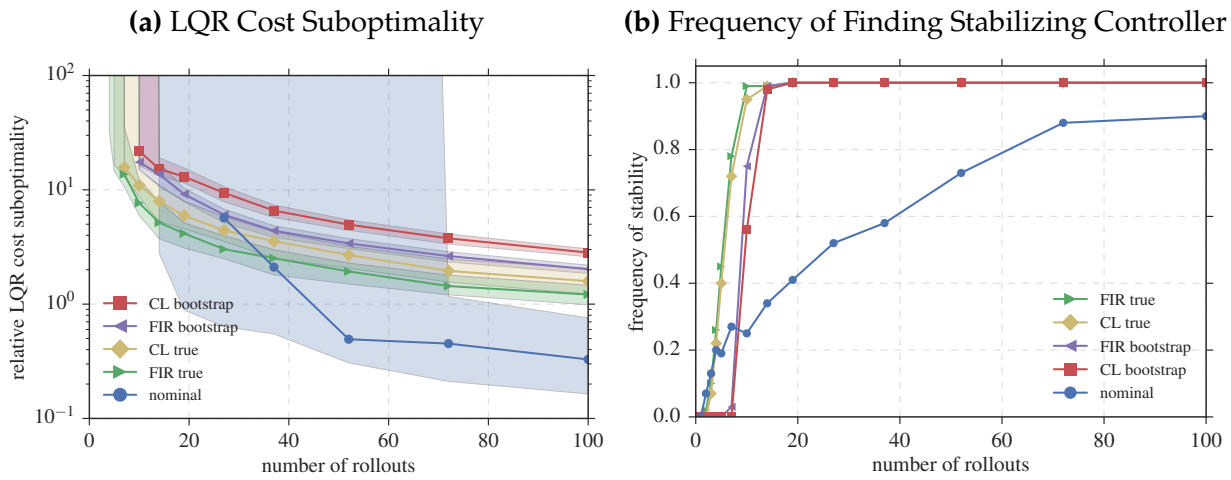

Figure 3.1: The performance of controllers synthesized on the results of 100 identification experiments is plotted against the number of rollouts. Controllers are synthesized nominally, using FIR truncation, and using the common Lyapunov (CL) relaxation. In (a), the median sub-optimality of nominal and robustly synthesized controllers are compared, with shaded regions displaying quartiles, which go off to infinity in the case that a stabilizing controller was not found. In (b), the frequency that the synthesis methods found stabilizing controllers.

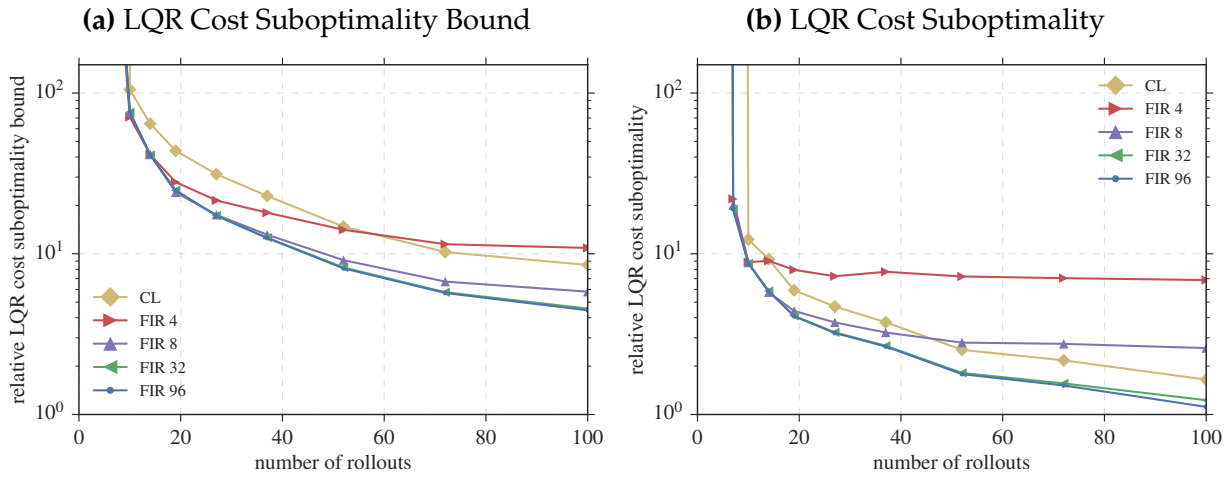

Figure 3.2: The performance of controllers synthesized with varying FIR filter lengths on the results of 10 of the identification experiments using true errors. The median sub-optimality of robustly synthesized controllers does not appear to change for FIR lengths greater than 32, and the common Lyapunov (CL) synthesis tracks the performance in both upper bound and actual cost.

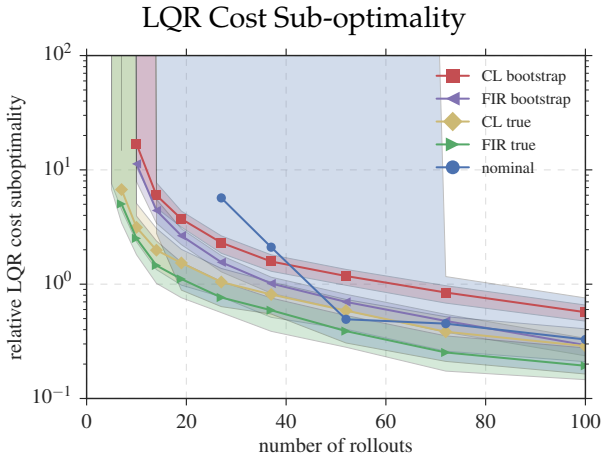

Figure 3.3: The performance of controllers synthesized on the results of 100 identification experiments is plotted against the number of rollouts. The plot compares the median sub-optimality of nominal controllers with fixed-γ robustly synthesized controllers ($\gamma = 0.999$).

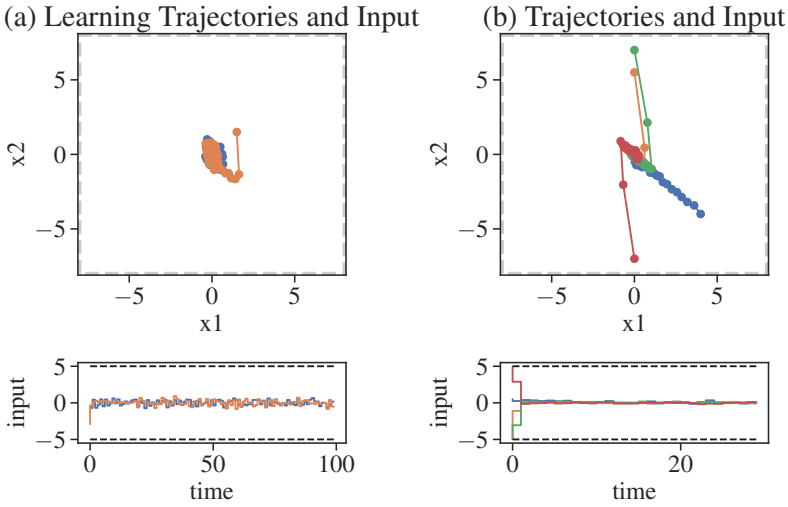

Figure 3.4: Safe learning trajectories synthesized with coarse initial estimates (a), then robust execution with reduced model errors (b).

length $L = 15$, and for larger initial conditions, the system is warm-started with a finite-time robust controller with horizon 20 to reduce the initial condition.

Figure 3.4 displays safe trajectories and input sequences for several example initial conditions. Figure 3.4a plots trajectories that we use for learning: the controller both regulates and excites the system (η_t is uniform on $[-0.5, 0.5]$), and is robust to initial uncertainties. Figure 3.4b demonstrates an ability to operate closer to the constraints when there is less uncertainty: in this case, there is no input excitation ($\eta_t = 0$) and the system estimates are better specified ($\varepsilon_\infty = 0.001$), so larger initial conditions are feasible.

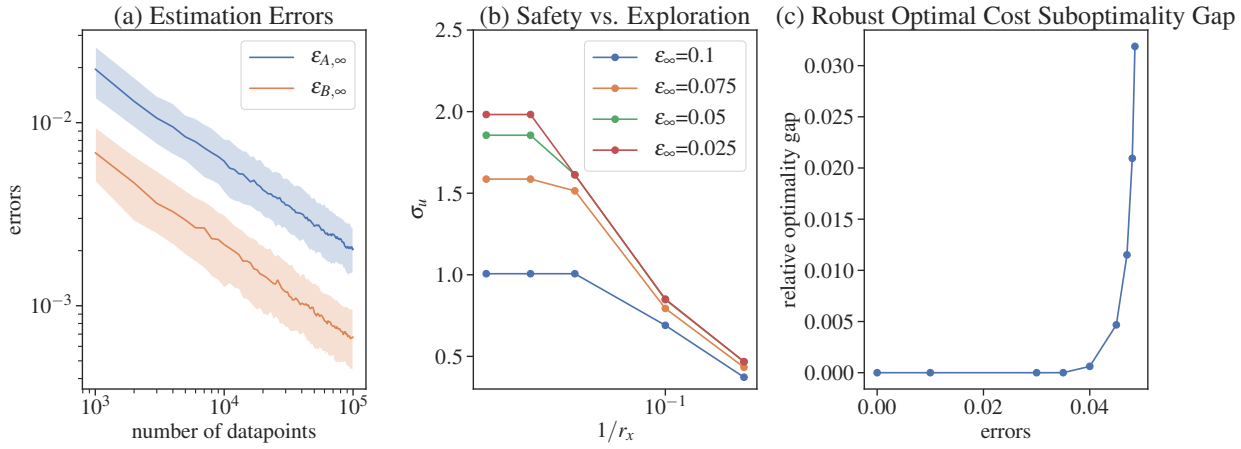

Figure 3.5: Over time, estimation errors decrease (a). As safety requirements increase, the maximum feasible excitation decreases (b). The robust cost sub-optimality gap M_ζ displays an abrupt transition from feasibility to near-optimality with error size (c).

Figure 3.5a displays the decreasing estimation errors over time, demonstrating learning. Shaded areas represent quartiles over 400 trials. Figure 3.5b displays the trade-off between safety and exploration by showing the largest value of r_η for which the robust synthesis is feasible, given constraint sets of the form

$$\mathcal{X} = \{x \mid \|x\|_\infty \leq r_x\} .$$

We leave $x_0 = 0$ and examine a variety of sizes of estimation error in the dynamics. As the uncertainties in the dynamics decrease, higher levels of both safety and exploration are achievable. Finally, Figure 3.5c shows the value of the relative robust sub-optimality gap M_ε for the given example. We see a sharp transition from infeasibility for $\varepsilon \geq 0.05$ to near-optimality for $\varepsilon \leq 0.03$. This indicates that the gap may be most significant as a feasibility condition for our sub-optimality guarantees to hold.

3.6 Conclusion and Open Problems

Coarse-ID control is a straightforward approach to merging non-asymptotic methods from system identification with contemporary approaches to robust control. Indeed, many of the principles of Coarse-ID control were well established in the 90s [CG00; CN93; HJN91], but fusing together an end-to-end result required contemporary analysis of random matrices and a new system level perspective on controller synthesis. Another benefit of our SLS approach is that it handles constraints on the state and input in a straightforward manner, allowing for the design of linear controllers that guarantee safety during and after data collection. Our results can be extended in a variety of directions,

and we close this chapter with a discussion of some of the short-comings of our approach and of several open questions.

Is it possible to reduce the conservatism? We use a very coarse characterization of the uncertainty to bound the quantity $\hat{\mathbf{\Delta}}$ to yield a tractable optimization problem. In fact, the only property we use about the error between our nominal system and the true system is that the maps

$$x \mapsto (A - \hat{A})x \text{ and } u \mapsto (B - \hat{B})u$$

are contractions. Stronger bounds that account for the fact that these are static linear operators could be engineered, for example, using the theory of IQCs [MR97]. Alternatively, incorporating known problem structure, like sparsity, could allow us to characterize the uncertainty $\hat{\mathbf{\Delta}}$ with more precision than just a norm bound. Such tighter bounds could yield considerably less conservative control schemes in both theory and practice.

Furthermore, our numerical experiments suggest that optimizing a nominal cost subject to robust stability constraints, as opposed to directly optimizing the cost upper bound, leads to better empirical performance. Is this a phenomenological observation specific to the systems used in our experiments, or is there a deeper principle at play?

Follow-up work on the unconstrained LQR problem suggests that the conservatism in robust control is detrimental once enough data has been collected. For small enough estimation errors ε, Riccati perturbation theory can guarantee that the sub-optimality of the nominal controller scales as $\tilde{\mathcal{O}}(\varepsilon^2)$ [MTR19]. Our robust synthesis, on the other hand, guarantees a scaling of $\tilde{\mathcal{O}}(\varepsilon)$. This is advantageous when errors are moderate to large, meaning that robust control is more valuable in low data regimes, as suggested by our numerical experiments. A similar understanding of this trade-off does not yet exist for systems that are required to satisfy safety constraints on the states and inputs.

What is the most efficient way to collect data? It would be of interest to determine system identification algorithms that are tuned to particular control tasks. In the Coarse-ID control approach, the estimation and control are completely decoupled. However, it may be beneficial to inform the identification algorithm about the desired cost, resulting in improved sample complexity. One direction would be to connect this work to experiment design literature, perhaps by replacing the objective in the synthesis problem (3.3.13) with an exploration inspired cost function for the data collecting controller $\mathbf{K}_0$. Recent investigation by Wagenmaker, Simchowitz, and Jamieson [WSJ21] along this line suggests that the certainty equivalent controller is in some sense optimal for the unconstrained LQR problem. However, we know of no such work on the constrained LQR problem.

An alternative way to reason about data collection is the through exploration-exploitation trade-offs studied in the setting of online learning. Casting LQR in this setting, one seeks to minimize cost at all times, including during learning. Recent work has shown that a simple ϵ-greedy extension of the certainty equivalent controller achieves optimal scaling

for unconstrained LQR [MTR19; SF20]. Developing an online algorithm for constrained LQR would require an analysis of recursive feasibility, to understand the transition that occurs when controllers are updated based on refined system estimates. It would also likely require a finer analysis on the performance loss characterized by by the constraints cost gap M_ε.

Can we quantify the difficulty of constraints? The nature of the unconstrained and constrained LQR problems are vastly different. While the optimal policy for unconstrained LQR is a static linear controller that can be efficiently computed, the optimal policy in the presence of constraints is piecewise affine and suffers from the curse of dimensionality [BMDP02]. Our results develop sub-optimality guarantees with respect to linear baselines; it would be interesting to characterize the performance gap between the optimal linear and nonlinear controllers.

We suspect that the nature of exploration is different for constrained problems as well. While certainty equivalent control is efficient in the online setting [MTR19; SF20], achieving a regret bounded by $\tilde{O}(\sqrt{T})$, it is unlikely that such an approach would be valid when constraint satisfaction is required. Follow up work extending our robust Coarse-ID control framework to the online setting results in a regret bounded by $\tilde{O}(T^{2/3})$ for unconstrained LQR [DMMRT18]. Robustness is necessary to guarantee constraint satisfaction. Is it possible to achieve $\tilde{O}(\sqrt{T})$ regret for constrained LQR?

To answer such questions, it is necessary to find lower bounds for constrained control problems when the model is unknown. Such bounds would offer a reasonable benchmark for how well one could ever expect to do.

Can these tools be extended to nonlinear control? Even when the underlying system is linear, when state and input constraints are present, the optimal control strategy is nonlinear. For example, in Model Predictive Control (MPC), controller synthesis problems are approximately solved on finite time horizons, one step is taken, and then this process is repeated [BBM17]. MPC is an effective solution which substitutes fast optimization solvers for clever, complex control design. The Coarse-ID paradigm has been successfully extended to MPC [CWMPM20], although for finite-time control problems, many other methods for incorporating robustness are possible [BZTB20; BRSB21].

Providing sample complexity results like the ones presented here is challenging for nonlinear control. Once the controller is nonlinear, the closed-loop system obeys nonlinear dynamics, and therefore many of the properties that enabled our sub-optimality analysis no longer apply. Recent work by Ho [Ho20] extends the SLS framework to nonlinear dynamics, which could be a promising way forward. However, even when the dynamics are fully known, optimality guarantees are scarce in nonlinear control due to the nonconvex nature of the problem. It may thus be more fruitful to turn away from optimality towards goals like ensuring stability and safety [SRSSP20; TDD+20].

3.7 Omitted Proofs

Estimation Proofs

Proof of Lemma 3.2.7. Recall that we want to show

$$\mathbb{P}\{|a+X| \geq \sqrt{\theta \mathbb{E}[X^2]}\} \geq (1-\theta)^2/\max\{4, 3C\} .$$

for $X \in \mathbb{R}$ zero-mean with $\mathbb{E}[X^4] \leq C(\mathbb{E}[X^2])^2$, a a fixed scalar, and and $\theta \in (0,1)$.

First, we note that we can assume $a \geq 0$ without loss of generality, since we can perform a change of variables $X \leftarrow -X$. We have that $\mathbb{E}[(a+X)^4] = a^4 + 6a^2\mathbb{E}[X^2] + 4a\mathbb{E}[X^3] + \mathbb{E}[X^4]$. Therefore, by the Paley-Zygmund inequality and Young's inequality, we have that

$$\begin{aligned}
\mathbb{P}\{|a+X| \geq \sqrt{\theta \mathbb{E}[X^2]}\} &\geq \mathbb{P}\{|a+X| \geq \sqrt{\theta \mathbb{E}[(a+X)^2]}\} \\
&= \mathbb{P}\{(a+X)^2 \geq \theta \mathbb{E}[(a+X)^2]\} \\
&\geq (1-\theta)^2 \frac{(\mathbb{E}[(a+X)^2])^2}{\mathbb{E}[(a+X)^4]} \\
&= (1-\theta)^2 \frac{a^4 + 2a^2\mathbb{E}[X^2] + (\mathbb{E}[X^2])^2}{a^4 + 8a^2\mathbb{E}[X^2] + 3\mathbb{E}[X^4]} .
\end{aligned}$$

Now define the function $f(a)$ for $a \geq 0$ as

$$f(a) = \frac{a^4 + 2a^2\mu + \mu^2}{a^4 + 8a^2\mu + 3\beta} .$$

Clearly $f(0) = \mu^2/(3\beta)$ and $f(\infty) = 1$. Since by Jensen's inequality we know that $\mu^2 \leq \beta$, this means $f(0) < f(\infty)$. On the other hand,

$$\frac{d}{da} f(a) = \frac{4a\left(a^2+\mu\right)\left(\mu\left(3a^2-4\mu\right)+3\beta\right)}{\left(a^4+8a^2\mu+3\beta\right)^2} .$$

Assume that $\mu \neq 0$ (otherwise the claim is trivially true). The only critical points of the function f on non-negative reals is at $a = 0$ and $a = \sqrt{\frac{4\mu^2-3\beta}{3\mu}}$ if $4\mu^2 \geq 3\beta$. If $4\mu^2 < 3\beta$, then $a = 0$ is the only critical point of f, so $f(a) \geq f(0) = \mu^2/(3\beta) \geq 1/(3C)$. On the other hand, if $4\mu^2 \geq 3\beta$ holds, Some algebra yields that

$$f(\sqrt{(4\mu^2-3\beta)/(3\mu)}) = \frac{7\mu^2-3\beta}{16\mu^2-3\beta} \geq \inf_{\gamma \in [3,4]} \frac{7-\gamma}{16-\gamma} = 1/4 .$$

The inequality above holds since $3\beta \in [3\mu^2, 4\mu^2]$. Hence,

$$f(a) \geq \min\{1/4, 1/(3C)\} = 1/\max\{4, 3C\} .$$

□

Proof of Lemma 3.2.8. Recall that we want to show that $z_t = [x_t^\top, u_t^\top]^\top$ satisfies the

$$\left(1, \frac{\sigma_\eta}{2C_u}, \frac{1}{C \cdot C_w}\right) \text{ BMSB condition.}$$

For all $t \geq 1$, denote

$$\begin{aligned}\xi_t &= u_t - \eta_t - \Phi_u(1) w_{t-1} \\ &= \Phi_u(t+1)x_0 + \sum_{k=0}^{t-2} \Phi_u(t-k)(B\eta_k + w_k) + \Phi_u(1)B\eta_{t-1}.\end{aligned}$$

Therefore, we have

$$\begin{bmatrix} x_{t+1} \\ u_{t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} Ax_t + Bu_t \\ \xi_{t+1} \end{bmatrix}}_{\mu_{z,t}} + \underbrace{\begin{bmatrix} I_n & 0 \\ \Phi_u(1) & I_m \end{bmatrix}}_{M_z} \begin{bmatrix} w_t \\ \eta_{t+1} \end{bmatrix}.$$

Notice that $\nu = \begin{bmatrix} w_t^\top & \eta_{t+1}^\top \end{bmatrix}^\top \in \mathbb{R}^{n+m}$ is a random vector with a finite fourth moment and each coordinate ν_i is independent and zero-mean. Then for any $q \in \mathbb{R}^{n+m}$, by Rosenthal's inequality, we have that $\mathbb{E}[|\langle q, \nu\rangle|^4] \leq C \cdot C_\nu \mathbb{E}[|\langle q, \nu\rangle|^2]^2$ for an absolute constant—we can take $C = 4$ (see e.g. [IS02]).

We now apply Lemma 3.2.7 with $a = \langle v, \mu_{z,t}\rangle$, $X = \langle M_z^\top v, \nu\rangle$, $C = 4C_\nu$, and $\theta = 1/2$. Then, for any fixed $v \in \mathbb{R}^{n+m}$,

$$\mathbb{P}_\nu\left(|\langle v, \mu_{z,t} + M_z \nu\rangle| \geq \sqrt{\lambda_{\min}(M_z \Sigma_\nu M_z^\top)/2}\right) \geq \frac{1}{C \cdot C_\nu}.$$

Recall that we denote $z_t = [x_t^\top, u_t^\top]^\top$ and define

$$\Sigma_z := \begin{bmatrix} \sigma_w^2 I_n & \sigma_w^2 \Phi_u(1)^\top \\ \sigma_w^2 \Phi_u(1) & \sigma_w^2 \Phi_u(1)\Phi_u(1)^\top + \sigma_\eta^2 I_m \end{bmatrix}.$$

Then we can write the expression as

$$\mathbb{P}\left(|\langle v, z_t\rangle| \geq \sqrt{\lambda_{\min}(\Sigma_z)}\right) \geq \frac{1}{C \cdot C_\nu}.$$

Since $\Phi_u \in \frac{1}{z}\mathcal{RH}_\infty(C_u, \rho)$ we have $\|\Phi_u(1)\| \leq C_u$. Then, by a simple argument based on a Schur complement it follows that

$$\lambda_{\min}(\Sigma_z) \geq \sigma_\eta^2 \min\left(\frac{1}{2}, \frac{\sigma_w^2}{2\sigma_w^2 C_u^2 + \sigma_\eta^2}\right).$$

The conclusion follows since $C_u \geq 1$. □

Proof of Lemma 3.2.9. Recall that we want to bound the covariance $\sum_{t=0}^{T-1} \mathbf{Tr}\left(\mathbb{E} z_t z_t^\top\right)$. Note that

$$\mathbb{E} z_t z_t^\top = \begin{bmatrix} \Phi_x(t+1) \\ \Phi_u(t+1) \end{bmatrix} x_0 x_0^\top \begin{bmatrix} \Phi_x(t+1) \\ \Phi_u(t+1) \end{bmatrix}^\top + \begin{bmatrix} 0 & 0 \\ 0 & \sigma_\eta^2 I_m \end{bmatrix} + \sum_{k=0}^{t-1} \begin{bmatrix} \Phi_x(t-k) \\ \Phi_u(t-k) \end{bmatrix} (\sigma_\eta^2 BB^\top + \sigma_w^2 I_n) \begin{bmatrix} \Phi_x(t-k) \\ \Phi_u(t-k) \end{bmatrix}.$$

Since for all $k \geq 1$ we have $\|\Phi_x(k)\| \leq C_x \rho^k$ and $\|\Phi_u(k)\| \leq C_u \rho^k$, we obtain

$$\mathbf{Tr}\, \mathbb{E} z_t z_t^\top \leq m\sigma_\eta^2 + (nC_x^2 + mC_u^2) \left(\rho^{2t+2} \|x_0\|_2^2 + (\sigma_w^2 + \sigma_\eta^2 \|B\|^2) \sum_{k=1}^{t} \rho^{2k} \right).$$

Therefore, we get that

$$\sum_{t=0}^{T-1} \mathbf{Tr}\, \mathbb{E} z_t z_t^\top \leq m\sigma_\eta^2 T + \frac{\rho^2 T}{1-\rho^2}(nC_x^2 + mC_u^2)(\sigma_w^2 + \sigma_\eta^2 \|B\|^2) + \frac{\rho^2}{1-\rho^2}(nC_x^2 + mC_u^2)\|x_0\|_2^2,$$

and the conclusion follows by simple algebra. □

Sub-optimality Proofs

Proof of Lemma 3.4.4. The proposed feasible solution is

$$\tilde{\mathbf{\Phi}}_{\mathbf{x}} = \mathbf{\Phi}_{\mathbf{x}}^c (I - \mathbf{\Delta})^{-1}, \ \tilde{\mathbf{\Phi}}_{\mathbf{u}} = \mathbf{\Phi}_{\mathbf{u}}^c (I - \mathbf{\Delta})^{-1}, \ \ \tilde{\gamma} = \frac{\sqrt{2}\zeta}{1 - \sqrt{2}\zeta}, \ \ \tilde{\tau} = \frac{2\zeta_\infty}{1 - 2\zeta_\infty},$$

where $\mathbf{\Delta} = -\begin{bmatrix} \Delta_A & \Delta_B \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_{\mathbf{x}}^c \\ \mathbf{\Phi}_{\mathbf{u}}^c \end{bmatrix}$. First, notice that $\|\mathbf{\Delta}\|_{\mathcal{H}_\infty} \leq \sqrt{2}\zeta < 1$ and $\|\mathbf{\Delta}\|_{\mathcal{L}_1} \leq 2\zeta_\infty < 1$ by Proposition 3.3.2 and the assumptions on ζ and ζ_∞. Then by construction, $\tilde{\mathbf{\Phi}}_{\mathbf{x}}$ and $\tilde{\mathbf{\Phi}}_{\mathbf{u}}$ satisfy the equality constraints.

Checking the $\mathcal{H}_\infty$ norm constraint,

$$\sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \tilde{\mathbf{\Phi}}_{\mathbf{x}} \\ \varepsilon_B \tilde{\mathbf{\Phi}}_{\mathbf{u}} \end{bmatrix} \right\|_{\mathcal{H}_\infty} = \sqrt{2} \left\| \begin{bmatrix} \varepsilon_A \mathbf{\Phi}_{\mathbf{x}}^c \\ \varepsilon_B \mathbf{\Phi}_{\mathbf{x}}^c \end{bmatrix} (I - \mathbf{\Delta})^{-1} \right\|_{\mathcal{H}_\infty} \leq \sqrt{2}\zeta \frac{1}{1 - \|\mathbf{\Delta}\|_{\mathcal{H}_\infty}} \leq \frac{\sqrt{2}\zeta}{1 - \sqrt{2}\zeta} = \tilde{\gamma}.$$

Similarly, for the $\mathcal{L}_1$ norm constraint,

$$2 \left\| \begin{bmatrix} \varepsilon_{A,\infty} \tilde{\mathbf{\Phi}}_{\mathbf{x}} \\ \varepsilon_{B,\infty} \tilde{\mathbf{\Phi}}_{\mathbf{u}} \end{bmatrix} \right\|_{\mathcal{L}_1} \leq 2\zeta_\infty \frac{1}{1 - \|\mathbf{\Delta}\|_{\mathcal{L}_1}} \leq \frac{2\zeta_\infty}{1 - 2\zeta_\infty} = \tilde{\tau}.$$

Then it remains to show that the tightened state and input constraints are satisfied. For compactness we will write $c_0 = \max(1, \frac{1}{r_w}\|x_0\|_\infty)$. Recall that the robust constraint functions have the form

$$G_x^\tau(\tilde{\mathbf{\Phi}}_\mathbf{x}; t)_j = F_{x,j}^\top \tilde{\Phi}_x(t+1)x_0 + r_w \|F_{x,j}^\top \tilde{\Phi}_x[t:1]\|_1 + \frac{\tau r_w c_0}{1-\tau}\|F_{x,j}^\top \tilde{\Phi}_x[t+1:1]\|_1 .$$

We have that $\tilde{\mathbf{\Phi}}_x = \mathbf{\Phi}_\mathbf{x}^c + \mathbf{\Phi}_\mathbf{x}^c \mathbf{\Delta}(I-\mathbf{\Delta})^{-1}$. Define the frequency response elements of $\mathbf{\Delta}(I-\mathbf{\Delta})^{-1}$ by $D(t)$ and the notation $D(1:t)$ to be the vertical concatenation of $D(1), \ldots, D(t)$ in a matrix and

$$\mathrm{Toep}_t(D) := \begin{bmatrix} D(1) & & \\ \vdots & \ddots & \\ D(t) & \ldots & D(1) \end{bmatrix} .$$

Then, by manipulating system response variables, we can show that

$$\tilde{\Phi}(t) = \Phi_x^c(t) + \Phi_x^c[t:1]D(1:t) , \;\; \tilde{\Phi}[t:1] = \Phi_x^c[t:1] + \Phi_x^c[t:1]\mathrm{Toep}_t(D) .$$

Before considering the constraint functions, we show a general fact for any vector v,

$$\|v^\top \mathrm{Toep}_k(D)\|_1 \le \frac{2\zeta_\infty}{1-2\zeta_\infty}\|v\|_1 . \tag{3.7.1}$$

This is true by the following manipulations:

$$\begin{aligned} \|v^\top \mathrm{Toep}_k(D)\|_1 &= \|\mathrm{Toep}_k(D)^\top v\|_1 \le \|\mathrm{Toep}_k(D)^\top\|_1 \|v\|_1 = \|\mathrm{Toep}_k(D)\|_\infty \|v\|_1 \\ &\le \|\mathbf{\Delta}(1-\mathbf{\Delta})^{-1}\|_{\mathcal{L}_1}\|v\|_1 \le \frac{2\zeta_\infty}{1-2\zeta_\infty}\|v\|_1 . \end{aligned}$$

Above, we make use of the fact that the ℓ_1 and ℓ_∞ norms are duals, and therefore $\|A\|_\infty = \|A^\top\|_1$. The second inequality holds because $\mathrm{Toep}_k(D)$ is a truncation of the semi-infinite Toeplitz matrix associated with the operator $\mathbf{\Delta}(1-\mathbf{\Delta})^{-1}$. The final decomposition is valid because $\|\mathbf{\Delta}\|_{\mathcal{L}_1} \le 2\zeta_\infty < 1$.

Now we are ready to consider the state constraint indexed by j and t,

$$\begin{aligned} G_x^{\tilde{\tau}}(\tilde{\mathbf{\Phi}}_\mathbf{x}; t)_j = F_{x,j}^\top &(\Phi_x^c(t+1) + \Phi_x^c[t+1:1]D(1:t+1))x_0 \\ &+ r_w \|F_{x,j}^\top \Phi_x^c[t:1](I + \mathrm{Toep}_t(D))\|_1 \\ &+ \frac{\tilde{\tau}}{1-\tilde{\tau}} r_w c_0 \|F_{x,j}^\top \Phi_x^c[t+1:1](I + \mathrm{Toep}_{t+1}(D))\|_1 . \end{aligned}$$

Considering each term individually,

$$\begin{aligned} F_{x,j}^\top &(\Phi_x^c(t+1) + \Phi_x^c[t+1:1]D(1:t+1))x_0 \\ &= F_{x,j}^\top \Phi_x^c(t+1)x_0 + F_{x,j}^\top \Phi_x^c[t+1:1]D(1:t+1)x_0 . \end{aligned}$$

Then defining E_1 to contain an identity in the first block and zeros elsewhere,

$$\begin{aligned} F_{x,j}^\top \Phi_x^c[t+1:1]D(1:t+1)x_0 &\le \|F_{x,j}^\top \Phi_x^c[t+1:1]\mathrm{Toep}_{t+1}(D)E_1\|_1 \|x_0\|_\infty \\ &\le \|F_{x,j}^\top \Phi_x^c[t+1:1]\|_1 \frac{2\zeta_\infty}{1-2\zeta_\infty}\|x_0\|_\infty \,. \end{aligned}$$

The first inequality is Hölder's inequality and the second by (3.7.1). Next, the second term:

$$r_w \|F_{x,j}^\top \Phi_x^c[t:1](I + \mathrm{Toep}_k(D))\|_1 \le r_w \frac{1}{1-2\zeta_\infty}\|F_{x,j}^\top \Phi_x^c[t:1]\|_1 \,.$$

Finally, the last term,

$$\frac{\tilde{\tau}}{1-\tilde{\tau}} r_w c_0 \|F_{x,j}^\top \Phi_x^c[t+1:1](I + \mathrm{Toep}_{t+1}(D))\|_1 \le \frac{2\zeta_\infty}{1-4\zeta_\infty} r_w c_0 \frac{1}{1-2\zeta_\infty}\|F_{x,j}^\top \Phi_x^c[t+1:1]\|_1$$

where we use (3.7.1) and plug in the definition of $\tilde{\tau}$:

$$\frac{\tilde{\tau}}{1-\tilde{\tau}} = \frac{2\zeta_\infty}{(1-2\zeta_\infty)(1-\frac{2\zeta_\infty}{1-2\zeta_\infty})} = \frac{2\zeta_\infty}{1-4\zeta_\infty} \,.$$

The resulting sum is

$$\begin{aligned} G_x^{\tilde{\tau}}(\tilde{\mathbf{\Phi}}_x; t)_j \le F_{x,j}^\top \Phi_x^c(t+1)x_0 &+ \frac{r_w}{1-2\zeta_\infty}\|F_{x,j}^\top \Phi_x^c[t:1]\|_1 \\ &+ \left(\frac{2\zeta_\infty}{1-2\zeta_\infty}\|x_0\|_\infty + \frac{2\zeta_\infty}{1-4\zeta_\infty}\frac{c_0 r_w}{1-2\zeta_\infty}\right)\|F_{x,j}^\top \Phi_x^c[t+1:1]\|_1 \,. \end{aligned}$$

Then considering constants around the final term,

$$\frac{2\zeta_\infty}{1-2\zeta_\infty}\|x_0\|_\infty + \frac{2\zeta_\infty}{1-4\zeta_\infty}\frac{c_0 r_w}{1-2\zeta_\infty} \le \frac{1}{1-2\zeta_\infty}\left(2\zeta_\infty + \frac{2\zeta_\infty}{1-4\zeta_\infty}\right) c_0 r_w = \frac{4\zeta_\infty}{1-4\zeta_\infty} c_0 r_w \,.$$

Thus, we see that $G_x^{\tilde{\tau}}(\tilde{\mathbf{\Phi}}_\mathbf{x}; t)_j \le \bar{G}_x^{\zeta}(\mathbf{\Phi}_\mathbf{x}^c; t)_j \le b_j$ due to the constraints on $\mathbf{\Phi}_\mathbf{x}^c$. A similar computation with the input constraints shows the same result. Therefore, the proposed solution is feasible. □

Proof of Theorem 3.4.3. Using (3.3.8) along with the bound in Proposition 3.3.2 and the constraints in optimization problem (3.4.1),

$$J(A, B, \widehat{\mathbf{K}}) \le \frac{1}{1-\widehat{\gamma}} J(\widehat{A}, \widehat{B}, \widehat{\mathbf{K}}) \,.$$

Next, we connect the optimal system response to the estimated system using the feasible solution to the robust optimization problem, constructed in Lemma 3.4.4:

$$\frac{1}{1-\widehat{\gamma}} J(\widehat{A}, \widehat{B}, \widehat{\mathbf{K}}) \leq \frac{1}{1-\tilde{\gamma}} J(\widehat{A}, \widehat{B}, \mathbf{K}_c) \,.$$

This is true because $(\widehat{\mathbf{K}}, \widehat{\gamma})$ is the optimal solution to (3.4.1), so objective function with feasible $(\tilde{\mathbf{\Phi}}_u \tilde{\mathbf{\Phi}}_x^{-1} = \mathbf{K}_c, \tilde{\gamma})$ is an upper bound. Then we have

$$\begin{aligned} J(A, B, \widehat{\mathbf{K}}) &\leq \frac{1}{1-\tilde{\gamma}} J(\widehat{A}, \widehat{B}, \mathbf{K}_c) \\ &\leq \frac{1}{1-\tilde{\gamma}} \frac{1}{1-\|\mathbf{\Delta}\|_2} J(A_\star, B_\star, \mathbf{K}_c) \\ &\leq \frac{1}{1-\frac{\sqrt{2}\zeta}{1-\sqrt{2}\zeta}} \frac{1}{1-\sqrt{2}\zeta} (1+M) J(A_\star, B_\star, \mathbf{K}_\star) \\ &\leq \left(1 + 4\sqrt{2}\zeta(1+M) + M\right) J(A_\star, B_\star, \mathbf{K}_\star) \,. \end{aligned}$$

The second inequality follows follows by the argument used to derive (3.3.8) with the roles of the nominal and true systems switched. The final follows from bounding $\|\mathbf{\Delta}\|_2$ by $\sqrt{2}\zeta$ and noticing that $\frac{x}{1-x} \leq 2x$ for $0 \leq x \leq \frac{1}{2}$, where we set $x = 2\sqrt{2}\zeta$. Finally, we note that

$$\zeta = \left\| \begin{bmatrix} \varepsilon_A I \\ \varepsilon_B \mathbf{K}_\star \end{bmatrix} \mathbf{\Phi}_x^\star \right\|_{\mathcal{H}_\infty} \leq (\varepsilon_A + \varepsilon_B \|\mathbf{K}_\star\|_{\mathcal{H}_\infty}) \|\mathbf{\Phi}_x^\star\|_{\mathcal{H}_\infty} \,.$$

□

Chapter 4

Perception-Based Control for Complex Observations

4.1 Introduction

In this chapter, we examine the problem of using perceptual information in feedback control loops. Whereas the previous chapter was devoted to proving safety and performance guarantees for learning-based controllers applied to systems with unknown dynamics, we now focus on the practical scenario where the underlying dynamics of a system are well understood, and it is instead the interaction with a perceptual sensor that is the limiting factor. This chapter uses material first presented in papers coauthored with Nikolai Matni, Benjamin Recht, and Vickie Ye [DMRY20; DR21].

We consider controlling a known linear dynamical system for which partial state information can only be extracted from high dimensional observations. Our approach is to design a *virtual sensor* by learning a perception map, i.e., a map from high dimensional observations to a subset of the state, and crucially to quantify its errors. The analysis of this approach combines contemporary techniques from statistical learning theory and robust control. To characterize the closed-loop behavior, it is *necessary* to bound the errors of the perception map pointwise to guarantee robustness. We consider two settings in which pointwise bounds can be derived: robust and certainty equivalent.

In the robust setting, we show that under suitable smoothness assumptions, bounded errors can be guaranteed within a neighborhood of the training data. This model of uncertainty allows us to synthesize a robust controller that ensures that the system does not deviate too far from states visited during training. Our main result shows that the perception and robust control loop is able to robustly generalize under adversarial noise models. In the certainty equivalent setting, we analyze a non-parametric approach to learning the perception map. Under a dynamically feasible dense sampling scheme and stochastic noise, the learned map converges uniformly. This allows us to show that certainty equivalent control, which treats the perception map as if it is true, has bounded

sub-optimality.

Problem Setting

Consider the dynamical system

$$x_{t+1} = Ax_t + Bu_t + Hw_t \,, \tag{4.1.1}$$
$$z_t = q(x_t) \,, \tag{4.1.2}$$

with system state $x_t \in \mathbb{R}^n$, control input $u_t \in \mathbb{R}^m$, disturbance $w_t \in \mathbb{R}^w$, and observation $z_t \in \mathbb{R}^N$. We take the dynamics matrices (A, B, H) to be known. The observation process is determined by the unknown *generative model* or *appearance map* q, which is nonlinear and potentially quite high dimensional. As an example, consider a camera affixed to the dashboard of a car tasked with driving along a road. Here, the observations $\mathbf{z}$ are the captured images and the map q generates these images as a function of position and velocity. We remark that a nondeterministic appearance map q may be of interest for modeling phenomena like noise and environmental uncertainty. While this chapter focuses on the deterministic case, many of the results can be extended in a straightforward manner to any noise class for which the perception map has a bounded response. For example, many computer vision algorithms are robust to random Gaussian pixel noise, gamma corrections, or sparse scene occlusions.

Motivated by such vision based control systems, our goal is to solve the optimal control problem

$$\begin{aligned} \text{minimize}_{(\gamma_t)_{t\geq 0}} \quad & c(\mathbf{x}, \mathbf{u}) \\ \text{subject to} \quad & \text{dynamics (4.1.1) and measurement (4.1.2)} \\ & u_t = \gamma_t(z_{0:t}), \end{aligned} \tag{4.1.3}$$

where $c(\mathbf{x}, \mathbf{u})$ is a suitably chosen cost function and γ_t is a measurable function of the image history $z_{0:t}$. This problem is made challenging by the nonlinear, high dimensional, and unknown generative model.

Suppose instead that there exists a *perception map* p that imperfectly predicts partial state information; that is $p(z_t) = Cx_t + e_t$ for $C \in \mathbb{R}^{d\times n}$ a known matrix and error $e_t \in \mathbb{R}^d$. Such a matrix C might be specified to encode, for example, that camera images provide good signal on position, but not velocity or acceleration. We therefore define a new measurement model in which the map p plays the role of a noisy sensor:

$$y_t = p(z_t) = Cx_t + e_t. \tag{4.1.4}$$

This allows us to reformulate problem (4.1.3) as a *linear* optimal control problem, where the measurements are defined by (4.1.4) and the control law $u_t = \mathbf{K}(y_{0:t})$ is a *linear* function of the outputs of past measurements $y_{0:t}$:

$$\begin{aligned} \text{minimize}_{\mathbf{K}} \quad & c(\mathbf{x}, \mathbf{u}) \\ \text{subject to} \quad & x_{t+1} = Ax_t + Bu_t + Hw_t \\ & y_t = Cx_t + e_t \\ & u_t = \mathbf{K}(y_{0:t}). \end{aligned} \tag{4.1.5}$$

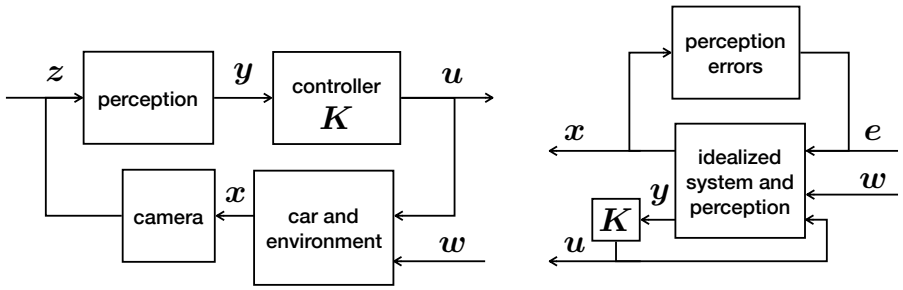

Figure 4.1: (Left) A diagram of the proposed perception-based control pipeline. (Right) The conceptual rearrangement of the closed-loop system permitted through our perception error characterization.

As illustrated in Figure 4.1, guarantees of performance, safety, and robustness require designing a controller which suitably responds to system disturbance and sensor error.

For linear optimal control problems, a variety of cost functions and noise models have well understood solutions; recall those reviewed in Chapter 2. Perhaps the most well known is the combination of *Kalman filtering* with static state feedback, which arises as the solution to the linear quadratic Gaussian (LQG) problem. However, the perception errors $\mathbf{e}$ do not necessarily obey assumptions about measurement noise made in traditional optimal control, and must be handled carefully.

In what follows, we overload the norm $\|\cdot\|$ so that it applies equally to elements x_t, signals $\mathbf{x}$, and linear operators $\mathbf{\Phi}$. For any element norm, we define the signal and linear operator norms as

$$\|\mathbf{x}\| = \sup_{t \geq 0} \|x_t\|, \qquad \|\mathbf{\Phi}\| = \sup_{\|\mathbf{w}\| \leq 1} \|\mathbf{\Phi}\mathbf{w}\|$$

We primarily focus on the triple $(\|x_k\|_\infty, \|\mathbf{x}\|_\infty, \|\mathbf{\Phi}\|_{\mathcal{L}_1})$, though many of our results could be extended.

Related Work

Vision based estimation, planning, and control. There is a rich body of work, spanning several research communities, that integrates complex sensing modalities into estimation, planning, and control loops. The robotics community has focused mainly on integrating camera measurements with inertial odometry via an Extended Kalman Filter (EKF) [JS11; KS11; HKBR14]. Similar approaches have also been used as part of Simultaneous Localization and Mapping (SLAM) algorithms in both ground [LSB+15] and aerial [LAWCS13] vehicles. We note that these works focus solely on the estimation component, and do not consider downstream use of the state estimate in control loops. In contrast, work by Loianno et al. [LBMK16], Tang, Wüest, and Kumar [TWK18], and Lin et al. [LGQ+18]

demonstrate techniques that use camera measurements to aid inertial position estimates in order to enable aggressive control maneuvers in unmanned aerial vehicles.

The machine learning community has taken a more data-driven approach. The earliest such example is likely [Pom89], in which a 3-layer neural-network is trained to infer road direction from images. Modern approaches to vision based planning, typically relying on deep neural networks, include learning maps from image to trail direction [GGC+15], learning Q-functions for indoor navigation using 3D CAD images [SL17], and using images to specify waypoints for indoor robotic navigation [BTGMT20]. Moving from planning to low-level control, end-to-end learning for vision based control has been achieved through imitation learning from training data generated via human [BDD+16] and model predictive control [PCS+18]. The resulting policies map raw image data directly to low-level control tasks. Codevilla et al. [CMLKD18] propose a method for mapping high level navigational commands, images, and other sensor measurements to control actions via imitation learning. Similarly, Williams et al. [WDGRT18] map image and inertial data to a cost landscape, which is then optimized via a path integral based sampling algorithm. More closely related to our approach is from work by Lambert et al. [LSRLB18], where a deep neural network is used to learn a map from image to system state—we note that such a perception module is naturally incorporated into our proposed pipeline. To the best of our knowledge, none of the aforementioned results provide safety or performance guarantees.

The observation of a linear system through a static nonlinearity is classically studied in the controls community as a *Weiner system* model [ST17]. While there are identification results for Weiner systems, they apply only to single-input-single-output systems, and often require assumptions that do not apply to the motivation of cameras [Has87; Wig94; Gre97; TS14]. More flexible identification schemes have been proposed [LB01; SK16], but they lack theoretical guarantees. Furthermore, these approaches focus on identifying the full forward model, rather than an inverse model as we do here.

Learning, robustness, and control. There is much related and recent work at the intersection of learning and control theory. Similar in spirit is a line of work on the Linear Quadratic Regulator which focuses on issues of system identification and sub-optimality [DMMRT20; DMMRT18; AS11]. This style of sample complexity analysis has allowed for illuminating comparisons between model based and model free policy learning approaches [TR19; ALS19]. Mania, Tu, and Recht [MTR19] and Simchowitz and Foster [SF20] show that the simple strategy of model based certainty equivalent control is efficient, though the argument is specialized to linear dynamics and quadratic cost. For nonlinear systems, analyses of learning often focus on ensuring safety over identification or sub-optimality [TDLYA19; BTSK17; WZ18; COMB19], and rely on underlying smoothness for their guarantees [LSCAY20; NLS+20]. An exception is a recent result by Mania, Jordan, and Recht [MJR20] which presents finite sample guarantees for parametric nonlinear system identification.

While the majority of work on learning for control focuses on settings with full state observation, output feedback is receiving growing attention for linear systems [SBR19; SSH20] and for safety-critical systems [LCT20]. Recent work in closely related problem settings includes Mhammedi et al. [MFS+20], who develop sample complexity guarantees for LQR with nonlinear observations and [MHKL20], who leverage representation learning in Block MDPs. However, neither address issues of stability due to their focus on finite horizon problems.

Our theoretical contributions in the robust setting are similar in spirit to those of the online learning community, in that we provide generalization guarantees under adversarial noise models [AHMS13; AHM15; KM16; HK01; YP18; ABHKS19]. The statistical analysis used in the certainty equivalent setting focuses on nonparametric pointwise error bounds over a compact set. Distinct from mean-error generalization arguments most common in learning theory, our analysis is directly related to classical statistical results on uniform convergence [Dev78; Lie89; Han08].

4.2 Bounded Errors via Robust Generalization

While it is typical in the machine learning community to consider mean error generalization bounds, we show in the following example that it is necessary to quantify the uncertainty pointwise to guarantee robustness. This motivating example shows how errors within sets of vanishingly small measure can cause systems to exit bounded regions of well-characterized perception and lead to instability. For a simple setting, we construct a disturbance sequence that leads to instability, illustrating that feedback control objectives require very strong function approximation error bounds.

Example 4.1. Consider a one dimensional linear system that is open-loop unstable ($|a| > 1$) with an arbitrary linear controller and adversarial disturbances:

$$x_{t+1} = a x_t + u_t + w_t, \quad u_t = \sum_{k=0}^{t} K_k \widehat{x}_{t-k}.$$

Further suppose that the perception map has been characterized within the bounded interval $[-\Gamma, \Gamma]$ and is imperfect only around the point $\bar{x} \in [-\Gamma, \Gamma]$:

$$\widehat{x} = p(q(x)) = \begin{cases} 0 & |x| \geq \Gamma \text{ or } x = \bar{x} \\ x & \text{otherwise} \end{cases}.$$

Then consider the disturbance sequence $w_0 = \bar{x} - (a + K_0)x_0$, $w_1 = \Gamma - a\bar{x} - K_1 x_0$, and $w_t = 0$ for all $t \geq 2$. The perception error at $\bar{x}$ causes the unstable system trajectory

$$|x_t| = |a^{t-1}\bar{x}| \to \infty \text{ as } t \to \infty .$$

Furthermore, by carefully choosing x_0 and $\bar{x}$, it is possible to construct scenarios where this instability occurs for disturbances of arbitrarily small norm.

Data-dependent perception error

In this section, we consider a general class of perception maps and introduce a procedure to estimate regions for which a the perception map can be used safely during operation. We first suppose access to initial training data $\mathcal{S} = \{(x_i, z_i)\}_{i=1}^T$, used to learn a perception map via any of the wide variety of traditional supervised methods. We then estimate safe regions around the training data, under an assumption of *slope boundedness*.

Definition 4.1 (Slope boundedness)**.** A fuction f is locally S-slope bounded for a radius γ around x_0 if for $x \in B_\gamma(x_0)$, $\|f(x) - f(x_0)\| \leq S\|x - x_0\|$, where the ball $B_\gamma(x_0) = \{x : \|x - x_0\| \leq \gamma\}$.

This allows us to characterize how quickly the learned perception map degrades as we move away from the initial training data. We will describe the regions of the state space within which the sensing is reliable using a safe set which approximates sub-level sets of the error function $e(x) = p(q(x)) - Cx$. We make this precise in the following lemma, defining a safe set which is valid under an assumption that the error function is locally slope bounded around training data.

Lemma 4.2.1 (Closeness implies generalization)**.** *Suppose that the error function $p \circ q - C$ is locally S-slope bounded with a radius of γ around training data points. Define the safe set*

$$\mathcal{X}_\varepsilon = \bigcup_{(x_i,z_i)\in\mathcal{S}} \{x \in B_\gamma(x_i) \; : \quad \|p(z_i) - Cx_i\| + S\|x - x_i\| \leq \varepsilon\} \,. \tag{4.2.1}$$

Then for any (x, z) with $x \in \mathcal{X}_\varepsilon$, the perception error is bounded: $\|p(z) - Cx\| \leq \varepsilon$.

Proof. The proof follows by a simple argument. For training data point (x_i, z_i) such that $x_i \in B_\gamma(x)$,

$$\begin{aligned}\|p(z) - Cx\| &= \|p(q(x)) - Cx - (p(q(x_i)) - Cx_i) + p(q(x_i)) - Cx_i\| \\ &\leq S\|x - x_i\| + \|p(q(x_i)) - Cx_i\| \,.\end{aligned}$$

The second line follows from the assumption about local slope boundedness. Then because $x \in B_\gamma(x_i)$, by the definition of $\mathcal{X}_\varepsilon$, we have

$$\|p(z) - Cx\| \leq S\|x - x_i\| + \|p(q(x_i)) - Cx_i\| \leq \varepsilon \,.$$

□

The safe set $\mathcal{X}_\varepsilon$ is defined in terms of a bound on the slope of the error function locally around the training data. Deriving the slope boundedness of the error function relies on the learned perception map as well as the underlying generative model. It is possible to also estimate a bound on S using an additional dataset composed of samples around each training data point [DMRY20]. However, the validity of this estimate will

depend on further assumptions about the error function $p \circ q - C$, like Lipschitz continuity. These assumptions are no more natural than assuming knowledge of the constant of slope boundedness S. Therefore, in this chapter, we treat S as a structural assumption about the problem setting. We remark that this notion of slope boundedness has connections to sector bounded nonlinearities, a classic setting for nonlinear system stability analysis [DV75].

Robust control for generalization

The local generalization result in Lemma 4.2.1 is useful only if the system remains close to states visited during training. We now show that robust control can ensure that the system will remain close to training data so long as the perception map generalizes well. By then enforcing that the composition of the two bounds is a contraction, a natural notion of controller robustness emerges that guarantees favorable behavior and generalization. In what follows, we adopt an adversarial noise model and exploit the fact that we can design system behavior to bound how far the system deviates from states visited during training.

Recall that for a state-observation pair (x, z), the perception error, defined as $e := p(z) - Cx$, acts as additive noise to the measurement model $y = p(z)$. While standard linear control techniques can handle uniformly bounded errors, more care is necessary to further ensure that the system remains with a safe region of the state space, as determined by the training data. Through a suitable convex reformulation of the safe region, this goal could be addressed through receding horizon strategies (e.g. [WK02; MRFA06]). While these methods are effective in practice, constructing terminal sets and ensuring a priori that feasible solutions exist is not an easy task. To make explicit connections between learning and control, we turn our analysis to a system level perspective on the closed-loop to characterize its sensitivity to noise.

Once the control input to dynamical system (4.1.1) is defined to be a linear function of the measurement (4.1.4), the closed-loop behavior is determined entirely by the process noise $\mathbf{w}$ and the measurement noise $\mathbf{e}$ (as illustrated in Figure 4.1). Therefore, we can write the system state and input directly as a linear function of the noise

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} \begin{bmatrix} H\mathbf{w} \\ \mathbf{e} \end{bmatrix} . \tag{4.2.2}$$

In what follows, we will state results in terms of these system response variables. Recall from Chapter 2 that the connection between these variables and the feedback control law $\mathbf{u} = \mathbf{K}\mathbf{y}$ that achieves the response (4.2.2) is formalized in the System Level Synthesis (SLS) framework [WMD19]. SLS states that there exists a linear feedback controller $\mathbf{K}$ that achieves the response $\mathbf{\Phi} = (\mathbf{\Phi_{xw}}, \mathbf{\Phi_{xv}}, \mathbf{\Phi_{uw}}, \mathbf{\Phi_{uv}})$ for any system response $\mathbf{\Phi}$ constrained to lie in the affine space $\mathcal{A}$ defined by the system dynamics. Finally, recall that the linear optimal control problem (4.1.5) can be written as

$$\text{minimize}_{\mathbf{\Phi}} \quad c(\mathbf{\Phi}) \;\; \text{subject to} \;\; \mathbf{\Phi} \in \mathcal{A}(A, B, C) \, , \tag{4.2.3}$$

where the cost function c is redefined to operate on system responses. Section 2.4 presents examples of control costs that can be written as system norms.

In what follows, we specialize controller design concerns to our perception-based setting, and develop further conditions on the closed-loop response $\mathbf{\Phi}$ to incorporate into the synthesis problem.

Lemma 4.2.2 (Generalization implies closeness)**.** *For a perception map p with errors $\mathbf{e} = p(\mathbf{z}) - C\mathbf{x}$, let the system responses $(\mathbf{\Phi_{xw}}, \mathbf{\Phi_{xv}}, \mathbf{\Phi_{uw}}, \mathbf{\Phi_{uv}})$ lie in the affine space defined by dynamics (A, B, C), and let $\mathbf{K}$ be the associated controller. Then the state trajectory $\mathbf{x}$ achieved by the control law $\mathbf{u} = \mathbf{K}p(\mathbf{z})$ and driven by noise process $\mathbf{w}$, satisfies, for any target trajectory $\bar{\mathbf{x}}$,*

$$\|\mathbf{x} - \bar{\mathbf{x}}\| \leq \|\widehat{\mathbf{x}} - \bar{\mathbf{x}}\| + \|\mathbf{\Phi_{xv}}\|\|\mathbf{e}\|. \tag{4.2.4}$$

where we define the nominal closeness $\|\widehat{\mathbf{x}} - \bar{\mathbf{x}}\| = \|\mathbf{\Phi_{xw}} H\mathbf{w} - \bar{\mathbf{x}}\|$ *to be the deviation from the target trajectory in the absence of measurement errors.*

Proof. Notice that over the course of a trajectory, we have system outputs $\mathbf{y} = p(\mathbf{z}) = C\mathbf{x}+\mathbf{e}$. Then recalling that the system responses are defined such that $\mathbf{x} = \mathbf{\Phi_{xw}} H\mathbf{w} + \mathbf{\Phi_{xv}}\mathbf{e}$, we have that

$$\|\mathbf{x} - \bar{\mathbf{x}}\| = \|\mathbf{\Phi_{xw}} H\mathbf{w} + \mathbf{\Phi_{xv}}\mathbf{e} - \bar{\mathbf{x}}\| \leq \|\mathbf{\Phi_{xw}} H\mathbf{w} - \bar{\mathbf{x}}\| + \|\mathbf{\Phi_{xv}}\|\|\mathbf{e}\| \,.$$

□

Letting $\bar{\mathbf{x}}$ represent a trajectory of training data, the terms in bound (4.2.4) capture different generalization properties. The first is small if we plan to visit states during operation that are similar to those seen during training. The second term is a measure of the robustness of the closed-loop system to the error $\mathbf{e}$.

We are now in a position to state the main result on robust generalization and stability, which shows that under an additional robustness condition, Lemmas 4.2.1 and 4.2.2 combine to define a control invariant set around the training data within which the perception errors, and consequently the performance, are bounded.

Theorem 4.2.3. *Let the assumptions of Lemmas 4.2.1 and 4.2.2 hold and, for simplicity of presentation, suppose that the training error is bounded: $\|p(z_i) - Cx_i\| \leq \varepsilon_{\mathcal{S}}$ for all $(x_i, z_i) \in \mathcal{S}$. Define the signal $\mathbf{x}_{\mathcal{S}}$ to have elements populated by states in the training dataset $\mathcal{S}$. Then as long as*

$$\|\mathbf{\Phi_{xv}}\| \leq \frac{1 - \frac{1}{\gamma}\|\widehat{\mathbf{x}} - \mathbf{x}_{\mathcal{S}}\|}{S + \frac{\varepsilon_{\mathcal{S}}}{\gamma}}\,, \tag{4.2.5}$$

the perception errors remain bounded

$$\|p(\mathbf{z}) - C\mathbf{x}\| \leq \frac{\|\widehat{\mathbf{x}} - \mathbf{x}_{\mathcal{S}}\| + \varepsilon_{\mathcal{S}}}{1 - S\|\mathbf{\Phi_{xv}}\|} =: \varepsilon, \tag{4.2.6}$$

and the closed-loop trajectory lies within $\mathcal{X}_\varepsilon$.

Proof. Recall that as in the proof of Lemma 4.2.1, as long as $x_t \in B_\gamma(x_{i_t})$ for all t and some $x_{i_t} \in \mathcal{S}$,

$$\|e_t\| \le S\|x_t - x_{i_t}\| + \|e_{i_t}\| \implies \|\mathbf{e}\| = \max_t \|e_t\| \le S\|\mathbf{x} - \mathbf{x}_\mathcal{S}\| + \|\mathbf{e}_\mathcal{S}\|, \tag{4.2.7}$$

where we define the data signal $\mathbf{x}_\mathcal{S} = (x_{i_t})_{t\ge 0}$ and error signal $\mathbf{e}_\mathcal{S} = (e_{i_t})_{t\ge 0}$ corresponding to $(x_t)_{t\ge 0}$.

By assumption, $\|\mathbf{e}_\mathcal{S}\| \le \varepsilon_\mathcal{S}$. Substituting this expression into the result of Lemma 4.2.2, we see that as long as $\|x_t - x_{i_t}\| \le \gamma$ for all t,

$$\|\mathbf{x} - \mathbf{x}_\mathcal{S}\| \le \|\widehat{\mathbf{x}} - \mathbf{x}_\mathcal{S}\| + \|\mathbf{\Phi}_{\mathbf{xv}}\|(S\|\mathbf{x} - \mathbf{x}_\mathcal{S}\| + \|\mathbf{e}_\mathcal{S}\|) \iff \|\mathbf{x} - \mathbf{x}_\mathcal{S}\| \le \frac{\|\widehat{\mathbf{x}} - \mathbf{x}_\mathcal{S}\| + \varepsilon_\mathcal{S}\|\mathbf{\Phi}_{\mathbf{xv}}\|}{1 - S\|\mathbf{\Phi}_{\mathbf{xv}}\|}.$$

Next, to ensure that the the radius γ is bounded, first note that by the definition of the ℓ_∞ norm, $\max_t \|x_t - x_{i_t}\| \le \gamma$ if and only if $\|\mathbf{x} - \mathbf{x}_\mathcal{S}\| \le \gamma$. A sufficient condition for this is given by

$$\frac{\|\widehat{\mathbf{x}} - \mathbf{x}_\mathcal{S}\| + \varepsilon_\mathcal{S}\|\mathbf{\Phi}_{\mathbf{xv}}\|}{1 - S\|\mathbf{\Phi}_{\mathbf{xv}}\|} \le \gamma \iff \|\mathbf{\Phi}_{\mathbf{xv}}\| \le \frac{1 - \frac{1}{\gamma}\|\widehat{\mathbf{x}} - \mathbf{x}_\mathcal{S}\|}{S + \frac{\varepsilon_\mathcal{S}}{\gamma}},$$

and thus we arrive at the robustness condition. This validates the the bound on $\|\mathbf{x} - \mathbf{x}_\mathcal{S}\|$, which we now use to bound $\|\mathbf{e}\|$, starting with (4.2.7) and rearranging,

$$\|\mathbf{e}\| \le \frac{\|\widehat{\mathbf{x}} - \mathbf{x}_\mathcal{S}\| + \varepsilon_\mathcal{S}}{1 - S\|\mathbf{\Phi}_{\mathbf{xv}}\|}.$$

□

Theorem 4.2.3 shows that the bound (4.2.5) should be used during controller synthesis to ensure generalization. Feasibility of the synthesis problem depends on the controllability and observability of the system (A, B, C), which impose limits on how small $\|\mathbf{\Phi}_{\mathbf{xv}}\|$ can be made to be, and on the planned deviation from training data as captured by the quantity $\|\widehat{\mathbf{x}} - \mathbf{x}_\mathcal{S}\|$. In the following section, we show how to simplify this term for a reference tracking problem.

4.3 Robust Perception-Based Control

Reference Tracking

We consider a reference tracking problem with the goal of ensuring that the system tracks a series of arbitrary waypoints that are sequentially revealed. Recall from Example 2.3 that the reference tracking objective is given by:

$$c(\boldsymbol{\xi}, \mathbf{u}) = \sup_{\substack{\|r_{t+1} - r_t\| \le \Delta_{\text{ref}}, \\ t \ge 0}} \left\| \begin{matrix} \bar{Q}^{1/2}(\xi_t - r_t) \\ R^{1/2} u_t \end{matrix} \right\|,$$

for a reference signal r_t and a linear system with state ξ_t such that

$$\xi_{t+1} = \bar{A}\xi_t + \bar{B}u_t \ .$$

We suppose that observations are generated as $z_t = q(\xi_t)$ and that the perception map $p(z_t) = \bar{C}\xi_t + e_t$ satisfies the slope-bounded assumption.

Further recall that we can reformulate the problem by defining the state as the concatenation with the next waypoint, $x_t := [\xi_t; r_t]$, the disturbance as the change in reference, $w_t := r_{t+1} - r_t$, and

$$A = \begin{bmatrix} \bar{A} & 0 \\ 0 & I \end{bmatrix}, \ B = \begin{bmatrix} \bar{B} \\ 0 \end{bmatrix}, \ C = \begin{bmatrix} \bar{C} & 0 \end{bmatrix}, \ H = \begin{bmatrix} 0 \\ I \end{bmatrix}, \ Q^{1/2} = \begin{bmatrix} \bar{Q}^{1/2} & -\bar{Q}^{1/2} \end{bmatrix} \ .$$

Under this formulation, the optimal control problem has the same form as (4.1.5) and can therefore be written in terms of system response variables as in (4.2.3) with the cost $c(\mathbf{\Phi})$ given by a weighted $\mathcal{L}_1$ system norm.

Robust Synthesis

We now show how to simplify the term $\|\widehat{\mathbf{x}} - \mathbf{x}_{\mathcal{S}}\|$ in the case of waypoint tracking. The expression allows us to propose a robust control synthesis problem.

Proposition 4.3.1. *Consider reference tracking problem and suppose that for all $t = 0, 1, \ldots$, the reference trajectory has bounded differences $\|r_{t+1} - r_t\| \leq \Delta_{\text{ref}}$ and is within a ball of radius γ_{ref} from the training data:*

$$\min_{\xi_i \in \mathcal{S}} \|r_t - \xi_i\| \leq \gamma_{\text{ref}} \ \ \forall \ t \geq 0 \ .$$

Then, the closed-loop deviation from training data is bounded by:

$$\|\widehat{\mathbf{x}} - \mathbf{x}_{\mathcal{S}}\| \leq \Delta_{\text{ref}} \| \begin{bmatrix} I & -I \end{bmatrix} \mathbf{\Phi}_{\mathbf{xw}} H \| + \gamma_{\text{ref}} \ .$$

Proof. Because the appearance and perception maps depend on the augmented state x_t only through ξ_t, the errors do not depend the reference signal r_t. We can arbitrarily define the second element of lifted training data so that

$$\|\widehat{\mathbf{x}} - \mathbf{x}_{\mathcal{S}}\| = \|\widehat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{\mathcal{S}}\| \leq \|\boldsymbol{\xi} - \mathbf{r}\| + \|\mathbf{r} - \boldsymbol{\xi}_{\mathcal{S}}\| \ .$$

Therefore, noting that $\boldsymbol{\xi} - \mathbf{r} = \begin{bmatrix} I & -I \end{bmatrix} \widehat{\mathbf{x}}$,

$$\begin{aligned} \|\widehat{\mathbf{x}} - \mathbf{x}_{\mathcal{S}}\| &\leq \| \begin{bmatrix} I & -I \end{bmatrix} \mathbf{\Phi}_{\mathbf{xw}} \mathbf{w} \| + \|\mathbf{r} - \boldsymbol{\xi}_{\mathcal{S}}\| \\ &\leq \Delta_{\text{ref}} \| \begin{bmatrix} I & -I \end{bmatrix} \mathbf{\Phi}_{\mathbf{xw}} \| + \gamma_{\text{ref}} \ . \end{aligned}$$

□

Using the expression in Proposition 4.3.1 it is possible to add the robustness condition from Theorem 4.2.5 to a control synthesis problem as a constraint on the norm of system response variables

$$\left(S + \frac{\varepsilon_{\mathcal{S}}}{\gamma}\right) \|\mathbf{\Phi_{xv}}\| + \frac{\Delta_{\text{ref}}}{\gamma} \| \begin{bmatrix} I & -I \end{bmatrix} \mathbf{\Phi_{xw}} H \| + \frac{\gamma_{\text{ref}}}{\gamma} \leq 1 \, .$$

This constraint immediately implies a trade-off between the size of different parts of the system response: $\mathbf{\Phi_{xv}}$, which is the closed-loop response to measurements, and $\mathbf{\Phi_{xw}}$, which is the response to waypoints. Because the system responses must lie in an affine space, they cannot both become arbitrarily small. Therefore, the synthesis problem must trade off between sensitivity to measurement errors and tracking fidelity. These trade-offs are mediated by the quality of the perception map, the ambition of the control task, and the comprehensiveness of the training data. High quality perception maps will have small training error $\varepsilon_{\mathcal{S}}$, small slope bound S, and large radius γ within which the slope bound holds. Ambitious control tasks will design for large jumps between waypoints Δ_{ref}. Large deviations from training data γ_{ref} will be required for more ambitious control tasks or less comprehensive training datasets. Therefore, the higher quality the perception map and the more comprehensive the training data, the more ambitious the control tasks that are possible.

This robustness constraint is enough to guarantee a finite cost, but it does not guarantee a small cost. To achieve this, we incorporate the constraint along with the perception error bound from Theorem 4.2.5 to arrive at the following robust synthesis procedure:

$$\begin{aligned} \text{minimize}_{\mathbf{\Phi}, \varepsilon > 0} \quad & \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} \begin{bmatrix} \Delta_{\text{ref}} H \\ \varepsilon I \end{bmatrix} \right\| \\ \text{subject to} \quad & \mathbf{\Phi} \in \mathcal{A}(A, B, C), \\ & (S + \tfrac{\varepsilon_{\mathcal{S}}}{\gamma}) \|\mathbf{\Phi_{xv}}\| + \tfrac{\Delta_{\text{ref}}}{\gamma} \| \begin{bmatrix} I & 0 \end{bmatrix} \mathbf{\Phi_{xw}} H \| + \tfrac{\gamma_{\text{ref}}}{\gamma} \leq 1 \\ & \Delta_{\text{ref}} \| \begin{bmatrix} I & 0 \end{bmatrix} \mathbf{\Phi_{xw}} H \| + \gamma_{\text{ref}} + \varepsilon_{\mathcal{S}} \leq \varepsilon (1 - S \|\mathbf{\Phi_{xv}}\|) \, . \end{aligned}$$

This can be formulated into a convex program for fixed ε. There is some lower limit ε_0 for which $\varepsilon < \varepsilon_0$ will be infeasible. In the limit as ε grows large, the cost is increasing in ε. Therefore, the full problem can be approximately solved via one-dimensional search over a bounded range of values.

Necessity of Robustness

Robust control is notoriously conservative, and our main result in Theorem 4.2.3 relies heavily on small gain-like arguments. Can the conservatism inherent in this approach be generally reduced? In this section, we answer in the negative by describing a class of examples for which the robustness condition in Theorem 4.2.3 is necessary. For simplicity, we specialize to the goal of regulating a system to the origin, rather than an arbitrary

waypoint tracking problem. Let $z_0 = q(0)$ and assume that $(0, z_0)$ is in the training set and that the perception map is perfect at the origin: $p(z_0) = 0$.

We consider the following optimal control problem

$$\text{minimize}_{\mathbf{\Phi}} \quad c(\mathbf{\Phi}) \;\text{ subject to }\; \mathbf{\Phi} \in \mathcal{A}(A, B, C), \;\; \|\mathbf{\Phi_{xv}}\| \leq \alpha \; .$$

and define $\bar{\mathbf{\Phi}}$ as a minimizing argument in the absence of the inequality constraint. For simplicity, we consider only systems in which the closed-loop is strictly proper and has a state-space realization.

To avoid notational collisions while discussing frequency domain quantities, we use ζ as the frequency domain variable (rather than z, as introduced in Chapter 2).

Proposition 4.3.2. *Suppose that the frequency response at $\zeta = 1$ satisfies*

$$\|\bar{\mathbf{\Phi}}_{\mathbf{xv}}\|_{\mathcal{L}_1} = \|\bar{\mathbf{\Phi}}_{\mathbf{xv}}(1)\|_{\infty\to\infty} \; .$$

Then there exists a differentiable error function with slope bound S such that the origin $x = 0$ is an asymptotically stable equilibrium if and only if $\alpha < \frac{1}{S}$.

Proof. Sufficiency follows from our main analysis, or alternatively from a simple application of the small gain theorem. We therefore focus on showing necessity, which follows by construction. We use a combination of classic nonlinear instability arguments with properties of real stability radii.

Recall that $y_t = Cx_t + (p(q(x_t)) - Cx_t)$. Define ξ_t to represent the internal state of the controller. Then we can write the closed-loop behavior of the system, ignoring the effect of process noise because it does not affect the stability analysis, as the following nonlinear recursion:

$$\begin{bmatrix} x_{t+1} \\ \xi_{t+1} \end{bmatrix} = A_{\mathrm{CL}} \begin{bmatrix} x_t \\ \xi_t \end{bmatrix} + B_{\mathrm{CL}} e(x_t) \; .$$

Suppose that the error function $e(x) = p(q(x)) - Cx$ is differentiable at $x = 0$ with derivative J. Then the stability of the origin depends on the linearized closed-loop system $A_{\mathrm{CL}}(J) := A_{\mathrm{CL}} + B_{\mathrm{CL}} J C_{\mathrm{CL}}$ where $C_{\mathrm{CL}} = [I \; 0]$ picks out the relevant component of the closed-loop state. If any eigenvalues of $A_{\mathrm{CL}}(J)$ lie outside of the unit disk, then $x = 0$ is not an asymptotically stable equilibrium.

Switching gears, we return to the system response variables to construct J that results in a lack of stability. We set $S = 1/\|\bar{\mathbf{\Phi}}_{\mathbf{xv}}\|$ and notice that if $\alpha \geq \frac{1}{S}$, then $\bar{\mathbf{\Phi}}$ is also the solution to the constrained problem. By assumption, $\|\bar{\mathbf{\Phi}}_{\mathbf{xv}}\|_{\mathcal{L}_1} = \|\bar{\mathbf{\Phi}}_{\mathbf{xv}}(1)\|_{\infty\to\infty}$. Thus, there exist real w, v such that $v^\top \bar{\mathbf{\Phi}}_{\mathbf{xv}}(1) w = \|\bar{\mathbf{\Phi}}_{\mathbf{xv}}\|_{\mathcal{L}_1}$. We set $J = Swv^\top$.

We now argue that for this choice of J, $A_{\mathrm{CL}}(J)$ has an eigenvalue on the unit disk. Recall that the frequency response $\bar{\mathbf{\Phi}}_{\mathbf{xv}}(\zeta) = C_{\mathrm{CL}}(\zeta I - A_{\mathrm{CL}})^{-1} B_{\mathrm{CL}}$, so

$$\bar{\mathbf{\Phi}}_{\mathbf{xv}}(1) = C_{\mathrm{CL}}(I - A_{\mathrm{CL}})^{-1} B_{\mathrm{CL}}$$

This makes the connection between $\bar{\mathbf{\Phi}}_{\mathbf{xv}}(1)$ and the matrices making up $A_{\mathrm{CL}}(J)$. Then by a classic result for stability radii (e.g. Remark 4.2 and Corollary 4.5 in Hinrichsen and Pritchard [HP90]), $A_{\mathrm{CL}}(Swv^\top)$ has an eigenvalue on the unit disk. Thus, for any error function with derivative $J = Swv^\top$ at zero, the robust condition is necessary as well as sufficient. One such error function is simply $e(x) = Jx$. □

We now present a simple example in which this condition is satisfied, and construct the corresponding error function which results in an unstable system.

Example 4.2. Consider the double integrator

$$x_{t+1} = \begin{bmatrix} 1 & dt \\ 0 & 1 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_t, \quad y_t = p(z_t) = x_t + e(x_t)\,, \tag{4.3.1}$$

and the control law resulting from the $\mathcal{H}_2$ optimal control problem with identity weight matrices along with the robustness constraint $\|\mathbf{\Phi}_{\mathbf{xv}}\| \leq \alpha$. Notice that the solution to the unconstrained problem would be the familiar LQG combination of Kalman filtering and static feedback on the estimated state. We denote the optimal unconstrained system response by $\bar{\mathbf{\Phi}}$.

Consider an error function with slope bounded by $S = 1/\|\bar{\mathbf{\Phi}}_{\mathbf{xv}}\|$ and derivative at $x = 0$ equal to

$$J = \begin{bmatrix} -1/\|\bar{\mathbf{\Phi}}_{\mathbf{xv}}\| & 0 \\ -1/\|\bar{\mathbf{\Phi}}_{\mathbf{xv}}\| & 0 \end{bmatrix},$$

Calculations confirm that the $\mathcal{L}_1$ norm satisfies the relevant property. Therefore, the origin is not a stable fixed point if the synthesis constraint $\alpha \geq \frac{1}{S}$.

4.4 Bounded Errors via Uniform Convergence

We now turn instead to perception maps learned via nonparametric regression, and show how sufficiently dense data implies uniformly bounded errors. We make several further assumptions to the problem setting introduced in Section 4.1. First, we assume that (A, B, C) is output controllable, meaning that for any initial condition $x_0 \in \mathbb{R}^n$ and arbitrary $y \in \mathbb{R}^d$, there exists a sequence of inputs $u_0, \ldots, u_{n-1}$ that, when applied to the noiseless system, ensure $Cx_n = y$. We also assume that for each t, the process noise w_t is zero mean, bounded by σ_w, and independent across time. Next, we assume that that observations depend on the state only through Cx, so that

$$z_t = q(Cx_t)\,.$$

The matrix C therefore determines the state subspace that affects the observation, which we will refer to as the *measurement subspace*. This emulates a natural setting in which observations from the system (e.g. camera images) are complex but encode a subset of the

state (e.g. position). We further assume that q is continuous and that there is a continuous inverse function $p_\star : \mathbb{R}^N \to \mathbb{R}^d$ with $p_\star(q(y)) = y$. For the example of a dashboard mounted camera, such an inverse exists whenever each camera pose corresponds to a unique image, which is a reasonable assumption in sufficiently feature rich environments.

We pose a learning problem focused on the unknown inverse function $p_\star$. We suppose that during the training phase, there is an additional system output,

$$y_t^{\text{train}} = Cx_t + v_t \tag{4.4.1}$$

where for each t, the noise v_t is zero mean and has independent entries bounded by σ_v. This assumption corresponds to using a simple but noisy sensor for characterizing the complex sensor. The noisy system output will both supervise the learning problem and allow for the execution of a sampling scheme where the system is driven to sample particular parts of the state space.

We suppose that the control cost c is given by the system $\mathcal{L}_1$ norm. Notice that due to its noisiness, using the training sensor would be suboptimal compared using transformed observations.

Nonparametric Regression

Since uniform error bounds are necessary for robust guarantees, we now introduce a method to learn perception maps with such bounds. For simplicity of analysis and exposition, we focus on Nadarya-Watson estimators. We expect that our insights will generalize to more complex techniques, and we demonstrate similarities with additional regressors in simulation experiments presented in Section 4.6.

The Nadarya-Watson regression estimators with training data $\{(z_t, y_t^{\text{train}})\}_{t=0}^T$, bandwidth $\gamma \in \mathbb{R}_+$, and metric $d_z : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}_+$ have the form

$$p(z) = \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} y_t^{\text{train}}, \quad s_T(z) = \sum_{t=0}^{T} \kappa_\gamma(z_t, z), \quad \kappa_\gamma(z_t, z) = \kappa\left(\tfrac{d_z(z_t,z)}{\gamma}\right), \tag{4.4.2}$$

with $p(z) = 0$ when $s_T(z) = 0$ and $\kappa : \mathbb{R}_+ \to [0, 1]$ is a kernel function. We assume that the kernel function is Lipschitz with parameter L_κ and that $\kappa(u) = 0$ for $u > 1$, and define the quantity $V_\kappa = \int_{\mathbb{R}_+^p} \kappa\left(\|y\|_\infty\right) dy$.

Thus, predictions are made by computing a weighted average over the labels y_t^{train} of training data points whose corresponding observations z_t are close to the current observation, as measured by the metric d_z. We assume the functions q and $p_\star$ are Lipschitz continuous with respect to d_z, i.e. for some L_q and L_p

$$d_z(q(y), q(y')) \le L_q \|y - y'\|, \quad \|p_\star(z) - p_\star(z')\| \le L_p d_z(z, z'). \tag{4.4.3}$$

While our final sub-optimality results depend on L_q and L_p, the perception map and synthesized controller do not need direct knowledge of these parameters.

For an arbitrary z with $s_T(z) \neq 0$, the prediction error can be decomposed as

$$\|p(z) - p_\star(z)\| \leq \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} (Cx_t - Cx) \right\| + \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} v_t \right\|. \tag{4.4.4}$$

The first term is the approximation error due to finite sampling, even in the absence of noisy labels. This term can be bounded using the continuity of the true perception map $p_\star$. The second term is the error due to measurement noise. We use this decomposition to state a pointwise error bound, which can be used to provide tight data-dependent estimates on error.

Lemma 4.4.1. *For a learned perception map of the form* (4.4.2) *with training data as in* (4.4.1) *collected during closed-loop operation of a system with appearance map satisfying* (4.4.3)*, we have with probability at least* $1 - \delta$ *that for a fixed* z *with* $s_T(z) \neq 0$*,*

$$\|p(z) - p_\star(z)\| \leq \gamma L_p + \frac{\sigma_v}{\sqrt{s_T(z)}} \sqrt{\log\left(d^2 \sqrt{s_T(z)}/\delta\right)}. \tag{4.4.5}$$

We present the proof of this result in the Section 4.8.

The expression illustrates that there is a tension between having a small bandwidth γ and ensuring that the coverage term $s_T(z)$ is large. Notice that most of the quantities in this upper bound can be readily computed from the training data; only L_p, which quantifies the continuity of the map from observation to state, is difficult to measure. We remark that while useful for building intuition, the result in Lemma 4.4.1 is only directly applicable for bounding error at a finite number of points. Since our main results handle stability over infinite horizons, they rely on a modified bound introduced in Section 4.8 which is closely tied to continuity properties of the estimated perception map p and the sampling scheme we propose in the next section.

Dense Sampling

We now propose a method for collecting training data and show a uniform, sample-independent bound on perception errors under the proposed scheme. This strategy relies on the structure imposed by the continuous and bijective map q, which ensures that driving the system along a dense trajectory in the measurement subspace corresponds to collecting dense samples from the space of possible observations. In what follows, we provide a method for driving the system along such a trajectory.

We assume that during training, the system state can be reset according to a distribution $\mathcal{D}_0$ which has has support bounded by σ_0. We do not assume that these states are observed. Between resets, an affine control law drives the system to evenly sample the measurement subspace with a combination of a stabilizing output feedback controller and affine inputs:

$$u_t = \mathbf{K}_{\text{train}}(y_{0:t}^{\text{train}}) + u_t^{\text{ref}}. \tag{4.4.6}$$

Algorithm 2 Uniform Sampling with Resets

1: Input system matrices B and C, stabilizing controller $\mathbf{K}_{\text{train}}$ and corresponding system response $\mathbf{\Phi}_{\mathbf{xw}}^{\text{train}}$, sampling radius $\bar{\Gamma}$, target dataset size T.
2: **for** ℓ from 1 to T **do**
3: reset $x_{0,\ell} \sim \mathcal{D}_0$ and sample $y_\ell^{\text{ref}} \sim \text{Unif}(B_{\bar{\Gamma}}(0_d))$
4: design inputs $u_{0:n-1,\ell}^{\text{ref}} := (C\mathbf{\Phi}_{\mathbf{xw}}^{\text{train}}B)[1:n]^\dagger y_\ell^{\text{ref}}$
5: **for** k from 0 to $n-1$ **do**
6: apply $u_{k,\ell} = \mathbf{K}_{\text{train}}(y_{0:k,t}^{\text{train}}) + u_{k,\ell}^{\text{ref}}$
7: **end for**
8: **end for**
9: Return uniformly sampled training data $\{(z_{n,\ell}, y_{n,\ell}^{\text{train}})\}_{\ell=1}^T =: \{(z_t, y_t^{\text{train}})\}_{t=1}^T$

The stabilizing feedback controller prevents the accumulation of errors resulting from the process noise and the unobserved reset state. Recalling the SLS machinery developed in Chapter 2, the closed-loop trajectories resulting from this controller can be written as linear functions of the reference input u_t^{ref} and the noise w_t, v_t in terms of the system response variables $\left(\mathbf{\Phi}_{\mathbf{xw}}^{\text{train}}, \mathbf{\Phi}_{\mathbf{xv}}^{\text{train}}, \mathbf{\Phi}_{\mathbf{uw}}^{\text{train}}, \mathbf{\Phi}_{\mathbf{uv}}^{\text{train}}\right)$, so long as $\mathbf{K}$ is stabilizing. Defining $\tilde{w}_{-1} = x_0$ and $\tilde{w}_t = Hw_t + Bu_t^{\text{ref}}$ for $t \geq 0$, we can write the trajectory of the system through the measurement subspace as

$$C\mathbf{x} = C\mathbf{\Phi}_{\mathbf{xw}}^{\text{train}}\tilde{\mathbf{w}} + C\mathbf{\Phi}_{\mathbf{xv}}^{\text{train}}\mathbf{v}$$

Because the feedback control law $\mathbf{K}$ is chosen such that the closed-loop system is stable, the system response variables decay. We therefore assume that for some $M \geq 1$ and $0 \leq \rho < 1$, the projection of the system response elements into the measurement subspace satisfies the decay condition

$$C\mathbf{\Phi}_{\mathbf{xw}}^{\text{train}}, C\mathbf{\Phi}_{\mathbf{xv}}^{\text{train}} \in \mathcal{RH}_\infty(M, \rho)\,. \tag{4.4.7}$$

We are then free to design u_t^{ref} to ensure even sampling. Since the triple (A, B, C) is output controllable, these reference inputs can drive the system to any state within n steps. Algorithm 2 leverages this fact to construct control sequences which drive the system to points uniformly sampled from the measurement subspace. The use of system resets ensures independent samples; we note that since the closed-loop system is stable, such a "reset" can approximately be achieved by waiting long enough with zero control input.

As a result of the unobserved reset states, process noise, and the noisy sensor, the states visited while executing Algorithm 2 do not exactly follow the desired uniform distribution. They can be decomposed into two terms:

$$Cx_{n,\ell} = \underbrace{\sum_{k=1}^{n} C\Phi_{xw}(k)Bu_{n-k,\ell}^{\text{ref}}}_{y_\ell^{\text{ref}}} + \underbrace{C\Phi_{xw}(n+1)x_0 + \sum_{k=1}^{n} C\Phi_{xw}(k)Hw_{n-k,\ell} + C\Phi_{xn}(k)v_{n-k,\ell}}_{\eta_\ell}$$

where y_ℓ^{ref} is uniformly sampled from $B_{\bar{\Gamma}}(0_d)$, and the noise variable η_ℓ is bounded:

$$\begin{aligned}\|\eta_\ell\| &\le \|C\Phi_{xw}(\ell+1)\|\|x_0\| + \sum_{\ell=1}^{n} \|C\Phi_{xw}(\ell)\|\|Hw_{n-\ell}\| + \|C\Phi_{xn}(\ell)\|\|v_{n-\ell}\| \\ &\le M\big(\sigma_0\rho^{n+1} + (\sigma_w+\sigma_v)\sum_{\ell=1}^{n}\rho^\ell\big) \le \frac{M\max\{\sigma_0,\sigma_w,\sigma_v\}}{1-\rho}\,.\end{aligned}$$

The following lemma shows that uniform samples corrupted with independent and bounded noise ensure dense sampling of the measurement subspace. In particular, the lemma provides a high probability lower bound on the coverage $s_T(z)$.

Lemma 4.4.2. *Suppose training data satisfying* (4.4.1) *is collected by the sampling scheme in Algorithm 2 with radius* $\bar{\Gamma} \ge \Gamma + \frac{M\max\{\sigma_0,\sigma_w,\sigma_v\}}{1-\rho} + \frac{\gamma}{L_q}$ *and a stabilizing controller whose system response variables satisfy* (4.4.7)*. If the appearance map satisfies continuity properties* (4.4.3)*, then for all* z *observed from a state* x *satisfying* $\|Cx\|_\infty \le \Gamma$*,*

$$s_T(z) \ge \frac{1}{2}\sqrt{TV_\kappa}\left(\frac{\gamma}{\bar{\Gamma}L_q}\right)^{\frac{d}{2}}$$

with probability at least $1-\delta$ *as long as* $T \ge 8V_\kappa^{-1}\log(1/\delta)(\bar{\Gamma}L_pL_q^2)^d\gamma^{-d}$*.*

We use this coverage property of the training data and the error decomposition presented in (4.4.4) to show our main uniform convergence result.

Theorem 4.4.3. *Suppose training data satisfying* (4.4.1) *is collected by the sampling scheme in Algorithm 2 with radius* $\bar{\Gamma} = \sqrt{2}\Gamma$ *and a stabilizing controller whose system response variables satisfy* (4.4.7)*. If the appearance map satisfies continuity properties* (4.4.3)*, then as long as the state remains within the set* $\{x \mid \|Cx\| \le \Gamma\}$*, the Nadarya-Watson regressor* (4.4.2) *will have bounded perception error for every observation* z*:*

$$\|p(z) - p_\star(z)\| \le \gamma L_p + \frac{\sigma_v}{T^{\frac{1}{4}}}\left(\frac{L_q\sqrt{2}\Gamma}{\gamma}\right)^{\frac{d}{4}}\left(\sqrt{d\log(T^2/\delta)}+1\right) \tag{4.4.8}$$

with probability at least $1-\delta$ *as long as* $\gamma \le L_q((\sqrt{2}-1)\Gamma - M\max\{\sigma_0,\sigma_w,\sigma_v\}(1-\rho)^{-1})$ *and*

$$T \ge \max\left\{8dV_\kappa^{-1}(\sqrt{2}L_pL_q^2)^d(\Gamma/\gamma)^d\log(T^2/\delta),\ \ V_\kappa^{-\frac{1}{3}}(24L_\kappa L_p)^{\frac{4}{3}}L_q^{\frac{d}{3}}(\Gamma/\gamma)^{\frac{d+4}{3}}\right\}\,.$$

Proofs are presented in Section 4.8.

4.5 Certainty Equivalent Perception-Based Control

The previous section shows that nonparametric regression can be successful for learning the perception map within a bounded region of the state space. We now show how to translate these bounded errors into closed-loop performance guarantees.

Suppose that we apply a linear controller to our perception estimates,

$$\mathbf{u} = \widehat{\pi}(\mathbf{z}) = \mathbf{K}p(\mathbf{z}) \,. \tag{4.5.1}$$

This is the certainty equivalent controller, which treats the learned perception map as if it is true. We will compare this controller with $\pi_\star(\mathbf{z}) = \mathbf{K}p_\star(\mathbf{z})$, the result of perfect perception. We suppose that $\|C\mathbf{x}\|$ is bounded by $\Gamma_{\max}(\mathbf{\Phi})$ under the optimal control law for all possible disturbance signals.

Proposition 4.5.1. *Let* $(\mathbf{\Phi_{xw}}, \mathbf{\Phi_{xv}}, \mathbf{\Phi_{uw}}, \mathbf{\Phi_{uv}})$ *denote the system responses induced by the controller* $\mathbf{K}$*, and let* $c(\pi_\star)$ *denote the cost associated with the policy* $\pi_\star(\mathbf{z}) = \mathbf{K}p_\star(\mathbf{z})$*. Then for a perception component with error bounded by* ε_p *within the set* $\{x \mid \|Cx\|_\infty \leq \Gamma\}$*, the sub-optimality of the certainty equivalent controller* (4.5.1) *is bounded by*

$$c(\widehat{\pi}) - c(\pi_\star) \leq \varepsilon_p \left\| \begin{bmatrix} Q^{1/2}\mathbf{\Phi_{xv}} \\ R^{1/2}\mathbf{\Phi_{uv}} \end{bmatrix} \right\| .$$

as long as the sampled region is large enough and the errors are small enough, $\varepsilon_p \leq \frac{\Gamma - \Gamma_{\max}(\mathbf{\Phi})}{\|C\mathbf{\Phi_{xv}}\|}$.

Proof. The main difference between the certainty equivalent and optimal closed-loop systems is the measurement noise signal. For the certainty equivalent closed-loop, we have the perception errors $\mathbf{e}$. For the optimal closed-loop, there is no measurement noise.

Therefore, we write the cost:

$$\begin{aligned}
c(\widehat{\pi}) &= \sup_{\mathbf{w}} \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} \begin{bmatrix} H\mathbf{w} \\ \mathbf{e} \end{bmatrix} \right\| \\
&\leq \sup_{\mathbf{w}} \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xw}} & \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uw}} & \mathbf{\Phi_{uv}} \end{bmatrix} \begin{bmatrix} H\mathbf{w} \\ 0 \end{bmatrix} \right\| + \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uv}} \end{bmatrix} \mathbf{e} \right\| \\
&\leq c(\pi_\star) + \sup_{\mathbf{w}} \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uv}} \end{bmatrix} \mathbf{e} \right\| \\
&\leq c(\pi_\star) + \left\| \begin{bmatrix} Q^{1/2} & \\ & R^{1/2} \end{bmatrix} \begin{bmatrix} \mathbf{\Phi_{xv}} \\ \mathbf{\Phi_{uv}} \end{bmatrix} \right\| \cdot \sup_{\mathbf{w}} \|p(\mathbf{z}) - p_\star(\mathbf{z})\|
\end{aligned}$$

Recall the uniform error bound on the learned perception map. As long as $\|C\mathbf{x}\| \leq \Gamma$, we can guarantee that $\|p(\mathbf{z}) - p_\star(\mathbf{z})\| \leq \varepsilon_p$. Notice that we have

$$\|C\mathbf{x}\| \leq \sup_{\mathbf{w}} \|C\mathbf{\Phi_{xw}}H\mathbf{w} + C\mathbf{\Phi_{xv}}\mathbf{e}\| \leq \Gamma_{\max}(\mathbf{\Phi}) + \varepsilon_p \|C\mathbf{\Phi_{xv}}\| \,,$$

by the definition of $\Gamma_{\max}(\mathbf{\Phi})$. Therefore, the result follows as long as $\varepsilon_p \leq \frac{\Gamma - \Gamma_{\max}(\mathbf{\Phi})}{\|C\mathbf{\Phi_{xv}}\|}$. □

Thus, the optimal closed-loop system's sensitivity to measurement noise is closely related to the sub-optimality. We now state our main result. The proof is presented in Section 4.8.

Corollary 4.5.2. *Suppose that training data satisfying* (4.4.1) *is collected with a stabilizing controller whose system response variables satisfy* (4.4.7) *according to Algorithm 2 with* $\bar{\Gamma} = 2\Gamma_{\max}(\mathbf{\Phi}) \geq \max\{1, M\frac{\max\{\sigma_0,\sigma_w,\sigma_v\}}{1-\rho}\}$ *from a system with appearance map satisfying* (4.4.3)*, and that the Nadarya-Watson regressor* (4.4.2) *uses bandwidth* γ *chosen to minimize the upper bound in* (4.4.8)*. Then the overall sub-optimality of the certainty equivalent controller* (4.5.1) *is bounded by*

$$c(\widehat{\pi}) - c(\pi_\star) \leq 4L_q L_p \Gamma_{\max}(\mathbf{\Phi}) \left(\frac{4d^2\sigma_v^4}{T}\right)^{\frac{1}{d+4}} \left\| \begin{bmatrix} Q^{1/2}\mathbf{\Phi_{xv}} \\ R^{1/2}\mathbf{\Phi_{uv}} \end{bmatrix} \right\| \sqrt{\log(T^2/\delta)} \tag{4.5.2}$$

with probability greater than $1-\delta$ *for large enough* $T \geq 4d^2\sigma_v^4 \left(10L_qL_p\|C\mathbf{\Phi_{xv}}\|\sqrt{\log(T^2/\delta)}\right)^{d+4}$.

4.6 Experiments

To illustrate our results and to probe their limits, we perform experiments in three simulated environments: a synthetic example, a simplified unmanned aerial vehicle (UAV) with a downward facing camera, and an autonomous driving example with a dashboard mounted camera. The latter two settings make use of complex graphics simulation. Code necessary for reproducing experiments is available at:

- `https://github.com/modestyachts/robust-control-from-vision`
- `https://github.com/modestyachts/certainty_equiv_perception_control`

In all cases, system dynamics are defined using a hovercraft model, where positions along east-west and north-south axes evolve independently according to double integrator dynamics. The hovercraft model is specified by

$$A = \begin{bmatrix} 1 & 0.1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{4.6.1}$$

so that $x^{(1)}$ and $x^{(3)}$ respectively represent position along east and north axes, while $x^{(2)}$ and $x^{(4)}$ represent the corresponding velocities.

Figure 4.2: In (a), visual inputs $\{z_t\}$ for the synthetic (left) and CARLA (right) examples. In (b), the nominal trajectory the synthetic circle (left) and simulated vehicle (right) are driven to follow.

Robust Controllers

We begin with an evaluation of robust control methods. Here, we compare a synthetic example to a complex autonomous driving example. The synthetic example uses generated 64×64 pixel images of a moving blurry white circle on a black background; the complex example uses 800×600 pixel dashboard camera images of a vehicle in the CARLA simulator platform [DRCLK]. Figure 4.2(a) shows representative images seen by the controllers.

Training and validation trajectories are generated by driving the system with an optimal state feedback controller (i.e. where measurement $\mathbf{y} = \mathbf{x}$) to track a desired reference trajectory $\mathbf{r} + \boldsymbol{\eta}$, where $\mathbf{r}$ is a nominal reference, and $\boldsymbol{\eta}$ is a random norm bounded random perturbation satisfying $\|\eta_t\| \leq 0.1$.

We consider three different perception maps: a linear map for the synthetic example and both visual odometry and a convolutional neural net for the CARLA example. For the CNN, we collect a dense set of training examples around the reference to train the model. We use the approach proposed by Coates and Ng [CN12] to learn a convolutional representation of each image: each resized and scaled image is passed through a single convolutional, ReLU activation, and max pooling layer. We then fit a linear map of these learned image features to position and heading of the camera pose. We require approximately $30,000$ training points. During operation, pixel-data z is passed through the CNN, and the position estimates y are used by the controller. We note that other more sophisticated architectures for feature extraction would also be reasonable to use in our control framework; we find that this one is conceptually simple and sufficient for tracking around our fixed reference.

To perform visual odometry, we first collect images from known poses around the desired reference trajectory. We then use ORB SLAM [MMT15] to build a global map of visual features and a database of reference images with known poses. This is the "training" phase. We use one trajectory of 200 points; the reference database is approximately this size. During operation, an image z is matched with an image z_i in the database. The re-projection error between the matched features in z_i with known pose x_i and their corresponding features in z is then minimized to generate a pose estimate. For more details on standard visual odometry methods, see [SF11]. We highlight that modern

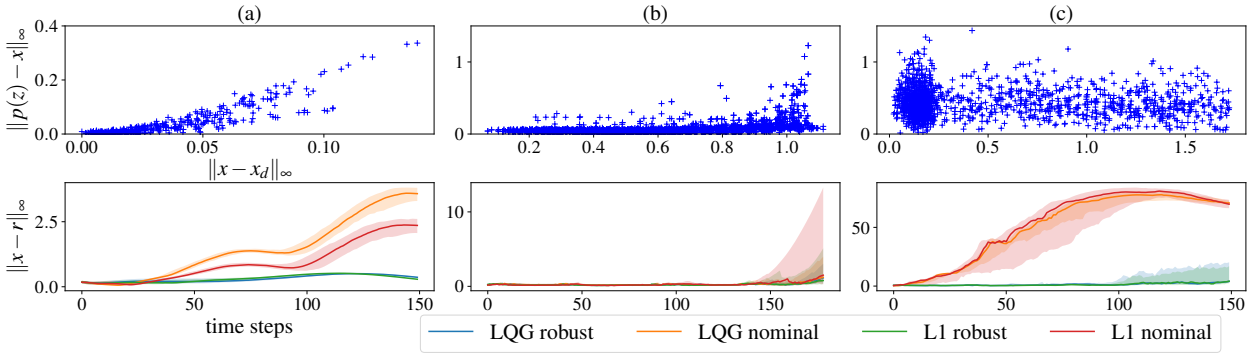

Figure 4.3: (top) Test perception error $\|p(z) - Cx\|$ vs. distance to the nearest training point $\|x - x_i\|$; and (bottom) median tracking error for 200 rollouts using the (a) linear map on synthetic images, (b) SLAM and (c) CNN on CARLA dashboard images. Error bars show upper and lower quartiles.

visual SLAM algorithms incorporate sophisticated filtering and optimization techniques for localization in previously unseen environments with complex dynamics; we use a simplified algorithm under this training and testing paradigm in order to better isolate the data dependence.

We then estimate safe regions for all three maps. In the top panels of Figure 4.3 we show the error profiles as a function of distance to the nearest training point for the linear (left), SLAM (middle), and CNN (right) maps. We see that these data-dependent localization schemes exhibit the local slope bounded property posited in Section 4.2.

We compare robust synthesis to the behavior of *nominal controllers* that do not take into account the nonlinearity in the measurement model. In particular, we compare the performance of naively synthesized LQG and $\mathcal{L}_1$ optimal controllers with controllers designed with the robustness condition (4.2.5). To make the synthesis computationally tractable, we take a finite approximation, restricting the system responses $(\mathbf{\Phi_{xw}}, \mathbf{\Phi_{xv}}, \mathbf{\Phi_{uw}}, \mathbf{\Phi_{uv}})$ to be finite impulse response (FIR) transfer matrices of length $T = 200$, i.e., we enforce that $\mathbf{\Phi}(T) = 0$. We then solve the resulting optimization problem with MOSEK under an academic license [ApS19]. For further details on formulating control synthesis problems, refer to Chapter 2.

The bottom panels of Figure 4.3 show that the robustly synthesized controllers remain within a bounded neighborhood around the training data. On the other hand, the nominal controllers drive the system away from the training data, leading to a failure of the perception and control loop. We note that although visual odometry may not satisfy smoothness assumptions when the feature detection and matching fails, we nevertheless observe safe system behavior, suggesting that using our robust controller, no such failures occur.

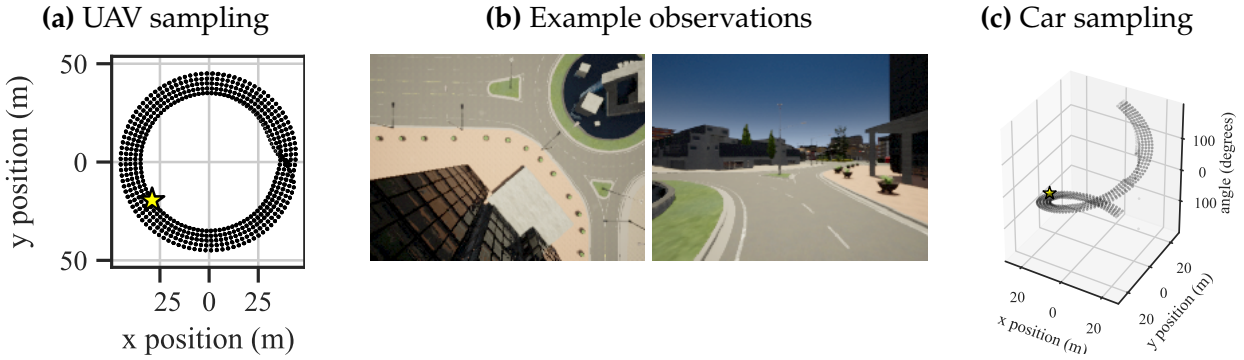

Figure 4.4: Coverage of training data for (a) UAV and (c) autonomous driving settings. In (b), example observations taken from positions indicated by yellow stars.

Certainty Equivalent Control

We now turn to an evaluation of certainty equivalent control. We compare the UAV and car environments. In both cases, observations are 300×400 pixel RGB images generated using the CARLA simulator [DRCLK]. For the UAV setting, we fix the height and orientation of the camera so that it faces downward from an elevation of 40m. For the autonomous driving setting, the height is fixed to ground level and the orientation is determined by the car's velocity. Figure 4.4b shows example observations.

For the UAV, the rendered image depends only on the position, with a scale factor of 20 to convert into meters in CARLA coordinates. For the car example, observations are determined as a function of vehicle pose, and thus additionally depend on the heading angle. We defined the heading to be $\arctan(x^{(4)}/x^{(2)})$, so the rendered image depends on the position as well as velocity state variables. For the driving setting, the scale factor is 11.

We construct training trajectories by applying a static reference tracking controller to trace circles of varying radius. For the training phase, the noisy sensor as gives rise to measurements $y_t^{\text{train}} = Cx_t + v_t$ where v_t is generated by clipping a Gaussian with standard deviation 0.01 between -1 and 1. We use the static reference tracking controller:

$$u_t = K(\widehat{x}_t - C^{\dagger} y_t^{\text{ref}}),$$

where the state estimate is computed as in (2.2.6). The feedback parameter K is generated as the optimal LQR controller when costs are specified as $Q = C^{\top}C$ and $R = I$ while the estimation parameter L is generated as the optimal infinite horizon Kalman filter for process noise with covariance $W = I$ and measurement noise with covariance $V = 0.1 \cdot I$. We use the periodic reference

$$y_t^{\text{ref}} = \begin{bmatrix} a_t \sin(2\pi t/100) & a_t \cos(2\pi t/100) \end{bmatrix}^{\top}, \quad a_t = 1.75 + 0.125\,(\lfloor t/100 \rfloor \mod 4)$$

which traces circles counterclockwise, ranging in radius from 1.75 to 2.25, or 35-45m for the UAV and 19.25-24.75m for the car. We collect training data for $0 \le t \le T =$

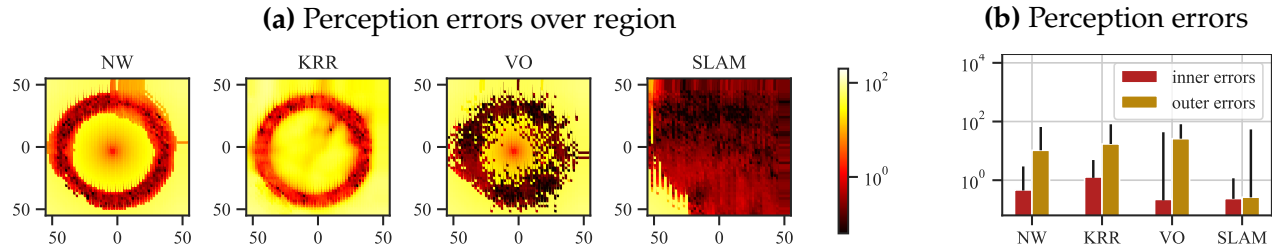

Figure 4.5: In (a), heat maps illustrate perception errors. In (b), median and 99th percentile errors within the inner (37-42m radius) and outer (25-55m radius, excluding inner) regions of training data.

2,000. Figure 4.4a and 4.4c display the positions from which training observations and measurements are collected. Notice that for the car, this strategy results in a sparsely sampled measurement subspace.

We consider four types of perception predictors:

- **Nadarya-Watson (NW):** The estimated position is computed based on training data as in (4.4.2) where $d_z(z, z_t) = \|z - z_t\|_2$ is the ℓ_2 distance on raw pixels. The only hyperparameter is the bandwidth γ.

  We investigated a few additional metrics based on ℓ_2 distance between *transformed pixels*, but did not observe any benefits. The transformations we considered were: pretrained Resnet-18 features,[1] Gaussian blur,[2] and edge detection-type filters.[3]

- **Kernel Ridge Regression (KRR):** The estimated position is computed as

$$p(z) = \begin{bmatrix}(y_0^{\text{train}})^\top & \dots & (y_T^{\text{train}})^\top\end{bmatrix} (\lambda I + K)^{-1} \begin{bmatrix}k(z_0, z) & \dots & k(z_T, z)\end{bmatrix}^\top$$

  where the kernel matrix K is defined from the training data with $K_{ij} = k(z_i, z_j)$ and we use the radial basis function kernel, $k(z, z') = e^{-\alpha\|z-z'\|_2^2}$. The hyperparameters are regularization λ and spread α.

- **Visual Odometry (VO):** This structured method relies on a database of training images with known poses, which we construct using ORB-SLAM [MMT15], and calibrate to the world frame using position measurements $\{y_t^{\text{train}}\}$ which determine scale and orientation. We use only the first 200 training data points to initialize this database.

  New observations z are matched with an image z_t in the training data based on "bag of words" image descriptor comparison. Then, the camera pose is estimated

[1] `https://github.com/christiansafka/img2vec`

[2] `https://scikit-image.org/docs/stable/api/skimage.filters.html#skimage.filters.gaussian`

[3] `https://scikit-image.org/docs/stable/api/skimage.filters.html`

to minimize the re-projection error between matched features in z and z_t, and from this we extract the position estimate $p(z)$.

- **Simultaneous Localization and Mapping (SLAM):** This structured method proceeds very similarly to the VO method described above, with two key differences. First, all new observations z are added to the database along with training images. Second, pose estimates are initialized based on the previous frame, resulting in a temporal component. We implement this method by running ORB-SLAM online.

These additional methods allow for comparison with a classical nonparametric approach, a classical computer vision approach, and a non-static state-of-the-art approach.

We evaluate the learned perception maps on a grid of $2,500$ points, with $-2.5 \leq x^{(1)}, x^{(3)} \leq 2.5$. For the car, we set the angle of the grid points as $\arctan(-x^{(1)}/x^{(3)})$ to mimic counter-clockwise driving, which results in a best-case evaluation of the learned perception components. For SLAM evaluation, the ordering of each evaluated point matters. We visit the grid points by scanning vertically, alternating upwards and downwards traversal. For NW and KRR, we choose hyperparameters which result in low errors within the region covered by training samples. In the UAV setting, we used $\gamma = 25,000$, $\alpha = 10^{-9}$, and $\lambda = 0.111$. In the car setting, we used $\gamma = 16666.67$.

The resulting errors are displayed for the UAV in Figure 4.5a and for the car in Figure 4.6a. The error heat maps are relatively similar for the three static regressors, with small errors within the training data coverage and larger errors outside of it. Though VO has very small errors at many points, its heat map looks much noisier. The large errors come from failures of feature matching within a database of key frames from the training data; in contrast, NW and KRR predictions are closely related to ℓ_2 distance between pixels, leading to smoother errors. Because SLAM performs mapping online, it can leverage the unlabeled evaluation data to build out good perception away from training samples. In the UAV setting, SLAM has high errors only due to a tall building obstructing the camera view, visible in Figure 4.4b. Similarly, in the driving setting, a wall and a statue obstruct the view in the locations that SLAM exhibits large errors. Figure 4.5b and 4.6b summarize the evaluations by plotting the median and 99th percentile errors in the *inner region* of training coverage compared with the *outer region*.

Finally, we evaluate the closed-loop performance of the predictors by using them for a reference tracking task. We consider reference trajectories of the form

$$y_t^{\text{ref}} = \begin{bmatrix} a\sin(2\pi k/100) & 2\cos(2\pi k/100) \end{bmatrix}^\top \tag{4.6.2}$$

and a static reference tracking controller which has the same parameters as the one used to collect training data.

We first examine how changes to the reference trajectory lead to failures in Figure 4.7. For the UAV setting, NW with dense sampling is sufficient to allow for good tracking of the reference with $a = 1.9$, but tracking fails when trying to track a reference outside the sampled region with $a = 1.6$. As the reference signal drives the system into a sparsely

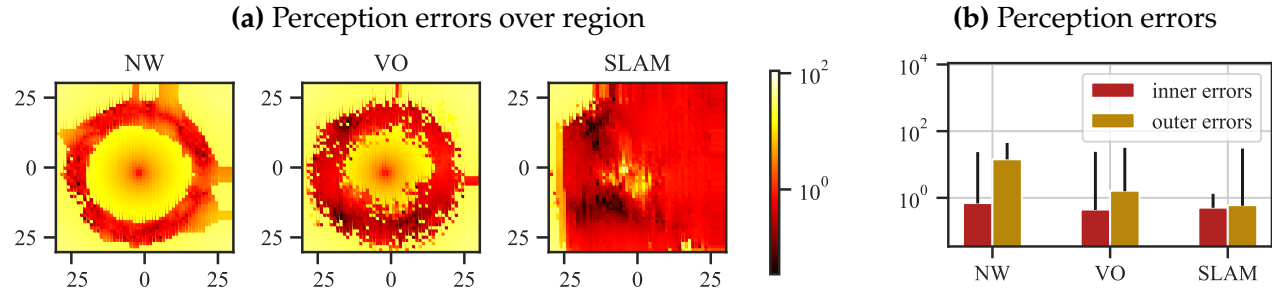

Figure 4.6: In (a), heat maps illustrate perception errors. In (b), median and 99th percentile errors within the inner (20.35-23.1m radius) and outer (13.75-30.25m radius, excluding inner) regions of training data.

sampled region, errors increase until eventually the UAV exits the region covered by training data. Once this occurs, the system loses stability due to the perception failure. Despite briefly regaining perception, the system does not recover.

The autonomous driving setting illustrates failure modes that arise for nonparametric predictors when training data is not dense. The trajectories in Figure 4.8 show that even though the reference is contained within the well-sampled region ($a = 1.9$), NW experiences failures due to neglecting to densely sample with respect to angle. Errors cause the system to deviate from the reference, and eventually the system exits the region covered by training data, losing stability due to the perception failure. SLAM does not have this problem.

4.7 Conclusion and Open Problems

Though standard practice is to treat the output of a perception module as an ordinary signal, this may only be justifiable in high-data and low-error settings. We have demonstrated both in theory and experimentally that accounting for the inherent uncertainty of perception-based sensors can improve the performance of the resulting control loop. We

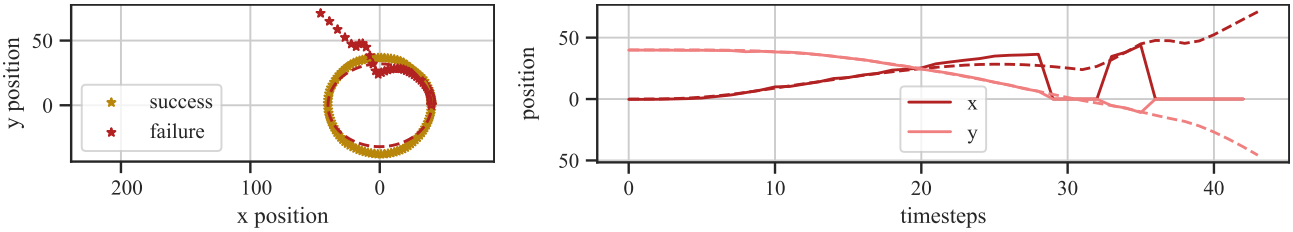

Figure 4.7: Two different reference trajectories for the UAV with NW perception lead to a success and a failure. Left, reference and actual trajectories. Right, predicted (solid) and actual (dashed) positions for the failed tracking example.

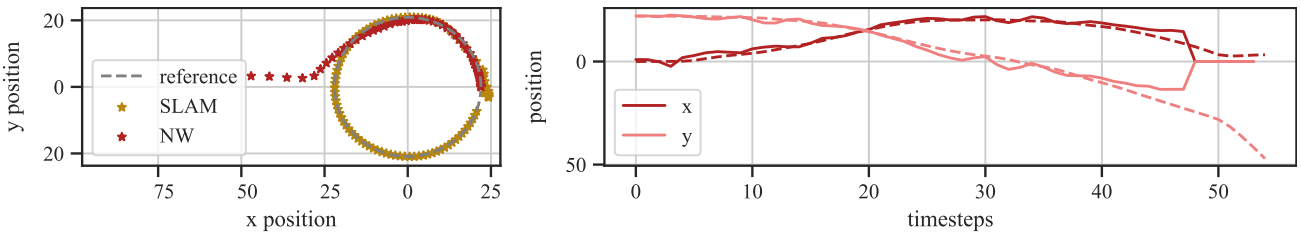

Figure 4.8: Two different predictors lead to a success and a failure of reference tracking for the car. Left, reference and actual trajectories for NW and SLAM. Right, predicted (solid) and actual (dashed) positions for NW.

developed a sample complexity bound for the case that the perception map is learned via nonparametric regression, and showed that using it in a control loop without accounting for its errors is valid for sufficiently dense training data. Our result depends exponentially on the dimension of the measurement space, highlighting that closed-loop stability requires onerous learning guarantees. The setting we consider incorporates both a lack of direct state observation and nonlinearity, making it relevant to modern robotic systems.

We hope that future work will continue to probe this problem setting to rigorously explore relevant trade-offs. One direction for future work is to contend with the sampling burden by collecting data in a more goal-oriented manner, perhaps with respect to a target task or the continuity of the observation map. It is also of interest to consider extensions which do not rely on the supervision of a noisy sensor or make clearer use of the structure induced by the dynamics on the observations. One path forward would be to assume that the training labels were the result of a state estimation process. Making closer connections with modern computer vision methods like SLAM could lead to insights about unsupervised and semi-supervised learning, particularly when data has known structure. Another direction is to relax the assumption that the dynamics are known a priori, making use of recent results for linear system identification.

We further hope to highlight the challenges involved in adapting learning-theoretic notions of generalization to the setting of controller synthesis. First note that if we collect data using one controller, and then use this data to build a new controller, there will be a distribution shift in the observations seen between the two controllers. Any statistical generalization bounds on performance must necessarily account for this shift. Second, from a more practical standpoint, most generalization bounds require knowing instance specific quantities governing properties of the class of functions we use to fit a predictor. Hence, they will include constants that are not measurable in practice. This issue can perhaps be mitigated using some sort of bootstrap technique for post-hoc validation. However, we note that the sort of bounds we aim to bootstrap are worst case, not average case. Indeed, the bootstrap typically does not even provide a consistent estimate of the maximum of independent random variables, see for instance [BF81], and Ch 9.3 in [Che11]. Other measures such as conditional value at risk require billions of samples to guarantee five 9s of reliability [RU+00]. We highlight these issues only to point out that adapting

statistical generalization to robust control is an active area with many open challenges to be considered in future work.

4.8 Omitted Proofs

In this section, we provide proofs and intermediate lemmas. To begin, recall the pointwise error bound presented in Lemma 4.4.1. Its proof relies on the following concentration result, which allows us to handle training data collected from the closed-loop system. Care is needed because u_t depends on $y_{0:t}$, and a result, $z_t = g(x_t)$ will depend on previous noise variables v_k for $k < t$.

Lemma 4.8.1 (adapted from Corollary 1 in [APS11])**.** *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration, let $(V_t)_{t\geq 0}$ be a real-valued stochastic process adapted to $(\mathcal{F}_t)$, and let $(W_t)_{t\geq 1}$ be a real-valued martingale difference process adapted to $(\mathcal{F}_t)$. Assume that $W_t \mid \mathcal{F}_{t-1}$ is conditionally sub-Gaussian with parameter σ. Then for all $T \geq 0$, with probability $1-\delta$,*

$$\left\| \sum_{k=1}^{T} V_t W_t \right\|_2^2 \leq 2\sigma^2 \log\left(\frac{\sqrt{\sum_{t=1}^{T} V_t^2}}{\delta} \right) \left(1 + \sum_{t=1}^{T} V_t^2 \right).$$

Proof of Lemma 4.4.1. Recall that the error can be bounded by an approximation term and a noise term:

$$\begin{aligned} \|p(z) - p_\star(z)\| &= \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} (Cx_t + v_t) - Cx \right\| \\ &\leq \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} (Cx_t - Cx) \right\| + \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} v_t \right\|. \end{aligned}$$

The approximation error is bounded by the weighted average value of $\|Cx_t - Cx\|$ for points in the training set that are close to the current observation. Using the continuity of the map p from observations to labels, we have that

$$\kappa_\gamma(z_t, z) > 0 \implies d_z(z_t, z) \leq \gamma \implies \|Cx_t - Cx\| \leq \gamma L_p .$$

This provides an upper bound on the average.

Turning to the measurement error term, we begin by noting that

$$P\left(\left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} v_t \right\| \geq s \right) \leq \sum_{i=1}^{d} P\left(\frac{1}{s_T(z)} \left| \sum_{t=0}^{T} \kappa_\gamma(z_t, z) v_{t,i} \right| \geq s \right). \tag{4.8.1}$$

We then apply Lemma 4.8.1 with $\mathcal{F}_t = x_{0:t}$, $V_t = \kappa_\gamma(z_t, z)$, and $W_t = v_{t,i}$. Note that $v_{t,i}$ is $\sigma_v/2$ sub-Gaussian due to the fact that it is bounded. Therefore, with probability at least $1 - \delta/d^2$,

$$\frac{1}{s_T(z)}\Big|\sum_{t=0}^{T}\kappa_\gamma(z_t,z)v_{t,i}\Big| \le \frac{\sigma_v}{2s_T(z)}\sqrt{2\log\left(\frac{d^2}{\delta}\sqrt{\sum_{k=0}^{T}\kappa_\gamma(z_t,z)^2}\right)(1+\sum_{k=0}^{T}\kappa_\gamma(z_t,z)^2)}$$
$$\overset{(a)}{\le} \frac{\sigma_v}{2}\sqrt{2\log\left(\frac{d^2}{\delta}\sqrt{s_T(z)}\right)}\frac{\sqrt{1+s_T(z)}}{s_T(z)} \overset{(b)}{\le} \sigma_v\sqrt{\frac{\log\left(d^2\sqrt{s_T(z)}/\delta\right)}{s_T(z)}},$$

where (a) holds since $\kappa_\gamma \le 1$, and (b) since $s_T(z) \ge 1$ implies that $\frac{\sqrt{1+s_T(z)}}{s_T(z)} \le \sqrt{\frac{2}{s_T(z)}}$. Then with probability $1-\delta$,

$$\Big\|\sum_{t=0}^{T}\frac{\kappa_\gamma(z_t-z)}{s_T(z)}v_t\Big\| \le \sigma_v\sqrt{\frac{\log\left(d^2\sqrt{s_T(z)}/\delta\right)}{s_T(z)}}. \tag{4.8.2}$$

□

Since Lemma 4.4.1 is only useful for bounding errors for a finite collection of points, we now prove a similar bound that holds over all points. While it is possible to use the following result in a data-driven way (with $z_i = z_t$ and $H = T$), we will use it primarily for our data independent bound.

Lemma 4.8.2. *Consider the setting of Lemma 4.4.1. Let $\{z_i\}_{i=1}^H$ be any arbitrary set of observations with $s_T(z_i) \neq 0$. Then for all z, the learned perception map has bounded errors with probability at least $1-\delta$:*

$$\|p(z) - p_\star(z)\| \le \gamma L_p + \min_i \frac{\sigma_v}{\sqrt{s_T(z_i)}}\sqrt{\log\left(d^2H\sqrt{s_T(z_i)}/\delta\right)} + \frac{2\sigma_v T}{s_T(z_i)}\frac{L_\kappa}{\gamma}d_z(z,z_i). \tag{4.8.3}$$

Proof. To derive a statement which holds for uncountably many points, we adapt the proof of Lemma 4.4.1. Previously, we used the fact that the error can be bounded by an approximation term and a noise term. Now, we additionally consider a continuity term

so that for any $1 \leq i \leq H$,

$$\begin{aligned}\|p(z) - p_\star(z)\| &\leq \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)}(Cx_t - Cx) \right\| + \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z_i)}{s_T(z_i)} v_t \right\| \\ &\qquad + \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} v_t - \frac{\kappa_\gamma(z_t, z_i)}{s_T(z_i)} v_t \right\| \\ &\leq \gamma L_p + \frac{\sigma_v}{\sqrt{s_T(z_i)}} \sqrt{\log\left(d^2 H \sqrt{s_T(z_i)}/\delta\right)} + \left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} v_t - \frac{\kappa_\gamma(z_t, z_i)}{s_T(z_i)} v_t \right\|.\end{aligned}$$

The final line holds for all z_i with probability at least $1 - H \cdot \delta/H$, following the logic of the proof of Lemma 4.4.1 along with a union bound. Now, consider the third term and the boundedness of the noise:

$$\left\| \sum_{t=0}^{T} \frac{\kappa_\gamma(z_t, z)}{s_T(z)} v_t - \frac{\kappa_\gamma(z_t, z_i)}{s_T(z_i)} v_t \right\| \leq \sigma_v \sum_{t=0}^{T} \left| \frac{\kappa_\gamma(z_t, z)}{s_T(z)} - \frac{\kappa_\gamma(z_t, z_i)}{s_T(z_i)} \right|.$$

Using the fact that each term in the expression is nonnegative,

$$\begin{aligned}\left| \frac{\kappa_\gamma(z_t, z)}{s_T(z)} - \frac{\kappa_\gamma(z_t, z_i)}{s_T(z_i)} \right| &= \frac{1}{s_T(z_i)} \left| \frac{s_T(z_i)}{s_T(z)} \kappa_\gamma(z_t, z) - \kappa_\gamma(z_t, z_i) \right| \\ &\leq \frac{1}{s_T(z_i)} \left(\left| \kappa_\gamma(z_t, z) - \kappa_\gamma(z_t, z_i) \right| + \left| \tfrac{s_T(z_i)}{s_T(z)} - 1 \right| \kappa_\gamma(z_t, z) \right)\end{aligned}$$

Using the smoothness of the kernel, we have that

$$|\kappa_\gamma(z_t, z) - \kappa_\gamma(z_t, z_i)| = \left| \kappa\left(\tfrac{d_z(z_t, z)}{\gamma}\right) - \kappa\left(\tfrac{d_z(z_t, z_i)}{\gamma}\right) \right| \leq L_\kappa \left| \tfrac{d_z(z_t, z) - d_z(z_t, z_i)}{\gamma} \right| \leq \frac{L_\kappa}{\gamma} d_z(z, z_i)$$

where the final inequality uses the fact that d_z is a metric. Additionally, we have that

$$\begin{aligned}\left| 1 - \tfrac{s_T(z_i)}{s_T(z)} \right| &= \frac{1}{s_T(z)} |s_T(z) - s_T(z_i)| \\ &\leq \frac{1}{s_T(z)} \sum_{t=0}^{T} |\kappa_\gamma(z_t, z) - \kappa_\gamma(z_t, z_i)| \\ &\leq \frac{1}{s_T(z)} \frac{L_\kappa}{\gamma} T d_z(z, z_i)\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{t=0}^{T}\left|\frac{\kappa_\gamma(z_t,z)}{s_T(z)}-\frac{\kappa_\gamma(z_t,z_i)}{s_T(z_i)}\right| &\le \sum_{t=0}^{T}\frac{1}{s_T(z_i)}\left(\frac{L_\kappa}{\gamma}+\frac{1}{s_T(z)}\frac{L_\kappa}{\gamma}T\kappa_\gamma(z_t,z_i)\right)d_z(z,z)\\ &= \frac{1}{s_T(z_i)}\left(T\frac{L_\kappa}{\gamma}+\frac{\sum_{t=0}^{T}\kappa_\gamma(z_t,z)}{s_T(z)}\frac{L_\kappa}{\gamma}T\right)d_z(z,z_i)\\ &= \frac{2T}{s_T(z_i)}\frac{L_\kappa}{\gamma}d_z(z,z_i)\end{aligned}$$

The result follows. □

Proof of Lemma 4.4.2. We begin by establishing some properties of the training data generated by Algorithm 2.

$$\begin{aligned}h(z_t) = h(z_{n,\ell}) &= Cx_{n,\ell}\\ &= \sum_{k=1}^{n} C\Phi_{xw}(k)Bu^{\text{ref}}_{n-k,\ell} + C\Phi_{xw}(n+1)x_0 + \sum_{k=1}^{n} C\Phi_{xw}(k)Hw_{n-k,\ell} + C\Phi_{xn}(k)\eta_{n-k,\ell}\\ &= \begin{bmatrix}C\Phi_{xw}(1)B & \dots & C\Phi_{xw}(n)\end{bmatrix} B \begin{bmatrix}C\Phi_{xw}(1)B & \dots & C\Phi_{xw}(n)B\end{bmatrix}^\dagger y^{\text{ref}}_\ell + \eta_\ell\\ &= y^{\text{ref}}_\ell + \eta_\ell\end{aligned}$$

where in the second line, we use the fact that (A, B, C) is output controllable. In this expression, y^{ref}_ℓ is sampled by Algorithm 2 uniformly from $\{y \mid \|y\| \le \bar{\Gamma}\}$ and

$$\begin{aligned}\|\eta_\ell\| &\le \|C\Phi_{xw}(\ell+1)\|\|x_0\| + \sum_{\ell=1}^{n}\|C\Phi_{xw}(\ell)\|\|Hw_{n-\ell}\| + \|C\Phi_{xn}(\ell)\|\|v_{n-\ell}\|\\ &\le M\big(\sigma_0\rho^{n+1} + (\sigma_w+\sigma_v)\sum_{\ell=1}^{n}\rho^\ell\big) \le \frac{M\sigma}{1-\rho}\,.\end{aligned}$$

where we let $\sigma = \max\{\sigma_0, \sigma_w, \sigma_v\}$

Now, we reason about the density of the training data over the measurement subspace. Above, we see that $Cx_{n,\ell} = Cx_t$ are generated as the sum of two independent random variables. The first is a uniform random variable which has density $f_{\text{Unif}}(y) = (2\bar{\Gamma})^{-d}\mathbf{1}\{\|y\| \le \bar{\Gamma}\}$. The second is a bounded random variable, so the support of its density $f_\eta(y)$ is contained within the set $\{\|y\| \le \frac{M\sigma}{1-\rho}\}$.

We now upper and lower bound the density of each Cx_t, which we denote as f.

$$\begin{aligned} f(y) &= \int_{-\infty}^{\infty} (2\bar{\Gamma})^{-d} \mathbf{1}\{\|u\| \le \bar{\Gamma}\} f_\eta(u-y)du \\ &= (2\bar{\Gamma})^{-d} \int_{\|u+y\| \le \bar{\Gamma}} f_\eta(u)du \\ &\le (2\bar{\Gamma})^{-d} . \end{aligned}$$

$$\begin{aligned} f(y) &= (2\bar{\Gamma})^{-d} \int_{\|u+y\| \le \bar{\Gamma}} f_\eta(u)du \\ &\ge (2\bar{\Gamma})^{-d} \mathbf{1}\{\|y\| \le \bar{\Gamma} - \tfrac{M\sigma}{1-\rho}\} \int_{\|u\| \le \frac{M\sigma}{1-\rho}} f_\eta(u)du \\ &= (2\bar{\Gamma})^{-d} \mathbf{1}\{\|y\| \le \bar{\Gamma} - \tfrac{M\sigma}{1-\rho}\} . \end{aligned}$$

The inequality follows by noting that $\{\|u\| \le \frac{M\sigma}{1-\rho}\} \subseteq \{\|u+y\| \le \bar{\Gamma}\}$ whenever $\|y\| \le \bar{\Gamma} - \frac{M\sigma}{1-\rho}$.

We now turn our attention to finding a lower bound on $s_T(z) = \sum_{t=1}^{T} \kappa_\gamma(z, q(Cx_t))$. We will use Bennett's inequality [Ben62], which requires computing the expectation and second moment of each $\kappa_\gamma(z, q(Cx_t))$. We begin by lower bounding the expected value.

$$\begin{aligned} \mathbb{E}_{Cx_t}[\kappa_\gamma(z, q(Cx_t))] &= \int K\left(\tfrac{d_z(q(y),q(u))}{\gamma}\right) f(p_\star(u))du \\ &\overset{(a)}{\ge} (2\bar{\Gamma})^{-d} \int_{\{\|u\| \le \bar{\Gamma} - \frac{M\sigma}{1-\rho}\}} \kappa\left(\tfrac{d_z(z,q(u))}{\gamma}\right) du \\ &\overset{(b)}{\ge} (2\bar{\Gamma})^{-d} \int_{\{\|u\| \le \bar{\Gamma} - \frac{M\sigma}{1-\rho}\}} \kappa\left(\tfrac{L_q\|p_\star(z)-u\|}{\gamma}\right) du \\ &\overset{(c)}{=} (2\bar{\Gamma})^{-d} \int \kappa\left(\tfrac{L_q\|p_\star(z)-u\|}{\gamma}\right) du \\ &\overset{(d)}{=} \frac{\gamma^d}{(2\bar{\Gamma}L_q)^d} \int \kappa(\|u\|) du \\ &\overset{(e)}{=} \left(\frac{\gamma}{2\bar{\Gamma}L_q}\right)^d 2^d V_\kappa \end{aligned}$$

where (a) follows by the lower bound on $f(y)$, (b) follows by the smoothness of g, (c) holds as long as $\|p_\star(z)\| \le \bar{\Gamma} - \frac{M\sigma}{1-\rho} - \frac{\gamma}{L_q}$ which is guaranteed for $\|p_\star(z)\| \le \Gamma$ by assumption, (d) is a result of a change of variables, and (e) follows by the assumption on κ. Turning to the

second moment,

$$\begin{aligned}\mathbb{E}_{Cx_t}[\kappa_\gamma(z, q(Cx_t))^2] &= \int \kappa\left(\tfrac{d_z(q(y),q(u))}{\gamma}\right)^2 f(p_\star(u))du \\ &\overset{(a)}{\leq} (2\bar{\Gamma})^{-d} \int \kappa\left(\tfrac{d_z(z,q(u))}{\gamma}\right)^2 du \\ &\overset{(b)}{\leq} (2\bar{\Gamma})^{-d} \int \kappa\left(\tfrac{\|p_\star(z)-u\|}{L_p\gamma}\right) du \\ &\overset{(c)}{=} \left(\frac{\gamma L_p}{2\bar{\Gamma}}\right)^d \int \kappa\left(\|u\|\right) du = \left(\frac{\gamma L_p}{\bar{\Gamma}}\right)^d V_\kappa\end{aligned}$$

where (a) follows by the upper bound on $f(y)$, (b) follows by the smoothness of $p_\star$ and the fact that $\kappa \in [0,1]$, and (c) by a change of variables.

We now bound the deviation

$$\begin{aligned}P\left(s_T(z) \leq T\left(\tfrac{\gamma}{\bar{\Gamma}L_q}\right)^d V_\kappa - \varepsilon\right) &= P\left(\sum_{t=1}^T \left(\left(\tfrac{\gamma}{\bar{\Gamma}L_q}\right)^d V_\kappa - \kappa_\gamma(z,z_t)\right) \geq \varepsilon\right) \\ &\overset{(a)}{\leq} P\left(\sum_{t=1}^T \mathbb{E}[\kappa_\gamma(z,z_t)] - \kappa_\gamma(z,z_t) \geq \varepsilon\right) \\ &\overset{(b)}{\leq} e^{-\varepsilon^2/(2T(\frac{\gamma L_p}{\bar{\Gamma}})^d V_\kappa)}\end{aligned}$$

where (a) follows by our lower bound on the expectation and (b) follows by a one-side version of Bennet's inequality for nonnegative random variables, which uses the fact that $\kappa_\gamma(z,z_t) \perp \kappa_\gamma(z,z_k)$ for $k \neq t$.

Therefore, with probability at least $1-\delta$,

$$\begin{aligned}s_T(z) &\geq TV_\kappa\left(\frac{\gamma}{\bar{\Gamma}L_q}\right)^d - \sqrt{2V_\kappa T\log(1/\delta)}\left(\frac{\gamma L_p}{\bar{\Gamma}}\right)^{\frac{d}{2}} \\ &\geq \frac{1}{2}\sqrt{TV_\kappa}\left(\frac{\gamma}{\bar{\Gamma}L_q}\right)^{\frac{d}{2}},\end{aligned}$$

where the second inequality holds under the assumption that $T \geq 8V_\kappa^{-1}\log(1/\delta)\left(\frac{\bar{\Gamma}L_pL_q^2}{\gamma}\right)^d$. □

We are now ready to show the main result by combining Lemma 4.8.2 with Lemma 4.4.2.

Proof of Theorem 4.4.3. We begin by specifying the set of $\{z_i\}$ to be used in the application of Lemma 4.8.2. By selecting $z_i = q(y_i)$ for $\{y_i\}_{i=1}^H$ defined as an ε grid of $\{y \mid \|y\| \le \Gamma\}$, we have that for any z, $\min_i d_z(z, z_i) \le \min_i L_p \|p_\star(z) - y_i\| \le L_p \varepsilon$. Then notice that an ε grid requires $H \ge (2\Gamma/\varepsilon)^d$ points, or rearranging, that $\varepsilon \le 2\Gamma H^{-1/d}$. Therefore, we have that with probability at least $1 - \delta/2$

$$\|p(z) - p_\star(z)\| \le \gamma L_p + \frac{\sigma_v}{\sqrt{s_T(z_{i_\star})}} \sqrt{\log\left(2d^2 H\sqrt{T}/\delta\right)} + 4\sigma_v \frac{T}{s_T(z_{i_\star})} \frac{L_\kappa}{\gamma} L_p \Gamma H^{-1/d} , \tag{4.8.4}$$

where $i_\star$ is the index of the closest element of $\{z_i\}_{i=1}^H$ to z.

Next, we use Lemma 4.4.2 to show that each $s_T(z_{i_\star})$ is bounded below. First, notice that for $\bar{\Gamma} = \sqrt{2}\Gamma$ the condition in Lemma 4.4.2 is met by the assumption that $\gamma \le L_q((\sqrt{2} - 1)\Gamma - \frac{M\sigma}{1-\rho})$. Therefore for any $1 \le i \le H$, with probability at least $1 - H \cdot \delta/(2H)$, as long as $T \ge 8V_\kappa^{-1} \log(2H/\delta) \left(\frac{\bar{\Gamma} L_p L_q^2}{\gamma}\right)^d$,

$$s_T(z_i) \ge \frac{1}{2}\sqrt{TV_\kappa} \left(\frac{\gamma}{\bar{\Gamma} L_q}\right)^{\frac{d}{2}} =: \underline{s}_T \gamma^{\frac{d}{2}} . \tag{4.8.5}$$

We can therefore select H to balance terms, with

$$H^{1/d} = \frac{4TL_\kappa L_p \Gamma}{\sqrt{\underline{s}_T}\gamma^{\frac{d+4}{4}}} = \frac{4\sqrt{2}T^{3/4} L_\kappa L_p \Gamma^{\frac{d+4}{4}} L_q^{\frac{d}{4}}}{V_\kappa^{1/4} \gamma^{\frac{d+4}{4}}} . \tag{4.8.6}$$

For this choice of H, we have by union bound that with probability at least $1 - \delta$,

$$\|p(z) - p_\star(z)\| \le \gamma L_p + \frac{\sigma_v}{\sqrt{\underline{s}_T}\gamma^{d/4}} \left(\sqrt{\log\left(2d^2 H\sqrt{T}/\delta\right)} + 1\right) . \tag{4.8.7}$$

Bounding the logarithmic term,

$$\begin{aligned}
\log\left(2d^2 H\sqrt{T}/\delta\right) &= d\log\left(2d^{2/d}\left(\frac{4\sqrt{2}T^{3/4} L_\kappa L_p \Gamma^{\frac{d+4}{4}} L_q^{\frac{d}{4}}}{V_\kappa^{1/4}\gamma^{\frac{d+4}{4}}}\right) T^{1/2d}/\delta^{1/d}\right) \\
&\le d\log\left(24 V_\kappa^{-\frac{1}{4}} L_\kappa L_p L_q^{\frac{d}{4}} \left(\frac{\Gamma}{\gamma}\right)^{\frac{d+4}{4}} T^{\frac{3d+2}{4d}}/\delta^{\frac{1}{d}}\right) \\
&\le d\log\left(T^2/\delta\right) .
\end{aligned}$$

where the first inequality follows because $d^{2/d} \le 3/\sqrt{2}$ and the final inequality follows due to the assumption that $(24L_\kappa L_p)^{\frac{4}{3}} V_\kappa^{-\frac{1}{3}} \left(\frac{\Gamma}{\gamma}\right)^{\frac{d+4}{3}} L_q^{\frac{d}{3}} \le T$.

The result follows. It only remains to note that the condition on T in Lemma 4.4.2 is satisfied due to the assumption that

$$T \geq 8V_\kappa^{-1} d \log(T^2/\delta) \left(\frac{\bar{\Gamma} L_p L_q^2}{\gamma} \right)^d$$

by noting that $\log(2H/\delta) \leq \log(2d^2 H\sqrt{T}/\delta) \leq d \log(T^2/\delta)$. □

Proof of Corollary 4.5.2. To show this result, we combine the expressions in Theorem 4.4.3 and Proposition 4.5.1 for a carefully chosen value of γ. To balance the terms, we set (recalling that $\bar{\Gamma} = \sqrt{2}\Gamma = 2\Gamma_{\max}(\mathbf{\Phi})$)

$$\gamma^{(d+4)/4} = \frac{d^{1/2}\sigma_v}{L_p \sqrt{\underline{S}_T}} = \frac{\sqrt{2} d^{1/2} \sigma_v (2\Gamma_{\max}(\mathbf{\Phi}) L_q)^{\frac{d}{4}}}{L_p (T V_\kappa)^{\frac{1}{4}}}$$

Returning to the unsimplified bound (4.8.7), this choice of γ results in

$$\|p(z) - p_\star(z)\| \leq (2\Gamma_{\max}(\mathbf{\Phi}) L_q L_p)^{\frac{d}{d+4}} \left(\frac{4d^2\sigma_v^4}{T} \right)^{\frac{1}{d+4}} \left(\sqrt{\frac{1}{d} \log\left(2d^2 H\sqrt{T}/\delta\right)} + \frac{1}{\sqrt{d}} + 1 \right). \quad (4.8.8)$$

We begin by simplifying the first term. Note that since $p_\star \circ q$ is an identity map, $L_p L_q \geq 1$, so $(2\Gamma_{\max}(\mathbf{\Phi}) L_q L_p)^{\frac{d}{d+4}} \leq 2\Gamma_{\max}(\mathbf{\Phi}) L_q L_p$ since $\Gamma_{\max}(\mathbf{\Phi}) \geq 1/2$ by assumption.

Next, considering the the logarithmic term:

$$\begin{aligned}
\log\left(2d^2 H\sqrt{T}/\delta\right) &= d \log\left(2d^{2/d} \left(\frac{4TL_\kappa L_p (2\Gamma_{\max}(\mathbf{\Phi}))}{\sqrt{\underline{S}_T} \frac{d^{1/2}\sigma_v}{L_p\sqrt{\underline{S}_T}}} \right) T^{1/2d}/\delta^{1/d} \right) \\
&= d \log\left(2d^{2/d - 1/2} \left(\frac{8 L_\kappa L_p^2 \Gamma_{\max}(\mathbf{\Phi})}{\sigma_v} \right) T^{1+1/2d}/\delta^{1/d} \right) \\
&\leq d \log\left(3 \cdot 8 L_\kappa L_p^2 \Gamma_{\max}(\mathbf{\Phi}) \sigma_v^{-1} T^{3/2}/\delta^{1/d} \right) \\
&\leq d \log\left(T^2/\delta \right)
\end{aligned}$$

where the first inequality follows because $d^{2/d-1/2} \leq 1.5$ and the final inequality is true as long as $(24 L_\kappa L_p^2 \Gamma_{\max}(\mathbf{\Phi}) \sigma_v^{-1})^2 \leq T$. Further simplifying,

$$\sqrt{\frac{1}{d} \log\left(2d^2 H\sqrt{T}/\delta\right)} + \frac{1}{\sqrt{d}} + 1 \leq \sqrt{\log\left(T^2/\delta\right)} + \frac{1}{\sqrt{d}} + 1 \leq 2\sqrt{\log\left(T^2/\delta\right)}.$$

The resulting error bound is

$$\|p(z) - p_\star(z)\| \leq 4L_q L_p \Gamma_{\max}(\mathbf{\Phi}) \left(\frac{4d^2\sigma^4}{T}\right)^{\frac{1}{d+4}} \sqrt{\log(T^2/\delta)} =: \varepsilon_p \,. \tag{4.8.9}$$

Then the result follows by Proposition 4.5.1 as long as $T \geq \max\{T_1, T_2, T_3, T_4\}$, where we ensure that the simplification in the logarithm above is valid with

$$T_1 = (24L_\kappa L_p^2 \Gamma_{\max}(\mathbf{\Phi})\sigma_v^{-1})^2,$$

we ensure that $\gamma \leq L_q((2-\sqrt{2})\Gamma_{\max}(\mathbf{\Phi}) - \frac{M\sigma}{1-\rho})$ (Theorem 4.4.3) with

$$T_2 = \frac{(4d\sigma_v^2)^2(2\Gamma_{\max}(\mathbf{\Phi}))^d L_q^4}{V_\kappa((2-\sqrt{2})\Gamma_{\max}(\mathbf{\Phi}) - M\frac{\sigma}{1-\rho})^{d+4} L_p^4},$$

we ensure $T \geq 8V_\kappa^{-1}\log(2H/\delta)\left(\frac{\bar{\Gamma}L_p L_q^2}{\gamma}\right)^d$ (the usage of Lemma 4.4.2 in Theorem 4.4.3) with

$$T_3 = V_\kappa \left(8(L_q L_p)^d d \log(T^2/\delta)\right)^{\frac{d+4}{4}} \left(\frac{L_q L_p \Gamma_{\max}(\mathbf{\Phi})}{\sqrt{2}d^{1/2}\sigma_v}\right)^d,$$

and finally we ensure that $\varepsilon_p \leq \frac{(\sqrt{2}-1)\Gamma_{\max}(\mathbf{\Phi})}{\|C\mathbf{\Phi}_{\mathbf{xv}}\|}$ (Proposition 4.5.1) with

$$T_4 = 4d^2\sigma_v^4 \left(10L_q L_p \|C\mathbf{\Phi}_{\mathbf{xv}}\| \sqrt{\log(T^2/\delta)}\right)^{d+4}.$$

For large enough values of T, T_4 will dominate.

□

Part II

Feedback in Social-Digital Systems

Chapter 5

Fairness and Wellbeing in Consequential Decisions

5.1 Introduction

In this chapter, we turn our attention from the continuous control of linear systems to making consequential decisions about individuals. We formally examine the *impact* of these decisions on populations, and ask whether fairness criteria promote the long-term well-being of disadvantaged groups as measured in terms of a temporal variable of interest. Going beyond the standard classification setting, we introduce a one-step feedback model of decision making that exposes how decisions change the underlying population over time. We further consider *welfare-aware machine learning* as an inherently multi-objective problem that requires explicitly balancing decision outcomes against the primary objective. This chapter uses material first presented in papers coauthored with Daniel Björkegren, Joshua Blumenstock, Moritz Hardt, Lydia T. Liu, Esther Rolf, and Max Simchowitz [LDRSH18; RSD+20].

Our running example is a hypothetical lending scenario. There are two groups in the population with features described by a summary statistic, such as a *credit score*, whose distribution differs between the two groups. The bank can choose thresholds for each group at which loans are offered. While group-dependent thresholds may face legal challenges [RY06], they are generally inevitable for some of the criteria we examine. The impact of a lending decision has multiple facets. A default event not only diminishes profit for the bank, it also worsens the financial situation of the borrower as reflected in a subsequent decline in credit score. A successful lending outcome leads to profit for the bank and also to an increase in credit score for the borrower.

When thinking of one of the two groups as disadvantaged, it makes sense to ask what lending policies (choices of thresholds) lead to an expected improvement in the score distribution within that group. An unconstrained bank would maximize profit, choosing thresholds that meet a break-even point above which it is profitable to give out loans. One

frequently proposed fairness criterion, sometimes called demographic parity, requires the bank to lend to both groups at an equal rate [e.g. CKP09; ZVRG17]. Subject to this requirement the bank would continue to maximize profit to the extent possible. Another criterion, originally called equality of opportunity, equalizes the *true positive rates* between the two groups, thus requiring the bank to lend in both groups at an equal rate among individuals who repay their loan [HPS16]. Other criteria are natural, but for clarity we restrict our attention to these three.

We further study a natural class of selection policies that balance multiple objectives (e.g., private profit and public welfare). We show that this class of score-based policies has a natural connection to classifiers constrained to satisfy demographic parity and its ϵ-fair analog [EJJ+19].

Related Work

A growing literature highlights the inability of any one fairness definition to solve more general concerns of social equity [CG18]. The impossibility of satisfying all desirable criteria [KMR17; Cho17] and the unintended consequences of enforcing parity constraints based on sensitive attributes [KNRW18] indicate that existing approaches to fairness in machine learning are not a panacea for these adverse effects. Recent work by Hu and Chen [HC20] contend that while social welfare is of primary concern in many applications, common fairness constraints may be at odds with the relevant notion of welfare.

A line of recent work considers long term outcomes in settings inspired by education and the labor market [HC18a; MOS19; LWH+20]. The equilibrium analysis of the dynamics arising from repeated decisions and adaptions allows for specific conclusions relating fairness criteria to long term outcomes. This style of analysis is reminiscent of economic arguments for affirmative action [FV92], setting which has also been extensively studied in the social sciences [see e.g., KBSW85; KDK06]. Complementary to this type of domain specific analysis, our framework is general: by focusing on a single decision, we avoid the need to model dynamics.

Ensign et al. [EFNSV18] consider feedback loops in predictive policing, where the police more heavily monitor high crime neighborhoods, thus further increasing the measured number of crimes in those neighborhoods. While the work addresses an important temporal phenomenon using the theory of urns, it is rather different from our one-step feedback model both conceptually and technically. A growing literature on fairness in the "bandits" setting of learning [see JKMR16, *et sequelae*] deals with online decision making that ought not to be confused with our one-step feedback setting. Fuster et al. [FGRW17] consider the problem of fairness in credit markets from a different perspective. Their goal is to study the effect of machine learning on interest rates in different groups at an equilibrium, under a static model without feedback. Ultimately, we advocate for a view towards long term outcomes and the design of decision policies that prioritize impact from the outset.

The definition and measurement of welfare is an important and complex problem that has received considerable attention in the social science literature [cf. Dea80; Dea16; SSF09]. There, a standard approach is to sum up individual measures of welfare, to obtain an aggregate measure of societal welfare. The separability assumption (independent individual scores) is a standard simplifying assumption [e.g. Flo14] that appears in the foundational work of Pigou [Pig20], as well as Burk [Bur38], Samuelson [Sam47], Arrow [Arr63] and Sen [Sen73]. However, alternative forms of social welfare functions exist [e.g. CO96] and may be of interest for future work.

Problem Setting

We consider two *groups* A and B, which comprise a g_{A} and $g_{\mathsf{B}} = 1 - g_{\mathsf{A}}$ fraction of the total population, and an *institution* which makes a binary decision for each individual in each group, called *selection*. Individuals in each group are assigned *scores* in $\mathcal{X} := [C]$, and the scores for group $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$ are distributed according $\mathcal{D}_{\mathsf{j}} \in \text{Simplex}^{C-1}$. The institution selects a *policy* $\pi := (\pi_{\mathsf{A}}, \pi_{\mathsf{B}}) \in [0,1]^{2C}$, where $\pi_{\mathsf{j}}(x)$ corresponds to the probability the institution selects an individual in group j with score x. One should think of a score as an abstract quantity which summarizes how well an individual is suited to being selected; examples are provided at the end of this section.

We assume that the institution is utility-maximizing, but may impose certain constraints to ensure that the policy π is *fair*, in a sense described in Section 5.2. We assume that there exists a function $\boldsymbol{u} : \mathcal{X} \to \mathbb{R}$, such that the institution's expected utility for a policy π is given by

$$\mathcal{U}(\pi) = \mathbb{E}[\boldsymbol{u}(x)] = \textstyle\sum_{\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \sum_{x \in \mathcal{X}} \pi_{\mathsf{j}}(x) \mathcal{D}_{\mathsf{j}}(x) \boldsymbol{u}(x). \tag{5.1.1}$$

Institutions will design policies which maximize $\mathcal{U}$, subject to fairness constraints.

Novel to this work, we focus on the effect of the selection policy π on the groups A and B. We quantify these *outcomes* in terms of an average effect that a policy π_{j} has on group j. Formally, for a function $\boldsymbol{\Delta}(x) : \mathcal{X} \to \mathbb{R}$, we define the average change of the mean score μ_{j} for group j

$$\Delta\mu_{\mathsf{j}}(\pi) := \mathbb{E}_{\mathsf{j}}[\boldsymbol{\Delta}(x)] = \textstyle\sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x) \pi_{\mathsf{j}}(x) \boldsymbol{\Delta}(x)\,. \tag{5.1.2}$$

We remark that many of our results also go through if $\Delta\mu_{\mathsf{j}}(\pi)$ simply refers to an abstract change in well-being, not necessarily a change in the mean score. Furthermore, it is possible to modify the definition of $\Delta\mu_{\mathsf{j}}(\pi)$ such that it directly considers outcomes of those who are not selected.[1] Lastly, we assume that the *success* of an individual is independent

[1] If we consider functions $\boldsymbol{\Delta}_1(x) : \mathcal{X} \to \mathbb{R}$ and $\boldsymbol{\Delta}_0(x) : \mathcal{X} \to \mathbb{R}$ to represent the average effect of selection and non-selection respectively, then $\Delta\mu_{\mathsf{j}}(\pi) := \sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x) \left(\pi_{\mathsf{j}}(x)\boldsymbol{\Delta}_1(x) + (1 - \pi_{\mathsf{j}}(x))\boldsymbol{\Delta}_0(x)\right)$. This model corresponds to replacing $\boldsymbol{\Delta}(x)$ in the original outcome definition with $\boldsymbol{\Delta}_1(x) - \boldsymbol{\Delta}_0(x)$, and adding a offset $\sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x)\boldsymbol{\Delta}_0(x)$. Under the assumption that $\boldsymbol{\Delta}_1(x) - \boldsymbol{\Delta}_0(x)$ increases in x, this model gives rise to outcomes curves resembling those in Figure 5.1 up to vertical translation. All presented results hold unchanged under the further assumption that $\Delta\mu(\beta^{\texttt{MaxUtil}}) \geq 0$.

of their group given the score; that is, the score summarizes all relevant information about the success event, so there exists a function $\rho : \mathcal{X} \to [0, 1]$ such that individuals of score x succeed with probability $\rho(x)$.

We now introduce the specific domain of credit scores as a running example in the rest of the chapter, after which we present two more examples showing the general applicability of our formulation to many domains.

Example 5.1 (Credit scores)**.** In the setting of loans, scores $x \in [C]$ represent credit scores, and the bank serves as the institution. The bank chooses to grant or refuse loans to individuals according to a policy π. Both bank and personal utilities are given as functions of loan repayment, and therefore depend on the success probabilities $\rho(x)$, representing the probability that any individual with credit score x can repay a loan within a fixed time frame. The expected utility to the bank is given by the expected return from a loan, which can be modeled as an affine function of $\rho(x)$: $\boldsymbol{u}(x) = u_+\rho(x)+u_-(1-\rho(x))$, where u_+ denotes the profit when loans are repaid and u_- the loss when they are defaulted on. Individual outcomes of being granted a loan are based on whether or not an individual repays the loan, and a simple model for $\Delta(x)$ may also be affine in $\rho(x)$: $\Delta(x) = c_+\rho(x) + c_-(1 - \rho(x))$, modified accordingly at boundary states. The constant c_+ denotes the gain in credit score if loans are repaid and c_- is the score penalty in case of default.

Example 5.2 (Advertising)**.** A second illustrative example is given by the case of advertising agencies making decisions about which groups to target. An individual with product interest score x responds positively to an ad with probability $\rho(x)$. The ad agency experiences utility $\boldsymbol{u}(x)$ related to click-through rates, which increases with $\rho(x)$. Individuals who see the ad but are uninterested may react negatively (becoming less interested in the product), and $\Delta(x)$ encodes the interest change. If the product is a positive good like education or employment opportunities, interest can correspond to well-being. Thus the advertising agency's incentives to only show ads to individuals with extremely high interest may leave behind groups whose interest is lower on average. A related historical example occurred in advertisements for computers in the 1980s, where male consumers were targeted over female consumers, arguably contributing to the current gender gap in computing.

Example 5.3 (College Admissions)**.** The scenario of college admissions or scholarship allotments can also be considered within our framework. Colleges may select certain applicants for acceptance according to a score x, which could be thought encode a "college preparedness" measure. The students who are admitted might "succeed" (this could be interpreted as graduating, graduating with honors, finding a job placement, etc.) with some probability $\rho(x)$ depending on their preparedness. The college might experience a utility $\boldsymbol{u}(x)$ corresponding to alumni donations, or positive rating when a student succeeds; they might also show a drop in rating or a loss of invested scholarship money when a student is unsuccessful. The student's success in college will affect their later success, which could be modeled generally by $\Delta(x)$. In this scenario, it is challenging to ensure

that a single summary statistic x captures enough information about a student; it may be more appropriate to consider x as a vector as well as more complex forms of $\boldsymbol{\rho}(x)$.

While a variety of applications are modeled faithfully within our framework, there are limitations to the accuracy with which real-life phenomenon can be measured by strictly binary decisions and success probabilities. Such binary rules are necessary for the definition and execution of existing fairness criteria, (see Sec. 5.2) and as we will see, even modeling these facets of decision making as binary allows for complex and interesting behavior.

5.2 Delayed Impact of Fair Decisions

The Outcome Curve

We now introduce important outcome regimes, stated in terms of the change in average group score. A policy $(\pi_{\mathsf{A}}, \pi_{\mathsf{B}})$ is said to cause *active harm* to group j if $\Delta\mu_{\mathsf{j}}(\pi_{\mathsf{j}}) < 0$, *stagnation* if $\Delta\mu_{\mathsf{j}}(\pi_{\mathsf{j}}) = 0$, and *improvement* if $\Delta\mu_{\mathsf{j}}(\pi_{\mathsf{j}}) > 0$. Under our model, `MaxUtil` policies can be chosen in a standard fashion which applies the same threshold policy $\pi^{\texttt{MaxUtil}}$ for both groups, and is agnostic to the distributions $\mathcal{D}_{\mathsf{A}}$ and $\mathcal{D}_{\mathsf{B}}$. Hence, if we define

$$\Delta\mu_{\mathsf{j}}^{\texttt{MaxUtil}} := \Delta\mu_{\mathsf{j}}(\pi^{\texttt{MaxUtil}}) \tag{5.2.1}$$

we say that a policy causes *relative harm* to group j if $\Delta\mu_{\mathsf{j}}(\pi_{\mathsf{j}}) < \Delta\mu_{\mathsf{j}}^{\texttt{MaxUtil}}$, and *relative improvement* if $\Delta\mu_{\mathsf{j}}(\pi_{\mathsf{j}}) > \Delta\mu_{\mathsf{j}}^{\texttt{MaxUtil}}$. In particular, we focus on these outcomes for a disadvantaged group, and consider whether imposing a fairness constraint improves their outcomes relative to the `MaxUtil` strategy. From this point forward, we take `A` to be disadvantaged group.

Figure 5.1 displays the important outcome regimes in terms of *selection rates*

$$\beta_{\mathsf{j}} := \sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x)\pi_{\mathsf{j}}(x)\,.$$

This succinct characterization is possible when considering decision rules based on (possibly randomized) score thresholding, in which all individuals with scores above a threshold are selected. In Section 5.4, we justify the restriction to such *threshold policies* by showing it preserves optimality. We further show that the outcome curve is concave, thus implying that it takes the shape depicted in Figure 5.1. To explicitly connect selection rates to decision policies, we define the rate function $r_{\mathcal{D}}(\pi_{\mathsf{j}})$ which returns the proportion of group j selected by the policy. We show that this function is invertible for a suitable class of threshold policies, and in fact the outcome curve is precisely the graph of the map from selection rate to outcome $\beta \mapsto \Delta\mu_{\mathsf{A}}(r_{\mathcal{D}_{\mathsf{A}}}^{-1}(\beta))$. Next, we define the values of β that mark boundaries of the outcome regions.

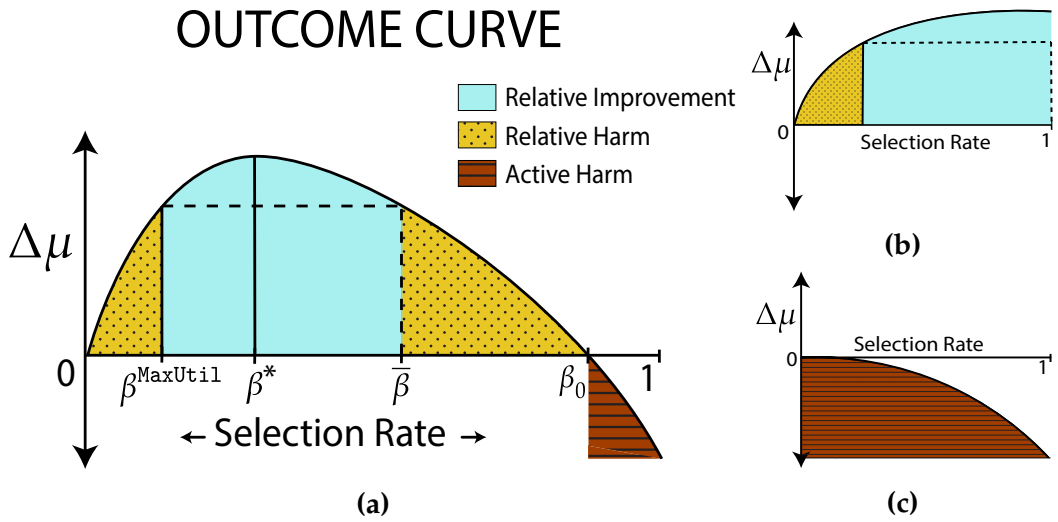

Figure 5.1: The above figure shows the *outcome curve*. The horizontal axis represents the selection rate for the population; the vertical axis represents the mean change in score. (a) depicts the full spectrum of outcome regimes, and colors indicate regions of active harm, relative harm, and no harm. In (b): a group that has much potential for gain, in (c): a group that has no potential for gain.

Definition 5.1 (Selection rates of interest)**.** Given the disadvantaged group A, the following selection rates are of interest in distinguishing between qualitatively different classes of outcomes (Figure 5.1). We define:

- $\beta^{\texttt{MaxUtil}}$ as the selection rate for A under `MaxUtil`
- β_0 as the harm threshold, such that $\Delta\mu_{\mathsf{A}}(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta_0)) = 0$
- β^* as the selection rate such that $\Delta\mu_{\mathsf{A}}$ is maximized
- $\overline{\beta}$ as the outcome-complement of the `MaxUtil` selection rate, $\Delta\mu_{\mathsf{A}}(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\overline{\beta})) = \Delta\mu_{\mathsf{A}}(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta^{\texttt{MaxUtil}}))$ with $\overline{\beta} > \beta^{\texttt{MaxUtil}}$

Decision Rules and Fairness Criteria

We will consider policies that maximize the institution's total expected utility, potentially subject to a constraint: $\pi \in C \subset [0,1]^{2C}$ which enforces some notion of "fairness". Formally, the institution selects $\pi^\star \in \text{argmax}\ \mathcal{U}(\pi)$ s.t. $\pi \in C$. We consider the three following constraints:

Definition 5.2 (Fairness criteria)**.** The *maximum utility* (`MaxUtil`) policy corresponds to the null-constraint $C = [0,1]^{2C}$, so that the institution is free to focus solely on utility.

The *demographic parity* (`DemParity`) policy results in equal selection rates between both groups. Formally, the constraint is $C = \{(\pi_{\mathsf{A}}, \pi_{\mathsf{B}}) : \sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{A}}(x)\pi_{\mathsf{A}} = \sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{B}}(x)\pi_{\mathsf{B}}\}$.

The *equal opportunity* (`EqOpt`) policy results in equal true positive rates (TPR) between both group, where TPR is defined as

$$\mathrm{TPR}_{\mathsf{j}}(\pi) := \frac{\sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x)\boldsymbol{\rho}(x)\pi(x)}{\sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x)\boldsymbol{\rho}(x)} .$$

`EqOpt` ensures that the conditional probability of selection given that the individual will be successful is independent of the population, formally enforced by the constraint $C = \{(\pi_{\mathsf{A}}, \pi_{\mathsf{B}}) : \mathrm{TPR}_{\mathsf{A}}(\pi_{\mathsf{A}}) = \mathrm{TPR}_{\mathsf{B}}(\pi_{\mathsf{B}})\}$.

Just as the expected outcome $\Delta\mu$ can be expressed in terms of selection rate for threshold policies, so can the total utility $\mathcal{U}$. In the unconstrained cause, $\mathcal{U}$ varies independently over the selection rates for group A and B; however, in the presence of fairness constraints the selection rate for one group determines the allowable selection rate for the other. The selection rates must be equal for `DemParity`, but for `EqOpt` we can define a *transfer function*, $G^{(\mathsf{A} \to \mathsf{B})}$, which for every loan rate β in group A gives the loan rate in group B that has the same true positive rate. Therefore, when considering threshold policies, decision rules amount to maximizing functions of single parameters. This idea is expressed in Figure 5.2, and underpins the results to follow.

In order to clearly characterize the outcome of applying fairness constraints, we make the following assumption.

Assumption 5.1 (Institution utilities)**.** *The institution's individual utility function is more stringent than the expected score changes,* $\boldsymbol{u}(x) > 0 \implies \boldsymbol{\Delta}(x) > 0$. *(For the linear form presented in Example 5.1,* $\frac{u_-}{u_+} < \frac{c_-}{c_+}$ *is necessary and sufficient.)*

This simplifying assumption quantifies the intuitive notion that institutions take a greater risk by accepting than the individual does by applying. For example, in the credit setting, a bank loses the amount loaned in the case of a default, but makes only interest in case of a payback. Using Assumption 5.1, we can restrict the position of `MaxUtil` on the outcome curve in the following sense.

Proposition 5.2.1 (`MaxUtil` does not cause active harm)**.** *Under Assumption 5.1,*

$$0 \leq \Delta\mu^{\texttt{MaxUtil}} \leq \Delta\mu^{*} .$$

We defer the proof to Section 5.8, and the proofs of all subsequent results presented in this section. The following results are corollaries to theorems presented in Section 5.5.

Prospects and Pitfalls of Fairness Criteria

We begin by characterizing general settings under which fairness criteria act to improve outcomes over unconstrained `MaxUtil` strategies. For this result, we will assume that

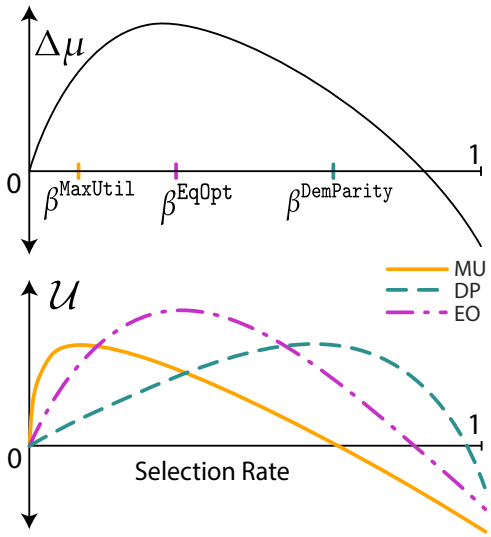

Figure 5.2: Both outcomes $\Delta\mu$ and institution utilities $\mathcal{U}$ can be plotted as a function of selection rate for one group. The maxima of the utility curves determine the selection rates resulting from various decision rules.

group A is disadvantaged in the sense that the `MaxUtil` acceptance rate for B is large compared to relevant acceptance rates for A.

Corollary 5.2.2 (Fairness Criteria can cause Relative Improvement)**.** *(a) Under the assumption that* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \overline{\beta}$ *and* $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$*, there exist population proportions* $g_0 < g_1 < 1$ *such that, for all* $g_{\mathsf{A}} \in [g_0, g_1]$, `DemParity` *causes relative improvement, i.e.* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{DemParity}} < \overline{\beta}$.

(b) Under the assumption that there exist $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta < \beta' < \overline{\beta}$ *such that* $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta), G^{(\mathsf{A}\to\mathsf{B})}(\beta')$*, there exist population proportions* $g_2 < g_3 < 1$ *such that, for all* $g_{\mathsf{A}} \in [g_2, g_3]$, `EqOpt` *causes relative improvement, i.e.* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta_{\mathsf{A}}^{\texttt{EqOpt}} < \overline{\beta}$.

This result gives the conditions under which we can guarantee the existence of settings in which fairness criteria cause improvement relative to `MaxUtil`. Relying on machinery proved in Section 5.5, the result follows from comparing the position of optima on the utility curve to the outcome curve. Figure 5.2 displays a illustrative example of both the outcome curve and the institutions' utility $\mathcal{U}$ as a function of the selection rates in group A. In the utility function (5.1.1), the contributions of each group are weighted by their population proportions g_{j}, and thus the resulting selection rates are sensitive to these proportions.

As we see in the remainder of this section, fairness criteria can achieve nearly any position along the outcome curve under the right conditions. This fact comes from the potential mismatch between the outcomes, controlled by $\boldsymbol{\Delta}$, and the institution's utility $\mathcal{U}$, controlled by $\boldsymbol{u}$.

The next result implies that DemParity can be bad for long term well-being of the disadvantaged group by being over-eager, under the mild assumption that $\Delta\mu_{\mathsf{A}}(\beta_{\mathsf{B}}^{\texttt{MaxUtil}}) < 0$:

Corollary 5.2.3 (DemParity can cause harm by being over-eager)**.** *Fix a selection rate β. Assume that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$. Then, there exists a population proportion g_0 such that, for all $g_{\mathsf{A}} \in [0, g_0]$, $\beta_{\mathsf{A}}^{\texttt{DemParity}} > \beta$. In particular, when $\beta = \beta_0$,* DemParity *causes active harm, and when $\beta = \overline{\beta}$,* DemParity *causes relative harm.*

The assumption $\Delta\mu_{\mathsf{A}}(\beta_{\mathsf{B}}^{\texttt{MaxUtil}}) < 0$ implies that a policy which selects individuals from group A at the selection rate that MaxUtil would have used for group B necessarily lowers average score in A. This is one natural notion of disadvantaged group A's 'disadvantage' relative to group B. In this case, DemParity penalizes the scores of group A even more than a naive MaxUtil policy, as long as group proportion g_{A} is small enough. Again, small g_{A} is another notion of group disadvantage.

Using credit scores as an example, Corollary 5.2.3 tells us that an overly aggressive fairness criterion will give too many loans to individuals in a disadvantaged group who cannot pay them back, hurting the group's credit scores on average. In the following theorem, we show that an analogous result holds for EqOpt.

Corollary 5.2.4 (EqOpt can cause harm by being over-eager)**.** *Suppose that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$ and $\beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$. Then, there exists a population proportion g_0 such that, for all $g_{\mathsf{A}} \in [0, g_0]$, $\beta_{\mathsf{A}}^{\texttt{EqOpt}} > \beta$. In particular, when $\beta = \beta_0$,* EqOpt *causes active harm, and when $\beta = \overline{\beta}$,* EqOpt *causes relative harm.*

We remark that in Corollary 5.2.4, we rely on the *transfer function*, $G^{(\mathsf{A}\to\mathsf{B})}$, which for every loan rate β in group A gives the loan rate in group B that has the same true positive rate. Notice that if $G^{(\mathsf{A}\to\mathsf{B})}$ were the identity function, Corollary 5.2.3 and Corollary 5.2.4 would be exactly the same. Indeed, our framework (detailed in Section 5.5) unifies the analyses for a large class of fairness constraints that includes DemParity and EqOpt as specific cases, and allows us to derive results about impact on $\Delta\mu$ using general techniques.

Fairness Under Measurement Error

Next, consider the implications of an institution with imperfect knowledge of scores. Under a simple model in which the estimate of an individual's score $X \sim \mathcal{D}$ is prone to errors $e(X)$ such that $X + e(X) =: \widehat{X} \sim \widehat{\mathcal{D}}$. Constraining the error to be negative results in the setting that scores are systematically *underestimated*. In this setting, it is equivalent to consider the CDF of underestimated distribution $\widehat{\mathcal{D}}$ to be *dominated* by the CDF true distribution $\mathcal{D}$, that is $\sum_{x\geq c} \widehat{\mathcal{D}}(x) \leq \sum_{x\geq c} \mathcal{D}(x)$ for all $c \in [C]$. Then we can compare the institution's behavior under this estimation to its behavior under the truth.

Proposition 5.2.5 (Underestimation causes under-selection)**.** *Fix the distribution of* B *as $\mathcal{D}_{\mathsf{B}}$ and let β be the acceptance rate of* A *when the institution makes the decision using perfect knowledge*

of the distribution $\mathcal{D}_{\mathsf{A}}$. Denote $\widehat{\beta}$ as the acceptance rate when the group is instead taken as $\widehat{\mathcal{D}}_{\mathsf{A}}$. Then $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} > \widehat{\beta}_{\mathsf{A}}^{\texttt{MaxUtil}}$ and $\beta_{\mathsf{A}}^{\texttt{DemParity}} > \widehat{\beta}_{\mathsf{A}}^{\texttt{DemParity}}$. If the errors are further such that the true TPR dominates the estimated TPR, it is also true that $\beta_{\mathsf{A}}^{\texttt{EqOpt}} > \widehat{\beta}_{\mathsf{A}}^{\texttt{EqOpt}}$.

Because fairness criteria encourage a higher selection rate for disadvantaged groups (Corollary 5.2.2), systematic underestimation widens the regime of their applicability. Furthermore, since the estimated `MaxUtil` policy under-loans, the region for relative improvement in the outcome curve (Figure 5.1) is larger, corresponding to more regimes under which fairness criteria can yield favorable outcomes. Thus the potential for measurement error should be a factor when motivating these criteria.

5.3 Impact-Aware Decisions

Outcome-Based Alternative

We have seen how fairness criteria may actively harm disadvantaged groups. It is thus natural to consider a modified decision rule which involves the explicit maximization of $\Delta\mu_{\mathsf{A}}$. In this case, imagine that the institution's primary goal is to aid the disadvantaged group, subject to a limited profit loss compared to the maximum possible expected profit $\mathcal{U}^{\texttt{MaxUtil}}$. The corresponding problem is as follows.

$$\max_{\pi_{\mathsf{A}}} \Delta\mu_{\mathsf{A}}(\pi_{\mathsf{A}}) \;\text{ s.t. }\; \mathcal{U}_{\mathsf{A}}^{\texttt{MaxUtil}} - \mathcal{U}_{\mathsf{A}}(\pi) < \delta\,. \tag{5.3.1}$$

Unlike the fairness constrained objective, this objective no longer depends on group B and instead depends on our model of the mean score change in group A, $\Delta\mu_{\mathsf{A}}$.

Proposition 5.3.1 (Outcome-based solution)**.** *In the above setting, the optimal policy π_{A} is a threshold policy with selection rate $\beta = \min\{\beta^*, \beta^{\max}\}$, where β^* is the outcome-optimal loan rate and $\beta^{\max}$ is the maximum loan rate under the bank's "budget".*

The above formulation's advantage over fairness constraints is that it directly optimizes the outcome of A and can be approximately implemented given reasonable ability to predict outcomes. Importantly, this objective shifts the focus to outcome modeling, highlighting the importance of domain specific knowledge.

Welfare-Aware Decisions

We now briefly describe and examine a framework to more explicitly consider tradeoffs between the profit gained by an institution and the impact of decisions on a population. We will see that fairness constraints are a special case within this broader framework. Consider two simultaneous objectives: to maximize the private return, generically referred to as *profit*; and to improve a measure of wellbeing (such as social welfare or user health), referred to as *welfare*.

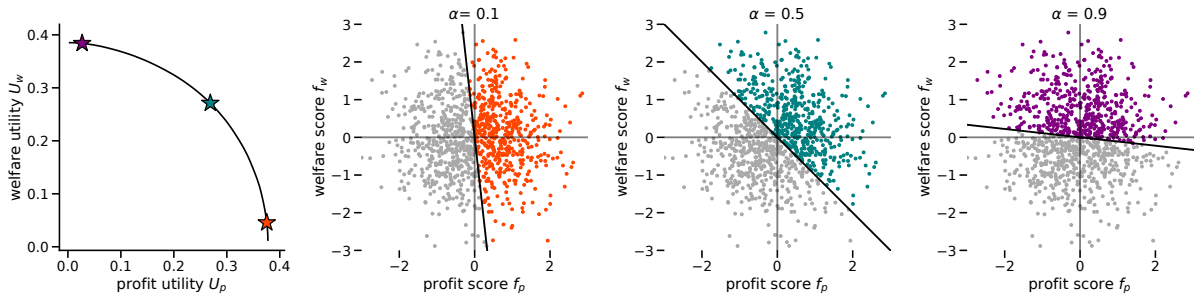

Figure 5.3: Illustration of a Pareto curve (bottom left) and the decision boundaries induced by three different trade-off parameters α. Colored (darker in gray scale) points indicate selected individuals.

This framework eschews the perspective based on groups, and instead associates to each individual a value p representing the expected profit to be garnered from approving this individual and w encoding the change in welfare. The profit score p generalizes the individual utility $\boldsymbol{u}(x)$ and the welfare score w generalizes the change in wellbeing $\boldsymbol{\Delta}(x)$. Rather than relying on an underlying score x, we suppose that p and w can be defined within a population arbitrarily, so that their values can vary independently. The profit and welfare objectives are thus expectations over the joint distribution of (p, w). We generalize the notation of $\mathcal{U}(\pi)$ and $\Delta\mu(\pi)$ to $\mathcal{U}_{\mathsf{P}}(\pi)$ and $\mathcal{U}_{\mathsf{W}}(\pi)$ to be explicit about this framing.

Given two objectives, one can no longer define a unique optimal policy π. Instead, we focus on policies π which are *Pareto-optimal* [Par06], in the sense that they are not strictly dominated by any alternative policy, i.e. there is no π' such that both $\mathcal{U}_{\mathsf{P}}$ and $\mathcal{U}_{\mathsf{W}}$ are strictly larger under π'. For a general set of policy classes, it is equivalent to consider policies that maximize a weighted combination of both objectives. We can thus parametrize the Pareto-optimal policies by $\alpha \in [0, 1]$:

Definition 5.3 (Pareto-optimal policies)**.** An α-Pareto-optimal policy (for $\alpha \in [0, 1]$) satisfies:

$$\pi^\star_\alpha \in \operatorname{argmax} \mathcal{U}_\alpha(\pi),$$
$$\mathcal{U}_\alpha(\pi) := (1-\alpha)\mathcal{U}_{\mathsf{P}}(\pi) + \alpha\mathcal{U}_{\mathsf{W}}(\pi).$$

In the definition above, the maximization of π is taken over the class of randomized policies. This weighted optimization is analogous in some sense to the outcome-based alternative described in the previous subsection. Supposing that the profit and welfare scores are exactly known, the expected weighted contribution from accepted individuals is $(1-\alpha)p + \alpha w$. Therefore, one can show [RSD+20] that the optimal policy is given by a threshold policy on this composite. We will write this Pareto-optimal policy as

$$\pi^\star_\alpha(p, w) = \mathbf{1}\{(1-\alpha)p + \alpha w \geq 0\}. \tag{5.3.2}$$

Note that in general, some care should be taken to define a suitably randomized threshold policy when the score distributions are discrete. Section 5.4 makes this precise. In the remained of this section, we will suppose taht the score distributions are continuous to simplify the exposition.

Though they are all Pareto-optimal, the policies $\pi^\star_\alpha$ induce different trade-offs between the two objectives. The parameter α determines this trade-off, tracing the *Pareto frontier*:

$$\mathcal{P} := \{(\mathcal{U}_\mathsf{P}(\pi^\star_\alpha), \mathcal{U}_\mathsf{W}(\pi^\star_\alpha)) : \alpha \in [0, 1]\}$$

Figure 5.3 plots an example of this curve (bottom-left panel) and the corresponding decision rules for three points along it. We note the concave shape of this curve, a manifestation of *diminishing marginal returns*: as a decision policy forgoes profit to increase total welfare, less welfare is gained for the same amount of profit forgone.

Rather than a fairness constraint, the weight α directly encodes the importance of a secondary objective concerned with social welfare. Notice that the larger the value of α, the more emphasis placed on the welfare objective, and the higher the achieved value of $\mathcal{U}_\mathsf{W}$. This framework rests heavily on the impact or welfare model which determines the scores w. Rolf et al. [RSD+20] examine the implications when the welfare and profit scores are not perfectly known.

Fairness as Welfare

Though it may seem that the weighted optimization of profit and welfare in this framework is completely distinct from our previous discussions of policies constrained to be fair, we now demonstrate that fairness can be represented as a special case. In particular, we show that profit maximization with group fairness constraints corresponds to multi-objective optimization over profit and welfare for an induced definition of welfare. This connection illustrates that fairness constraints are, in some sense, a special case of a broader definition of welfare-aware machine learning.

We begin with a general connection that applies to a wide class of constrained utility maximization problems. Consider a population partitioned into subgroups $\mathsf{j} \in \Omega$ and a classifier which has access to the profit score p of each individual. In this case, the policies decompose over groups such that $\pi = (\pi_\mathsf{j})_{\mathsf{j}\in\Omega}$. The fairness-constrained problem maximizes $\mathcal{U}_\mathsf{P}(\pi)$ subject to a fairness constraint.

For a large class of fairness criteria including `DemParity` and `EqOpt`, we can restrict our attention to threshold policies $\pi_\mathsf{j}(p) = \mathbf{1}\{p \geq t_\mathsf{j}\}$ where t_j are group-dependent thresholds. Notice that the unconstrained solution would simply be $\pi^{\texttt{MaxUtil}}(p) = \mathbf{1}\{p \geq 0\}$ for all groups. For this reason, we consider groups with $t_\mathsf{j} < 0$ as comparatively *disadvantaged* (since their threshold increases in the absence of fairness constraints) and $t_\mathsf{j} > 0$ as *advantaged*.

In this setting, there exist fixed welfare scores w such that a multi-objective framework would arrive at the same solution policy for any population.

Proposition 5.3.2. *Any fairness-constrained threshold policy giving rise to thresholds* $(t_{\mathsf{j}}^\star)_{\mathsf{j}\in\Omega}$ *is equivalent to a set of* α*-Pareto policies for* $\alpha \in (0,1)$ *in* (5.3.2) *with welfare scores fixed within each group and defined as*

$$w_{\mathsf{j}} = -\frac{1-\alpha}{\alpha} t_{\mathsf{j}}^\star \ .$$

In particular, w_{j} *and* $t_{\mathsf{j}}^\star$ *have opposite signs for all settings of* $\alpha \in (0,1)$*, and any relative scale between them achieved by some choice of* α*.*

Proof. The equivalence follows by comparing the policies

$$\pi_\alpha(w,p) = \mathbf{1}\{\alpha w + (1-\alpha)p \geq 0\} \quad \text{and} \quad \pi_{\text{fair},\mathsf{j}}(p) = \mathbf{1}\{p \geq t_{\mathsf{j}}^\star\} \ .$$

Restricting the choice to a fixed score within each group yields the expression

$$w_{\mathsf{j}} = -\frac{1-\alpha}{\alpha} t_{\mathsf{j}}^\star =: -c t_{\mathsf{j}}^\star \ .$$

Thus we have that $w_{\mathsf{j}} \propto -t_{\mathsf{j}}^*$ for all $_{\mathsf{j}}$. Further, notice that for any $c > 0$ there exists some $\alpha \in (0,1)$ achieving that c with $\alpha = \frac{1}{1+c}$. □

Notice that disadvantaged groups (with negative thresholds) are assigned a positive welfare weight, while advantaged groups (with positive thresholds) are assigned a negative welfare weight. This highlights a particular interpretation of fairness constraints as assigning a social value to individuals of a group which is disadvantaged.

Tradeoffs Between Fairness and Profit

While the result presented above is valid for a wide class of policies, it does not shed light on the trade-off between profit and fairness. We now consider soft fairness constraints, restricting our attention to two groups under `DemParity` for simplicity:

$$\begin{aligned} \pi_{\text{fair}}^\epsilon \in \underset{\substack{\pi=(\pi_{\mathsf{A}},\pi_{\mathsf{B}})\\ \beta=(\beta_{\mathsf{A}},\beta_{\mathsf{B}})}}{\text{argmax}} \ & \mathcal{U}_{\mathsf{P}}(\pi) \\ \text{s.t. } & \mathbb{E}_{\mathsf{j}}[\pi_{\mathsf{j}}(p)] = \beta_{\mathsf{j}} \ \ \mathsf{j} \in \{\mathsf{A},\mathsf{B}\}, \ \ |\beta_{\mathsf{A}} - \beta_{\mathsf{B}}| \leq \epsilon \end{aligned} \tag{5.3.3}$$

where $\mathbb{E}_{\mathsf{j}}$ is the expectation taken over all members of group j. We note that with additional mild assumptions, our arguments extend naturally to other criteria, including equal opportunity.

We now show how this inexact fairness setting can be modeled equivalently by the multi-objective framework. Recall that we assume that the distribution of the profit score p has continuous support within these populations. This is a departure from the previous

setting, but it simplifies the argument considerably. The following proposition shows that the that solution to the constrained profit maximization problem (5.3.3) changes monotonically with the fairness parameter ϵ. Its proof is deferred to Section 5.8.

Proposition 5.3.3. *Suppose that the unconstrained selection rate in group* A *is less than or equal to the unconstrained selection rate in group* B*. Then the policy* $\pi^\epsilon_{\text{fair}} = (\pi^\epsilon_{\mathsf{A}}, \pi^\epsilon_{\mathsf{B}})$ *that optimizes eq.* (5.3.3) *is equivalent to randomized group-dependent threshold policies with thresholds* t^ϵ_{A} *and* t^ϵ_{B} *satisfying the following:*

- $t^\epsilon_{\mathsf{A}} \leq 0$ *for all* $\epsilon \geq 0$ *and* t^ϵ_{A} *is increasing in* ϵ *,*
- $t^\epsilon_{\mathsf{B}} \geq 0$ *for all* $\epsilon \geq 0$ *and* t^ϵ_{B} *is decreasing in* ϵ *.*

Notice that the unconstrained selection rate in group A being less than the unconstrained selection rate in group B is equivalent to A being disadvantaged compared with B. Thus we see that as ϵ increases, the group-dependent optimal thresholds shrink toward the unconstrained profit maximizing solution, where $t_{\mathsf{A}} = t_{\mathsf{B}} = 0$. We present the proof of this result in the next section.

We define the map $\epsilon_{\mathsf{A}}(p) := \epsilon$ s.t. $t^\epsilon_{\mathsf{A}} = p$ for $p \in [t^0_{\mathsf{A}}, 0]$. Proposition 5.3.3 implies that $\epsilon_{\mathsf{A}}(p)$ is increasing in p.

Theorem 5.3.4. *Under the conditions of Proposition 5.3.3, the family of policies* $\pi^\epsilon_{\text{fair}}$ *parametrized by* ϵ *corresponds to a family of* α*-Pareto policies for a fixed choice of group-dependent welfare weightings. In particular, denoting the associated thresholds as* t^ϵ_{A} *and* t^ϵ_{B} *and defining for each individual in* A *with profit score* p*,*

$$w_{\mathsf{A}} = \begin{cases} -\frac{p}{t^{\epsilon_{\mathsf{A}}(p)}_{\mathsf{B}}} & t^0_{\mathsf{A}} \leq p \leq 0 \\ 0 & \textit{otherwise} \end{cases}$$

and for all individuals in B*,*

$$w_{\mathsf{B}} = \begin{cases} -1 & 0 \leq p \leq t^0_{\mathsf{B}} \\ 0 & \textit{otherwise} \end{cases},$$

then for each $\pi^\epsilon_{\text{fair}}$ *there exists an equivalent* $\alpha(\epsilon)$*-Pareto policy* $\pi_{\alpha(\epsilon)}$ *where the trade-off parameter* $\alpha(\epsilon)$ *decreases in* ϵ*.*

Proof. By Proposition 5.3.3, the policy π^ϵ is equivalent to a threshold policy with group dependent thresholds denoted t^ϵ_{A} and t^ϵ_{B}. The group dependent threshold policy $\mathbf{1}\{p \geq t^\epsilon_{\mathsf{j}}\}$ is equivalent to an α-Pareto optimal policy (for some definition of welfare score w) if and only if for all values of p:

$$\mathbf{1}\{p \geq t^\epsilon_{\mathsf{j}}\} = \mathbf{1}\{\alpha(\epsilon)w + (1 - \alpha(\epsilon))p \geq 0\} .$$

It is sufficient to restrict our attention to welfare scores w that depend on profit score and group membership, which we denote as w_{j}^p. Starting with group B, we have that for $0 \le p \le t_{\mathsf{B}}^0$, $w_{\mathsf{B}}^p = -1$, so

$$\pi_{\alpha(\epsilon)} = \mathbf{1}\{-\alpha(\epsilon) + (1-\alpha(\epsilon))p \ge 0\} = \mathbf{1}\{p \ge \frac{\alpha(\epsilon)}{1-\alpha(\epsilon)}\}\,.$$

Thus, equivalence is achieved for this case if $\frac{\alpha(\epsilon)}{1-\alpha(\epsilon)} = t_{\mathsf{B}}^\epsilon$, or equivalently,

$$\alpha(\epsilon) = \frac{t_{\mathsf{B}}^\epsilon}{1 + t_{\mathsf{B}}^\epsilon}\,. \tag{5.3.4}$$

We will use this definition for $\alpha(\epsilon)$ moving forward, and verify that the proposed welfare score definitions work.

We now turn to group A in the case that $t_{\mathsf{A}}^0 \le p \le 0$. We have $w_{\mathsf{A}}^p = \frac{p}{t_{\mathsf{B}}^{\epsilon_{\mathsf{A}}(p)}}$, so

$$\pi_{\alpha(\epsilon)} = \mathbf{1}\{-\frac{t_{\mathsf{B}}^\epsilon}{t_{\mathsf{B}}^{\epsilon_{\mathsf{A}}(p)}}\frac{p}{1+t_{\mathsf{B}}^\epsilon} + \frac{p}{1+t_{\mathsf{B}}^\epsilon} \ge 0\}\,.$$

Because $1 + t_{\mathsf{B}}^\epsilon \ge 0$ and $p \le 0$, the indicator will be one if and only if $t_{\mathsf{B}}^\epsilon \ge t_{\mathsf{B}}^{\epsilon_{\mathsf{A}}(p)}$. By Proposition 5.3.3, this is true if and only if $\epsilon \le \epsilon_{\mathsf{A}}(p)$, which is true if any only if $t_{\mathsf{A}}^\epsilon \le t_{\mathsf{A}}^{\epsilon_{\mathsf{A}}(p)} = p$. This is exactly the condition for $\pi_{\text{fair},\mathsf{A}}^\epsilon$, as desired.

Then finally we consider the remaining cases. In the case that $p \le t_{\mathsf{A}}^0$ in A or $p \le 0$ in B, we have that $\pi_{\text{fair},\mathsf{j}}^\epsilon = 0$ for all ϵ by Proposition 5.3.3. Then as desired, $0 + (1-\alpha(\epsilon))p \le 0$ in this case. In the case that $p \ge 0$ in A or $p \ge t_{\mathsf{B}}^0$ in B, we have that $\pi_{\text{fair},\mathsf{j}}^\epsilon = 1$ for all ϵ. Then as desired, $0 + (1-\alpha(\epsilon))p \ge 0$ in this case.

Finally, we remark on the form of $\alpha(\epsilon)$ in (5.3.4). By proposition 5.3.3, $t_{\mathsf{B}}^\epsilon \ge 0$ and is decreasing in ϵ, so $\alpha(\epsilon)$ is decreasing in ϵ. □

Note that the presented construction of induced welfare scores is not unique. In fact, simply switching the roles of A and B in the proof verifies the alternate definitions,

$$w_{\mathsf{A}} = \begin{cases} 1 & t_{\mathsf{A}}^0 \le p \le 0 \\ 0 & \text{otherwise} \end{cases}, \qquad w_{\mathsf{B}} = \begin{cases} -\frac{p}{t_{\mathsf{A}}^{\epsilon_{\mathsf{B}}(p)}} & 0 \le p \le t_{\mathsf{B}}^0 \\ 0 & \text{otherwise} \end{cases}, \tag{5.3.5}$$

in which case we define $\epsilon_{\mathsf{B}}(p)$ to be the value of ϵ such that $p = t_{\mathsf{B}}^\epsilon$. This construction generalizes in a straightforward manner to multiple groups, where functions similar to $\epsilon_{\mathsf{B}}(p)$ would be defined for each group.

Theorem 5.3.4 shows that it is possible to define fixed welfare scores such that the family of inexact fair policies parametrized by any $\epsilon \ge 0$ in (5.3.3) corresponds to a family

of Pareto-optimal policies parametrized by $\alpha(\epsilon)$. The group-dependent welfare scores are such that $w \geq 0$ for all individuals in the disadvantaged group and $w \leq 0$ in the advantaged group. Furthermore, the induced trade-off parameter $\alpha(\epsilon)$ increases as ϵ decreases.

Fairness constraints can be seen as encoding implicit group-dependent welfare scores for individuals, where members of disadvantaged groups are assigned positive welfare scores and members of advantaged groups assigned negative scores. This correspondence is related to the analysis of welfare scores in Hu and Chen [HC18b], however, our perspective focuses on trade-offs between welfare and profit objectives, in contrast to pure welfare maximization. It is also related to the analysis by Kasy and Abebe [KA21], who use this perspective to argue that fairness criteria suffer from limitations. While some applications may directly call for statistical parity as a criterion, our results emphasize the inevitability of fairness constraints as trade-offs between multiple objectives, and frames these trade-offs explicitly in terms of welfare measures.

5.4 Optimality of Threshold Policies

Now, we move towards statements of the main theorems underlying the results presented in Section 5.2. We begin by establishing notation which we shall use throughout. Recall that $\circ$ denotes the Hadamard product between vectors. We identify functions mapping $\mathcal{X} \to \mathbb{R}$ with vectors in $\mathbb{R}^C$. We also define the group-wise utilities

$$\mathcal{U}_{\mathsf{j}}(\pi_{\mathsf{j}}) := \sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x)\pi_{\mathsf{j}}(x)\boldsymbol{u}(x) \,, \tag{5.4.1}$$

so that for $\pi = (\pi_{\mathsf{A}}, \pi_{\mathsf{B}})$, $\mathcal{U}(\pi) := g_{\mathsf{A}}\mathcal{U}_{\mathsf{A}}(\pi_{\mathsf{A}}) + g_{\mathsf{B}}\mathcal{U}_{\mathsf{B}}(\pi_{\mathsf{B}})$.

First, we formally describe threshold policies, and rigorously justify why we may always assume without loss of generality that the institution adopts policies of this form.

Definition 5.4 (Threshold selection policy)**.** A single group selection policy $\pi \in [0,1]^C$ is called a *threshold policy* if it has the form of a randomized threshold on score:

$$\pi_{c,\gamma} = \begin{cases} 1, & x > c \\ \gamma, & x = c \\ 0, & x < c \end{cases} \text{, for some } c \in [C] \text{ and } \gamma \in (0,1] \,. \tag{5.4.2}$$

As a technicality, if no members of a population have a given score $x \in \mathcal{X}$, there may be multiple threshold policies which yield equivalent selection rates for a given population. To avoid redundancy, we introduce the notation $\pi_{\mathsf{j}} \cong_{\mathcal{D}_{\mathsf{j}}} \pi'_{\mathsf{j}}$ to mean that the set of scores on which π_{j} and π'_{j} differ has probability 0 under $\mathcal{D}_{\mathsf{j}}$. For any distribution $\mathcal{D}_{\mathsf{j}}$, $\cong_{\mathcal{D}_{\mathsf{j}}}$ is an equivalence relation. Moreover, if $\pi_{\mathsf{j}} \cong_{\mathcal{D}_{\mathsf{j}}} \pi'_{\mathsf{j}}$, then π_{j} and π'_{j} both provide the same utility for the institution, induce the same outcomes for individuals in group j, and have the

same selection and true positive rates. Hence, if $(\pi_{\mathsf{A}}, \pi_{\mathsf{B}})$ is an optimal solution to any of `MaxUtil`, `EqOpt`, or `DemParity`, so is any $(\pi'_{\mathsf{A}}, \pi'_{\mathsf{B}})$ for which $\pi_{\mathsf{A}} \cong_{\mathcal{D}_{\mathsf{A}}} \pi'_{\mathsf{A}}$ and $\pi_{\mathsf{B}} \cong_{\mathcal{D}_{\mathsf{B}}} \pi'_{\mathsf{B}}$.

For threshold policies in particular, their equivalence class under $\cong_{\mathcal{D}_{\mathsf{j}}}$ is uniquely determined by the selection rate function,

$$r_{\mathcal{D}_{\mathsf{j}}}(\pi_{\mathsf{j}}) := \sum_{x \in \mathcal{X}} \mathcal{D}_{\mathsf{j}}(x)\pi_{\mathsf{j}}(x) \,, \tag{5.4.3}$$

which denotes the fraction of group j which is selected. Remark that the inverse $r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta_{\mathsf{j}})$ is an equivalence class rather than a single policy. However, $\mathcal{D}_{\mathsf{j}} \circ r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\pi_{\mathsf{j}})$ is well defined, since $\mathcal{D}_{\mathsf{j}} \circ \pi_{\mathsf{j}} = \mathcal{D}_{\mathsf{j}} \circ \pi'_{\mathsf{j}}$ for any two policies in the same equivalence class. All quantities of interest will only depend on policies π_{j} through $\mathcal{D}_{\mathsf{j}} \circ \pi_{\mathsf{j}}$, it does not matter *which* representative of $r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta_{\mathsf{j}})$ we pick. One choice is the set of all threshold policies $\pi_{c,\gamma}$ such that, $\gamma = 1$ if $\mathcal{D}_{\mathsf{j}}(c) = 0$ and $\mathcal{D}_{\mathsf{j}}(c-1) > 0$ if $\gamma = 1$ and $c > 1$.

It turns out the policies which arise in this away are always optimal in the sense that, for a given loan rate β_{j}, the threshold policy $r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta_{\mathsf{j}})$ is the (essentially unique) policy which maximizes both the institution's utility and the utility of the group. We have the following result:

Proposition 5.4.1 (Threshold policies are preferable)**.** *Suppose that $\boldsymbol{u}(x)$ and $\boldsymbol{\Delta}(x)$ are strictly increasing in x. Given any loaning policy π_{j} for population with distribution $\mathcal{D}_{\mathsf{j}}$, then the policy $\pi^{\text{thresh}}_{\mathsf{j}} := r^{-1}_{\mathcal{D}_{\mathsf{j}}}(r_{\mathcal{D}_{\mathsf{j}}}(\pi_{\mathsf{j}})) \in \mathcal{T}_{\text{thresh}}(\mathcal{D}_{\mathsf{j}})$ satisfies*

$$\Delta\mu_{\mathsf{j}}(\pi^{\text{thresh}}_{\mathsf{j}}) \geq \Delta\mu_{\mathsf{j}}(\pi_{\mathsf{j}}) \text{ and } \mathcal{U}_{\mathsf{j}}(\pi^{\text{thresh}}_{\mathsf{j}}) \geq \mathcal{U}_{\mathsf{j}}(\pi_{\mathsf{j}}) \,. \tag{5.4.4}$$

Moreover, both inequalities hold with equality if and only if $\pi_{\mathsf{j}} \cong_{\mathcal{D}_{\mathsf{j}}} \pi^{\text{thresh}}_{\mathsf{j}}$.

The map $\pi_{\mathsf{j}} \mapsto r^{-1}_{\mathcal{D}_{\mathsf{j}}}(r_{\mathcal{D}_{\mathsf{j}}}(\pi_{\mathsf{j}}))$ can be thought of transforming an arbitrary policy π_{j} into a threshold policy with the same selection rate. In this language, the above proposition states that this map never reduces institution utility or individual outcomes. We can also show that optimal `MaxUtil` and `DemParity` policies are threshold policies, as well as all `EqOpt` policies under an additional assumption:

Proposition 5.4.2 (Existence of optimal threshold policies under fairness constraints)**.** *Suppose that $\boldsymbol{u}(x)$ is strictly increasing in x. Then all optimal* `MaxUtil` *policies $(\pi_{\mathsf{A}}, \pi_{\mathsf{B}})$ satisfy $\pi_{\mathsf{j}} \cong_{\mathcal{D}_{\mathsf{j}}} r^{-1}_{\mathcal{D}_{\mathsf{j}}}\left(r_{\mathcal{D}_{\mathsf{j}}}(\pi_{\mathsf{j}})\right)$ for $\mathsf{j} \in \{\mathsf{A}, \mathsf{B}\}$. The same holds for all optimal* `DemParity` *policies, and if in addition $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ is increasing, the same is true for all optimal* `EqOpt` *policies.*

To prove Proposition 5.4.1, we invoke the following general lemma which is proved using standard convex analysis arguments (in Section 5.8):

Lemma 5.4.3. *Let $\boldsymbol{v} \in \mathbb{R}^C$, and let $\boldsymbol{w} \in \mathbb{R}^C_{>0}$, and suppose either that $\boldsymbol{v}(x)$ is increasing in x, and $\boldsymbol{v}(x)/\boldsymbol{w}(x)$ is increasing or, $\forall x \in \mathcal{X}$, $\boldsymbol{w}(x) = 0$. Let $\mathcal{D} \in \text{Simplex}^{C-1}$ and fix $t \in [0, \sum_{x\in\mathcal{X}} \mathcal{D}(x) \cdot \boldsymbol{w}(x)]$. Then any*

$$\pi^* \in \arg\max_{\pi\in[0,1]^C} \langle \boldsymbol{v} \circ \mathcal{D}, \pi\rangle \quad \text{s.t.} \quad \langle \mathcal{D} \circ \boldsymbol{w}, \pi\rangle = t \tag{5.4.5}$$

satisfies $\pi^* \cong_{\mathcal{D}} r^{-1}_{\mathcal{D}}(r_{\mathcal{D}}(\pi^*))$. Moreover, at least one threshold policy maximizer π^* exists.

Proof of Proposition 5.4.1. We will first prove Proposition 5.4.1 for the function $\mathcal{U}_\text{j}$. Given our nominal policy π_j, let $\beta_j = r_{\mathcal{D}_\text{j}}(\pi_\text{j})$. We now apply Lemma 5.4.3 with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = 1$. For this choice of $\boldsymbol{v}$ and $\boldsymbol{w}$, $\langle \boldsymbol{v}, \pi\rangle = \mathcal{U}_\text{j}(\pi)$ and that $\langle \mathcal{D}_j \circ \boldsymbol{w}, \pi = r_{\mathcal{D}_\text{j}}(\pi)$. Then, if $\pi_\text{j} \in \arg\max_\pi \mathcal{U}_\text{j}(\pi)$ s.t. $r_{\mathcal{D}_\text{j}}(\pi) = \beta_\text{j}$, Lemma 5.4.5 implies that $\pi_\text{j} \cong_{\mathcal{D}_\text{j}} r^{-1}_{\mathcal{D}_\text{j}}(r_{\mathcal{D}_\text{j}}(\pi_\text{j}))$.

On the other hand, assume that $\pi_\text{j} \cong_{\mathcal{D}_\text{j}} r^{-1}_{\mathcal{D}_\text{j}}\left(r_{\mathcal{D}_\text{j}}(\pi_\text{j})\right)$. We show that $r^{-1}_{\mathcal{D}_\text{j}}(r_{\mathcal{D}_\text{j}}(\pi_\text{j}))$ is a maximizer; which will imply that π_j is a maximizer since $\pi_\text{j} \cong_{\mathcal{D}_\text{j}} r^{-1}_{\mathcal{D}_\text{j}}(r_{\mathcal{D}_\text{j}}(\pi_\text{j}))$ implies that $\mathcal{U}_\text{j}(\pi_\text{j}) = \pi_\text{j} \cong_{\mathcal{D}_\text{j}} r^{-1}_{\mathcal{D}_\text{j}}(r_{\mathcal{D}_\text{j}}(\pi_\text{j}))$. By Lemma 5.4.3 there exists a maximizer π^*_j, which means that $\pi^*_\text{j} = r^{-1}_{\mathcal{D}_\text{j}}(r_{\mathcal{D}_\text{j}}(\pi^*_\text{j}))$. Since π^*_j is feasible, we must have $r_{\mathcal{D}_\text{j}}(\pi^*_\text{j}) = r_{\mathcal{D}_\text{j}}(\pi_\text{j})$, and thus $\pi^*_\text{j} = r^{-1}_{\mathcal{D}_\text{j}}(r_{\mathcal{D}_\text{j}}(\pi_\text{j}))$, as needed. The same argument follows verbatim if we instead choose $\boldsymbol{v}(x) = \boldsymbol{\Delta}(x)$, and compute $\langle \boldsymbol{v}, \pi\rangle = \Delta\mu_\text{j}(\pi)$. □

We now argue Proposition 5.4.2 for `MaxUtil`, as it is a straightforward application of Lemma 5.4.3. We will prove Proposition 5.4.2 for `DemParity` and `EqOpt` separately in Section 5.5.

Proof of Proposition 5.4.2 for `MaxUtil`**.** `MaxUtil` follows from lemma 5.4.3 with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$, and $t = 0$ and $\boldsymbol{w} = \boldsymbol{0}$.

□

Quantiles and Concavity of the Outcome Curve

To further our analysis, we now introduce left and right quantile functions, allowing us to specify thresholds in terms of both selection rate and score cutoffs.

Definition 5.5 (Upper quantile function)**.** Define Q to be the upper quantile function corresponding to $\mathcal{D}$, i.e.

$$\text{Q}_\text{j}(\beta) = \text{argmax}\{c : \sum_{x=c}^{C} \mathcal{D}_\text{j}(x) > \beta\} \quad \text{and} \quad \text{Q}^+_\text{j}(\beta) := \text{argmax}\{c : \sum_{x=c}^{C} \mathcal{D}_\text{j}(x) \geq \beta\}\,. \tag{5.4.6}$$

Crucially $\text{Q}(\beta)$ is continuous from the right, and $\text{Q}^+(\beta)$ is continuous from the left. Further, $\text{Q}(\cdot)$ and $\text{Q}^+(\cdot)$ allow us to compute derivatives of key functions, like the mapping from selection rate β to the group outcome associated with a policy of that rate, $\Delta\mu(r^{-1}_\pi(\beta))$.

Because we take $\mathcal{D}$ to have discrete support, all functions in this work are *piecewise linear*, so we shall need to distinguish between the left and right derivatives, defined as follows

$$\partial_- f(x) := \lim_{t \to 0^-} \frac{f(x+t) - f(x)}{t} \quad \text{and} \quad \partial_+ f(y) := \lim_{t \to 0^+} \frac{f(y+t) - f(y)}{t} \,. \tag{5.4.7}$$

For f supported on $[a, b]$, we say that f is left- (resp. right-) differentiable if $\partial_- f(x)$ exists for all $x \in (a, b]$ (resp. $\partial_+ f(y)$ exists for all $y \in [a, b)$). We now state the fundamental derivative computation which underpins the results to follow:

Lemma 5.4.4. *Let $\boldsymbol{e}_x$ denote the vector such that $\boldsymbol{e}_x(x) = 1$, and $\boldsymbol{e}_x(x') = 0$ for $x' \neq x$. Then $\mathcal{D}_{\mathsf{j}} \circ r_{\mathcal{D}_{\mathsf{j}}}^{-1}(\beta) : [0,1] \to [0,1]^{\mathsf{C}}$ is continuous, and has left and right derivatives*

$$\partial_+ \left(\mathcal{D}_{\mathsf{j}} \circ r_{\mathcal{D}_{\mathsf{j}}}^{-1}(\beta) \right) = \boldsymbol{e}_{\mathrm{Q}(\beta)} \quad \textit{and} \quad \partial_- \left(\mathcal{D}_{\mathsf{j}} \circ r_{\mathcal{D}_{\mathsf{j}}}^{-1}(\beta) \right) = \boldsymbol{e}_{\mathrm{Q}^+(\beta)} \,. \tag{5.4.8}$$

We defer the proof to Section 5.8. Moreover, Lemma 5.4.4 implies that the outcome curve is concave under the assumption that $\boldsymbol{\Delta}(x)$ is monotone:

Proposition 5.4.5. *Let $\mathcal{D}$ be a distribution over C states. Then $\beta \mapsto \Delta\mu(r_{\mathcal{D}}^{-1}(\beta))$ is concave. In fact, if $\boldsymbol{w}(x)$ is any non-decreasing map from $\mathcal{X} \to \mathbb{R}$, $\beta \mapsto \langle \boldsymbol{w}, r_{\mathcal{D}}^{-1}(\beta) \rangle$ is concave.*

Proof. Recall that a univariate function f is concave (and finite) on $[a, b]$ if and only (a) f is left- and right-differentiable, (b) for all $x \in (a, b)$, $\partial_- f(x) \geq \partial_+ f(x)$ and (c) for any $x > y$, $\partial_- f(x) \leq \partial_+ f(y)$.

Observe that $\Delta\mu(r_{\mathcal{D}}^{-1}(\beta)) = \langle \boldsymbol{\Delta}, \mathcal{D} \circ r_{\mathcal{D}}^{-1}(\beta) \rangle$. By Lemma 5.4.4, $\mathcal{D} \circ r_{\mathcal{D}}^{-1}(\beta)$ has right and left derivatives $\boldsymbol{e}_{\mathrm{Q}(\beta)}$ and $\boldsymbol{e}_{\mathrm{Q}^+(\beta)}$. Hence, we have that

$$\partial_+ \Delta\mu(\beta_{\mathsf{B}}) = \boldsymbol{\Delta}(\mathrm{Q}(\beta_{\mathsf{B}})) \quad \text{and} \quad \partial_- \Delta\mu(\beta_{\mathsf{B}}) = \boldsymbol{\Delta}(\mathrm{Q}^+(\beta_{\mathsf{B}})) \,. \tag{5.4.9}$$

Using the fact that $\boldsymbol{\Delta}(x)$ is monotone, and that $\mathrm{Q} \leq \mathrm{Q}^+$, we see that $\partial_+ \Delta\mu(f_{\mathcal{D}}^{-1}(\beta_{\mathsf{B}})) \leq \partial_- \Delta\mu(f_{\mathcal{D}}^{-1}(\beta_{\mathsf{B}}))$, and that $\partial \Delta\mu(f_{\mathcal{D}}^{-1}(\beta_{\mathsf{B}}))$ and $\partial_+ \Delta\mu(f_{\mathcal{D}}^{-1}(\beta_{\mathsf{B}}))$ are non-increasing, from which it follows that $\Delta\mu(f_{\mathcal{D}}^{-1}(\beta_{\mathsf{B}}))$ is concave. The general concavity result holds by replacing $\boldsymbol{\Delta}(x)$ with $\boldsymbol{w}(x)$. □

5.5 Main Characterization Results

We are now ready to present and prove theorems that characterize the selection rates under fairness constraints, namely `DemParity` and `EqOpt`. These characterizations are crucial for proving the results in Section 5.2. We also show that these computations generalize readily to other linear constraints.

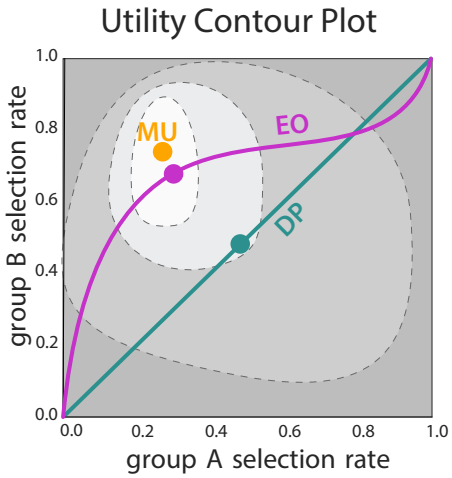

Figure 5.4: Considering the utility as a function of selection rates, fairness constraints correspond to restricting the optimization to one-dimensional curves. The `DemParity` (DP) constraint is a straight line with slope 1, while the `EqOpt` (EO) constraint is a curve given by the graph of the transfer function $G^{(\mathsf{A}\to\mathsf{B})}$. The derivatives considered throughout Section 5.5 are taken with respect to the selection rate β_A (horizontal axis); projecting the EO and DP constraint curves to the horizontal axis recovers concave utility curves such as those shown in the lower panel of Figure 5.2 (where `MaxUtil` in is represented by a horizontal line through the MU optimal solution).

A Characterization Theorem for `DemParity`

In this section, we provide a theorem that gives an explicit characterization for the range of selection rates β_A for A when the bank loans according to `DemParity`. Observe that the `DemParity` objective corresponds to solving the following linear program:

$$\max_{\pi=(\pi_\mathsf{A},\pi_\mathsf{B})\in[0,1]^{2C}} \mathcal{U}(\pi) \quad \text{s.t.} \quad \langle \mathcal{D}_\mathsf{A}, \pi_\mathsf{A}\rangle = \langle \mathcal{D}_\mathsf{B}, \pi_\mathsf{B}\rangle \,.$$

Let us introduce the auxiliary variable $\beta := \langle \mathcal{D}_\mathsf{A}, \pi_\mathsf{A}\rangle = \langle \mathcal{D}_\mathsf{B}, \pi_\mathsf{B}\rangle$ corresponding to the selection rate which is held constant across groups, so that all feasible solutions lie on the green DP line in Figure 5.4. We can then express the following equivalent linear program:

$$\max_{\substack{\pi=(\pi_\mathsf{A},\pi_\mathsf{B})\in[0,1]^{2C},\\ \beta\in[0,1]}} \mathcal{U}(\pi) \quad \text{s.t.} \quad \beta = \langle \mathcal{D}_\mathsf{j}, \pi_\mathsf{j}\rangle,\ \mathsf{j}\in\{\mathsf{A},\mathsf{B}\} \,.$$

This is equivalent because, for a given β, Proposition 5.4.2 says that the utility maximizing policies are of the form $\pi_\mathsf{j} = r^{-1}_{\mathcal{D}_\mathsf{j}}(\beta)$. We now prove this:

Proof of Proposition 5.4.2 for `DemParity`. Noting that $r_{\mathcal{D}_\mathsf{j}}(\pi_\mathsf{j}) = \langle \mathcal{D}_\mathsf{j}, \pi_\mathsf{j}\rangle$, we see that, by Lemma 5.4.3, under the special case where $v(x) = u(x)$ and $w(x) = 1$, the optimal solution

$(\pi^*_{\mathsf{A}}(\beta), \pi^*_{\mathsf{B}}(\beta))$ for fixed $r_{\mathcal{D}_{\mathsf{A}}}(\pi_{\mathsf{A}}) = r_{\mathcal{D}_{\mathsf{B}}}(\pi_{\mathsf{B}}) = \beta$ can be chosen to coincide with the threshold policies. Optimizing over β, the global optimal must coincide with thresholds.

□

Hence, any optimal policy is equivalent to the threshold policy $\pi = (r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta), r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta))$, where β solves the following optimization:

$$\max_{\beta \in [0,1]} \mathcal{U}\left(\left(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta), r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\right)\right) . \tag{5.5.1}$$

We shall show that the above expression is in fact a *concave* function in β, and hence the set of optimal selection rates can be characterized by first order conditions. This is presented formally in the following theorem:

Theorem 5.5.1 (Selection rates for `DemParity`)**.** *The set of optimal selection rates β^* satisfying* (5.5.1) *forms a continuous interval $[\beta^-_{\texttt{DemParity}}, \beta^+_{\texttt{DemParity}}]$, such that for any $\beta \in [0,1]$, we have*

$$\begin{aligned} \beta < \beta^-_{\texttt{DemParity}} \quad &\text{if} \quad g_{\mathsf{A}}\boldsymbol{u}\left(\mathrm{Q}_{\mathsf{A}}(\beta)\right) + g_{\mathsf{B}}\boldsymbol{u}\left(\mathrm{Q}_{\mathsf{B}}(\beta)\right) > 0 , \\ \beta > \beta^+_{\texttt{DemParity}} \quad &\text{if} \quad g_{\mathsf{A}}\boldsymbol{u}\left(\mathrm{Q}^+_{\mathsf{A}}(\beta)\right) + g_{\mathsf{B}}\boldsymbol{u}\left(\mathrm{Q}^+_{\mathsf{B}}(\beta)\right) < 0 . \end{aligned}$$

Proof. Note that we can write

$$\mathcal{U}\left(\left(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta), r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\right)\right) = g_{\mathsf{A}}\langle \boldsymbol{u}, \mathcal{D}_{\mathsf{A}} \circ r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta)\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \mathcal{D}_{\mathsf{B}} \circ r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\rangle .$$

Since $\boldsymbol{u}(x)$ is non-decreasing in x, Proposition 5.4.5 implies that $\beta \mapsto \mathcal{U}\left(\left(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta), r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\right)\right)$ is concave in β. Hence, all optimal selection rates β^* lie in an interval $[\beta^-, \beta^+]$. To further characterize this interval, let us us compute left- and right-derivatives.

$$\begin{aligned} \partial_+\mathcal{U}\left(\left(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta), r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\right)\right) &= \partial_+ g_{\mathsf{A}}\langle \boldsymbol{u}, \mathcal{D}_{\mathsf{A}} \circ r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta)\rangle + \partial_+ g_{\mathsf{B}}\langle \boldsymbol{u}, \mathcal{D}_{\mathsf{B}} \circ r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\rangle \\ &= g_{\mathsf{A}}\langle \boldsymbol{u}, \partial_+\left(\mathcal{D}_{\mathsf{A}} \circ r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta)\right)\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \partial_+\left(\mathcal{D}_{\mathsf{B}} \circ r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\right)\rangle \\ &\overset{\text{Lemma 5.4.4}}{=} g_{\mathsf{A}}\langle \boldsymbol{u}, \boldsymbol{e}_{\mathrm{Q}_{\mathsf{A}}(\beta)}\rangle + g_{\mathsf{B}}\langle \boldsymbol{u}, \boldsymbol{e}_{\mathrm{Q}_{\mathsf{B}}(\beta)}\rangle \\ &= g_{\mathsf{A}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) + g_{\mathsf{B}}\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) . \end{aligned}$$

The same argument shows that

$$\partial_-\mathcal{U}((r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta), r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta))) = g_{\mathsf{A}}\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{A}}(\beta)) + g_{\mathsf{B}}\boldsymbol{u}(\mathrm{Q}^+_{\mathsf{B}}(\beta)).$$

By concavity of $\mathcal{U}\left(\left(r^{-1}_{\mathcal{D}_{\mathsf{A}}}(\beta), r^{-1}_{\mathcal{D}_{\mathsf{B}}}(\beta)\right)\right)$, a positive right derivative at β implies that $\beta < \beta^*$ for all β^* satisfying (5.5.1), and similarly, a negative left derivative at β implies that $\beta > \beta^*$ for all β^* satisfying (5.5.1).

□

With a result of the above form, we can now easily prove statements such as that in Corollary 5.2.3 (see Section 5.8 for proofs), by fixing a selection rate of interest (e.g. β_0) and inverting the inequalities in Theorem 5.5.1 to find the exact population proportions under which, for example, `DemParity` results in a higher selection rate than β_0.

`EqOpt` and General Constraints

Next, we will provide a theorem that gives an explicit characterization for the range of selection rates β_{A} for A when the bank loans according to `EqOpt`. Observe that the `EqOpt` objective corresponds to solving the following linear program:

$$\max_{\pi=(\pi_{\mathsf{A}},\pi_{\mathsf{B}})\in[0,1]^{2C}} \mathcal{U}(\pi) \quad \text{s.t.} \quad \langle \boldsymbol{w}_{\mathsf{A}} \circ \mathcal{D}_{\mathsf{A}}, \pi_{\mathsf{A}} \rangle = \langle \boldsymbol{w}_{\mathsf{B}} \circ \mathcal{D}_{\mathsf{B}}, \pi_{\mathsf{B}} \rangle \ , \tag{5.5.2}$$

where $\boldsymbol{w}_{\mathsf{j}} = \frac{\boldsymbol{\rho}}{\langle \boldsymbol{\rho}, \mathcal{D}_{\mathsf{j}} \rangle}$. This problem is similar to the demographic parity optimization in (5.5.1), except for the fact that the constraint includes the weights. Whereas we parameterized demographic parity solutions in terms of the acceptance rate β in equation (5.5.1), we will parameterize equation (5.5.2) in terms of the true positive rate (TPR), $t := \langle \boldsymbol{w}_{\mathsf{A}} \circ \mathcal{D}_{\mathsf{A}}, \pi_{\mathsf{A}} \rangle$. Thus, (5.5.2) becomes

$$\max_{t\in[0,t_{\max}]} \max_{(\pi_{\mathsf{A}},\pi_{\mathsf{B}})\in[0,1]^{2C}} \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}}(\pi_{\mathsf{j}}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_{\mathsf{j}} \circ \mathcal{D}_{\mathsf{j}}, \pi_{\mathsf{j}} \rangle = t, \ \mathsf{j} \in \{\mathsf{A},\mathsf{B}\} \ , \tag{5.5.3}$$

where $t_{\max} = \min_{\mathsf{j}\in\{\mathsf{A},\mathsf{B}\}}\{\langle \mathcal{D}_{\mathsf{j}}, \boldsymbol{w}_{\mathsf{j}} \rangle\}$ is the largest possible TPR. The magenta EO curve in Figure 5.4 illustrates that feasible solutions to this optimization problem lie on a curve parametrized by t. Note that the objective function decouples for $\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}$ for the inner optimization problem,

$$\max_{\pi_{\mathsf{j}}\in[0,1]^C} \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}}(\pi_{\mathsf{j}}) \quad \text{s.t.} \quad \langle \boldsymbol{w}_{\mathsf{j}} \circ \mathcal{D}_{\mathsf{j}}, \pi_{\mathsf{j}} \rangle = t \ . \tag{5.5.4}$$

We will now show that all optimal solutions for this inner optimization problem are $\mathcal{D}_{\mathsf{j}}$-a.e. equal to a threshold policy, and thus can be written as $r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta_{\mathsf{j}})$, depending only on the resulting selection rate.

Proof of Proposition 5.4.2 for `EqOpt`**.** We apply Lemma 5.4.3 to the inner optimization in (5.5.4) with $\boldsymbol{v}(x) = \boldsymbol{u}(x)$ and $\boldsymbol{w}(x) = \frac{\boldsymbol{\rho}(x)}{\langle \boldsymbol{\rho}, \mathcal{D}_{\mathsf{j}} \rangle}$. The claim follows from the assumption that $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ is increasing by optimizing over t. □

This selection rate β_{j} is uniquely determined by the TPR t:

Lemma 5.5.2. *Suppose that $\boldsymbol{w}(x) > 0$ for all x. Then the function*

$$T_{\mathsf{j},\boldsymbol{w}_{\mathsf{j}}}(\beta) := \langle r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta), \mathcal{D}_{\mathsf{j}} \circ \boldsymbol{w}_{\mathsf{j}} \rangle$$

is a bijection from $[0,1]$ to $[0, \langle \mathcal{D}_{\mathsf{j}}, \boldsymbol{w} \rangle]$.

Hence, for any $t \in [0, t_{\max}]$, the mapping from TPR to acceptance rate, $T^{-1}_{\mathsf{j},w_\mathsf{j}}(t)$, is well defined and any solution to (5.5.4) is $\mathcal{D}_\mathsf{j}$-a.e. equal to the policy $r^{-1}_{\mathcal{D}_\mathsf{j}}(T^{-1}_{\mathsf{j},w_\mathsf{j}}(t))$. Thus (5.5.3) reduces to

$$\max_{t\in[0,t_{\max}]} \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j} \mathcal{U}_\mathsf{j} \left(r^{-1}_{\mathcal{D}_\mathsf{j}} \left(T^{-1}_{\mathsf{j},w_\mathsf{j}}(t) \right) \right) . \tag{5.5.5}$$

The above expression parametrizes the optimization problem in terms of a single variable. We shall show that the above expression is in fact a *concave* function in t, and hence the set of optimal selection rates can be characterized by first order conditions. This is presented formally in the following theorem:

Theorem 5.5.3 (Selection rates for `EqOpt`)**.** *The set of optimal selection rates β^* for group A satsifying* (5.5.3) *forms a continuous interval $[\beta^-_{\texttt{EqOpt}}, \beta^+_{\texttt{EqOpt}}]$, such that for any $\beta \in [0,1]$, we have*

$$\beta < \beta^-_{\texttt{EqOpt}} \quad \textit{if} \quad g_\mathsf{A} \frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{A}(\beta))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}_\mathsf{A}(\beta))} + g_\mathsf{B} \frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_w(\beta)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_w(\beta)))} > 0 ,$$

$$\beta > \beta^+_{\texttt{EqOpt}} \quad \textit{if} \quad g_\mathsf{A} \frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{A}(\beta))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}^+_\mathsf{A}(\beta))} + g_\mathsf{B} \frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_w(\beta)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}^+_\mathsf{B}(G^{(\mathsf{A}\to\mathsf{B})}_w(\beta)))} < 0 .$$

Here, $G^{(\mathsf{A}\to\mathsf{B})}_w(\beta) := T^{-1}_{\mathsf{B},w_\mathsf{B}}(T^{-1}_{\mathsf{A},w_\mathsf{A}}(\beta))$ denotes the (well-defined) map from selection rates β_A for A to the selection rate β_B for B such that the policies $\pi^*_\mathsf{A} := r^{-1}_{\mathcal{D}_\mathsf{A}}(\beta_\mathsf{A})$ and $\pi^*_\mathsf{B} := r^{-1}_{\mathcal{D}_\mathsf{B}}(\beta_\mathsf{B})$ satisfy the constraint in (5.5.2)*.*

Proof. Starting with the equivalent problem in (5.5.5), we use the concavity result of Lemma 5.8.2. Because the objective function is the positive weighted sum of two concave functions, it is also concave. Hence, all optimal true positive rates t^* lie in an interval $[t^-, t^+]$. To further characterize $[t^-, t^+]$, we can compute left- and right-derivatives, again using the result of Lemma 5.8.2.

$$\begin{aligned}
\partial_+ \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j} \mathcal{U}_\mathsf{j} \left(r^{-1}_{\mathcal{D}_\mathsf{j}}(T^{-1}_{\mathsf{j},w_\mathsf{j}}(t)) \right) &= g_\mathsf{A} \partial_+ \mathcal{U}_\mathsf{A} \left(r^{-1}_{\mathcal{D}_\mathsf{A}}(T^{-1}_{\mathsf{A},w_\mathsf{A}}(t)) \right) + g_\mathsf{A} \partial_+ \mathcal{U}_\mathsf{A} \left(r^{-1}_{\mathcal{D}_\mathsf{A}}(T^{-1}_{\mathsf{A},w_\mathsf{A}}(t)) \right) \\
&= g_\mathsf{A} \frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{A}(T^{-1}_{\mathsf{A},w_\mathsf{A}}(t)))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}_\mathsf{A}(T^{-1}_{\mathsf{A},w_\mathsf{A}}(t)))} + g_\mathsf{B} \frac{\boldsymbol{u}(\mathrm{Q}_\mathsf{B}(T^{-1}_{\mathsf{B},w_\mathsf{B}}(t)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}_\mathsf{B}(T^{-1}_{\mathsf{B},w_\mathsf{B}}(t)))}
\end{aligned}$$

The same argument shows that

$$\partial_- \sum_{j\in\{\mathsf{A},\mathsf{B}\}} g_\mathsf{j} \mathcal{U}_\mathsf{j} \left(r^{-1}_{\mathcal{D}_\mathsf{j}}(T^{-1}_{\mathsf{j},w_\mathsf{j}}(t)) \right) = g_\mathsf{A} \frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{A}(T^{-1}_{\mathsf{A},w_\mathsf{A}}(t)))}{\boldsymbol{w}_\mathsf{A}(\mathrm{Q}^+_\mathsf{A}(T^{-1}_{\mathsf{A},w_\mathsf{A}}(t)))} + g_\mathsf{B} \frac{\boldsymbol{u}(\mathrm{Q}^+_\mathsf{B}(T^{-1}_{\mathsf{B},w_\mathsf{B}}(t)))}{\boldsymbol{w}_\mathsf{B}(\mathrm{Q}^+_\mathsf{B}(T^{-1}_{\mathsf{B},w_\mathsf{B}}(t)))}.$$

By concavity, a positive right derivative at t implies that $t < t^*$ for all t^* satisfying (5.5.5), and similarly, a negative left derivative at t implies that $t > t^*$ for all t^* satisfying (5.5.5).

Finally, by Lemma 5.5.2, this interval in t uniquely characterizes an interval of acceptance rates. Thus we translate directly into a statement about the selection rates β for group A by seeing that $T^{-1}_{\mathsf{A},\boldsymbol{w}_\mathsf{A}}(t) = \beta$ and $T^{-1}_{\mathsf{B},\boldsymbol{w}_\mathsf{B}}(t) = G^{(\mathsf{A}\to\mathsf{B})}_{\boldsymbol{w}}(\beta)$. □

Lastly, we remark that the results derived in this section go through verbatim for any linear constraint of the form $\langle \boldsymbol{w}, \mathcal{D}_\mathsf{A} \circ \pi_\mathsf{A} \rangle = \langle \boldsymbol{w}, \mathcal{D}_\mathsf{B} \circ \pi_\mathsf{B} \rangle$, as long as $\boldsymbol{u}(x)/\boldsymbol{w}(x)$ is increasing in x, and $\boldsymbol{w}(x) > 0$.

5.6 Simulations

We examine the outcomes induced by fairness constraints in the context of FICO scores for two race groups. FICO scores are a proprietary classifier widely used in the United States to predict credit worthiness. Our FICO data is based on a sample of 301,536 TransUnion TransRisk scores from 2003 [US 07], preprocessed by Hardt, Price, and Srebro [HPS16]. These scores, corresponding to x in our model, range from 300 to 850 and are meant to predict credit risk. Empirical data labeled by race allows us to estimate the distributions $\mathcal{D}_\mathsf{j}$, where j represents race, which is restricted to two values: Black and White non-Hispanic (labeled "White" in figures). In this dataset 12% of the population is Black while 88% is White.

Individuals were labeled as defaulted if they failed to pay a debt for at least 90 days on at least one account in the ensuing 18-24 month period; we use this data to estimate the success probability given score, $\boldsymbol{\rho}_\mathsf{j}(x)$, which we allow to vary by group to match the empirical data (see Figure 5.5). Our outcome curve framework allows for this relaxation; however, this discrepancy can also be attributed to group-dependent mis-measurement of score, and adjusting the scores accordingly would allow for a single $\boldsymbol{\rho}(x)$. We use the success probabilities to define the affine utility and score change functions defined in Example 5.1. We model individual penalties as a score drop of $c_- = -150$ in the case of a default, and in increase of $c_+ = 75$ in the case of successful repayment.

Delayed Impact of Fairness Criteria

In Figure 5.6, we display the empirical CDFs along with selection rates resulting from different loaning strategies for two different settings of bank utilities. In the case that the bank experiences a loss/profit ratio of $\frac{u_-}{u_+} = -10$, no fairness criteria surpass the active harm rate β_0; however, in the case of $\frac{u_-}{u_+} = -4$, `DemParity` over-loans, in line with the statement in Corollary 5.2.3.

These results are further examined in Figure 5.7, which displays the normalized outcome curves and the utility curves for both the Black and the White group. To plot the `MaxUtil` utility curves, the group that is not on display has selection rate fixed at $\beta^{\texttt{MaxUtil}}$.

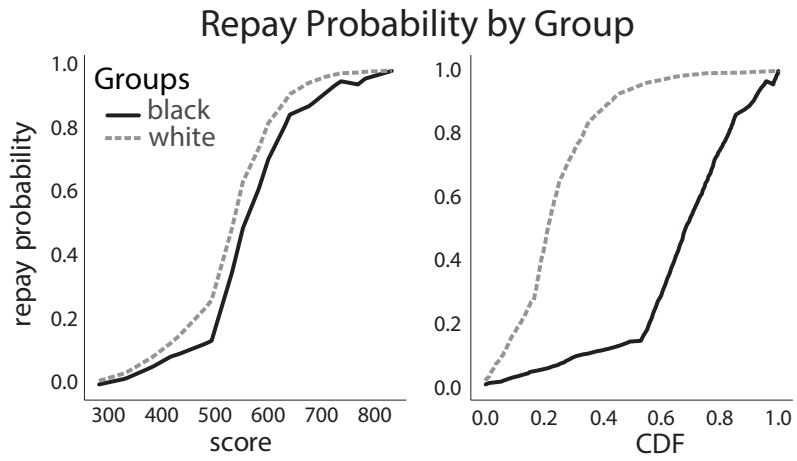

Figure 5.5: The empirical payback rates as a function of credit score and CDF for both groups from the TransUnion TransRisk dataset.

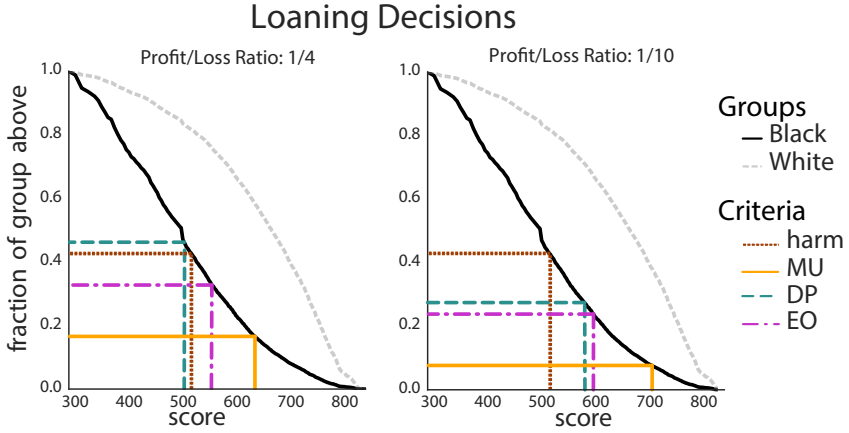

Figure 5.6: The empirical CDFs of both groups are plotted along with the decision thresholds resulting from `MaxUtil`, `DemParity`, and `EqOpt` for a model with bank utilities set to (a) $\frac{u_-}{u_+} = -4$ and (b) $\frac{u_-}{u_+} = -10$. The threshold for active harm is displayed; in (a) `DemParity` causes active harm while in (b) it does not. `EqOpt` and `MaxUtil` never cause active harm.

In this figure, the top panel corresponds to the average change in credit scores for each group under different loaning rates β; the bottom panels shows the corresponding *total* utility $\mathcal{U}$ (summed over both groups and weighted by group population sizes) for the bank.

Figure 5.7 highlights that the position of the utility optima in the lower panel determines the loan (selection) rates. In this specific instance, the utility and change ratios are fairly close, $\frac{u_-}{u_+} = -4$, and $\frac{c_-}{c_+} = -2$, meaning that the bank's profit motivations align with individual outcomes to some extent. Here, we can see that `EqOpt` loans much closer to optimal than `DemParity`.

Although one might hope for decisions made under fairness constraints to positively affect the Black group, we observe the opposite behavior. The `MaxUtil` policy (solid orange line) and the `EqOpt` policy result in similar expected credit score change for the Black group. However, `DemParity` (dashed green line) causes a negative expected credit score change in the Black group, corresponding to active harm. For the White group, the bank utility curve has almost the same shape under the fairness criteria as it does under `MaxUtil`, the main difference being that fairness criteria lowers the total expected profit from this group.

This behavior stems from a discrepancy in the outcome and profit curves for each population. While incentives for the bank and positive results for individuals are somewhat aligned for the majority group, under fairness constraints, they are more heavily misaligned in the minority group, as seen in graphs (left) in Figure 5.7. We remark that in other settings where the *unconstrained* profit maximization is misaligned with individual outcomes (e.g., when $\frac{u_-}{u_+} = -10$), fairness criteria may perform more favorably for the minority group by pulling the utility curve into a shape consistent with the outcome curve.

Connections to Welfare

We demonstrate the induced welfare scores in the context of a credit lending scenario. In this context, we define the profit score as the expected gain from lending to an individual,

$$p = u_+ \cdot \rho + u_- \cdot (1 - \rho) \,,$$

where ρ is the individual's probability of repayment and we set $u_+ = 1$ and $u_- = -4$. The empirical cumulative density functions are displayed in Figure 5.8a.

We solve the relaxed fairness problem (5.3.3) using a two dimensional grid over thresholds. Figure 5.8b shows the thresholds $(t^\epsilon_{\mathsf{A}}, t^\epsilon_{\mathsf{B}})$ for various values of ϵ. As predicted by Proposition 5.3.3, they are generally shrinking in magnitude towards $p = 0$, and the threshold is negative for the Black group and positive for the White group. Due to the discrete support of the empirical distributions, the monotonicity of these thresholds is not perfect.

Lastly, we use these threshold values to compute induced welfare scores using the construction given in Theorem 5.3.4. Figure 5.8c shows how welfare scores are assigned

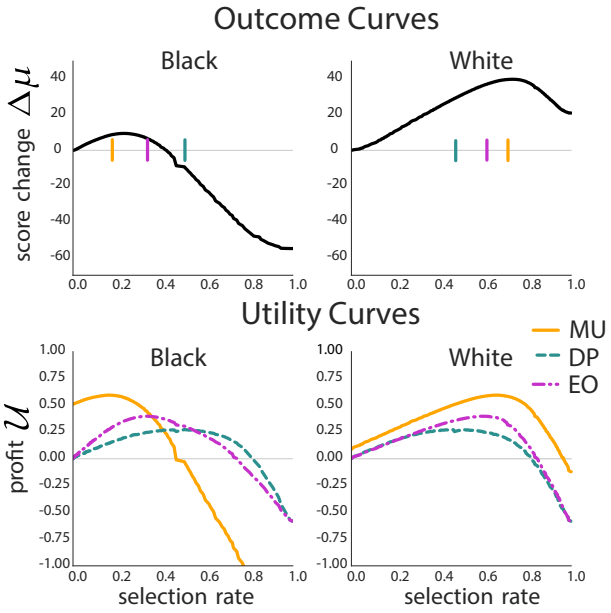

Figure 5.7: The outcome and utility curves are plotted for both groups against the group selection rates. The relative positions of the utility maxima determine the position of the decision rule thresholds. We hold $\frac{u_-}{u_+} = -4$ as fixed.

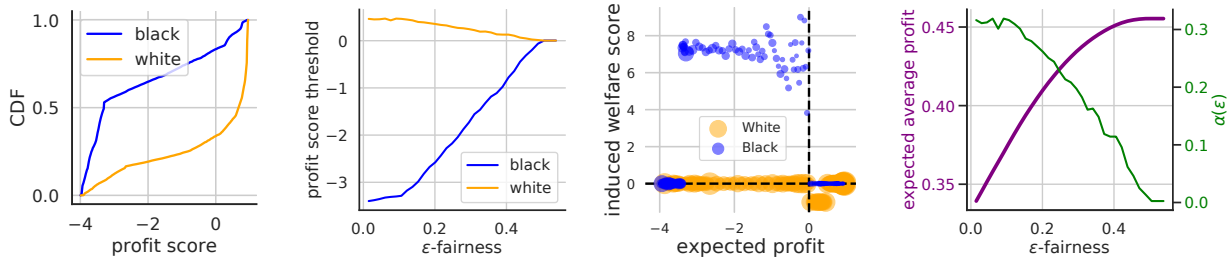

Figure 5.8: Empirical example of how trade-offs between profit and fairness in lending can be equivalently encoded by a multi-objective framework. In (a), cumulative density of profit scores group by race, where 88% of the population is White while 12% is Black. In (b), thresholds for ϵ-fair policies. As ϵ increases, the magnitude of the thresholds decrease. In (c), distribution of profit and induced welfare scores. Marker size corresponds to population sizes. In (d), the fairness parameter ϵ determines the profit trade-off and corresponds to the welfare weight α.

depending on group and on profit score. The fact that only some individuals have nonzero welfare scores highlights the limitations of fairness criteria to affect only individuals within a bounded interval of the max profit solution. Individuals with sufficiently low repayment probabilities will never be considered by "fair" policies. Figure 5.8d shows the welfare weight $\alpha(\epsilon)$, which is generally decreasing in ϵ, though not monotonically.

5.7 Conclusion and Discussion

We focus on the impact of a selection policy over a single epoch. The motivation is that the designer of a system usually has an understanding of the time horizon after which the system is evaluated and possibly redesigned. Formally, nothing prevents us from repeatedly applying our model and tracing changes over multiple epochs. Indeed, follow up work by D'Amour et al. [DSA+20] analyzes the equilibrium behavior of the model we propose. In reality, however, it is plausible that over greater time periods, economic background variables might dominate the effect of selection.

In this chapter, we advocate for a view toward long-term outcomes in the discussion of "fair" machine learning. We argue that without a careful model of delayed outcomes, we cannot foresee the impact a fairness criterion would have if enforced as a constraint on a classification system. However, if such an accurate outcome model is available, we show that there are more direct ways to optimize for positive outcomes than via existing fairness criteria. The extent to which such a solution could form the basis of successful regulation depends on the accuracy of the available outcome model. We further draw connections to a more general framework, in which the simultaneous optimization of profit and welfare are considered. This allows us to more broadly frame the goals of decisions, illustrating how to design policies that prioritize the social impact of an algorithmic decision from the outset, rather than as an afterthought.

A pertinent question is whether our model of outcomes is rich enough to faithfully capture realistic phenomena. By focusing on the impact that selection has on individuals at a given score, we model the effects for those *not* selected as zero-mean. For example, not getting a loan in our model has no negative effect on the credit score of an individual. This does not mean that wrongful rejection (i.e., a false negative) has no visible manifestation in our model. If a classifier has a higher false negative rate in one group than in another, we expect the classifier to increase the disparity between the two groups (under natural assumptions). In other words, in our outcome-based model, the harm of denied opportunity manifests as growing disparity between the groups. The cost of a false negative could also be incorporated directly into the outcome-based model by a simple modification (see Footnote 1). This may be fitting in some applications where the immediate impact of a false negative to the individual is not zero-mean, but significantly reduces their future success probability.

In essence, the formalism we propose requires us to understand the two-variable causal mechanism that translates decisions to outcomes. This can be seen as relaxing the require-

ments compared with recent work on avoiding discrimination through causal reasoning that often required stronger assumptions [KLRS17; NS18; KRP+17]. In particular, these works required knowledge of how sensitive attributes (such as gender, race, or proxies thereof) causally relate to various other variables in the data. Our model avoids the delicate modeling step involving the sensitive attribute, and instead focuses on an arguably more tangible economic mechanism. Nonetheless, depending on the application, such an understanding might necessitate greater domain knowledge and additional research into the specifics of the application. This is consistent with much scholarship that points to the context-sensitive nature of fairness in machine learning.

5.8 Omitted Proofs

Proofs of Main Results

We remark that the proofs in this section rely crucially on the characterizations of the optimal fairness-constrained policies developed in Section 5.5. We first define the notion of CDF domination, which is referred to in a few of the proofs. Intuitively, it means that for any score, the fraction of group B above this is higher than that for group A. It is realistic to assume this if we keep with our convention that group A is the disadvantaged group relative to group B.

Definition 5.6 (CDF domination)**.** $\mathcal{D}_{\mathsf{A}}$ is said to be *dominated by* $\mathcal{D}_{\mathsf{B}}$ if $\forall a \geq 1, \sum_{x>a} \mathcal{D}_{\mathsf{A}}(x) < \sum_{x>a} \mathcal{D}_{\mathsf{B}}(x)$. We denote this as $\mathcal{D}_{\mathsf{A}} \prec \mathcal{D}_{\mathsf{B}}$.

Frequently, we shall use the following lemma:

Lemma 5.8.1. *Suppose that $\mathcal{D}_{\mathsf{A}} \prec \mathcal{D}_{\mathsf{B}}$. Then, for all $\beta > 0$, it holds that $\mathrm{Q}_{\mathsf{A}}(\beta) \leq \mathrm{Q}_{\mathsf{B}}(\beta)$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) \leq \boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))$*

Proof. The fact that $\mathrm{Q}_{\mathsf{A}}(\beta) \leq \mathrm{Q}_{\mathsf{B}}(\beta)$ follows directly from the definition of monotonicty of $\boldsymbol{u}$ implies that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) \leq \boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))$. □

Proof of Proposition 5.2.1. The `MaxUtil` policy for group j solves the optimization

$$\max_{\pi_{\mathsf{j}} \in [0,1]^C} \mathcal{U}_{\mathsf{j}}(\pi_{\mathsf{j}}) = \max_{\beta_{\mathsf{j}} \in [0,1]} \mathcal{U}_{\mathsf{j}}(r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta_{\mathsf{j}}))\,.$$

Computing left and right derivatives of this objective yields

$$\partial_+ \mathcal{U}_{\mathsf{j}}(r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta_{\mathsf{j}})) = \boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}(\beta)), \qquad \partial_- \mathcal{U}_{\mathsf{j}}(r^{-1}_{\mathcal{D}_{\mathsf{j}}}(\beta_{\mathsf{j}})) = \boldsymbol{u}(\mathrm{Q}^+_{\mathsf{j}}(\beta))\,.$$

By concavity, solutions β^* satisfy

$$\begin{aligned} \beta < \beta^* \quad &\text{if} \quad \boldsymbol{u}(\mathrm{Q}_{\mathsf{j}}(\beta)) > 0\,, \\ \beta > \beta^* \quad &\text{if} \quad \boldsymbol{u}(\mathrm{Q}^+_{\mathsf{j}}(\beta)) < 0\,. \end{aligned} \tag{5.8.1}$$

Therefore, we conclude that the `MaxUtil` policy loans only to scores x s.t. $\boldsymbol{u}(x) > 0$, which implies $\boldsymbol{\Delta}(x) > 0$ for all scores loaned to. Therefore we must have that $0 \leq \Delta\mu^{\texttt{MaxUtil}}$. By definition $\Delta\mu^{\texttt{MaxUtil}} \leq \Delta\mu^{*}$. □

Proof of Corollary 5.2.2. We begin with proving part (a), which gives conditions under which `DemParity` cases relative improvement. Recall that $\overline{\beta}$ is the largest selection rate for which $\Delta\mu(\overline{\beta}) = \Delta\mu(\beta_{\mathsf{A}}^{\texttt{MaxUtil}})$. First, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{DemParity}}$ from below. Fix an acceptance rate β such that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta < \min\{\beta_{\mathsf{B}}^{\texttt{MaxUtil}}, \overline{\beta}\}$. By Theorem 5.5.1, we have that `DemParity` selects to group A with rate higher than β as long as

$$g_{\mathsf{A}} \leq g_1 := \frac{1}{1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}}\,.$$

By (5.8.1) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) > 0$, so $0 < g_1 < 1$.

Next, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{DemParity}}$ from above. First, consider the case that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < \overline{\beta}$, and fix β' such that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < \beta' < \overline{\beta}$. Then `DemParity` selects group A at a rate $\beta_{\mathsf{A}} < \beta'$ for any proportion g_{A}. This follows from applying Theorem 5.5.1 since we have that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta')) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(\beta')) < 0$ by (5.8.1) and the monotonicity of $\boldsymbol{u}$.

Instead, in the case that $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \overline{\beta}$, fix β' such that $\overline{\beta} < \beta' < \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$. Then `DemParity` selects group A at a rate less than β' as long as

$$g_{\mathsf{A}} \geq g_0 := \frac{1}{1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta'))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(\beta'))}}\,.$$

By (5.8.1) and the monotonicity of $\boldsymbol{u}$, $0 < g_0 < g_1$. Thus for $g_{\mathsf{A}} \in [g_0, g_1]$, the `DemParity` selection rate for group A is bounded between β and β', and thus `DemParity` results in relative improvement.

Next, we prove part (b), which gives conditions under which `EqOpt` cases relative improvement. First, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{EqOpt}}$ from below. Fix an acceptance rate β such that $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} < \beta$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$. By Theorem 5.5.3, `EqOpt` selects group A at a rate higher than β as long as

$$g_{\mathsf{A}} > g_3 := \frac{1}{1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))}}\,.$$

By (5.8.1) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) > 0$, so $g_3 > 0$.

Next, we derive a condition which bounds the selection rate $\beta_{\mathsf{A}}^{\texttt{EqOpt}}$ from above. First, consider the case that there exists β' such that $\beta' < \overline{\beta}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} < G^{(\mathsf{A}\to\mathsf{B})}(\beta')$. Then `EqOpt` selects group A at a rate less than this β' for any g_{A}. This follows from Theorem 5.5.3

since we have that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta')) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta'))) < 0$ by (5.8.1) and the monotonicity of $\boldsymbol{u}$.

In the other case, fix β' such that $\beta < \beta' < \overline{\beta}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta')$. By Theorem 5.5.3, `EqOpt` selects group A at a rate lower than β' as long as

$$g_{\mathsf{A}} > g_2 := \frac{1}{1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta')))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}^{+}(G^{(\mathsf{A}\to\mathsf{B})}(\beta')))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta'))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}^{+}(\beta'))}}.$$

By (5.8.1) and the monotonicity of $\boldsymbol{u}$, $0 < g_2 < g_3$. Thus for $g_{\mathsf{A}} \in [g_2, g_3]$, the `EqOpt` selection rate for group A is bounded between β and β', and thus `EqOpt` results in relative improvement. □

Proof of Corollary 5.2.3. Recall our assumption that $\beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$ and $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > \beta$. As argued in the above proof of Corollary 5.2.2, by (5.8.1) and the monotonicity of $\boldsymbol{u}$, $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta)) > 0$. Applying Theorem 5.5.1, `DemParity` selects at a higher rate than β for any population proportion $g_{\mathsf{A}} \le g_0$, where $g_0 = 1/(1 - \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(\beta))}) \in (0, 1)$. In particular, if $\beta = \beta_0$, which we defined as the harm threshold (i.e. $\Delta\mu_{\mathsf{A}}(r_{\mathcal{D}_{\mathsf{A}}}^{-1}(\beta_0)) = 0$ and $\Delta\mu_{\mathsf{A}}$ is decreasing at β_0), then by the concavity of $\Delta\mu_{\mathsf{A}}$, we have that $\Delta\mu_{\mathsf{A}}(r_{\mathcal{D}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\texttt{DemParity}})) < 0$, that is, `DemParity` causes active harm. □

Proof of Corollary 5.2.4. By Theorem 5.5.3, `EqOpt` selects at a higher rate than β for any population proportion $g_{\mathsf{A}} \le g_0$, where $g_0 = 1/(1 - \frac{1}{\kappa} \cdot \frac{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))}{\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta)))} \frac{\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta))}{\boldsymbol{\rho}(\mathrm{Q}_{\mathsf{A}}(\beta))})$. Using our assumptions $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} > G^{(\mathsf{A}\to\mathsf{B})}(\beta)$ and $\beta > \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$, we have that $\boldsymbol{u}(\mathrm{Q}_{\mathsf{B}}(G^{(\mathsf{A}\to\mathsf{B})}(\beta))) > 0$ and $\boldsymbol{u}(\mathrm{Q}_{\mathsf{A}}(\beta)) < 0$, by (5.8.1) and the monotonicity of $\boldsymbol{u}$. This verifies that $g_0 \in (0, 1)$. In particular, if $\beta = \beta_0$, then by the concavity of $\Delta\mu_{\mathsf{A}}$, we have that $\Delta\mu_{\mathsf{A}}(r_{\mathcal{D}_{\mathsf{A}}}^{-1}(\beta_{\mathsf{A}}^{\texttt{EqOpt}})) < 0$, that is, `EqOpt` causes active harm. □

Proof of Proposition 5.2.5. Denote the upper quantile function under $\widehat{\mathcal{D}}$ as $\widehat{\mathrm{Q}}$. Since $\widehat{\mathcal{D}} \prec \mathcal{D}$, we have $\widehat{\mathrm{Q}}(\beta) \le \mathrm{Q}(\beta)$. The conclusion follows for `MaxUtil` and `DemParity` from Theorem 5.5.1 by the monotonicity of $\boldsymbol{u}$.

If we have that $\mathrm{TPR}_{\mathsf{A}}(\pi) > \widehat{\mathrm{TPR}_{\mathsf{A}}}(\pi)\ \forall\, \pi$, that is, the true TPR dominates estimated TPR, then by the monotonicity of $\boldsymbol{u}(x)/\boldsymbol{\rho}(x)$ and Theorem 5.5.3, we have that the conclusion follows for `EqOpt`.

□

Proof of Proposition 5.3.1. By Proposition 5.4.5, $\beta^* = \mathrm{argmax}_{\beta}\, \Delta\mu_{\mathsf{A}}(\beta)$ exists and is unique. $\beta^{\max} = \max\{\beta \in [\beta_{\mathsf{A}}^{\texttt{MaxUtil}}, 1] : \mathcal{U}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}) - \mathcal{U}_{\mathsf{A}}(\beta) \le \delta\}$ exists and is unique, by the continuity of $\Delta\mu_{\mathsf{A}}$ and Proposition 5.4.5. Either β^* is feasible, or $\beta^* > \beta^{\max}$.

□

Characterization of Fairness Solutions

Derivative computation for `EqOpt`. In this section, we prove Lemma 5.5.2. We will prove Lemma 5.5.2 in tandem with the following derivative computation which we applied in the proof of Theorem 5.5.3.

Lemma 5.8.2. *The function*

$$\mathcal{U}_{\mathrm{j}}(t; \boldsymbol{w}_{\mathrm{j}}) := \mathcal{U}_{\mathrm{j}}\left(r_{\mathcal{D}_{\mathrm{j}}}^{-1}\left(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)\right)\right)$$

is concave in t and has left and right derivatives

$$\partial_+ \mathcal{U}_{\mathrm{j}}(t; \boldsymbol{w}_{\mathrm{j}}) = \frac{\boldsymbol{u}(\mathrm{Q}_{\mathrm{j}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}_{\mathrm{j}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))} \quad \textit{and} \quad \partial_- \mathcal{U}_{\mathrm{j}}(t; \boldsymbol{w}_{\mathrm{j}}) = \frac{\boldsymbol{u}(\mathrm{Q}_{\mathrm{j}}^+(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}_{\mathrm{j}}^+(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}\,.$$

Proof of Lemmas 5.5.2 and 5.8.2. Consider a $\beta \in [0,1]$. Then, $\mathcal{D}_{\mathrm{j}} \circ r_{\mathcal{D}_{\mathrm{j}}}^{-1}(\beta)$ is continuous and left and right differentiable by Lemma 5.4.4, and its left and right derivatives are indicator vectors $\boldsymbol{e}_{\mathrm{Q}_{\mathrm{j}}(\beta)}$ and $\boldsymbol{e}_{\mathrm{Q}_{\mathrm{j}}^+(\beta)}$, respectively. Consequently, $\beta \mapsto \langle \boldsymbol{w}_{\mathrm{j}}, \mathcal{D}_{\mathrm{j}} \circ r_{\mathcal{D}_{\mathrm{j}}}^{-1}(\beta)\rangle$ has left and right derivatives $\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}(\beta))$ and $\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}^+(\beta))$, respectively; both of which are both strictly positive by the assumption $\boldsymbol{w}(x) > 0$. Hence, $T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\beta) = \langle \boldsymbol{w}_{\mathrm{j}}, \mathcal{D}_{\mathrm{j}} \circ r_{\mathcal{D}_{\mathrm{j}}}^{-1}(\beta)\rangle$ is strictly increasing in β, and so the map is injective. It is also surjective because $\beta = 0$ induces the policy $\pi_{\mathrm{j}} = \mathbf{0}$ and $\beta = 1$ induces the policy $\pi_{\mathrm{j}} = \mathbf{1}$ (up to $\mathcal{D}_{\mathrm{j}}$-measure zero). Hence, $T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\beta)$ is an order preserving bijection with left- and right-derivatives, and we can compute the left and right derivatives of its inverse as follows:

$$\partial_+ T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t) = \frac{1}{\partial_+ T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\beta)\big|_{\beta = T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)}} = \frac{1}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}_{\mathrm{j}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}\,,$$

and similarly, $\partial_- T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t) = \frac{1}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}^+(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}$. Then we can compute that

$$\begin{aligned}
\partial_+ \mathcal{U}_{\mathrm{j}}(r_{\mathcal{D}_{\mathrm{j}}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t))) &= \partial_+ \boldsymbol{\mathcal{U}}(r_{\mathcal{D}_{\mathrm{j}}}(\beta))\big|_{\beta = T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t))} \cdot \partial_+ T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(\sup(t)) \\
&= \frac{\boldsymbol{u}(\mathrm{Q}_{\mathrm{j}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}_{\mathrm{j}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}\,.
\end{aligned}$$

and similarly $\partial_- \mathcal{U}_{\mathrm{j}}(r_{\mathcal{D}_{\mathrm{j}}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}(t))) = \frac{\mathcal{U}(\mathrm{Q}_{\mathrm{j}}^+(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}{\boldsymbol{w}_{\mathrm{j}}(\mathrm{Q}_{\mathrm{j}}^+(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))}$. One can verify that for all $t_1 < t_2$, one has that $\partial_+ \mathcal{U}_{\mathrm{j}}(r_{\mathcal{D}_{\mathrm{j}}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t_1))) \geq \partial_- \mathcal{U}_{\mathrm{j}}(r_{\mathcal{D}_{\mathrm{j}}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t_2)))$, and that for all t, $\partial_+ \mathcal{U}_{\mathrm{j}}(r_{\mathcal{D}_{\mathrm{j}}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t))) \leq \partial_- \mathcal{U}_{\mathrm{j}}(r_{\mathcal{D}_{\mathrm{j}}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))$. These facts establish that the mapping $t \mapsto \mathcal{U}_{\mathrm{j}}(r_{\mathcal{D}_{\mathrm{j}}}(T_{\mathrm{j},\boldsymbol{w}_{\mathrm{j}}}^{-1}(t)))$ is concave. □

Optimality of Threshold Policies

Proof of Lemma 5.4.3. Given $\pi \in [0,1]^C$, we define the *normal cone* at π as $\mathrm{NC}(\pi) := \mathrm{ConicalHull}\{z : \pi + z \in [0,1]^C\}$. We can describe $\mathrm{NC}(\pi)$ explicitly as:

$$\mathrm{NC}(\pi) := \{z \in \mathbb{R}^C : z_i \leq 0 \text{ if } \pi_i = 0, z_i \geq 0 \text{ if } \pi_i = 1\} .$$

Immediately from the above definition, we have the following useful identity, which is that for any vector $g \in \mathbb{R}^C$,

$$\langle g, z\rangle \leq 0 \; \forall z \in \mathrm{NC}(\pi), \quad \text{if and only if} \quad \forall x \in \mathcal{X}, \begin{cases} \pi(x) = 0 & g(x) < 0 \\ \pi(x) = 1 & g(x) > 0 \\ \pi(x) \in [0,1] & g(x) = 0 \end{cases} . \tag{5.8.2}$$

Now consider the optimization problem (5.4.5). By the first order KKT conditions, we know that for any optimizer π_* of the above objective, there exists some $\widehat{\lambda} \in \mathbb{R}$ such that, for all $z \in \mathrm{NC}(\pi_*)$

$$\langle z, v \circ \mathcal{D} + \widehat{\lambda} \mathcal{D} \circ w\rangle \leq 0 .$$

By (5.8.2), we must have that

$$\pi_*(x) = \begin{cases} 0 & \mathcal{D}(x)(v(x) + \widehat{\lambda} w(x)) < 0 \\ 1 & \mathcal{D}(x)(v(x) + \widehat{\lambda} w(x)) > 0 \\ \in [0,1] & \mathcal{D}(x)(v(x) + \widehat{\lambda} w(x)) = 0 \end{cases} .$$

Now $\pi_*(x)$ is not necessarily a threshold policy. To conclude the theorem, it suffices to exhibit a threshold policy $\widetilde{\pi}_*$ such that $\pi_*(x) \cong_{\mathcal{D}} \widetilde{\pi}_*$. (Note that $\widetilde{\pi}_*(x)$ will also be feasible for the constraint, and have the same objective value; hence $\widetilde{\pi}_*$ will be optimal as well.)

Given π_* and $\widehat{\lambda}$, let $c_* = \min\{c \in \mathcal{X} : v(x) + \widehat{\lambda} w(x) \geq 0\}$. If either (a) $w(x) = 0$ for all $x \in \mathcal{X}$ and $v(x)$ is strictly increasing or (b) $v(x)/w(x)$ is strictly increasing, then the modified policy

$$\widetilde{\pi}_*(x) = \begin{cases} 0 & x < c_* \\ \pi_*(x) & x = c_* \\ 1 & x > c_* \end{cases} ,$$

is a threshold policy, and $\pi_*(x) \cong_{\mathcal{D}} \widetilde{\pi}_*$. Moreover, $\langle w, \widetilde{\pi}_*\rangle = \langle w, \widetilde{\pi}_*\rangle$ and $\langle \mathcal{D}, \widetilde{\pi}_*\rangle = \langle \mathcal{D}, \widetilde{\pi}_*\rangle$, which implies that $\widetilde{\pi}_*$ is an optimal policy for the objective in Lemma 5.4.3. □

Proof of Lemma 5.4.4 . We shall prove

$$\partial_+ \left(\mathcal{D}_{\mathrm{j}} \circ r^{-1}_{\mathcal{D}_{\mathrm{j}}}(\beta) \right) = e_{\mathrm{Q_j}(\beta)} , \tag{5.8.3}$$

where the derivative is with respect to β. The computation of the left-derivative is analogous. Since we are concerned with right-derivatives, we shall take $\beta \in [0, 1)$. Since $\mathcal{D}_j \circ r_{\mathcal{D}_j}^{-1}(\beta)$ does not depend on the choice of representative for $r_{\mathcal{D}_j}^{-1}$, we can choose a cannonical representation for $r_{\mathcal{D}_j}^{-1}$. The threshold policy $\pi_{Q_j(\beta),\gamma(\beta)}$ has acceptance rate β, where we define

$$\beta_+ = \sum_{x=Q_j(\beta)}^{C} \mathcal{D}(x) \text{ and } \beta_- = \sum_{x=Q_j(\beta)+1}^{C} \mathcal{D}(x)\,, \tag{5.8.4}$$

$$\gamma(\beta) = \frac{\beta - \beta_-}{\beta_+ - \beta_-}\,. \tag{5.8.5}$$

Note then that for each x, $\pi_{Q_j(\beta),\gamma(\beta)}(x)$ is piece-wise linear, and thus admits left and right derivatives. We first claim that

$$\forall x \in \mathcal{X} \setminus \{Q_j(\beta)\},\ \ \partial_+ \pi_{Q_j(\beta),\gamma(\beta)}(x) = 0\,. \tag{5.8.6}$$

To see this, note that $Q_j(\beta)$ is right continuous, so for all ϵ sufficiently small, $Q_j(\beta + \epsilon) = Q_j(\beta)$. Hence, for all ϵ sufficiently small and all $x \neq Q(\beta)$, we have $\pi_{Q_j(\beta+\epsilon),\gamma(\beta+\epsilon)}(x) = \pi_{Q_j(\beta+\epsilon),\gamma(\beta+\epsilon)}(x)$, as needed. Thus, Equation (5.8.6) implies that $\partial_+ \mathcal{D}_j \circ r_{\mathcal{D}_j}^{-1}(\beta)$ is supported on $x = Q_j(\beta)$, and hence

$$\partial_+ \mathcal{D}_j \circ r_{\mathcal{D}_j}^{-1}(\beta) = \partial_+ \mathcal{D}_j(x) \pi_{Q_j(\beta),\gamma(\beta)}(x)\big|_{x=Q_j(\beta)} \cdot \boldsymbol{e}_{Q_j(\beta)}\,.$$

To conclude, we must show that $\partial_+ \mathcal{D}_j(x) \pi_{Q_j(\beta),\gamma(\beta)}(x)\big|_{x=Q_j(\beta)} = 1$. To show this, we have

$$\begin{aligned}
1 &= \partial_+(\beta) \\
&= \partial_+(r_{\mathcal{D}_j}(\pi_{Q_j(\beta),\gamma(\beta)})) \quad \text{since} \qquad r_{\mathcal{D}_j}(\pi_{Q_j(\beta),\gamma(\beta)}) = \beta\ \forall \beta \in [0,1) \\
&= \partial_+ \left(\sum_{x \in \mathcal{X}} \mathcal{D}(x) \cdot \pi_{Q_j(\beta),\gamma(\beta)}(x) \right) \\
&= \partial_+ \mathcal{D}(x) \cdot \pi_{Q_j(\beta),\gamma(\beta)}(x)\big|_{x=Q_j(\beta)}\,, \text{ as needed.}
\end{aligned}$$

□

Connections to Welfare

We begin with the definition of important quantities. Recall that demographic parity constrains the selection rates of policies. Further recall that the rate function for each group $r_j(\pi)$. Because we focus on threshold policies, with some abuse of notation, we will write $r_j(t) := r_j(\mathbf{1}\{p \geq t\})$. This function is monotonic in the threshold t, and therefore its inverse maps acceptance rates to thresholds which achieve that rate, i.e. $r_j^{-1}(\beta) = t_j$.

Define $f_{\mathsf{A}}(\beta)$ and $f_{\mathsf{B}}(\beta)$ to be components of the objective function in (5.3.3) due to each group, that is,

$$f_{\mathsf{j}}(\beta) = g_{\mathsf{j}} \mathcal{U}_{\mathsf{j}}(r_{\mathsf{j}}^{-1}(\beta)) \,.$$

By Proposition 5.4.5, the functions f_{j} are concave. Therefore, the combined objective function is concave in each argument,

$$\mathcal{U}_{\mathsf{P}}(\beta_{\mathsf{A}}, \beta_{\mathsf{B}}) = \sum_{\mathsf{j} \in \{\mathsf{A},\mathsf{B}\}} f_{\mathsf{j}}(\beta_{\mathsf{j}}) \,. \tag{5.8.7}$$

We restrict our attention to the case that p has continuous support. In this case, the functions f_{j} are differentiable.

Lemma 5.8.3. *If the distribution of exact profit scores for groups* A *and* B *are such that that maximum profit selection rates* $\beta_{\mathsf{A}}^{\texttt{MaxUtil}} = r_{\mathsf{A}}(0)$ *and* $\beta_{\mathsf{B}}^{\texttt{MaxUtil}} = r_{\mathsf{B}}(0)$ *and* $r_{\mathsf{A}}(0) \leq r_{\mathsf{A}}(0)$*, then for any* $\epsilon \geq 0$*, the selection rates* β^{ϵ} *maximizing the optimization problem* (5.3.3) *under demographic parity satisfy the following:*

$$\beta_{\mathsf{A}}^{\texttt{MaxUtil}} \leq \beta_{\mathsf{A}}^{\epsilon} \leq \beta_{\mathsf{B}}^{\epsilon} \leq \beta_{\mathsf{B}}^{\texttt{MaxUtil}} \,.$$

Proof of Proposition 5.8.3. First, we argue that $\beta_{\mathsf{A}}^{\epsilon} \leq \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$. If it were that $\beta_{\mathsf{A}}^{\epsilon} > \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$, then the alternate solution $\beta_{\mathsf{A}} = \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ and $\beta_{\mathsf{B}} = \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ would be feasible for (5.3.3) and achieve a higher objective value by the concavity of (5.8.7).

Next we show that $\beta_{\mathsf{B}}^{\epsilon} \leq \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$. Assume for the sake of contradiction that $\beta_{\mathsf{B}}^{\epsilon} > \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$. Then since $\beta_{\mathsf{A}}^{\epsilon} \leq \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$, setting $\beta_{\mathsf{B}} = \beta_{\mathsf{B}}^{\texttt{MaxUtil}}$ achieves higher objective value without increasing $|\beta_{\mathsf{B}} - \beta_{\mathsf{A}}^{\epsilon}|$, and thus would be feasible for (5.3.3). By a similar argument, $\beta_{\mathsf{A}}^{\epsilon} \geq \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$.

Lastly, we show that for any optimal selection rates, $\beta_{\mathsf{A}}^{\epsilon} \leq \beta_{\mathsf{B}}^{\epsilon}$ for all $\epsilon \geq 0$. Suppose for the sake of contradiction that $\beta_{\mathsf{A}}^{\epsilon} > \beta_{\mathsf{B}}^{\epsilon}$. In this case, we can equivalently write that

$$\beta_{\mathsf{B}}^{\texttt{MaxUtil}} - \beta_{\mathsf{B}}^{\epsilon} > \beta_{\mathsf{B}}^{\texttt{MaxUtil}} - \beta_{\mathsf{A}}^{\epsilon} \text{ and/or } \beta_{\mathsf{A}}^{\epsilon} - \beta_{\mathsf{A}}^{\texttt{MaxUtil}} > \beta_{\mathsf{B}}^{\epsilon} - \beta_{\mathsf{A}}^{\texttt{MaxUtil}} \,.$$

In either case, setting $\beta_{\mathsf{A}}^{\epsilon} = \beta_{\mathsf{B}}^{\epsilon}$ would be a feasible solution which would achieve a higher objective function value, by the concavity of (5.8.7). This contradicts the assumption that $\beta_{\mathsf{A}}^{\epsilon} > \beta_{\mathsf{B}}^{\epsilon}$, and thus it must be that $\beta_{\mathsf{A}}^{\epsilon} \leq \beta_{\mathsf{B}}^{\epsilon}$. □

Lemma 5.8.4. *Under the conditions of Lemma 5.8.3, the maximizer* $(\beta_{\mathsf{A}}^{\epsilon}, \beta_{\mathsf{B}}^{\epsilon})$ *of the* ϵ*-demographic parity constrained problem in* (5.3.3) *is either satisfied with the maximum profit selection rates* $(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}, \beta_{\mathsf{B}}^{\texttt{MaxUtil}})$*, or* $\beta_{\mathsf{B}}^{\epsilon} - \beta_{\mathsf{A}}^{\epsilon} = \epsilon$ *(or the two conditions coincide).*

Proof of Lemma 5.8.4. Suppose that the `MaxUtil` selection rates are not feasible. If it were that $|\beta_{\mathsf{A}}^{\epsilon} - \beta_{\mathsf{B}}^{\epsilon}| = \gamma < \epsilon$ then we could construct an alternative solution using the remaining $\epsilon - \gamma$ slack in the constraint. This would achieve a higher objective function value, since the functions f_{j} are concave. Therefore, $|\beta_{\mathsf{A}}^{\epsilon} - \beta_{\mathsf{B}}^{\epsilon}| = \epsilon$. Furthermore, by Lemma 5.8.3, we have that $|\beta_{\mathsf{A}}^{\epsilon} - \beta_{\mathsf{B}}^{\epsilon}| = \beta_{\mathsf{B}}^{\epsilon} - \beta_{\mathsf{A}}^{\epsilon}$. □

This result implies that the complexity of the maximization (5.3.3) can be reduced to a single variable search:

$$\beta^\star = \operatorname*{argmax}_\beta f_\mathsf{A}(\beta) + f_\mathsf{B}(\beta+\epsilon), \quad \pi^\epsilon_{\text{fair}} = (\mathbf{1}\{p \geq r_\mathsf{j}^{-1}(\beta^\star)\}, \mathbf{1}\{p \geq r_\mathsf{j}^{-1}(\beta^\star+\epsilon)\}) \tag{5.8.8}$$

This expression holds when $|\beta_\mathsf{A}^{\texttt{MaxUtil}} - \beta_\mathsf{B}^{\texttt{MaxUtil}}| > \epsilon$, and otherwise the solution is given by $(\beta_\mathsf{A}^{\texttt{MaxUtil}}, \beta_\mathsf{B}^{\texttt{MaxUtil}})$.

Lemma 5.8.5. *Under the conditions of Lemma 5.8.3, as $\epsilon \geq 0$ decreases, the group-dependent selection rates $\beta_\mathsf{A}^\epsilon$ and $\beta_\mathsf{B}^\epsilon$ become closer to the profit maximizing selection rates for each group. That is, the functions $|\beta_\mathsf{A}^\epsilon - \beta_\mathsf{A}^{\texttt{MaxUtil}}|$ and $|\beta_\mathsf{B}^\epsilon - \beta_\mathsf{B}^{\texttt{MaxUtil}}|$ are both increasing in ϵ.*

Proof of Lemma 5.8.5. We show that for any $\epsilon' \geq \epsilon \geq 0$, it must be that $|\beta_\mathsf{j}^\epsilon - \beta_\mathsf{j}^{\texttt{MaxUtil}}| \leq |\beta_\mathsf{j}^{\epsilon'} - \beta_\mathsf{j}^{\texttt{MaxUtil}}|$. First, we remark that if $|\beta_\mathsf{A}^{\texttt{MaxUtil}} - \beta_\mathsf{B}^{\texttt{MaxUtil}}| \leq \epsilon$ or if $\epsilon \leq |\beta_\mathsf{A}^{\texttt{MaxUtil}} - \beta_\mathsf{B}^{\texttt{MaxUtil}}| \leq \epsilon'$, the claim holds by application of Lemma 5.8.3.

Otherwise, let the ϵ-demographic parity constrained solution be optimized by $(\beta, \beta+\epsilon)$ and the ϵ'-demographic parity constrained solution be optimized by $(\beta', \beta'+\epsilon')$. This is valid by Lemma 5.8.4. Equivalently, $\beta \in \operatorname{argmax}\{f_\mathsf{A}(\beta)+f_\mathsf{B}(\beta+\epsilon)\}$ and $\beta' \in \operatorname{argmax}\{f_\mathsf{A}(\beta')+f_\mathsf{B}(\beta'+\epsilon')\}$. Since f_A and f_B are concave and differentiable,

$$f'_\mathsf{A}(\beta) + f'_\mathsf{B}(\beta+\epsilon) = 0 \quad \text{and} \quad f'_\mathsf{A}(\beta') + f'_\mathsf{B}(\beta'+\epsilon') = 0 \,.$$

Assume for sake of contradiction that $\beta < \beta'$ and recall that by Lemma 5.8.3 we further have $\beta' > \beta \geq \beta_\mathsf{A}^{\texttt{MaxUtil}}$, so by the concavity of f_A,

$$f_\mathsf{A}(\beta) \geq f_\mathsf{A}(\beta') \quad \text{and} \quad f'_\mathsf{A}(\beta') \leq f'_\mathsf{A}(\beta) \,.$$

Analogously, we must have that $\beta_\mathsf{B}^{\texttt{MaxUtil}} \geq \beta'+\epsilon' > \beta+\epsilon$, so that

$$f_\mathsf{B}(\beta'+\epsilon') \geq f_\mathsf{B}(\beta+\epsilon) \quad \text{and} \quad f'_\mathsf{B}(\beta'+\epsilon') \geq f'_\mathsf{B}(\beta+\epsilon) \,.$$

Using the equations above, we have that

$$\begin{aligned} f'_\mathsf{B}(\beta+\epsilon) &= -f'_\mathsf{A}(\beta) \\ &\leq -f'_\mathsf{A}(\beta') \\ &= f'_\mathsf{B}(\beta'+\epsilon') \end{aligned}$$

Since f_B is concave and thus its derivative is decreasing, this statement implies that $\beta+\epsilon \geq \beta'+\epsilon'$, which is a contradiction. Thus, it must be that $\beta \geq \beta'$, i.e. $\beta_\mathsf{A}^\epsilon \geq \beta_\mathsf{A}^{\epsilon'}$. With an analogous proof by contradiction, one can show that $\beta_\mathsf{B}^{\epsilon'} \geq \beta_\mathsf{B}^\epsilon$.

Combining these two inequalities in Lemma 5.8.3 completes the proof of Lemma 5.8.5. □

Proof of Proposition 5.3.3. The proof makes use of Proposition 5.8.3 and Lemma 5.8.5.

First, we show that $t_{\mathsf{A}}^{\epsilon} \leq 0$ for all $\epsilon \geq 0$. This is a consequence of Lemma 5.8.3, which shows that $\beta_{\mathsf{A}}^{\epsilon} \geq \beta_{\mathsf{A}}^{\texttt{MaxUtil}}$. Since r_{A} is a decreasing function (and thus, r_{A}^{-1} is also a decreasing function), this implies that

$$t_{\mathsf{A}}^{\epsilon} = r_{\mathsf{A}}^{-1}(\beta_{\mathsf{A}}^{\epsilon}) \leq r_{\mathsf{A}}^{-1}(\beta_{\mathsf{A}}^{\texttt{MaxUtil}}) = 0$$

A similar argument holds to show that $t_{\mathsf{B}}^{\epsilon} \geq 0$ for all $\epsilon \geq 0$.

Now we show that $t_{\mathsf{A}}^{\epsilon}$ is increasing in ϵ and $t_{\mathsf{B}}^{\epsilon}$ is decreasing in ϵ to show that both are shrinking toward 0 as ϵ increases. Since $t_{\mathsf{j}} = r_{\mathsf{j}}^{-1}(\beta)$ is decreasing in β, Lemma 5.8.5 implies that the functions $|t_{\mathsf{A}}^{\epsilon}| = |t_{\mathsf{A}}^{\epsilon} - t_{\mathsf{A}}^{\texttt{MaxUtil}}|$ and $|t_{\mathsf{B}}^{\epsilon}| = |t_{\mathcal{B}}^{\epsilon} - t_{\mathsf{B}}^{\texttt{MaxUtil}}|$ are also decreasing in ϵ toward the max profit thresholds of $t_{\mathsf{A}}^{\texttt{MaxUtil}} = t_{\mathsf{B}}^{\texttt{MaxUtil}} = 0$. Since $t_{\mathsf{A}}^{\epsilon} \leq 0$ and $t_{\mathsf{B}}^{\epsilon} \geq 0$ for all $\epsilon \geq 0$, this concludes the proof of Proposition 5.3.3. □

Chapter 6

Reachability in Recommender Systems

6.1 Introduction

In this chapter, we consider the setting of recommendation systems. Rather than a single consequential decision, these systems make algorithmic selections from an overabundance of choices and characterized by their interactivity. This chapter uses material first presented in papers coauthored with Mihaela Curmei, Benjamin Recht, and Sarah Rich [DRR20; CDR21].

Recommendation systems influence the way information is presented to individuals for a wide variety of domains including music, videos, dating, shopping, and advertising. On one hand, the near-ubiquitous practice of filtering content by predicted preferences makes the digital information overload possible for individuals to navigate. By exploiting the patterns in ratings or consumption across users, preference predictions are useful in surfacing relevant and interesting content. On the other hand, this personalized curation is a potential mechanism for social segmentation and polarization. The exploited patterns across users may in fact encode undesirable biases which become self-reinforcing when used in feedback to make recommendations.

Recent empirical work shows that personalization on the Internet has a limited effect on political polarization [FGR16], and in fact it can increase the diversity of content consumed by individuals [NHHTK14]. However, these observations follow by comparison to non-personalized baselines of cable news or well known publishers. In a digital world where all content is algorithmically sorted by default, how do we articulate the trade-offs involved? In the past year, YouTube has come under fire for promoting disturbing children's content and working as an engine of radicalization [Tuf18; Nic18; Bri17]. This comes after a push on algorithm development towards the goal of reaching 1 billion hours of watchtime per day; over 70% of views now come from the recommended videos [Sol18].

The Youtube controversy is an illustrative example of potential pitfalls when putting large scale machine learning-based systems in feedback with people, and highlights the importance of creating analytical tools to anticipate and prevent undesirable behavior.

Such tools should seek to quantify the degree to which a recommender system will meet the information needs of its users or of society as a whole, where these "information needs" must be carefully defined to include notions like relevance, coverage, and diversity. An important approach involves the empirical evaluation of these metrics by simulating recommendations made by models once they are trained [ETA+18]. In this chapter, we develop a complementary approach which differs in two major ways: First, we investigate properties of the predictive model analytically, making it possible to understand underlying mechanisms. Second, our evaluation considers a range of possible user behaviors rather than a static snapshot.

Drawing conclusions about the likely effects of recommendations involves treating humans as a component within the system, and the validity of these conclusions hinges on modeling human behavior. We propose an alternative evaluation that favors the agency of individuals over the limited perspective offered by behavioral predictions. Our main focus is on questions of *possibility* and *access*: to what extent can someone be pigeonholed by their viewing history? What videos may they never see, even after a drastic change in viewing behavior? And how might a recommender system encode biases in a way that effectively limits the available library of content?

This perspective brings user agency into the center, prioritizing the the ability for models to be as adaptable as they are accurate, able to accommodate arbitrary changes in the interests of individuals. We adopt an *interventional* lens, which considers arbitrary and strategic user actions. This chapter develops a notion of reachability, which measures the ability of an individual to influence a recommender model to select a certain piece of content. Reachability provides an upper bound on the ability of individuals to discover specific content, thus isolating unavoidable biases within preference models from those due to user behavior. While there are many system-level or post-hoc approaches to incorporating user feedback, we focus directly on the machine learning model that powers recommendations.

We use reachability to define of metrics which capture the possible outcomes of a round of system interactions, including the *availability* of content and *discovery* possibilities for individuals. In Section 6.3, we show that they can be computed by solving a convex optimization problem for a class of relevant recommenders. In Section 6.4, we draw connections between the stochastic and deterministic settings. This perspective allows us to describe the relationship between agency and stochasticity and further to argue that there is not an inherent trade-off between reachability and model accuracy. Finally, we present an audit of recommendation systems using a variety of datasets and preference models. We explore how design decisions influence reachability and the extent to which biases in the training datasets are propagated.

Related Work

Much work on recommender systems focuses on the accuracy of the model. This encodes an implicit assumption that the primary information needs of users or society are

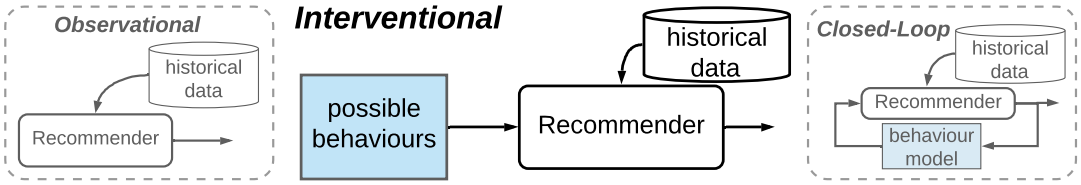

Figure 6.1: Conceptual framings of recommendation systems consider user behaviors to varying degrees. In this chapter we focus on evaluating interventional properties.

described by predictive performance. The recommender systems literature has long proposed a variety of other metrics for evaluation, including notions of novelty, serendipity, diversity, and coverage [HKTR04; VC11]. Alternative metrics are useful both for diagnosing biases and as objectives for post-hoc mitigating techniques such as calibration [Ste18] and re-ranking [SJ18]. There is also a long history of measuring and mitigating bias in recommendation systems [CDW+20]. Empirical investigations have found evidence of popularity and demographic bias in domains including movies, music, books, and hotels [AMBM19; ETA+18; ETKMK18; JLKJ15]. A inherent limitation of these approaches is that they focus on *observational* bias induced by preference models, i.e. examining the result of a single round of recommendations without considering individuals' behaviors. While certainly useful, they fall short of providing further understanding into the interactive nature of recommendation systems.

The behavior of recommendation systems over time and in *closed-loop* is still an open area of study. It is difficult to definitively link anecdotal evidence of radicalization [ROWAM20; FCF20] to proprietary recommendation algorithms. Empirical studies of human behavior find mixed results on the relationship between recommendation and content diversity [NHHTK14; FGR16]. Simulation studies [CSE18; YHT+21; KDZ+20] and theoretical investigations [DGL13] shed light on phenomena in simplified settings, showing how homogenization, popularity bias, performance, and polarization depend on assumed user behavior models. Even ensuring accuracy in sequential dynamic settings requires contending with closed-loop behaviors. Recommendation algorithms must mitigate biased sampling in order to learn underlying user preference models, using causal inference based techniques [SSSCJ16; YCX+18] or by balancing exploitation and exploration [KBKTC15; MGP15]. Reinforcement Learning algorithms contend with these challenges while considering a longer time horizon [CBC+19; IJW+19], implicitly using data to exploit user behavior.

Our work eschews behavior models in favor of an *interventional* framing which considers a variety of possible user actions (Figure 6.1). Giving users control over their recommendations has been found to have positive effects, while reducing agency has negative effects [HXK+15; LLZ+21]. Most similar to our work is a handful of papers focusing on decision systems through the lens of the agency of individuals. This chapter extends the notion of recourse proposed by Ustun, Spangher, and Liu [USL19] to multiclass classifica-

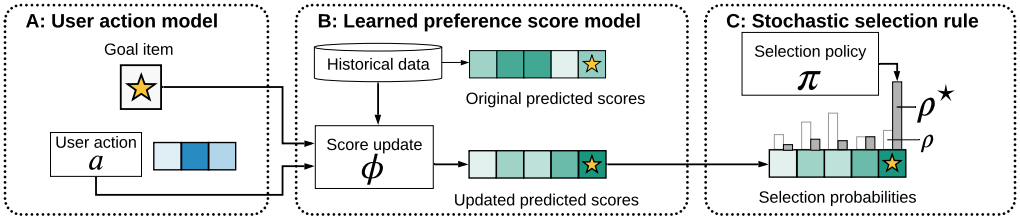

Figure 6.2: We audit recommender systems under a user action model (A), which represents possibilities for system interaction; a learned preference model (B), which scores items based on data and updates based on interactions; and a stochastic selection rule (C), which selects the next recommendation.

tion settings and specializes to concerns most relevant for information retrieval systems. Recourse in consequential decision making focuses on binary decisions, where users seek to change negative classification through modifications to their features. [KBSV20]. This work has connections to concepts in explainability and transparency via the idea of *counterfactual explanations* [Rus19; WMR17], which provide statements of the form: if a user had features X, then they would have been assigned alternate outcome Y. Work in strategic manipulation studies nearly the same problem with the goal of creating a decision system that is robust to malicious changes in features [HMPW16; MMDH19].

Applying these ideas to recommender systems is complex because while they can be viewed as classifiers or decision systems, there are as many outcomes as pieces of content. Computing precise action sets for recourse for every user-item pair is unrealistic; we don't expect a user to even become aware of the majority of items. Instead, we consider the "reachability" of items by users, drawing philosophically from the fields of formal verification and dynamical system analysis [BCHT17; OF06].

6.2 Recommenders and Reachability

Stochastic Recommender Setting

We consider systems composed of n individuals as well as a collection of m pieces of content. For consistency with the recommender systems literature, we refer to individuals as users, pieces of content as items, and expressed preferences as ratings. We will denote a rating by user u of item i as $r_{ui} \in \mathcal{R}$, where $\mathcal{R} \subseteq \mathbb{R}$ denotes the space of values which ratings can take. For example, ratings corresponding to the percentage of a video watched would have $\mathcal{R} = [0, 1]$ while discrete star ratings would have $\mathcal{R} = \{1, 2, 3, 4, 5\}$. The number of *observed ratings* will generally be much smaller than the total number of possible ratings, and we denote by $\Omega_u \subseteq \{1, \ldots, m\}$ the set of items seen by the user u. The goal of a recommendation system is to understand the preferences of users and recommend relevant content.

In this work, we focus on the common setting in which recommenders are the composition of a *scoring function* ϕ with *selection rule* π (Figure 6.2). The scoring function models the preferences of users. It is constructed based on historical data (e.g. observed ratings, user/item features) and returns a score for each user and item pair. For a given user u and item i, we denote $s_{ui} \in \mathbb{R}$ to be the associated score, and for user u we will denote by $\mathbf{s}_u \in \mathbb{R}^m$ the vector of scores for all items. A common example of a scoring function is a machine learning model which predicts future ratings based on historical data.

We will focus on the way that scores are updated after a round of user interaction. For example, if a user consumes and rates several new items, the recommender system should update the scores in response. Therefore, we define the score function as an update rule which takes as its argument the user action. The new score vector is $\mathbf{s}_u^+ = \phi_u(\mathbf{a})$, where $\mathbf{a} \in \mathcal{A}_u$ represents actions taken by user u and $\mathcal{A}_u$ represents the set of all possible actions. Thus ϕ_u encodes the historical data, the preference model class, and the update algorithm. The action space $\mathcal{A}_u$ represents possibilities for system interaction, encoding for example limitations due to user interface design. We define the form of the score update function and discuss the action space in more detail in Section 6.3.

The selection rule π is a policy which, for given user u and scores $\mathbf{s}_u$, selects one or more items from a set of specified *target items* $\Omega_u^t \subseteq \{1, \ldots, m\}$ as the next recommendation. The simplest selection rule is a top-1 policy, which is a deterministic rule that selects the item with the highest score for each user. A simple stochastic rule is the ϵ-greedy policy which with probability $1 - \epsilon$ selects the top scoring item and with probability ϵ chooses uniformly from the remaining items. Many additional approaches to recommendation can be viewed as the composition of a score function with a selection policy. This setting also encompasses implicit feedback scenarios, where clicks or other behaviors are defined as or aggregated into "ratings." Many recommendation algorithms, even those not specifically motivated by regression, include an intermediate score prediction step, e.g. point-wise approaches to ranking. Further assumptions in Section 6.3 will not capture the full complexity of other techniques such as pairwise ranking and slate-based recommendations. We leave such extensions to future work.

In this work, we are primarily interested in stochastic policies which select items according to a probability distribution on the scores $\mathbf{s}_u$ parametrized by a exploration parameter. Policies of this form are often used to balance exploration and exploration in online or sequential learning settings. A stochastic selection rule recommends an item i according to $\mathbb{P}\left(\pi(\mathbf{s}_u, \Omega_u^t) = i\right)$, which is 0 for all non-target items $i \notin \Omega_u^t$. For example, to select among items that have not yet been seen by the user, the target items are set as $\Omega_u^t = \Omega_u^c$ (recalling that Ω_u denotes the set of items seen by the user u). Deterministic policies are a special case of stochastic policies, with a degenerate distribution.

Stochastic policies have been proposed in the recommender system literature to improve diversity [CPNB15] or efficiently explore in a sequential setting [KBKTC15]. By balancing exploitation of items with high predicted ratings against explorations of items with lower predictions, preferences can be estimated so that future predicted ratings are more accurate. However, our work decidedly does not take a perspective based on

accuracy. Rather than supposing that users' reactions are predictable, we consider a perspective centered on agency and access.

Reachability

We define an item i to be *deterministically reachable* by a user u if there is some allowable action $\mathbf{a}$ that causes item to be recommended. In the setting where recommendations are made stochastically, we define an item i to be ρ *reachable* by a user u if there is some allowable action $\mathbf{a}$ such that the updated probability that item i is recommended after applying the action$\mathbb{P}(\pi(\phi_u(\mathbf{a}), \Omega_u^t) = i)$ is larger than ρ. The maximum ρ reachability for a user-item pair is defined as the solution to the following optimization problem:

$$\rho^\star(u, i) = \max_{\mathbf{a} \in \mathcal{A}_u} \; P(\pi(\phi_u(\mathbf{a}), \Omega_u^t) = i). \tag{6.2.1}$$

We will also refer to $\rho^\star(u, i)$ as "max reachability." For example, in the case of top-1 recommendation, $\rho^\star(u, i)$ is a binary indicator of whether item i is deterministically reachable to user u. In the case of ε-greedy policy, $\rho^\star(u, i) = 1 - \varepsilon$ if item i is deterministically reachable by user u, and is $\varepsilon/(|\Omega_u^t| - 1)$ otherwise.

By measuring the maximum achievable probability of recommending an item to a user, we are characterizing a granular metric of *access* within the recommender system. It can also be viewed as an upper bound on the likelihood of recommendation with minimal assumptions about user behavior. It may be illuminating to contrast this measure with a notion of expected reachability. Computing expected reachability would require specifying the distribution over user actions, which would amount to modeling human behavior. In contrast, max reachability requires specifying only the constraints arising from system design choices to define $\mathcal{A}_u$ (e.g. the user interface). By computing max reachability, we focus our analysis on the design of the recommender system, and avoid conclusions which are dependent on behavioral modeling choices.

Two related notions of user agency with respect to a target item i are *lift* and *rank gain*. The lift measures the ratio between the maximum achievable probability of recommendation and the baseline:

$$\lambda^\star(u, i) = \frac{\rho^\star(u, i)}{\rho_0(u, i)} \tag{6.2.2}$$

where the baseline $\rho_0(u, i)$ is defined to capture the default probability of recommendation in the absence of strategic behavior, e.g. $P\left(\pi\left(\mathbf{s}_u, \Omega_u^t\right) = i\right)$ for initial scores $\mathbf{s}_u$.

The rank gain for an item i is the difference in the ranked position of the item within the original list of scores $\mathbf{s}_u$ and its rank within the updated list of scores $\mathbf{s}_u^+$.

Lift and rank gain are related concepts, but ranked position is combinatorial in nature and thus difficult to optimize for directly. They both measure agency because they compare the default behavior of a system to its behavior under a strategic intervention by the user.

Given that recommenders are designed with personalization in mind, we view the ability of users to influence the model in a positive light. This is in contrast to much recent work in robust machine literature where strategic manipulation is undesirable.

Diagnosing System Limitations

The analysis of stochastic reachability can be used to audit recommender systems and diagnose systemic biases from an interventional perspective (Figure 6.1). Unlike studies of observational bias, these analyses take into account system interactivity. Unlike studies of closed-loop bias, there is no dependence on a behavior model. Because max reachability considers the best case over possible actions, it isolates structural biases from those caused in part by user behavior.

Max reachability is a metric defined for each user-item pair, and disparities across users and items can be detected through aggregations. Aggregating over target items gives insight into a user's ability to discover content, thus detecting users who have been "pigeonholed" by the algorithm. Aggregations over users can be used to compare how the system makes items available for recommendation.

We define the following user- and item-based aggregations:

$$D_u = \sum_{i \in \Omega_u^t} \frac{\mathbf{1}\{\rho_{ui} > \rho_t\}}{|\Omega_u^t|}, \quad A_i = \frac{\sum_u \rho_{ui} \mathbf{1}\{i \in \Omega_u^t\}}{\sum_u \mathbf{1}\{i \in \Omega_u^t\}} \tag{6.2.3}$$

The discovery D_u is the proportion of target items that have a high chance of being recommended, as determined by the threshold ρ_t. A natural threshold is the better-than-uniform threshold, $\rho_t = 1/|\Omega_u^t|$, recalling that Ω_u^t is the set of target items. When $\rho_{ui} = \rho_0(u, i)$, baseline discovery counts the number of items that will be recommended with better-than-uniform probability and is determined by the spread of the recommendation distribution. For deterministic top-1 recommendation, we have that $D_u = 1$. When $\rho_{ui} = \rho^\star(u, i)$, discovery counts the number of items that a user *could* be recommended with better-than-uniform probability in the best case. In the deterministic case, this is the proportion of target items deterministically reachable by the user. Low best-case discovery means that the recommender system inherently limits user access to content.

The item availability A_i is the average likelihood of recommendation over all users who have item i as a target. It can be thought of as the chance that a uniformly selected user will be recommended item i. When $\rho_{ui} = \rho_0(u, i)$, the baseline availability measures the prevalence of the item in the recommendations. When $\rho_{ui} = \rho^\star(u, i)$, availability measures the prevalence of an item in the best case. For deterministic top-1 recommendation, it is the proportion of eligible users who can reach the item. Low best-case availability means that the recommender system inherently limits the distribution of a given item.

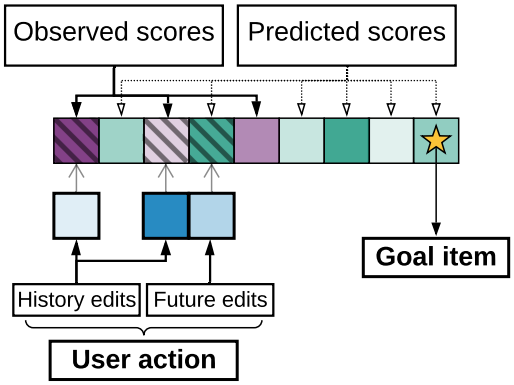

Figure 6.3: User action space: The shade represents the magnitude of historical (purple) or predicted (green) rating. The *action items* are marked with diagonal lines; they can be strategically modified to maximize the recommendation probability of the *goal item* (star). The value of the user action is shaded in blue.

6.3 Computation via Convex Optimization

Affine Recommendation

In this section, we consider a restricted class of recommender systems for which the max reachability problem can be efficiently solved via convex optimization.

User action model. We suppose that users interact with the system through expressed preferences, and thus actions are updates to the vector $\mathbf{r}_u \in \mathcal{R}^m$, a sparse vector of observed ratings. For each user, the action model is based on distinguishing between *action* and *immutable* items.

Let $\Omega_u^{\mathcal{A}}$ denote the set of items for which the ratings can be strategically modified by the user u. Then the action set $\mathcal{A}_u = \mathcal{R}^{|\Omega_u^{\mathcal{A}}|}$ corresponds to changing or setting the value of these ratings. Figure 6.3 provides an illustration. The action set should be defined to correspond to the interface through which a user interacts with the recommender system. For example, it could correspond to a display panel of "previously viewed" or "up next" items.

The updated rating vector $\mathbf{r}_u^+ \in \mathcal{R}^m$ is equal to $\mathbf{r}_u$ at the indices corresponding to immutable items and equal to the action $\mathbf{a}$ at the action items. Note the partition into action and immutable is distinct from earlier partition of items into observed and unobserved; action items can be both seen (history edits) and unseen (future reactions), as illustrated in Figure 6.2A. For the reachability problem, we will consider a set of target items Ω_u^t that does not intersect with the action items $\Omega_u^{\mathcal{A}}$. Depending on the specifics of the recommendation

setting, we may also require that it does not intersect with the previously rated items Ω_u.

We remark that additional user or item features used for scoring and thus recommendations could be incorporated into this framework as either mutable or immutable features. The only computational difficulty arises when mutable features are discrete or categorical.

Recommender model. The recommender model is composed of a scoring function ϕ and a selection function π, which we now specify. We consider *affine score update functions* where for each user, scores are determined by an affine function of the action: $\mathbf{s}_u^+ = \phi_u(\mathbf{a}) = B_u\mathbf{a} + \mathbf{c}_u$ where $B_u \in \mathbb{R}^{m \times |\Omega_u^{\mathcal{A}}|}$ and $\mathbf{c}_u \in \mathbb{R}^m$ are model parameters determined in part by historical data. Such a scoring model arises from a variety of preference models, as shown in the examples later in this section.

We now turn to the selection component of the recommender, which translates the score $\mathbf{s}_u$ into a probability distribution over target items. The stochastic policy we consider is:

Definition 6.1 (Soft-max selection)**.** For $i \in \Omega_u^t$, the probability of item selection is given by

$$P(\pi_\beta(\mathbf{s}_u, \Omega_u^t) = i) = \frac{e^{\beta s_{ui}}}{\sum_{j \in \Omega_u^t} e^{\beta s_{uj}}}.$$

This form of stochastic policy samples an item according to a Boltzmann distribution defined by the predicted scores (Figure 6.2C). Distributions of this form are common in machine learning applications, and are known as Boltzmann sampling in reinforcement learning or online learning settings [WXLGC17; CGLN17].

Convex Optimization

We now show that under affine score update models and soft-max selection rules, the maximum stochastic reachability problem can be solved by an equivalent convex problem. First notice that for a soft-max selection rule with parameter β, we have that

$$\log\left(P(\pi_\beta(\mathbf{s}_u, \Omega_u^t) = i)\right) = \beta s_{ui} - \operatorname*{LSE}_{j \in \Omega_u^t}\left(\beta s_{uj}\right)$$

where LSE is the log-sum-exp function.

Maximizing stochastic reachability is equivalent to minimizing its negative log-likelihood. Letting $\mathbf{b}_{ui} \in \mathbb{R}^{|\Omega_u^{\mathcal{A}}|}$ denote the ith row of the action matrix B_u and substituting the form of the score update rule, we have the equivalent optimization problem:

$$\min_{\mathbf{a} \in \mathcal{A}_u} \operatorname*{LSE}_{j \in \Omega_u^t}\left(\beta(\mathbf{b}_{uj}^\top \mathbf{a} + c_{uj})\right) - \beta(\mathbf{b}_{ui}^\top \mathbf{a} + c_{ui}) \tag{6.3.1}$$

If the optimal value to (6.3.1) is $\gamma^\star(u, i)$, then the optimal value for (6.2.1) is given by $\rho^\star(u, i) = e^{-\gamma^\star(u,i)}$.

The objective in (6.3.1) is convex because log-sum-exp is a convex function, affine functions are convex, and the composition of a convex and an affine function is convex. Therefore, whenever the action space $\mathcal{A}_u$ is convex, so is the optimization problem. The size of the decision variable scales with the dimension of the action, while the objective function relies on a matrix-vector product of size $|\Omega_u^t| \times |\mathcal{A}_u|$. Being able to solve the maximum reachability problem quickly is of interest, since auditing an entire system requires computing $\rho^\star$ for many user and item pairs.

In particular, the optimization problem (6.3.1) can be solved as an optimization over the exponential cone:

$$\begin{aligned} \min_{t,\mathbf{a},\mathbf{u}} \;\; & t - \beta(\mathbf{b}_{ui}^\top \mathbf{a} + c_{ui}) \\ \text{s.t.} \;\; & \mathbf{a} \in \mathcal{A}_u, \quad \sum_{j \in \Omega_u^t} u_j \leq 1, \\ & \left(u_j, 1, \beta(\mathbf{b}_{uj}^\top \mathbf{a} + c_{uj}) - t\right) \in \mathcal{K}_{exp} \;\; \forall \;\; j \in \Omega_u^t \end{aligned} \tag{6.3.2}$$

Examples

We review examples of common preference models and show how the score updates have an affine form.

Example 6.1 (MF-SGD)**.** Matrix factorization models compute scores as rating predictions so that $S = PQ^\top$, where $P \in \mathbb{R}^{n \times d}$ and $Q \in \mathbb{R}^{m \times d}$ are respectively user and item factors for some latent dimension d. They are learned via the optimization

$$\min_{P,Q} \sum_{u-1}^{n} \sum_{i \in \Omega_u} \|\mathbf{p}_u^\top \mathbf{q}_i - r_{ui}\|_2^2 \,.$$

Under a stochastic gradient descent minimization scheme with step size α, the one-step update rule for a user factor is

$$\mathbf{p}_u^+ = \mathbf{p}_u - \alpha \sum_{i \in \Omega_u^{\mathcal{A}}} (\mathbf{q}_i \mathbf{q}_i^\top \mathbf{p}_u - \mathbf{q}_i r_{ui}) \,,$$

Notice that this expression is affine in the action items. Therefore, we have an affine score function:

$$\phi_u(\mathbf{a}) = Q\mathbf{p}_u^+ = Q\left(\mathbf{p}_u - \alpha Q_{\mathcal{A}}^\top Q_{\mathcal{A}} \mathbf{p}_u - \alpha Q_{\mathcal{A}}^\top \mathbf{a}\right)$$

where we define $Q_{\mathcal{A}} = Q_{\Omega_u^{\mathcal{A}}} \in \mathbb{R}^{|\Omega_u^{\mathcal{A}}| \times d}$. Therefore,

$$B_u = -\alpha Q Q_{\mathcal{A}}^\top, \quad \mathbf{c}_u = Q\left(\mathbf{p}_u - \alpha Q_{\mathcal{A}}^\top Q_{\mathcal{A}} \mathbf{p}_u\right) \,.$$

Example 6.2 (Item-KNN)**.** Neighborhood models compute scores as rating predictions by a weighted average, with:

$$s_{ui} = \frac{\sum_{j \in \mathcal{N}_i} w_{ij} r_{uj}}{\sum_{j \in \mathcal{N}_i} |w_{ij}|}$$

where w_{ij} are weights representing similarities between items and $\mathcal{N}_i$ is a set of indices of previously rated items in the neighborhood of item i. Regardless of the details of how these parameters are computed, the predicted scores are a linear function of observed scores: $\mathbf{s}_u = W\mathbf{r}_u$.

Therefore, the score updates take the form

$$\phi_u(\mathbf{a}) = W\mathbf{r}_u^+ = \underbrace{W\mathbf{r}_u}_{\mathbf{c}_u} + \underbrace{WE_{\Omega_u^{\mathcal{A}}}}_{B_u}\mathbf{a}$$

where $E_{\Omega_u^{\mathcal{A}}}$ selects rows of W corresponding to action items.

Example 6.3 (SLIM and EASE)**.** For both SLIM [NK11] and EASE [Ste19], scores are computed as

$$s_{ui} = \mathbf{w}_i^\top \mathbf{r}_u$$

for $\mathbf{w}_i$ the row vectors of a weight matrix W. For SLIM, the sparse weights are computed as

$$\begin{aligned} \min_W \ & \frac{1}{2}\|R - RW\|_F^2 + \frac{\beta}{2}\|W\|_F^2 + \lambda\|W\|_1 \\ \text{s.t. } & W \geq 0, \operatorname{diag}(W) = 0 \end{aligned}$$

For EASE, the weights are computed as

$$\begin{aligned} \min_W \ & \frac{1}{2}\|R - RW\|_F^2 + \lambda\|W\|_F^2 \\ \text{s.t. } & \operatorname{diag}(W) = 0 \end{aligned}$$

In both cases, the score updates take the form

$$\phi_u(\mathbf{a}) = \underbrace{W\mathbf{r}_u}_{\mathbf{c}_u} + \underbrace{WE_{\Omega_u^{\mathcal{A}}}}_{B_u}\mathbf{a}\,.$$

In all these examples, the action matrices can be decomposed into two terms. The first is a term that depends only on the preference model (e.g. item factors Q or weights W), while the second is further dependent on the user action model (e.g. action item factors $Q_{\mathcal{A}}$ or action selector $E_{\Omega_u^{\mathcal{A}}}$).

For simplicity of presentation, the examples above leave out model bias terms, which are common in practice. Incorporating these model biases changes only the definition of the affine term in the score update expression. We now present the full action model derivation with biases.

Example 6.4 (Biased MF-SGD)**.** Biased matrix factorization models [KB15] compute scores as rating predictions with

$$s_{ui} = \mathbf{p}_u^\top \mathbf{q}_i + f_u + g_i + \mu$$

$P \in \mathbb{R}^{n\times d}$ and $Q \in \mathbb{R}^{m\times d}$ are respectively user and item factors for some latent dimension d, $\mathbf{f} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$ are respectively user and item biases, and $\mu \in \mathbb{R}$ is a global bias.

The parameters are learned via the regularized optimization

$$\min_{P,Q,\mathbf{f},\mathbf{g},\mu} \frac{1}{2}\sum_{u=1}^{n}\sum_{i\in\Omega_u} \|\mathbf{p}_u^\top \mathbf{q}_i + f_u + g_i + \mu - r_{ui}\|_2^2 + \frac{\lambda}{2}\|P\|_F^2 + \frac{\lambda}{2}\|Q\|_F^2 \,.$$

Under a stochastic gradient descent minimization scheme [Kor08] with step size α, the one-step update rule for a user factor is

$$\mathbf{p}_u^+ = \mathbf{p}_u - \alpha \sum_{i\in\Omega_u^{\mathcal{A}}} (\mathbf{q}_i\mathbf{q}_i^\top \mathbf{p}_u + \mathbf{q}_i(f_u + g_i + \mu) - \mathbf{q}_i r_{ui}) - \alpha\lambda\mathbf{p}_u \,.$$

User bias terms can be updated in a similar manner, but because the user bias is equal across items, it does not impact the selection of items.

Notice that this expression is affine in the mutable ratings. Therefore, we have an affine score function:

$$\phi_u(\mathbf{a}) = Q\mathbf{p}_u^+ = Q\left((1-\alpha\lambda)\mathbf{p}_u - \alpha Q_{\mathcal{A}}^\top (Q_{\mathcal{A}}\mathbf{p}_u + \mathbf{g}_{\mathcal{A}} + (\mu + f_u)\mathbf{1}) + \alpha Q_{\mathcal{A}}^\top \mathbf{a}\right)$$

where we define $Q_{\mathcal{A}} = Q_{\Omega_u^{\mathcal{A}}} \in \mathbb{R}^{|\Omega_u^{\mathcal{A}}|\times d}$ and $\mathbf{g}_{\mathcal{A}} = \mathbf{g}_{\Omega_u^{\mathcal{A}}} \in \mathbb{R}^{|\Omega_u^{\mathcal{A}}|}$. Therefore,

$$B_u = \alpha Q Q_{\mathcal{A}}^\top, \quad \mathbf{c}_u = Q\left((1+\lambda)\mathbf{p}_u - \alpha Q_{\mathcal{A}}^\top (Q_{\mathcal{A}}\mathbf{p}_u + \mathbf{g}_{\mathcal{A}} + (\mu + f_u)\mathbf{1})\right) \,.$$

Example 6.5 (Biased MF-ALS)**.** Rather than a stochastic gradient descent minimization scheme, we may instead update the model with an alternating least-squares strategy [ZWSP08]. In this case, the update rule is

$$\begin{aligned}\mathbf{p}_u^+ &= \arg\min_{\mathbf{p}} \sum_{i\in\Omega_u^{\mathcal{A}}\cap\Omega_u} \|\mathbf{p}^\top \mathbf{q}_i + f_u + g_i + \mu - r_{ui}\|_2^2 + \lambda\|\mathbf{p}\|_2^2 \\ &= (Q_u^\top Q_u + \lambda I)^{-1}(Q^\top \mathbf{r}_u + Q_{\mathcal{A}}^\top(\mathbf{g}_{\mathcal{A}} + (\mu + f_u)\mathbf{1}) + Q_{\mathcal{A}}^\top \mathbf{a})\end{aligned}$$

where we define $Q_u = Q_{\Omega_u^{\mathcal{A}}\cap\Omega_u}$. Similar to in the SGD setting, this is an affine expression, and therefore we end up with the affine score parameters

$$B_u = Q(Q_u^\top Q_u + \lambda I)^{-1} Q_{\mathcal{A}}^\top, \quad \mathbf{c}_u = Q(Q_u^\top Q_u + \lambda I)^{-1}(Q^\top \mathbf{r}_u + Q_{\mathcal{A}}^\top(\mathbf{g}_{\mathcal{A}} + (\mu + f_u)\mathbf{1})) \,.$$

Example 6.6 (Biased Item-KNN)**.** Biased neighborhood models [DK11] compute scores as rating predictions by a weighted average, with

$$s_{ui} = \mu + f_u + g_i + \frac{\sum_{j \in \mathcal{N}_i} w_{ij}(r_{uj} - \mu - f_u - g_i)}{\sum_{j \in \mathcal{N}_i} |w_{ij}|}$$

where w_{ij} are weights representing similarities between items, $\mathcal{N}_i$ is a set of indices which are in the neighborhood of item i, and $\mathbf{f}, \mathbf{g}, \mu$ are bias terms. Regardless of the details of how these parameters are computed, the predicted scores are an affine function of observed scores:

$$\mathbf{s}_u = W\mathbf{r}_u - W(\mathbf{g} + (\mu + f_u)\mathbf{1}) + \mathbf{g} + (\mu + f_u)\mathbf{1}$$

where we can define

$$W_{ij} = \begin{cases} \frac{w_{ij}}{\sum_{j \in \mathcal{N}_i} |w_{ij}|} & j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the score updates take the form

$$\phi_u(\mathbf{a}) = \underbrace{W(\mathbf{r}_u - \mathbf{g} + (\mu + f_u)\mathbf{1}) + \mathbf{g} + (\mu + f_u)\mathbf{1})}_{\mathbf{c}_u} + \underbrace{WE_{\Omega_u^{\mathcal{A}}}}_{B_u}\mathbf{a}\,.$$

6.4 Impact of Preference Model Geometry

In this section, we explore the connection between stochastic and deterministic reachability to illustrate how both randomness and agency contribute to discovery as defined by the max reachability metric. We make connections between agency and model geometry for matrix factorization models. This allows us to argue by example that it is possible to design preference models that guarantee deterministic reachability, and that doing so does not induce accuracy trade-offs.

Connection to Deterministic Recommendation

We now explore how the softmax style selection rule is a relaxation of top-1 recommendation. For larger values of β, the selection rule distribution becomes closer to the deterministic top-1 rule. This also means that the stochastic reachability problem can be viewed as a relaxation of the top-1 reachability problem.

In stochastic settings it is relevant to inquire the extent to which randomness impacts discovery and availability. In the deterministic setting, the reachability of an item to a user is closely tied to agency—the ability of a user to influence their outcomes. The addition of randomness induces exploration, but not in a way that is controllable by users. In the following result, we show how this trade-off manifests in the max reachability metric itself.

Proposition 6.4.1. *Consider the stochastic reachability problem for a β-softmax selection rule as $\beta \to \infty$. Then if an item $i \in \Omega_u^t$ is top-1 reachable by user u, $\rho^\star(u, i) \to 1$. In the opposite case that item i* is not *top-1 reachable, we have that $\rho^\star(u, i) \to 0$.*

Proof. Define

$$\gamma_\beta(\mathbf{a}) = \underset{j \in \Omega_u^t}{\text{LSE}} \left(\beta \phi_{uj}(\mathbf{a})\right) - \beta \phi_{ui}(\mathbf{a})$$

and see that $\rho_{ui}(\mathbf{a}) = e^{-\gamma_\beta(\mathbf{a})}$. Then we see that

$$\lim_{\beta \to \infty} \frac{1}{\beta} \gamma_\beta(\mathbf{a}) = \max_{j \in \Omega_u^t} \left(\phi_{uj}(\mathbf{a})\right) - \phi_{ui}(\mathbf{a})$$

yields a top-1 expression. If an item i is top-1 reachable for user u, then there is some $\mathbf{a}$ such that the above expression is equal to zero. Therefore, as $\beta \to \infty$, $\gamma^\star \to 0$, hence $\rho^\star \to 1$. In the opposite case when an item *is not* top-1 reachable we have that $\gamma^\star \to \infty$, hence $\rho^\star \to 0$. □

This connection yields insight into the relationship between max reachability, randomness, and agency in stochastic recommender systems. For items which are top-1 reachable, larger values of β result in larger $\rho^\star$, and in fact the largest possible max reachability is attained as $\beta \to \infty$, i.e. there is no randomness. On the other hand, if β is too large, then items which are not top-1 reachable will have small $\rho^\star$. There is some optimal finite $\beta \geq 0$ that maximizes $\rho^\star$ for top-1 unreachable items. Therefore, we see a delicate balance when it comes to ensuring access with randomness.

Viewed in another light, this result says that for a fixed $\beta \gg 1$, deterministic top-1 reachability ensures that $\rho^\star$ will be close to 1. We now explore this perspective. Specializing to affine score update models, we now highlight how parameters of the preference and action models play a role in determining max reachability.

Recall the definition of the convex hull.

Definition 6.2 (Convex hull)**.** The *convex hull* of a set of vectors $\mathcal{V} = \{\mathbf{v}_i\}_{i=1}^n$ is defined as

$$\text{conv}\,(\mathcal{V}) = \left\{ \sum_{i=1}^n w_i \mathbf{v}_i \mid \mathbf{w} \in \mathbb{R}_+^n, \quad \sum_{i=1}^n w_i = 1 \right\} .$$

A point $\mathbf{v}_j \in \mathcal{V}$ is a *vertex* of the convex hull if

$$\mathbf{v}_j \notin \text{conv}\left(\mathcal{V} \setminus \{\mathbf{v}_j\}\right) .$$

Proposition 6.4.2. *Consider the stochastic reachability problem for a β-softmax selection rule and affine score update model with action vectors $\{\mathbf{b}_{uj}\}_{j=1}^m$. Fix an $i \in \Omega_u^t$. If $\mathbf{b}_{ui}$ is a vertex on the convex hull of $\{\mathbf{b}_{uj}\}_{j \in \Omega_u^t}$ and actions are real-valued, then $\rho_{ui}^\star \to 1$ as $\beta \to \infty$ and the item i is top-1 reachable.*

Proof. We begin by showing that if $\mathbf{b}_{ui}$ is a vertex on the convex hull of $\mathcal{B} = \{\mathbf{b}_{uj}\}_{j \in \Omega_u^t}$, then item i is top-1 reachable.

Item i is top-1 reachable if there exists some $\mathbf{a} \in \mathbb{R}^{|\Omega_u^{\mathcal{A}}|}$ such that $\mathbf{b}_{ui}^\top \mathbf{a} + c_{ui} \geq \mathbf{b}_{uj}^\top \mathbf{a} + c_{uj}$ for all $j \neq i$. Therefore, top-1 reachability is equivalent to the feasibility of the following linear program

$$\begin{aligned} &\min\ 0^\top \mathbf{a} \\ &\text{s.t. } D_{ui}\mathbf{a} \geq \mathbf{f}_{ui} \end{aligned}$$

where D_{ui} has rows given by $\mathbf{b}_{ui} - \mathbf{b}_{uj}$ and $\mathbf{f}_{ui}$ has entries given by $c_{uj} - c_{ui}$ for all $j \in \Omega_u^t$ with $j \neq i$. Feasibility of this linear program is equivalent to boundedness of its dual:

$$\begin{aligned} &\max\ \mathbf{f}_{ui}^\top \lambda \\ &\text{s.t. } D_{ui}^\top \lambda = 0,\ \ \lambda \geq 0. \end{aligned}$$

We now show that if $\mathbf{b}_{ui}$ is a vertex on the convex hull of $\mathcal{B}$, then the dual is bounded because the only feasible solution is $\lambda = 0$. To see why, notice that

$$D_{ui}^\top \lambda = 0 \iff \mathbf{b}_{ui} \sum_{\substack{j \in \Omega_u^t \\ j \neq i}} \lambda_j = \sum_{\substack{j \in \Omega_u^t \\ j \neq i}} \lambda_j \mathbf{b}_{uj}\ .$$

If this expression is true for some $\lambda \neq 0$, then we can write

$$\mathbf{b}_{ui} = \sum_{\substack{j \in \Omega_u^t \\ j \neq i}} w_j \mathbf{b}_{uj}, \quad w_j = \frac{\lambda_j}{\sum_{\substack{j \in \Omega_u^t \\ j \neq i}} \lambda_j} \implies \mathbf{b}_{ui} \in \operatorname{conv}(\mathcal{B} \setminus \{\mathbf{b}_{ui}\})\ .$$

This is a contradiction, and therefore it must be that $\lambda = 0$ and therefore the dual is bounded and item i is top-1 reachable.

To finish the proof, we appeal to Proposition 6.4.1 to argue that since item i is top-1 reachable, then $\rho_{ui}^\star \to 1$ as $\beta \to \infty$. □

This result highlights how the geometry of the score model determines when it is preferable for the system to have minimal exploration, from the perspective of reachability.

Latent Geometry in MF Models

We now restrict our attention to top-1 recommendation policies with score models based on matrix factorization, as in Example 6.1. This allows for a re-interpretation of our results so far in terms of the latent space of item and user factors. The action vectors specialized to the matrix factorization case are given as, for each $i \in \{1, \ldots, m\}$

$$\mathbf{b}_{ui} = -\alpha Q_{\mathcal{A}} \mathbf{q}_i\ .$$

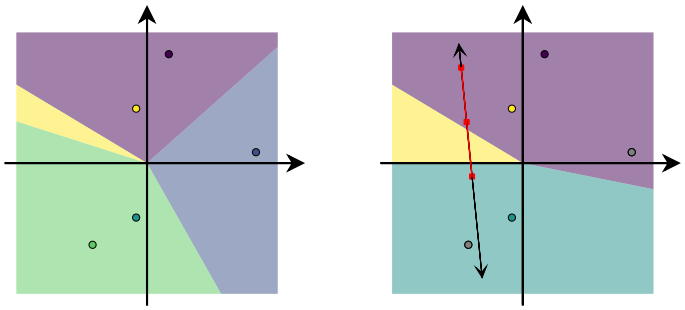

Figure 6.4: Left: Items in Latent Space. Right: Availability of Items to a User. An example of item factors (indicted by colored points) in $d = 2$. In (a), the top-1 *item regions* are indicated by color. The teal item is unavailable. In (b), reachability for a user who has seen the blue and the green items (now in gray) with the blue item's rating fixed. The black line indicates how the user's representation can change depending on their rating of the green item. The red region indicates the constraining effect of requiring bounded and integer-valued ratings, which affect the reachability of the yellow item.

The geometry of the action vectors is thus determined by the geometry of the item factors and the set of action items. These two can be decomposed by re-interpreting the score update model, described in Example 6.1, as consisting of two steps:

1. Update the user factor, depending on $Q_{\mathcal{A}}$ and actions $\mathbf{a}$.
2. Recompute scores, depending on the target item factors $\{\mathbf{q}_j \mid j \in \Omega_u^t\}$.

With this interpretation in mind, we can define the *item-region* for item i as

$$\mathcal{S}_i = \{\mathbf{p} \mid \mathbf{q}_i^\top \mathbf{p} > \mathbf{q}_j^\top \mathbf{p} \; \forall \; j \neq i \in \Omega_u^t\}$$

If a user's latent vector lies within this set, then the user will be recommended item i. We can therefore interpret reachability as the search for an action $\mathbf{a}$ that brings the user's updated latent vector $\mathbf{p}_u^+$ into the item region $\mathcal{S}_i$. There are thus two potential barriers to reachability: first, the region $\mathcal{S}_i$ could be empty; second, the action space may not be rich enough to allow for full movement. The first stems from the geometry of the learned preference model parameters, while the second from user interface design. This idea is illustrated in Figure 6.4 for a toy example with latent dimension $d = 2$.

User cold-start. Reachability yields a novel perspective on how to incorporate new users into a recommender system. The user cold-start problem is the challenge of selecting items to show a user who enters a system with no rating history from which to predict their preferences. This is a major issue with collaboratively filtered recommendations, and systems often rely on incorporating extraneous information [SPUP02]. These strategies

focus on presenting items which are most likely to be rated highly or to be most informative about user preferences [BLR17].

The idea of reachability offers an alternative point of view. Rather than evaluating a potential "onboarding set" only for its contribution to model accuracy, we can choose a set which additionally ensures that many types of items are reachable. We can evaluate an onboarding set by the geometry of the action vectors in latent space. Inspired by Proposition 6.4.2, in the case of matrix factorization, we should check for any candidate set of action items $\Omega_u^{\mathcal{A}}$, what proportion of the elements $\{Q_{\mathcal{A}}\mathbf{q}_j \mid j \in \Omega_u^t\}$ are vertices on the convex hull. Due to computational difficulties, we primarily propose this be used to distinguish between candidate onboarding sets, based on the ways these sets provide user agency. We leave to future work the task of generating candidate sets based on these reachability properties.

Reachability Without Sacrificing Accuracy

We now consider whether geometric properties of the preference model relevant to reachability are predetermined by the goal of accurate prediction. Is there a tension between ensuring reachability and accuracy? We answer in the negative by presenting a construction for the case of matrix factorization models. Our result shows that the item and user factors (P and Q) can be slightly altered such that all items become top-1 reachable at no loss of predictive accuracy. The construction expands the latent dimension of the user and item factors by one and relies on a notion of sufficient richness for action items.

Definition 6.3 (Rich actions)**.** For a set of item factors $\{\mathbf{q}_j\}_{j=1}^m$, let $C = \max_j \|\mathbf{q}_j\|_2$. Then a set of action items $\Omega_u^{\mathcal{A}} \subseteq \{1, \ldots, m\}$ is *sufficiently rich* if the vertical concatenation of their item factors and norms is full rank:

$$\operatorname{rank}\left(\begin{bmatrix}\mathbf{q}_i^\top & \sqrt{C^2 - \|\mathbf{q}_i\|_2^2}\end{bmatrix}_{i\in\Omega_u^{\mathcal{A}}}\right) = d + 1\,.$$

Notice that this can only be true if $|\Omega_u^{\mathcal{A}}| \geq d + 1$.

Proposition 6.4.3. *Consider the MF model with user factors $P \in \mathbb{R}^{n\times d}$ and item factors $Q \in \mathbb{R}^{m\times d}$. Further consider any user u with a sufficiently rich set of at least $d+1$ action items and real-valued actions. Then there exist $\tilde{P} \in \mathbb{R}^{n\times d+1}$ and $\tilde{Q} \in \mathbb{R}^{m\times d+1}$ such that $PQ^\top = \tilde{P}\tilde{Q}^\top$ and under this model, all target items $i \in \Omega_u^t$ are top-1 reachable.*

Proof. Let C be the maximum row norm of Q and define $\mathbf{v} \in \mathbb{R}^m$ such that $v_i^2 = C^2 - \|\mathbf{q}_i\|_2^2$. Then we construct modified item and user factors as

$$\tilde{Q} = \begin{bmatrix}Q & \mathbf{v}\end{bmatrix}, \quad \tilde{P} = \begin{bmatrix}P & \mathbf{0}\end{bmatrix}\,.$$

Therefore, we have that $\tilde{P}\tilde{Q}^\top = PQ^\top$.

Then notice that by construction, each row of $\tilde{Q}$ has norm C, so each $\tilde{\mathbf{q}}_i$ is on the boundary of an ℓ_2 ball in $\mathbb{R}^{d+1}$. As a result, each $\tilde{\mathbf{q}}_i$ is a vertex on the convex hull of $\{\tilde{\mathbf{q}}_j\}_{j=1}^n$ as long as all $\mathbf{q}_j$ are unique.

For an arbitrary user u, the score model parameters are given by $\tilde{\mathbf{b}}_{ui} = -\alpha \tilde{Q}_{\mathcal{A}} \tilde{\mathbf{q}}_i$. We show by contradiction that as long as the action items are sufficiently rich, each $\tilde{\mathbf{b}}_{ui}$ is a vertex on the convex hull of $\{\tilde{\mathbf{b}}_{uj}\}_{j=1}^n$. Supposing this is not the case for an arbitrary i,

$$\tilde{\mathbf{b}}_{ui} = \sum_{\substack{j=1 \\ j\neq i}}^{n} w_j \tilde{\mathbf{b}}_{uj} \iff \tilde{Q}_{\mathcal{A}} \tilde{\mathbf{q}}_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} w_j \tilde{Q}_{\mathcal{A}} \tilde{\mathbf{q}}_i \implies \tilde{\mathbf{q}}_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} w_j \tilde{\mathbf{q}}_i$$

where the final implication follows because the fact that $\tilde{Q}_{\mathcal{A}}$ is full rank (due to richness) implies that $\tilde{Q}_{\mathcal{A}}^\top \tilde{Q}_{\mathcal{A}}$ is invertible. This is a contradiction, and therefore we have that each $\tilde{\mathbf{b}}_{ui}$ must be a vertex on the convex hull of $\{\tilde{\mathbf{b}}_{uj}\}_{j=1}^n$. Finally, we appeal to Proposition 6.4.2. □

The existence of such a construction demonstrates that there is not an unavoidable trade-off between accuracy and reachability in recommender systems.

6.5 Audit Demonstration

Experimental Setup

Code is available at `www.github.com/modestyachts/stochastic-rec-reachability`.

Datasets. We evaluate max ρ reachability in settings based on three popular recommendation datasets: MovieLens 1M (ML-1M) [HK15], LastFM 360K [Cel10] and MIcrosoft News Dataset (MIND) [WQC+20]. ML-1M is a dataset of 1 through 5 explicit ratings of movies, containing over one million recorded ratings; we do not perform any additional preprocessing. LastFM is an impliciGangnamt rating dataset containing the number of times a user has listened to songs of an artist. We used the version of the LastFM dataset preprocessed by Shakespeare et al. [SPGC20]. For computational tractability, we select a random subset of 10% of users and 10% artists and define ratings as $r_{ui} = \log(\#\text{listens}(u, i) + 1)$ to ensure that rating matrices are well conditioned. MIND is an implicit rating dataset containing clicks and impressions data. We use a subset of 50K users and 40K news articles spanning 17 categories and 247 subcategories. We transform news level click data into subcategory level aggregation and define the rating associated with a user-subcategory pair as a function of the number of times that the user clicked on news from that subcategory: $r_{ui} = \log(\#\text{clicks}(u, i) + 1)$. Table 6.1 provides summary statistics and Section 6.7 contains further details about the datasets and preprocessing steps.

Table 6.1: Audit datasets

Data set	ML 1M	LastFM 360K	MIND
Users	6040	13698	50000
Items	3706	20109	247
Ratings	1000209	178388	670773
Density (%)	4.47%	0.065%	5.54%
LibFM rmse	0.716	1.122	0.318
KNN rmse	0.756	1.868	-

Preference models. We consider two preference models: one based on matrix factorization (MF) as well as a neighborhood based model (KNN). We use the LibFM SGD implementation [Ren12] for the MF model and use the item-based k-nearest neighbors model implemented by Krauth et al. [KDZ+20]. For each dataset and recommender model we perform hyper-parameter tuning using a 10%-90% test-train split. We report test performance in Table 6.1. See Section 6.7 for details about tuning. Prior to performing the audit, we retrain the recommender models with the full dataset.

Reachability experiments. We solve the conic program (6.3.2) using the MOSEK Python API under an academic license [ApS19]. To compute reachability, it is further necessary to specify additional elements of the recommendation pipeline: the user action model, the set of target items, and the soft-max selection parameter.

We consider three types of user action spaces: *History Edits*, *Future Edits*, and *Next K* in which users can strategically modify the ratings associated to K randomly chosen items from their history, K randomly chosen unobserved items, or the top-K items according to the baseline scores of the preference model. For each of the action spaces we consider a range of K values. We further constrain actions to lie in an interval corresponding to the rating range, using $[1, 5]$ for movies and $[0, 10]$ for music and news.

In the case of movies (ML-1M) we consider target items to be all items that are neither seen nor action items. In the case of music and news recommendations (LastFM & MIND), the target items are all the items with the exception of action items. This reflects an assumption that music created by a given artist or news within a particular subcategory can be consumed repeatedly, while movies are viewed once.

For each dataset and recommendation pipeline, we compute max reachability for soft-max selection rules parametrized by a range of β values. Due to the computational burden of large dense matrices, we compute metrics for a subset of users and target items sampled uniformly at random. For details about runtime, see Section 6.7.

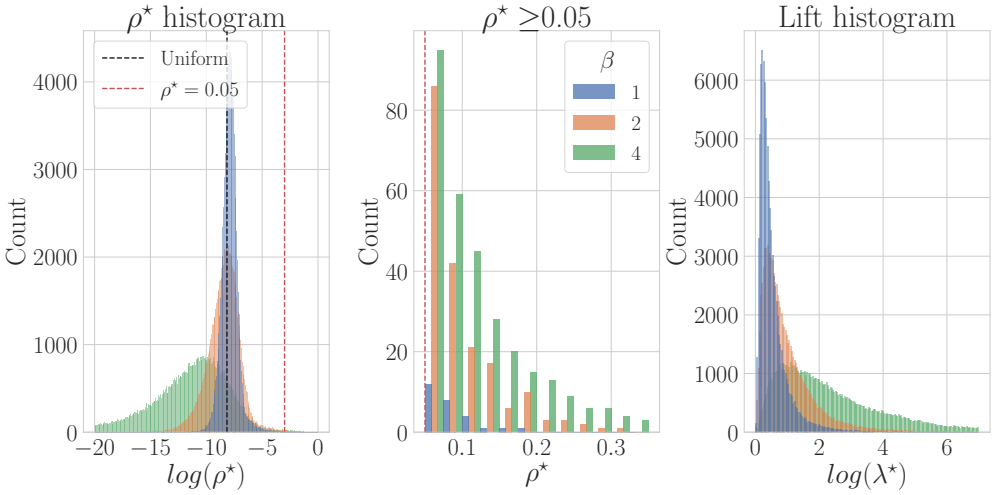

Figure 6.5: Left: Histogram of log max reachability values for $\beta = [1, 2, 4]$. Black dotted line denotes $\rho^\star$ for uniformly random recommender. Center: Histogram of $\rho^\star > 0.05$ (red dotted line). Right: Histogram of log-lifts. Reachability evaluated on ML-1M for $K = 5$ Random Future action space and a LibFM model.

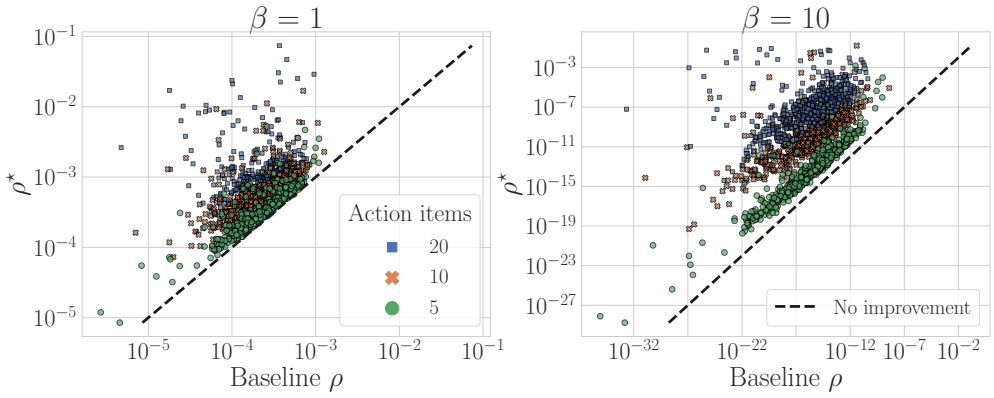

Figure 6.6: Log scale scatterplot of $\rho^\star$ values against baseline ρ for $K \in [5, 10, 20]$. Colors indicate action space size K. We compare low (left) and high (right) stochasticity. Reachability evaluated on ML-1M for Random Future action space and a LibFM model.

Impact of Recommender Pipeline

We begin by examining the role of recommender pipeline components: stochasticity of item selection, user action models, and choice of preference model. All presented experiments in this section use the ML-1M dataset.

These experiments show that more stochastic recommendations correspond to higher average max reachability values, whereas more deterministic recommenders have a more disparate impact, with a small number of items achieving higher $\rho^\star$. We also see that the impact of the user action space differs depending on the preference model. For neighborhood based preference models, strategic manipulations to the history are most effective at maximizing reachability, whereas manipulations of the items most likely to be recommended next are ineffective.

Role of stochasticity. We investigate the role of the β parameter in the item selection policy. Figure 6.5 illustrates the relationship between the stochasticity of the selection policy and max reachability. There are significantly more target items with better than random reachability for low values of β. However, higher values of β yield more items with high reachability potential ($> 5\%$ likelihood of recommendation). These items are typically items that are top-1 or close to top-1 reachable. While lower β values provide better reachability on average and higher β values provide better reachability at the "top", higher β uniformly out-performs lower β values in terms of the lift metric. This suggests that larger β corresponds to more user agency, since the relative effect of strategic behavior is larger. However, note that for very large values of β, high lift values are not so much the effect of improved reachability as they are due to very low baseline recommendation probabilities.

Role of user action model. We now consider different action space sizes. In Figure 6.6 we plot max reachability for target items of a particular user over varying levels of selection rule stochasticity and varying action space sizes. Larger action spaces correspond to improved item reachability for all values of β. However, increases in the number of action items have a more pronounced effect for larger β values.

While increasing the size of the action space uniformly improves reachability, the same cannot be said about the type of action space. For each user, we compute the average lift over target items as a metric for user agency in a recommender (Figure 6.7). For LibFM, the choice of action space does not strongly impact the average user lift, though *Next K* displays more variance across users than the other two. However, for Item KNN, there is a stark difference between *Next K* and and random action spaces.

Role of preference model. As Figure 6.7 illustrates, a system using LibFM provides more agency on average than one using KNN. We now consider how this relates to properties of the preference models. First, consider the fact that for LibFM, there is higher variance among user-level average lifts observed for *Next K* action space compared with random

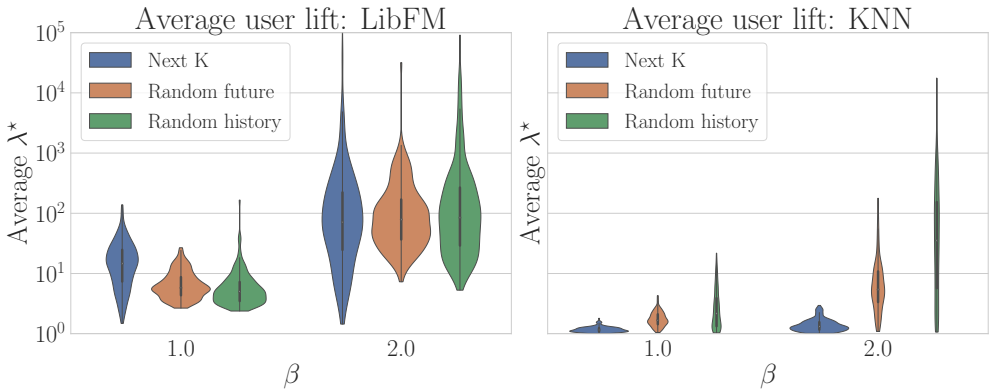

Figure 6.7: The distribution of average lifts (a notion of agency) over users. Colors indicate different user action spaces for LibFM (left) and KNN (right) on ML-1M.

action spaces. This can be understood as resulting from the user-specific nature of *Next K* recommended items. On the other hand, random action spaces are user independent, so it is not surprising that there is less variation across users.

In a neighborhood-based model users have leverage to increase the ρ reachability only for target items in the neighborhood of action items. In the case of KNN, the next items up for recommendation are in close geometrical proximity to each other. This limits the opportunity for discovery of more distant items for *Next K* action space. On the other hand, the action items are more uniformly over space of item ratings in random action models, thus contributing to much higher opportunities for discovery. Additionally, we see that *History Edits* displays higher average lift values than *Future Edits*. We posit that this is due to the fact that editing K items from the history leads to a larger ratio of strategic to non-strategic items.

Bias in Movie, Music, and News Recommendation

We further compare aggregated stochastic reachability against properties of user and items to investigate bias. We aggregate baseline and max reachability to compute user-level metrics of discovery and item-level metrics of availability. The audit demonstrates popularity bias for items with respect to baseline availability. This bias persists in the best case for neighborhood based recommenders and is thus unavoidable, whereas it could be mitigated for MF recommenders. User discovery aggregation reveals inconclusive results with weak correlations between the length of users' experience and their ability to access content.

Popularity bias. In Figure 6.8, we plot the baseline and best case item availability (as in (6.2.3)) to investigate popularity bias. We consider popularity defined by the average rating of an item in a dataset. Another possible definition of popularity is rating frequency,

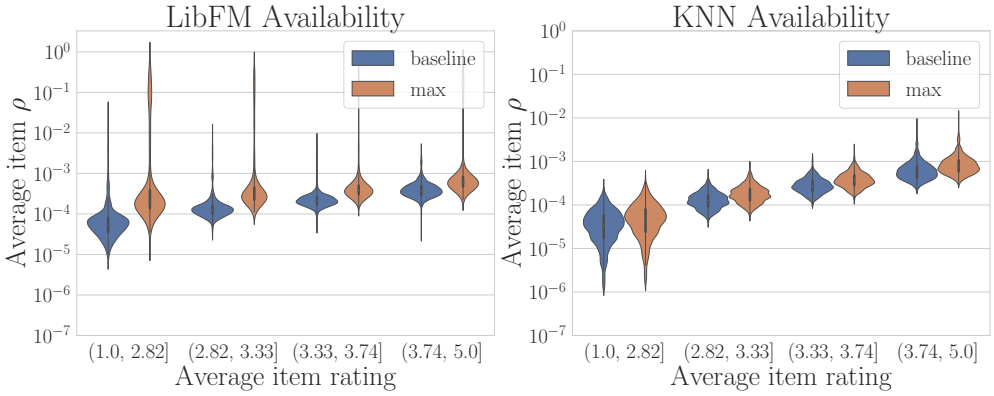

Figure 6.8: Comparison of baseline and best case availability of content, across four popularity categories for LibFM (left) and KNN (right) preference models. Reachability evaluated on ML-1M for *Next 10* action space with $\beta = 2$.

but for this definition we did not observe any discernible bias. For both LibFM and KNN models, the baseline availability displays a correlation with item popularity, with Spearman's rank-order correlations of $r_s = 0.87$ and $r_s = 0.95$. This suggests that as recommendations are made and consumed, more popular items will be recommended at disproportionate rates.

Furthermore, the best case availability for KNN displays a similar trend ($r_s = 0.94$), indicating that the propagation of popularity bias can occur independent of user behavior. This does not hold for LibFM, where the best case availability is less clearly correlated with popularity ($r_s = 0.57$). The lack of correlation for best case availability holds in the additional settings of music artist and news recommendation with the LibFM model (Figure 6.9). Our audit does not reveal an unavoidable systemic bias for LibFM recommender, meaning that any biases observed in deployment are due in part to user behavior. In contrast, we see a systematic bias for the KNN recommender, meaning that regardless of user actions, the popularity bias will propagate.

Experience bias. To consider the opportunities for discovery provided to users, we perform user level aggregations of max reachability values as in (6.2.3). We investigate experience bias by considering how the discovery metric changes as a function of the number of different items a user has consumed so far, i.e. their experience. Figure 6.10 illustrates that experience is weakly correlated with baseline discovery for movie recommendation ($r_s = 0.48$), but not so much for news recommendation ($r_s = 0.05$). The best case discovery is much higher, meaning that users have the opportunity to discover many of their target items. However, the weak correlation with experience remains for best case discovery of movies ($r_s = 0.53$).

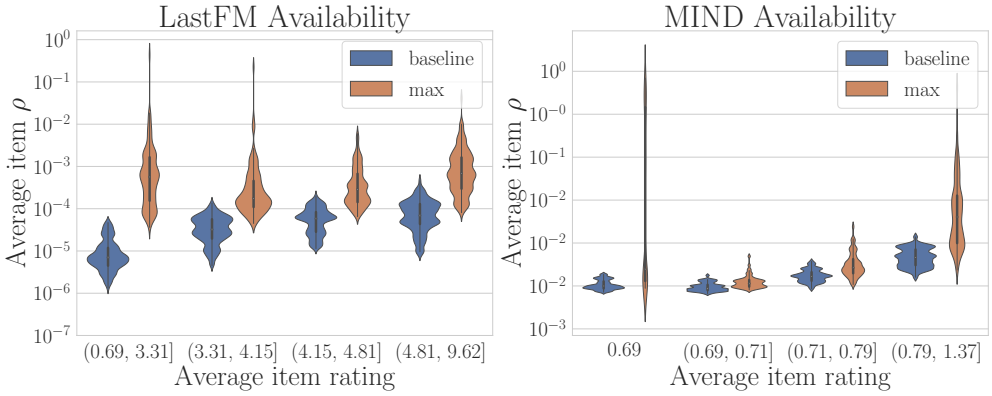

Figure 6.9: Comparison of baseline and best case availability of content for four popularity categories for LastFM (left) and MIND (right) with *Next 10* actions, LibFM model, and $\beta = 2$.

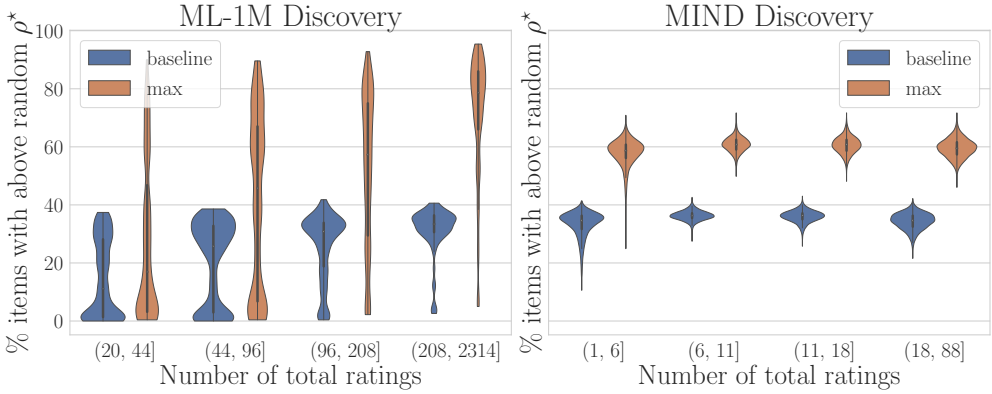

Figure 6.10: Comparison of baseline and best case fraction of items with better than random $\rho^\star$, grouped across four levels of user history length. Reachability evaluated on ML-1M (left) and MIND (right) for *Next 10* action space, $\beta = 2$, and LibFM model.

6.6 Conclusion and Open Problems

In this chapter, we consider the effects of using predicted user preferences to recommend content, a practice prevalent throughout the digital world today. By defining reachability for deterministic and stochastic recommenders, we provide a way to evaluate the impact of using these predictive models to mediate the discovery of content. In applying this lens to linear preference models, we see several interesting phenomena. The first is simple but worth stating: good predictive models, when used to moderate information, can unintentionally make portions of the content library inaccessible to users. This is illustrated in practice in our study of the MovieLens, LastFM, and MIND datasets. Our experiments illustrate the impact of system design choices and historical data on the

availability of content and users' opportunities for discovery, highlighting instances in which popularity bias is inevitable regardless of user behavior.

To some extent, the differences in the availability of items are related to their unpopularity within training data. Popularity bias is a well known phenomenon in which systems fail to personalize [Ste11], and instead over-recommend the most popular pieces of content. Empirical work shows connections between popularity bias and undesirable demographic biases, including the under-recommendation of female authors [ETKMK18]. YouTube was long known to have a popularity bias problem (known as the "Gangnam Style Problem"), until the recommendations began optimizing for predicted "watch time" over "number of views." Their new model has been criticized for its radicalization and political polarization [Tuf18; Nic18]. The choice of prediction target can have a large effect on the types of content users can or are likely to discover, motivating the use of analytic tools like the ones proposed here to reason about these trade-offs before deployment.

While the reachability criteria proposed in this work form an important basis for reasoning about the availability of content within a recommender system, they do not guarantee less biased behavior on their own. Because our audits primarily consider the feasibility, we can only confirm possible outcomes rather than distinguish between probable ones. This can illuminate limitations inherent in recommender systems for organizing and promoting content. Consequently, a reachability audit can lead to actionable insights: system designers can assess potential biases before releasing algorithmic updates into production. Moreover, as reachability depends on the choice of action space, such system-level insights might motivate user interface design: for example, a sidebar encouraging users to re-rate K items from their history. However, reachability alone will not fix problems of filter bubbles or toxic content. There is an important distinction between technically providing reachability and the likelihood that people will actually avail themselves of it.

With these limitations in mind, we mention several ways to extend the ideas presented in this work:

- **Generative theory of reachability.** Analyzing connections between training data and the reachability properties of the resulting model would to give context to empirical work showing how biases can be reproduces in the way items are recommended [ETKMK18; ETA+18].

- **Extended class of preference models.** Many models for rating predictions do not fall into the linear class described here, especially those that incorporate implicit feedback or perform different online update strategies for users. Not all simple models are linear—for example, subset based recommendations offer greater scrutability and thus user agency by design [BRA19].

- **Extended class of selection rules.** Slate-based and top-N recommendation is common, but it is computationally difficult to evaluate the reachability properties of these selection rules. One avenue for addressing more generic recommender poli-

cies might be an approach based on sampling, which could apply to even black-box recommender evaluation.

- **Reachability by design.** Our result on the lack of trade-off between accuracy and reachability is encouraging. Minimum one-step reachability conditions could be efficiently incorporated into learning algorithms for preference models. Alternatively, post-processing approaches to the recommender policy π could work with existing models to modify their reachability properties. For example, Steck [Ste18] proposed a method to ensure that the distribution of recommendations over genres remains the same despite model predictions.

- **Interaction over long horizons.** Directly considering multiple rounds of interactions between users and recommendation systems would shed light on how these models evolve over time. This is a path towards understanding phenomena like filter bubbles and polarization.

We highlight that the reachability lens presents a contrasting view to the popular line of work on robustness in machine learning. When human behaviors are the subject of classification and prediction, building "robustness" into a system may be at odds with ensuring agency. Because the goal of recommendation is personalization more than generalization, it would be appropriate to consider robust access over robust accuracy. This calls for questioning the current normative stance and critically examining system desiderata in light of usage context.

More broadly, we emphasize the importance of auditing systems with learning-based components in ways that directly consider the models' behavior when put into feedback with humans. In the field of formal verification, making guarantees about the behavior of complex dynamical systems over time has a long history. There are many existing tools [ABDM00], though they are generally specialized to the case of physical systems and suffer from the curse of dimensionality. We accentuate the abundance of opportunity for developing novel approximations and strategies for evaluating large scale machine learning systems.

6.7 Additional Experimental Details

Detailed Data Description

MovieLens 1M. ML-1M dataset was downloaded from Group Lens[1] via the RecLab [KDZ+20] interface[2]. It contains 1 through 5 rating data of 6040 users for 3706 movies. There are a total of 1000209 ratings (4.47% rating density). The original data is accompanied by additional user attributes such as age, gender, occupation and zip code. Our experiments

[1] https://grouplens.org/datasets/movielens/1m/

[2] https://github.com/berkeley-reclab/RecLab

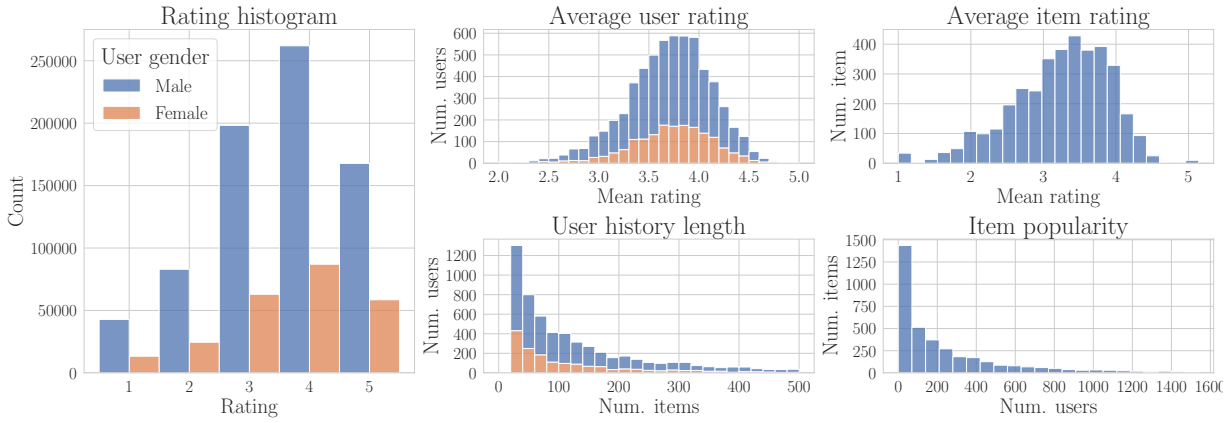

Figure 6.11: Descriptive statistics for the MovieLens 1M dataset split by user gender (28.3% female). The mean ratings of both users and items are roughly normally distributed while user's history length and item popularity display power law distributions.

didn't indicate observable biases across these attributes. We later show user discovery results split by gender.

Figure 6.11 illustrates descriptive statistics for the ML-1M dataset.

LastFM 360K. The LastFM 360K dataset preprocessed[3] by Shakespeare et al. [SPGC20] was loaded via the RecLab interface. It contains data on the number of times users have listened to various artists. We select a random subset of 10% users and a random subset of 10% items yielding 13698 users, 20109 items and 178388 ratings (0.056% rating density). The item ratings are not explicitly expressed by users as in the MovieLens case. For a user u and an artist i we define implicit ratings $r_{ui} = \log(\#\text{listens}(u, i) + 1)$. This data is accompanied by artist gender, an item attribute.

Figure 6.12 illustrates descriptive statistics for the LastFM dataset.

MIcrosoft News Dataset (MIND). MIND is a recently published impression dataset collected from logs of the Microsoft News website [4]. We downloaded the `MIND-small` dataset[5], which contains behavior log data for 50000 randomly sampled users. There are 42416 unique news articles, spanning 17 categories and 247 subcategories. We aggregate user interactions at the subcategory level and consider the problem of news subcategory recommendation. The implicit rating of a user u for subcategory i is defined as: $r_{ui} = \log(\#\text{clicks}(u, i) + 1)$. The resulting aggregated dataset contains 670773 ratings (5.54% rating density).

Figure 6.13 illustrates descriptive statistics for the MIND dataset.

[3] `https://zenodo.org/record/3964506#.XyE5N0FKg5n`

[4] `https://microsoftnews.msn.com/`

[5] `https://msnews.github.io/`

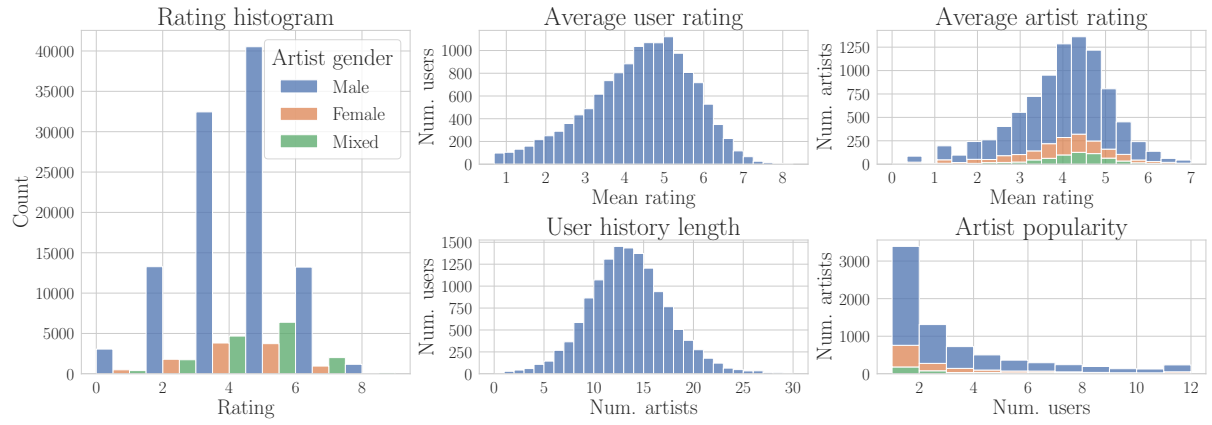

Figure 6.12: Descriptive statistics for the LastFM dataset split by artist gender (over 54% of artists have unknown gender, 36% are male, 6.5% are female and 3.5% are mixed gender). Unlike ML 1M, for the LastFM dataset the user history lengths are normally distributed around a mean of around 12 artists.

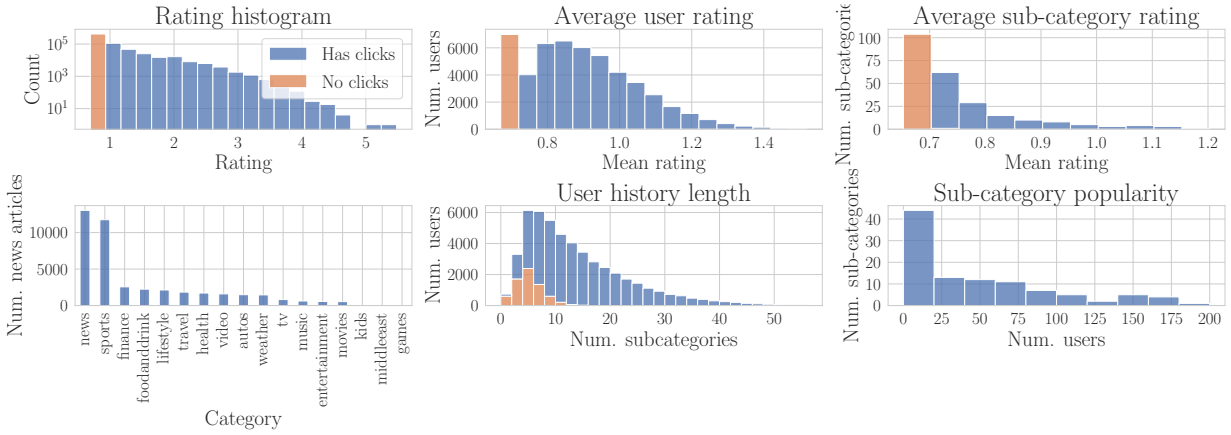

Figure 6.13: Descriptive statistics for the MIND dataset: The orange bars correspond to either user or items that have been displayed but have not clicked/ have not been clicked on. Unlike ML 1M and LastFM, the MIND ratings have strongly skewed distribution, with most user-subcategory ratings corresponding to users clicking on a small number of articles from the sub-category. There is a long tail of higher ratings that corresponds to most popular subcategories. The leftmost plot illustrates the unequal distribution of news articles across categories. The same qualitative behavior holds for sub-categories.

Table 6.2: Tuning results

	LibFM			
Dataset	LR	Reg.	Test RMSE	Run time (s)
ML 1M	0.0112	0.0681	0.716	2.76 ± 0.32
LastFM	0.0478	0.2278	1.122	0.78 ± 0.13
MIND	0.09	0.0373	0.318	3.23 ± 0.37
	KNN			
Dataset	Neigh. size	Shrinkage	Test RMSE	Run time (s)
ML 1M	100	22.22	0.756	0.34 ± 0.07

Model Tuning

For each dataset and recommender model we perform grid search for progressively finer meshes over the tunable hyper-parameters of the recommender. We use recommenders implemented by the RecLab library. For each dataset and recommender we evaluate hyperparameters on a 10% split of test data. The best hyper-parameters for each setting are presented in Table 6.2.

LibFM. We performed hyper-parameter tuning to find suitable learning rate and regularization parameter for each dataset. Following [DBCJ21] we consider $\texttt{lr} \in (0.001, 0.5)$ as the range of hyper-parameters for the learning rate and $\texttt{reg} \in (10^{-5}, 10^{0})$ for the regularization parameter. In all experimental settings we follow the setup of [RZK19] and use 64 latent dimensions and train with SGD for 128 iterations.

KNN. We perform hyperparameter tuning with respect to neighborhood size and shrinkage parameter. Following [DBCJ21] we consider the range $(5, 1000)$ for the neighborhood size and $(0, 1000)$ for the shrinkage parameter. We tune KNN only for the ML-1M dataset.

Experimental Infrastructure and Computational Complexity

All experiments were performed on a 64 bit desktop machine equipped with 20 CPUs (Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz) and a 62 GiB RAM. Average run times for training an instance of each recommender can be found in Table 6.2.

Experimental Setup for Computing Reachability

The parameters B_u and $\mathbf{c}_u$ are computed for each user based on the recommender model as described in the examples in Section 6.3. For the LibFM model, we consider user updates

with $\alpha = 0.1$ and $\lambda = 0$. Average run times for computing reachability of a user-item pair in various settings can be found in Table 6.3.

ML 1M. We compute max stochastic reachability for the LibFM and KNN preference model. We consider three types of user action spaces: *History Edits*, *Future Edits*, and *Next K* in which users can strategically modify the ratings associated to K randomly chosen items from their history, K randomly chosen items from that they have not yet seen, or the top-K unseen items according to the baseline scores of the preference model. For each of the action spaces we consider $K \in \{5, 10, 20\}$.

We perform reachability experiments on a random 3% subset of users (176). For each choice of preference model, action space type and action space size we sample for each user 500 random items that have not been previously rated and are not action items. For each user-item pair we compute reachability for a range of stochasticity parameters $\beta \in \{1, 2, 4, 10\}$. Note that across all experimental settings we compute reachability for the same subset of users, but different subsets of randomly selected target items.

We use the ML 1M dataset to primarily gain insights in the role that preference models, item selection stochasticity and strategic action spaces play in determining the maximum achievable degree of stochastic reachability in a recommender system.

LastFM. We run reachability experiment for LibFM recommender with *Next K = 10* action model and stochasticity parameter $\beta = 2$. We compute $\rho^\star$ values for 100 randomly sampled users and 500 randomly sampled items from the set of non-action items (target items can include previously seen items). Unlike the ML 1M dataset, the set of target items is shared among all users.

MIND. We run reachability experiments for LibFM recommender with *Next K = 10* action model and stochasticity parameter $\beta = 2$. We compute reachability for all items and users.

Reachability run times. In Table 6.3 we present the average clock time for computing reachability for a user-item pair in the settings described above. Due to internal representation of action spaces as matrices the runtime dependence on the dimension of the action space is fairly modest. We do not observe significant run time differences between different types of action spaces. We further add multiprocessing functionality to parallelize reachability computations over multiple target items.

Detailed Results: Impact of Recommender Design

We present further insights in the experimental settings studied in Section 6.5. For ML-1M, we replicate the log scale scatter plots of $\rho^\star$ against baseline ρ for all the action spaces (*Next K, Random Future, Random History*), the full range of $\beta \in \{1, 2, 4, 10\}$ and the two

Table 6.3: Reachability run times (in seconds).

Num. actions	ML 1M (LibFM)	ML 1M (KNN)	LastFM	MIND
K = 5	0.82 ± 0.04	9.8 ± 3.4	-	-
K = 10	0.87 ± 0.04	10.2 ± 6.1	4.91 ± 0.32	0.44 ± 0.01
K= 20	0.91 ± 0.05	11.4 ± 6.8	-	-

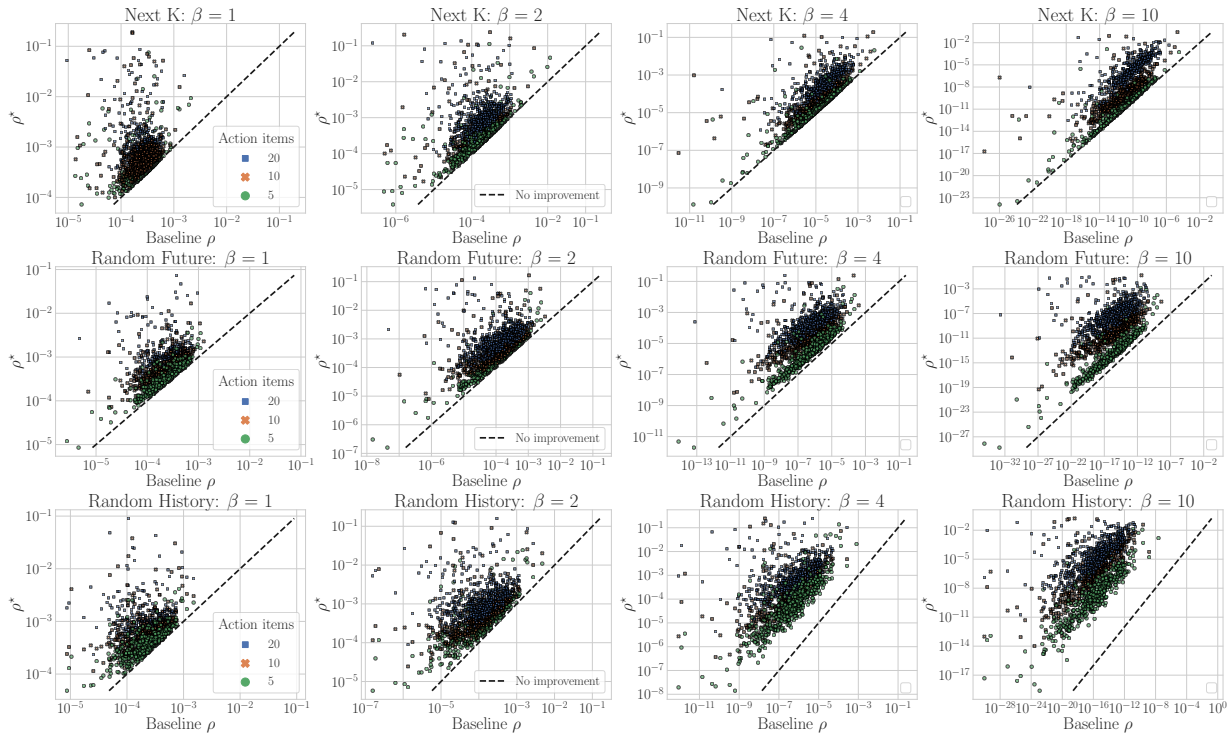

Figure 6.14: Log scale scatter plots of $\rho^\star$ against baseline ρ evaluated for the LibFM preference model.

preference models: LibFM (Figure 6.14) and KNN (Figure 6.15). We observe that for both KNN and LibFM, random history edits can lead to higher $\rho^\star$ values. We posit that this increased agency is partly due to the fact that when editing K items from the history a user edits a larger fraction of total ratings compared to editing K future items.

The most striking feature of KNN reachability results is the strong correlation between baseline ρ and $\rho^\star$. The correlations between baseline and max probability of recommendation is less strong in the case of LibFM. These insights are corroborated by Figure 6.16 which compares the average LibFM and KNN user lifts for different choices of action space, action size K, stochasticity parameter β.

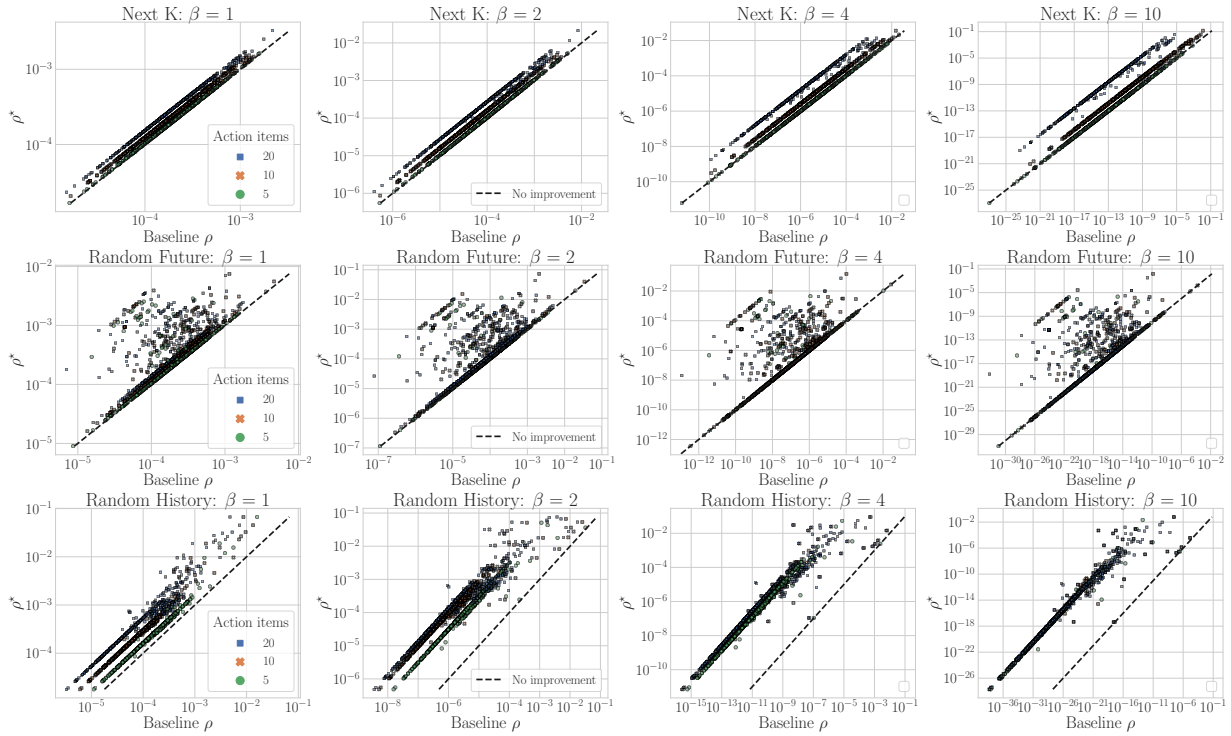

Figure 6.15: Log scale scatter plots of $\rho^\star$ against baseline ρ evaluated for the KNN preference model.

Detailed Results: Bias

We present further results on the settings studied in Section 6.5. We replicate the popularity bias results on ML-1M for different action spaces and plot the results in Figure 6.17. We see that the availability bias for KNN is dependent on the action space, with *Random History* displaying no or little correlation between popularity and max availability. This is not surprising given the results in Figure 6.7.

To systematically study the popularity bias, we compute the Spearman rank-order correlation coefficient to measure the presence of a monotonic relationship between popularity (as measured by average rating) and availability (either in the baseline or max case). We also compute the correlation between the popularity and the prevalence in the dataset, as measured by number of ratings.

The impact of user action spaces is displayed in Figure 6.18, which plots the correlation between popularity and max availability for different action spaces. For comparison, the correlation between popularity and baseline availability is just over 0.8 for all of these settings[6], while the correlation with dataset prevalence is 0.346. Table 6.4 shows these correlation values across datasets for a fixed action model. In all cases with the LibFM

[6]Due to variation in baseline actions, the baseline availability is not exactly the same.

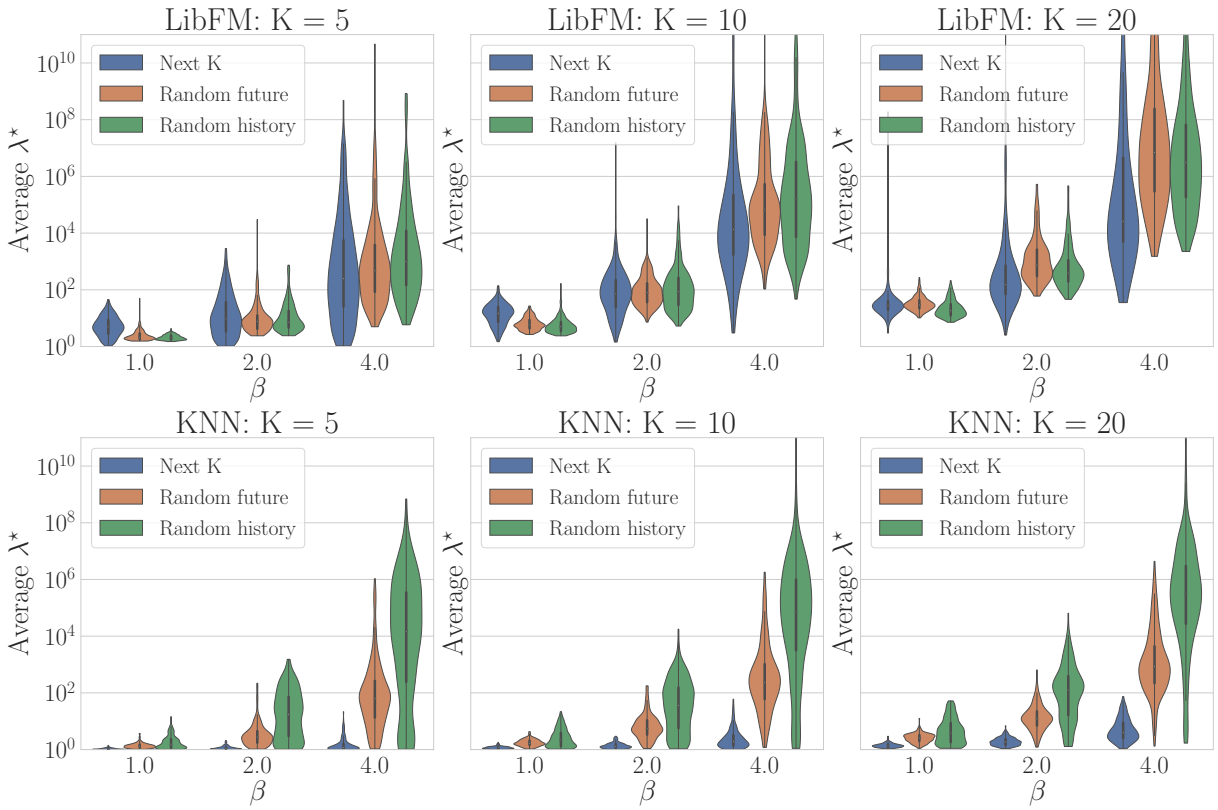

Figure 6.16: Side by side comparison of average user lifts for LibFM (top row)and KNN(bottom row).

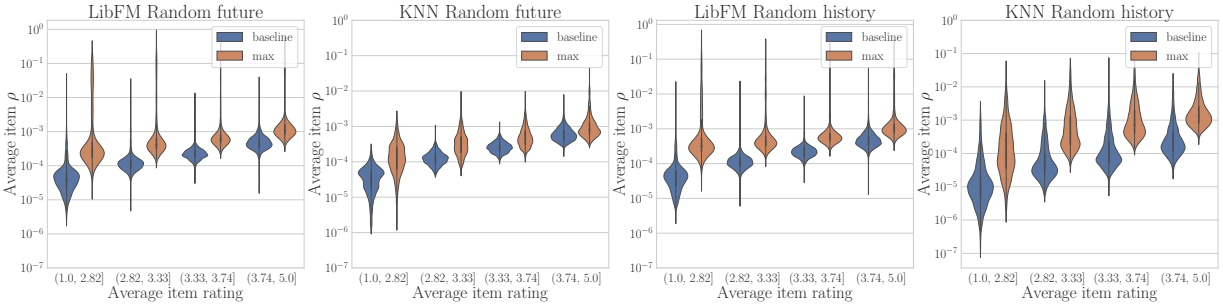

Figure 6.17: Side by side comparison of baseline and best-case availability of content, across four popularity categories. From left to right: LibFM preference model with *Random Future*, KNN preference model with *Random Future*, LibFM preference model with *Random History*, KNN preference model with *Random History*. Reachability evaluated on ML-1M for with $K = 10$ and $\beta = 2$.

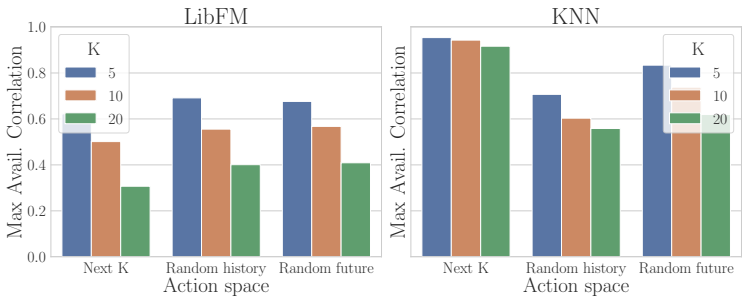

Figure 6.18: Comparison of Spearman's correlation between item popularity and max availability for different action spaces and models. Reachability evaluated on ML-1M with $\beta = 2$.

Table 6.4: Spearman's correlation with popularity for *Next K* with $K = 10$ and $\beta = 2$.

dataset	model	corr. with dataset prevalence	corr. with baseline availability	corr. with max availability
ml-1m	libfm	0.346280	0.827492	0.501316
ml-1m	knn	0.346280	0.949581	0.942986
mind	libfm	0.863992	0.825251	0.435212
lastfm	libfm	0.133318	0.671101	0.145949

model, the pattern that popularity is less correlated with max availability than baseline availability holds; however, the correlation with dataset prevalence varies.

To investigate experience bias, we similarly compute the Spearman rank-order correlation coefficient to measure the presence of a monotonic relationship between user experience (as measured by number of items rated) and discovery (either in the baseline or max case). We observe correlation values of varying sign across datasets and models, and none are particularly strong (Table 6.5).

Finally, we investigate gender bias. We compare discovery across user gender for ML-1M and availability across artist gender for LastFM (Figure 6.19). We do not observe any trends in either baseline or max values.

Table 6.5: Spearman's correlation with experience for *Next K* with $K = 10$ and $\beta = 2$.

dataset	model	corr. with baseline discovery	corr. with max discovery
ml-1m	libfm	0.475777	0.530359
ml-1m	knn	0.206556	-0.031929
mind	libfm	0.050961	0.112558
lastfm	libfm	-0.084130	-0.089226

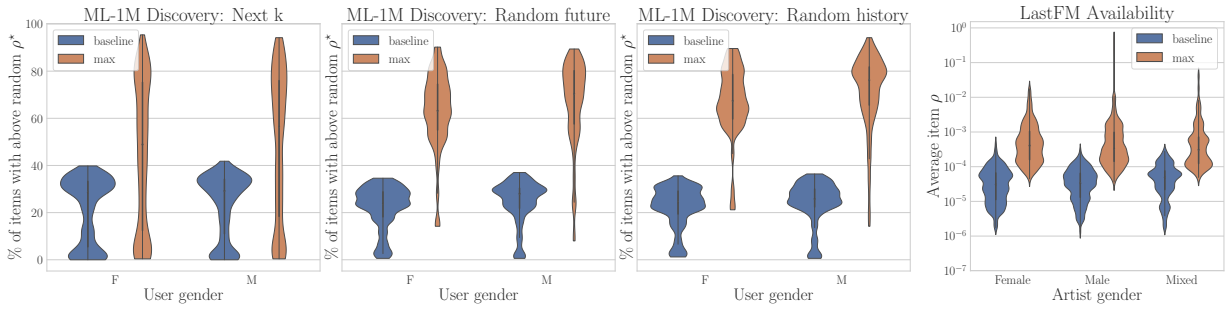

Figure 6.19: Side by side comparison of baseline and maximum discovery across user gender (left 3 panels) and availability across artist gender (rightmost panel). Reachability evaluated on ML-1M and LastFM with LibFM model, $K = 10$, different action spaces, and $\beta = 2$.

Bibliography

[ABC+20] OpenAI: Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, et al. "Learning dexterous in-hand manipulation". In: *The International Journal of Robotics Research* 39.1 (2020).

[ABDM00] Eugene Asarin, Olivier Bournez, Thao Dang, and Oded Maler. "Approximate reachability analysis of piecewise-linear dynamical systems". In: *International Workshop on Hybrid Systems: Computation and Control*. Springer. 2000.

[ABHKS19] Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. "Online control with adversarial disturbances". In: *International Conference on Machine Learning*. PMLR. 2019.

[ACST21] Amir Ali Ahmadi, Abraar Chaudhry, Vikas Sindhwani, and Stephen Tu. "Safely Learning Dynamical Systems from Short Trajectories". In: *Learning for Dynamics and Control*. PMLR. 2021.

[AGST13] Anil Aswani, Humberto Gonzalez, S Shankar Sastry, and Claire Tomlin. "Provably Safe and Robust Learning-Based Model Predictive Control". In: *Automatica* 49.5 (2013).

[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*. 2015.

[AHMS13] Oren Anava, Elad Hazan, Shie Mannor, and Ohad Shamir. "Online learning for time series prediction". In: *Conference on Learning Theory*. PMLR. 2013.

[AL17] Marc Abeille and Alessandro Lazaric. "Thompson Sampling for Linear-Quadratic Control Problems". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2017.

[AL18] Marc Abeille and Alessandro Lazaric. "Improved regret bounds for thompson sampling in linear quadratic control problems". In: *International Conference on Machine Learning*. PMLR. 2018.

[AL20] Marc Abeille and Alessandro Lazaric. "Efficient optimistic exploration in linear-quadratic regulators via lagrangian relaxation". In: *International Conference on Machine Learning*. PMLR. 2020.

[ALS19] Yasin Abbasi-Yadkori, Nevena Lazic, and Csaba Szepesvári. "Model-Free Linear Quadratic Control via Reduction to Expert Prediction". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2019.

[AM17] James Anderson and Nikolai Matni. "Structured state space realizations for SLS distributed controllers". In: *Allerton Conference on Communication, Control, and Computing*. IEEE. 2017.

[AMBM19] Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. "The impact of popularity bias on fairness and calibration in recommendation". In: *arXiv preprint arXiv:1910.05755* (2019).

[APS11] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. "Online Least Squares Estimation with Self-Normalized Processes: An Application to Bandit Problems". In: *arXiv preprint arXiv:1102.2670* (2011).

[ApS15] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.1 (Revision 25).* 2015. URL: `http://docs.mosek.com/8.1/toolbox/index.html`.

[ApS19] MOSEK ApS. *MOSEK Optimizer API for Python Release 9.0.88*. 2019. URL: `https://docs.mosek.com/9.0/pythonapi.pdf`.

[Arr63] Kenneth Joseph Arrow. *Social Choice and Individual Values*. 12. Yale University Press, 1963.

[AS11] Yasin Abbasi-Yadkori and Csaba Szepesvári. "Regret Bounds for the Adaptive Control of Linear Quadratic Systems". In: *Conference on Learning Theory*. PMLR. 2011.

[BBM17] Francesco Borrelli, Alberto Bemporad, and Manfred Morari. *Predictive control for linear and hybrid systems*. New York, NY, USA: Cambridge University Press, 2017.

[BCHT17] Somil Bansal, Mo Chen, Sylvia Herbert, and Claire J Tomlin. "Hamilton-Jacobi reachability: A brief overview and recent advances". In: *Conference on Decision and Control*. IEEE. 2017.

[BDD+16] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. "End to end learning for self-driving cars". In: *arXiv preprint arXiv:1604.07316* (2016).

[Ben62] George Bennett. "Probability inequalities for the sum of independent random variables". In: *Journal of the American Statistical Association* 57.297 (1962).

[Ber95] Dimitri P Bertsekas. *Dynamic programming and optimal control*. Vol. 1. 3. Athena scientific Belmont, MA, 1995. Chap. 5.

[BF81] Peter J Bickel and David A Freedman. "Some asymptotic theory for the bootstrap". In: *The annals of statistics* 9.6 (1981).

[BJRL15] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[Bla99] Franco Blanchini. "Set Invariance in Control". In: *Automatica* 35.11 (1999).

[BLR17] Sampoorna Biswas, Laks VS Lakshmanan, and Senjuti Basu Ray. "Combating the cold start user problem in model based collaborative filtering". In: *arXiv preprint arXiv:1703.00397* (2017).

[BM99] Alberto Bemporad and Manfred Morari. "Robust Model Predictive Control: A Survey". In: *Robustness in identification and control*. Springer, 1999, pp. 207–226.

[BMDP02] Alberto Bemporad, Manfred Morari, Vivek Dua, and Efstratios N Pistikopoulos. "The explicit linear quadratic regulator for constrained systems". In: *Automatica* 38.1 (2002).

[BMR18] Ross Boczar, Nikolai Matni, and Benjamin Recht. "Finite-data performance guarantees for the output-feedback control of an unknown system". In: *Conference on Decision and Control*. IEEE. 2018.

[Boa17] National Transportation Safety Board. *Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016*. Tech. rep. NTSB/HAR-17/02. Washington, DC, 2017.

[Boa20] National Transportation Safety Board. *Collision Between a Sport Utility Vehicle Operating With Partial Driving Automation and a Crash Attenuator Mountain View, California, March 23, 2018*. Tech. rep. NTSB/HAR-20/01. Washington, DC, 2020.

[BRA19] Krisztian Balog, Filip Radlinski, and Shushan Arakelyan. "Transparent, scrutable and explainable user models for personalized recommendation". In: *Conference on Research and Development in Information Retrieval*. ACM SIGIR. 2019.

[Bri17] James Bridle. "Something is wrong on the internet". In: *Medium* (Nov. 2017). URL: `https://medium.com/@jamesbridle/something-is-wrong-on-the-internet-c39c471271d2`.

[BRSB21] Monimoy Bujarbaruah, Ugo Rosolia, Yvonne R Stürz, and Francesco Borrelli. "A Simple Robust MPC for Linear Systems with Parametric and Additive Uncertainty". In: *arXiv preprint arXiv:2103.12351* (2021).

[BS15] Felix Berkenkamp and Angela P. Schoellig. "Safe and Robust Learning Control with Gaussian Processes". In: *European Control Conference (ECC)*. 2015.

[BS16] Solon Barocas and Andrew D. Selbst. "Big Data's Disparate Impact". In: *California Law Review* 104 (2016).

[BTGMT20] Somil Bansal, Varun Tolani, Saurabh Gupta, Jitendra Malik, and Claire Tomlin. "Combining optimal control and learning for visual navigation in novel environments". In: *Conference on Robot Learning*. PMLR. 2020.

[BTSK17] Felix Berkenkamp, Matteo Turchetta, Angela Schoellig, and Andreas Krause. "Safe Model-based Reinforcement Learning with Stability Guarantees". In: *Neural Information Processing Systems*. 2017.

[Bur38] Abram Burk. "A reformulation of certain aspects of welfare economics". In: *The Quarterly Journal of Economics* 52.2 (1938). Publisher: MIT Press.

[BYDM94] Richard P Braatz, Peter M Young, John C Doyle, and Manfred Morari. "Computational complexity of μ calculation". In: *IEEE Transactions on Automatic Control* 39.5 (1994).

[BZTB20] Monimoy Bujarbaruah, Xiaojing Zhang, Marko Tanaskovic, and Francesco Borrelli. "Adaptive Stochastic MPC Under Time-Varying Uncertainty". In: *IEEE Transactions on Automatic Control* 66.6 (2020).

[CA19] Yuxiao Chen and James Anderson. "System level synthesis with state and input constraints". In: *Conference on Decision and Control*. IEEE. 2019.

[CBC+19] Minmin Chen, Alex Beutel, Paul Covington, Sagar Jain, Francois Belletti, and Ed H Chi. "Top-k off-policy correction for a REINFORCE recommender system". In: *Conference on Web Search and Data Mining*. 2019.

[CDR21] Mihaela Curmei, Sarah Dean, and Benjamin Recht. "Quantifying Availability and Discovery in Recommender Systems via Stochastic Reachability". In: *International Conference on Machine Learning*. PMLR. 2021. eprint: `arXiv:2107.00833`.

[CDW+20] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. "Bias and Debias in Recommender System: A Survey and Future Directions". In: *arXiv preprint arXiv:2010.03240* (2020).

[Cel10] Oscar Celma. "Music recommendation". In: *Music recommendation and discovery*. Springer, 2010, pp. 43–85.

[CG00] Jie Chen and Guoxiang Gu. *Control-Oriented System Identification: An $\mathcal{H}_\infty$ Approach*. Wiley, 2000.

[CG18] Sam Corbett-Davies and Sharad Goel. "The measure and mismeasure of fairness: A critical review of fair machine learning". In: *arXiv preprint arXiv:1808.00023* (2018).

[CGLN17] Nicolò Cesa-Bianchi, Claudio Gentile, Gabor Lugosi, and Gergely Neu. "Boltzmann Exploration Done Right". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[Che11] Michael R Chernick. *Bootstrap methods: A guide for practitioners and researchers*. Vol. 619. John Wiley & Sons, 2011.

[Cho17] Alexandra Chouldechova. "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments". In: *Big data* 5.2 (2017).

[CKM19] Alon Cohen, Tomer Koren, and Yishay Mansour. "Learning Linear-Quadratic Regulators Efficiently with only $\sqrt{T}$ Regret". In: *International Conference on Machine Learning*. PMLR. 2019.

[CKP09] Toon Calders, Faisal Kamiran, and Mykola Pechenizkiy. "Building Classifiers with Independency Constraints". In: *International Conference on Data Mining Workshops*. IEEE. 2009.

[CMLKD18] Felipe Codevilla, Matthias Miiller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. "End-to-end driving via conditional imitation learning". In: *International Conference on Robotics and Automation*. IEEE. 2018.

[CN12] Adam Coates and Andrew Y Ng. "Learning feature representations with k-means". In: *Neural networks: Tricks of the trade*. Springer, 2012, pp. 561–580.

[CN93] Jie Chen and Carl N Nett. "The Caratheodory-Fejer problem and H_∞ identification: a time domain approach". In: *Conference on Decision and Control*. IEEE. 1993.

[CNDG18] Yinlam Chow, Ofir Nachum, Edgar Duenez-Guzman, and Mohammad Ghavamzadeh. "A Lyapunov-based Approach to Safe Reinforcement Learning". In: *arXiv preprint arXiv:1805.07708* (2018).

[CO96] Andrew E Clark and Andrew J Oswald. "Satisfaction and comparison income". In: *Journal of public economics* 61.3 (1996).

[COMB19] Richard Cheng, Gábor Orosz, Richard M. Murray, and Joel W. Burdick. "End-to-End Safe Reinforcement Learning through Barrier Functions for Safety-Critical Continuous Control Tasks". In: *Conference on Artificial Intelligence.* AAAI. 2019.

[CPNB15] Fabian Christoffel, Bibek Paudel, Chris Newell, and Abraham Bernstein. "Blockbusters and wallflowers: Accurate, diverse, and scalable recommendations with random walks". In: *Conference on Recommender Systems*. ACM. 2015.

[CSE18] Allison JB Chaney, Brandon M Stewart, and Barbara E Engelhardt. "How algorithmic confounding in recommendation systems increases homogeneity and decreases utility". In: *Conference on Recommender Systems*. ACM. 2018.

[CW02] Marco C Campi and Erik Weyer. "Finite sample properties of system identification methods". In: *IEEE Transactions on Automatic Control* 47.8 (2002).

[CWMPM20] Shaoru Chen, Han Wang, Manfred Morari, Victor M Preciado, and Nikolai Matni. "Robust closed-loop model predictive control via system level synthesis". In: *Conference on Decision and Control*. IEEE. 2020.

[DBCJ21] Maurizio Ferrari Dacrema, Simone Boglio, Paolo Cremonesi, and Dietmar Jannach. "A troubling analysis of reproducibility and progress in recommender systems research". In: *Transactions on Information Systems* 39.2 (2021).

[DCHSA16] Yan Duan, Xi Chen, Rein Houthooft, John Schulman, and Pieter Abbeel. "Benchmarking deep reinforcement learning for continuous control". In: *International Conference on Machine Learning*. PMLR. 2016.

[DD94] Munther A Dahleh and Ignacio J Diaz-Bobillo. *Control of uncertain systems: a linear programming approach*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1994.

[DDV+18] Gal Dalal, Krishnamurthy Dvijotham, Matej Vecerik, Todd Hester, Cosmin Paduraru, and Yuval Tassa. "Safe Exploration in Continuous Action Spaces". In: *arXiv preprint arXiv:1801.08757* (2018).

[Dea16] Angus Deaton. "Measuring and understanding behavior, welfare, and poverty". In: *American Economic Review* 106.6 (2016).

[Dea80] Angus Deaton. *The measurement of welfare: Theory and practical guidelines*. English. World Bank, Development Research Center, 1980.

[Dev78] Luc P Devroye. "The uniform convergence of the nadaraya-watson regression function estimate". In: *Canadian Journal of Statistics* 6.2 (1978).

[DGL13] Pranav Dandekar, Ashish Goel, and David Lee. "Biased assimilation, homophily, and the dynamics of polarization". In: *Proceedings of the National Academy of Sciences*. 2013.

[DK11] Christian Desrosiers and George Karypis. "A comprehensive survey of neighborhood-based recommendation methods". In: *Recommender systems handbook* (2011).

[DMMRT18] Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. "Regret bounds for robust adaptive control of the linear quadratic regulator". In: *Advances in Neural Information Processing Systems*. 2018. eprint: `arXiv:1805.09388`.

[DMMRT20] Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. "On the sample complexity of the linear quadratic regulator". In: *Foundations of Computational Mathematics* 20.4 (2020). eprint: `arXiv:1710.01688`.

[DMRY20] Sarah Dean, Nikolai Matni, Benjamin Recht, and Vickie Ye. "Robust guarantees for perception-based control". In: *Learning for Dynamics and Control*. PMLR. 2020. eprint: `arXiv:1907.03680`.

[Doy82] John Doyle. "Analysis of feedback systems with structured uncertainties". In: *IEEE Proceedings D - Control Theory and Applications* 129.6 (1982).

[DP87] M. Dahleh and J. Pearson. "ℓ^1-optimal feedback controllers for MIMO discrete-time systems". In: *IEEE Transactions on Automatic Control* 32.4 (Apr. 1987).

[DR21] Sarah Dean and Benjamin Recht. "Certainty equivalent perception-based control". In: *Learning for Dynamics and Control*. PMLR. 2021. eprint: `arXiv:2008.12332`.

[DRCLK] Alexey Dosovitskiy, Germán Ros, Felipe Codevilla, Antonio M. López, and Vladlen Koltun. "CARLA: An Open Urban Driving Simulator". In: *Conference on Robot Learning*. PMLR. eprint: `arXiv:1711.03938`.

[DRR20] Sarah Dean, Sarah Rich, and Benjamin Recht. "Recommendations and user agency: the reachability of collaboratively-filtered information". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2020. eprint: `arXiv:1912.10068`.

[DSA+20] Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, D Sculley, and Yoni Halpern. "Fairness is not static: deeper understanding of long term fairness via simulation studies". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2020.

[DTMR19] Sarah Dean, Stephen Tu, Nikolai Matni, and Benjamin Recht. "Safely learning to control the constrained linear quadratic regulator". In: *American Control Conference*. IEEE. 2019. eprint: `arXiv:1809.10121`.

[Dum07] Bogdan Dumitrescu. *Positive trigonometric polynomials and signal processing applications*. Vol. 103. Springer Science & Business Media, 2007.

[DV75] Charles A Desoer and Mathukumalli Vidyasagar. *Feedback systems: input-output properties*. Vol. 55. Siam, 1975.

[EFNSV18] Danielle Ensign, Sorelle A Friedler, Scott Nevlle, Carlos Scheidegger, and Suresh Venkatasubramanian. "Runaway Feedback Loops in Predictive Policing". In: *Conference on Fairness, Accountability and Transparency*. Vol. 81. 2018.

[Efr92] Bradley Efron. "Bootstrap methods: another look at the jackknife". In: *Breakthroughs in statistics*. Springer, 1992, pp. 569–593.

[EJJ+19] Hadi Elzayn, Shahin Jabbari, Christopher Jung, Michael Kearns, Seth Neel, Aaron Roth, and Zachary Schutzman. "Fair algorithms for learning in allocation problems". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2019.

[ETA+18] Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. "All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness". In: *Conference on Fairness, Accountability and Transparency*. ACM. 2018.

[ETKMK18] Michael D Ekstrand, Mucun Tian, Mohammed R Imran Kazi, Hoda Mehrpouyan, and Daniel Kluver. "Exploring author gender in book rating and recommendation". In: *Conference on Recommender Systems*. ACM. 2018.

[Exe16] Executive Office of the President. *Big data: A report on algorithmic systems, opportunity, and civil rights*. Tech. rep. White House, May 2016.

[FCF20] Marc Faddoul, Guillaume Chaslot, and Hany Farid. "A Longitudinal Analysis of YouTube's Promotion of Conspiracy Videos". In: *arXiv preprint arXiv:2003.03318* (2020).

[Fer97] E. Feron. "Analysis of Robust $\mathcal{H}_2$ Performance Using Multiplier Theory". In: *SIAM Journal on Control and Optimization* 35.1 (1997).

[FGKM18] Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. "Global Convergence of Policy Gradient Methods for the Linear Quadratic Regulator". In: *International Conference on Machine Learning*. Vol. 80. PMLR. 2018.

[FGR16] Seth Flaxman, Sharad Goel, and Justin M Rao. "Filter bubbles, echo chambers, and online news consumption". In: *Public opinion quarterly* 80.S1 (2016).

[FGRW17] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. "Predictably Unequal? The Effects of Machine Learning on Credit Markets". In: *SSRN* (2017).

[FGS16] Marcello Farina, Luca Giulioni, and Riccardo Scattolini. "Stochastic Linear Model Predictive Control with Chance Constraints–a Review". In: *Journal of Process Control* 44 (2016).

[Fie97] Claude-Nicolas Fiechter. "PAC Adaptive Control of Linear Systems". In: *Conference on Computational Learning Theory*. ACM. 1997.

[Flo14] Massimo Florio. *Applied welfare economics: Cost-benefit analysis of projects and policies*. Routledge, 2014.

[FTD91] M. K. H. Fan, A. L. Tits, and J. C. Doyle. "Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics". In: *IEEE Transactions on Automatic Control* 36.1 (1991).

[FV92] Dean P Foster and Rakesh V Vohra. "An economic argument for affirmative action". In: *Rationality and Society* 4.2 (1992).

[GGC+15] Alessandro Giusti, Jérôme Guzzi, Dan C Cireşan, Fang-Lin He, Juan P Rodríguez, Flavio Fontana, Matthias Faessler, Christian Forster, Jürgen Schmidhuber, Gianni Di Caro, et al. "A machine learning approach to visual perception of forest trails for mobile robots". In: *IEEE Robotics and Automation Letters* 1.2 (2015).

[GKM06] Paul J Goulart, Eric C Kerrigan, and Jan M Maciejowski. "Optimization over State Feedback Policies for Robust Control with Constraints". In: *Automatica* 42.4 (2006).

[Gol98] A. Goldenshluger. "Nonparametric estimation of transfer functions: rates of convergence and adaptation". In: *IEEE Transactions on Information Theory* 44.2 (Mar. 1998).

[Gre97] Włodzimierz Greblicki. "Nonparametric approach to Wiener system identification". In: *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 44.6 (1997).

[GZ01] Alexander Goldenshluger and Assaf Zeevi. "Nonasymptotic bounds for autoregressive time series modeling". In: *The Annals of Statistics* 29.2 (Apr. 2001).

[Han08] Bruce E Hansen. "Uniform convergence rates for kernel estimation with dependent data". In: *Econometric Theory* 24.3 (2008).

[Has87] Z Hasiewicz. "Identification of a linear system observed through zero-memory non-linearity". In: *International Journal of Systems Science* 18.9 (1987).

[HC18a] Lily Hu and Yiling Chen. "A Short-term Intervention for Long-term Fairness in the Labor Market". In: *The World Wide Web Conference*. 2018.

[HC18b] Lily Hu and Yiling Chen. "Welfare and Distributional Impacts of Fair Classification". In: *Workshop on Fairness, Accountability, and Transparency in Machine Learning*. 2018.

[HC20] Lily Hu and Yiling Chen. "Fair Classification and Social Welfare". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2020.

[HD19] Dimitar Ho and John C Doyle. "Robust model-free learning and control without prior knowledge". In: *2019 IEEE 58th Conference on Decision and Control*. IEEE. 2019.

[HJN91] A. J. Helmicki, C. A. Jacobson, and C. N. Nett. "Control oriented system identification: a worst-case/deterministic approach in $\mathcal{H}_\infty$". In: *IEEE Transactions on Automatic Control* 36.10 (Oct. 1991).

[HK01] Babak Hassibi and Thomas Kaliath. "H-∞ bounds for least-squares estimators". In: *IEEE Transactions on Automatic Control* 46.2 (2001).

[HK15] F Maxwell Harper and Joseph A Konstan. "The MovieLens datasets: History and context". In: *ACM transactions on interactive intelligent systems* 5.4 (2015).

[HKBR14] Joel A Hesch, Dimitrios G Kottas, Sean L Bowman, and Stergios I Roumeliotis. "Camera-IMU-based localization: Observability analysis and consistency improvement". In: *The International Journal of Robotics Research* 33.1 (2014).

[HKTR04] Jonathan Herlocker, Joseph Konstan, Loren Terveen, and John Riedl. "Evaluating collaborative filtering recommender systems". In: *Ransactions on Information Systems* 22.1 (2004).

[HLSZZ18] Elad Hazan, Holden Lee, Karan Singh, Cyril Zhang, and Yi Zhang. "Spectral filtering for general linear dynamical systems". In: *Advances in Neural Information Processing Systems*. 2018.

[HMPW16] Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. "Strategic classification". In: *Conference on Innovations in Theoretical Computer Science*. 2016.

[HMR18] Moritz Hardt, Tengyu Ma, and Benjamin Recht. "Gradient descent learns linear dynamical systems". In: *The Journal of Machine Learning Research* 19.1 (2018).

[Ho20] Dimitar Ho. "A system level approach to discrete-time nonlinear systems". In: *American Control Conference*. IEEE. 2020.

[HP90] Diederich Hinrichsen and Anthony J Pritchard. "Real and complex stability radii: a survey". In: *Control of uncertain systems*. Springer, 1990, pp. 119–162.

[HPS16] Moritz Hardt, Eric Price, and Nati Srebro. "Equality of Opportunity in Supervised Learning". In: *Advances in Neural Information Processing Systems*. 2016.

[HSZ17] Elad Hazan, Karan Singh, and Cyril Zhang. "Learning linear dynamical systems via spectral filtering". In: *Advances in Neural Information Processing Systems*. 2017.

[HXK+15] F Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. "Putting users in control of their recommendations". In: *Conference on Recommender Systems*. ACM. 2015.

[IJR12] Morteza Ibrahimi, Adel Javanmard, and Benjamin V. Roy. "Efficient Reinforcement Learning for High Dimensional Linear Quadratic Systems". In: *Advances in Neural Information Processing Systems*. 2012.

[IJW+19] Eugene Ie, Vihan Jain, Jing Wang, Sanmit Narvekar, Ritesh Agarwal, Rui Wu, Heng-Tze Cheng, Tushar Chandra, and Craig Boutilier. "SlateQ: A Tractable Decomposition for Reinforcement Learning with Recommendation Sets". In: *International Joint Conference on Artificial Intelligence*. 2019.

[IS02] R. Ibragimov and S. Sharakhmetov. "The Exact Constant in the Rosenthal Inequality for Random Variables with Mean Zero". In: *Theory of Probability & Its Applications* 46.1 (2002).

[JKMR16] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. "Fairness in Learning: Classic and Contextual Bandits". In: *Advances in Neural Information Processing Systems*. 2016.

[JLKJ15] Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, and Michael Jugovac. "What recommenders recommend: an analysis of recommendation biases and possible countermeasures". In: *User Modeling and User-Adapted Interaction* 25.5 (2015).

[JS11] Eagle S Jones and Stefano Soatto. "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach". In: *The International Journal of Robotics Research* 30.4 (2011).

[KA21] Maximilian Kasy and Rediet Abebe. "Fairness, equality, and power in algorithmic decision-making". In: *Conference on Fairness, Accountability, and Transparency*. 2021.

[KB15] Yehuda Koren and Robert Bell. "Advances in collaborative filtering". In: *Recommender systems handbook* (2015).

[KBKTC15] Jaya Kawale, Hung H Bui, Branislav Kveton, Long Tran-Thanh, and Sanjay Chawla. "Efficient thompson sampling for online matrix-factorization recommendation". In: *Advances in Neural Information Processing Systems*. 2015.

[KBM96] Mayuresh V Kothare, Venkataramanan Balakrishnan, and Manfred Morari. "Robust Constrained Model Predictive Control Using Linear Matrix Inequalities". In: *Automatica* 32.10 (1996).

[KBSV20] Amir-Hossein Karimi, Gilles Barthe, Bernhard Schölkopf, and Isabel Valera. "A survey of algorithmic recourse: definitions, formulations, solutions, and prospects". In: *arXiv preprint arXiv:2010.04050* (2020).

[KBSW85] Stephen N. Keith, Robert M. Bell, August G. Swanson, and Albert P. Williams. "Effects of Affirmative Action in Medical Schools". In: *New England Journal of Medicine* 313.24 (1985).

[KBTK18] Torsten Koller, Felix Berkenkamp, Matteo Turchetta, and Andreas Krause. "Learning-based model predictive control for safe exploration". In: *Conference on Decision and Control*. IEEE. 2018.

[KDK06] Alexandra Kalev, Frank Dobbin, and Erin Kelly. "Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies". In: *American Sociological Review* 71.4 (2006).

[KDZ+20] Karl Krauth, Sarah Dean, Alex Zhao, Wenshuo Guo, Mihaela Curmei, Benjamin Recht, and Michael I Jordan. "Do Offline Metrics Predict Online Performance in Recommender Systems?" In: *arXiv preprint arXiv:2011.07931* (2020).

[KLRS17] Matt J. Kusner, Joshua R. Loftus, Chris Russell, and Ricardo Silva. "Counterfactual Fairness". In: *Advances in Neural Information Processing Systems*. 2017.

[KM16] Vitaly Kuznetsov and Mehryar Mohri. "Time series prediction and online learning". In: *Conference on Learning Theory*. PMLR. 2016.

[KM17] Vitaly Kuznetsov and Mehryar Mohri. "Generalization bounds for non-stationary mixing processes". In: *Machine Learning* 106.1 (Jan. 2017).

[KMR17] Jon M. Kleinberg, Sendhil Mullainathan, and Manish Raghavan. "Inherent Trade-Offs in the Fair Determination of Risk Scores". In: *Conference on Innovations in Theoretical Computer Science* (2017).

[KNRW18] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. "Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness". In: *International Conference on Machine Learning*. PMLR. 2018.

[Kor08] Yehuda Koren. "Factorization meets the neighborhood: a multifaceted collaborative filtering model". In: *Conference on Knowledge Discovery and Data mining*. 2008.

[KRP+17] Niki Kilbertus, Mateo Rojas-Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. "Avoiding Discrimination through Causal Reasoning". In: *Advances in Neural Information Processing Systems*. 2017.

[KS11] Jonathan Kelly and Gaurav S Sukhatme. "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration". In: *The International Journal of Robotics Research* 30.1 (2011).

[KTR19] Karl Krauth, Stephen Tu, and Benjamin Recht. "Finite-time Analysis of Approximate Policy Iteration for the Linear Quadratic Regulator". In: *Advances in Neural Information Processing Systems* 32 (2019).

[LAC17] Matthias Lorenzen, Frank Allgöwer, and Mark Cannon. "Adaptive Model Predictive Control with Robust Constraint Satisfaction". In: *IFAC Proceedings Volumes* 50.1 (2017).

[LAHA20] Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. "Logarithmic Regret Bound in Partially Observable Linear Dynamical Systems". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020. eprint: `arXiv:2003.11227`.

[LAWCS13] Simon Lynen, Markus W Achtelik, Stephan Weiss, Margarita Chli, and Roland Siegwart. "A robust and modular multi-sensor fusion approach applied to mav navigation". In: *2013 IEEE/RSJ international conference on intelligent robots and systems*. IEEE. 2013.

[LB01] Seth L Lacy and Dennis S Bernstein. "Subspace identification for nonlinear systems that are linear in unmeasured states". In: *Conference on Decision and Control*. Vol. 4. IEEE. 2001.

[LBMK16] Giuseppe Loianno, Chris Brunner, Gary McGrath, and Vijay Kumar. "Estimation, control, and planning for aggressive flight with a small quadrotor with a single camera and IMU". In: *IEEE Robotics and Automation Letters* 2.2 (2016).

[LCT20] Forrest Laine, Chiu-Yuan Chiu, and Claire Tomlin. "Eyes-Closed Safety Kernels: Safety for Autonomous Systems Under Loss of Observability". In: *Robotics: Science and Systems* (2020).

[LDRSH18] Lydia T Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. "Delayed impact of fair machine learning". In: *International Conference on Machine Learning*. PMLR. 2018. eprint: `arXiv:1803.04383`.

[LFDA16] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. "End-to-end Training of Deep Visuomotor Policies". In: *Journal of Machine Learning Research* 17.1 (Jan. 2016).

[LGQ+18] Yi Lin, Fei Gao, Tong Qin, Wenliang Gao, Tianbo Liu, William Wu, Zhenfei Yang, and Shaojie Shen. "Autonomous aerial navigation using monocular visual-inertial fusion". In: *Journal of Field Robotics* 35.1 (2018).

[Lie89] Hannelore Liero. "Strong uniform consistency of nonparametric regression function estimates". In: *Probability theory and related fields* 82.4 (1989).

[Lju99] Lennart Ljung. *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1999.

[LLZ+21] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J Vera Liao, James Choi, Kaiyue Fan, Sean A Munson, and Alexis Hiniker. "How the Design of YouTube Influences User Sense of Agency". In: *Conference on Human Factors in Computing Systems*. 2021.

[Löf04] J. Löfberg. "YALMIP : A toolbox for modeling and optimization in MATLAB". In: *IEEE International Symposium on Computer Aided Control System Design*. 2004.

[LRH11] Christian A Larsson, Cristian R Rojas, and Håkan Hjalmarsson. "MPC Oriented Experiment Design". In: *IFAC Proceedings Volumes* 44.1 (2011).

[LSB+15] Simon Lynen, Torsten Sattler, Michael Bosse, Joel A Hesch, Marc Pollefeys, and Roland Siegwart. "Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization." In: *Robotics: Science and Systems*. 2015.

[LSCAY20] Anqi Liu, Guanya Shi, Soon-Jo Chung, Anima Anandkumar, and Yisong Yue. "Robust regression for safe exploration in control". In: *Learning for Dynamics and Control*. PMLR. 2020.

[LSRLB18] Alexander Lambert, Amirreza Shaban, Amit Raj, Zhen Liu, and Byron Boots. "Deep forward and inverse perceptual models for tracking and prediction". In: *International Conference on Robotics and Automation*. IEEE. 2018.

[LWH+20] Lydia T. Liu, Ashia Wilson, Nika Haghtalab, Adam Tauman Kalai, Christian Borgs, and Jennifer Chayes. "The Disparate Equilibria of Algorithmic Decision Making when Individuals Invest Rationally". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2020.

[LZBRS17] Tyler Lu, Martin Zinkevich, Craig Boutilier, Binz Roy, and Dale Schuurmans. "Safe Exploration for Identifying Linear Systems via Robust Optimization". In: *arXiv preprint arXiv:1711.11165* (2017).

[MFS+20] Zakaria Mhammedi, Dylan J Foster, Max Simchowitz, Dipendra Misra, Wen Sun, Akshay Krishnamurthy, Alexander Rakhlin, and John Langford. "Learning the Linear Quadratic Regulator from Nonlinear Observations". In: *Advances in Neural Information Processing Systems* 33 (2020).

[MGP15] Jérémie Mary, Romaric Gaudel, and Philippe Preux. "Bandits and recommender systems". In: *International Workshop on Machine Learning, Optimization and Big Data*. Springer. 2015.

[MHKL20] Dipendra Misra, Mikael Henaff, Akshay Krishnamurthy, and John Langford. "Kinematic state abstraction and provably efficient rich-observation reinforcement learning". In: *International conference on machine learning*. PMLR. 2020.

[MJR20] Horia Mania, Michael I. Jordan, and Benjamin Recht. "Active Learning for Nonlinear System Identification with Guarantees". In: *arXiv preprint arXiv:2006.10277* (2020).

[MKS+15] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *Nature* 518 (Feb. 2015).

[MMDH19] Smitha Milli, John Miller, Anca D Dragan, and Moritz Hardt. "The social cost of strategic classification". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2019.

[MMT15] Raúl Mur-Artal, J. M. M. Montiel, and Juan D. Tardós. "ORB-SLAM: a Versatile and Accurate Monocular SLAM System". In: *IEEE Transactions on Robotics* 31.5 (2015).

[MOS19] Hussein Mouzannar, Mesrob I Ohannessian, and Nathan Srebro. "From fair decision making to social equality". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2019.

[MR10] Mehryar Mohri and Afshin Rostamizadeh. "Stability Bounds for Stationary ϕ-mixing and β-mixing Processes". In: *Journal of Machine Learning Research* 11 (Mar. 2010).

[MR97] A. Megretski and A. Rantzer. "System analysis via integral quadratic constraints". In: *IEEE Transactions on Automatic Control* 42.6 (June 1997).

[MRFA06] David Q Mayne, SV Raković, Rolf Findeisen, and Frank Allgöwer. "Robust output feedback model predictive control of constrained linear systems". In: *Automatica* 42.7 (2006).

[MSR05] David Q Mayne, Maria M Seron, and SV Rakovic. "Robust Model Predictive Control of Constrained Linear Systems with Bounded Disturbances". In: *Automatica* 41.2 (2005).

[MSS17] Daniel J. McDonald, Cosma Rohilla Shalizi, and Mark Schervish. "Nonparametric Risk Bounds for Time-series Forecasting". In: *Journal of Machine Learning Research* 18.1 (Jan. 2017).

[MTR19] Horia Mania, Stephen Tu, and Benjamin Recht. "Certainty Equivalence is Efficient for Linear Quadratic Control". In: *Advances in Neural Information Processing Systems*. 2019.

[MWA17] N. Matni, Y. Wang, and J. Anderson. "Scalable system level synthesis for virtually localizable systems". In: *Conference on Decision and Control*. IEEE. 2017.

[MZSJ21] Hesameddin Mohammadi, Armin Zare, Mahdi Soltanolkotabi, and Mihailo R Jovanovic. "Convergence and sample complexity of gradient methods for the model-free linear quadratic regulator problem". In: *IEEE Transactions on Automatic Control* (2021).

[Nat17] National Transportation Safety Board. *Preliminary Report Highway HWY18MH010*. `https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf`. Online; accessed 14 July 2021. 2017.

[NHHTK14] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. "Exploring the filter bubble: the effect of using recommender systems on content diversity". In: *The World Wide Web Conference*. 2014.

[Nic18] Jack Nicas. "How YouTube Drives People to the Internet's Darkest Corners". In: *The Wall Street Journal* (Feb. 2018). URL: `https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478`.

[NK11] Xia Ning and George Karypis. "Slim: Sparse linear methods for top-n recommender systems". In: *2011 IEEE 11th International Conference on Data Mining*. IEEE. 2011.

[NLS+20] Yashwanth Kumar Nakka, Anqi Liu, Guanya Shi, Anima Anandkumar, Yisong Yue, and Soon-Jo Chung. "Chance-constrained trajectory optimization for safe exploration and learning of nonlinear systems". In: *IEEE Robotics and Automation Letters* 6.2 (2020).

[NS18] Razieh Nabi and Ilya Shpitser. "Fair inference on outcomes". In: *Conference on Artificial Intelligence*. Vol. 32. 1. AAAI. 2018.

[OF06] Stanley Osher and Ronald Fedkiw. *Level set methods and dynamic implicit surfaces*. Vol. 153. Springer Science & Business Media, 2006.

[OGJ17] Y. Ouyang, M. Gagrani, and R. Jain. "Control of unknown linear systems with Thompson sampling". In: *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. 2017.

[OJM08] Frauke Oldewurtel, Colin N Jones, and Manfred Morari. "A Tractable Approximation of Chance Constrained Stochastic MPC Based on Affine Disturbance Feedback". In: *Conference on Decision and Control*. IEEE. 2008.

[OO19] Samet Oymak and Necmiye Ozay. "Non-asymptotic identification of lti systems from a single trajectory". In: *American Control Conference*. IEEE. 2019.

[Pag95] Franco Paganini. "Necessary and sufficient conditions for robust $\mathcal{H}_2$ performance". In: *Conference on Decision and Control*. Vol. 2. IEEE. 1995.

[Par06] Vilfredo Pareto. *Manuale di economia politica*. Vol. 13. Societa Editrice, 1906.

[PBND17] R. Postoyan, L. Buşoniu, D. Nešić, and J. Daafouz. "Stability Analysis of Discrete-Time Infinite-Horizon Optimal Control With Discounted Cost". In: *IEEE Transactions on Automatic Control* 62.6 (June 2017).

[PCS+18] Yunpeng Pan, Ching-An Cheng, Kamil Saigol, Keuntaek Lee, Xinyan Yan, Evangelos Theodorou, and Byron Boots. "Agile autonomous driving using end-to-end deep imitation learning". In: *Robotics: Science and Systems* (2018).

[PD93] Andrew Packard and John Doyle. "The complex structured singular value". In: *Automatica* 29.1 (1993).

[Per17] Nathaniel Persily. "The 2016 US Election: Can democracy survive the internet?" In: *Journal of democracy* 28.2 (2017).

[Pig20] A. C. Pigou. *The economics of welfare*. English. Macmillan London, 1920.

[PIM10] José Pereira, Morteza Ibrahimi, and Andrea Montanari. "Learning networks of stochastic differential equations". In: *Advances in Neural Information Processing Systems*. 2010.

[Pom89] Dean A Pomerleau. "Alvinn: An autonomous land vehicle in a neural network". In: *Advances in Neural Information Processing Systems*. 1989.

[PSAB21] Juan C Perdomo, Max Simchowitz, Alekh Agarwal, and Peter Bartlett. "Towards a Dimension-Free Understanding of Adaptive Linear Control". In: *arXiv preprint arXiv:2103.10620* (2021).

[QBR+95] Li Qiu, B. Bernhardsson, A. Rantzer, E.J. Davison, P.M. Young, and J.C. Doyle. "A formula for computation of the real stability radius". In: *Automatica* 31.6 (1995).

[Ren12] Steffen Rendle. "Factorization Machines with libFM". In: *ACM Trans. Intell. Syst. Technol.* 3.3 (May 2012). ISSN: 2157-6904.

[ROWAM20] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio AF Almeida, and Wagner Meira Jr. "Auditing radicalization pathways on YouTube". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2020.

[RSD+20] Esther Rolf, Max Simchowitz, Sarah Dean, Lydia T Liu, Daniel Björkegren, Moritz Hardt, and Joshua Blumenstock. "Balancing competing objectives with noisy data: Score-based classifiers for welfare-aware machine learning". In: *International Conference on Machine Learning*. PMLR. 2020.

[RU+00] R Tyrrell Rockafellar, Stanislav Uryasev, et al. "Optimization of conditional value-at-risk". In: *Journal of risk* 2 (2000).

[Rus19] Chris Russell. "Efficient search for diverse coherent explanations". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2019.

[RVKOW18] Daniel J. Russo, Benjamin Van Roy, Abbas Kazerouni, Ian Osband, and Zheng Wen. "A Tutorial on Thompson Sampling". In: *Foundations and Trends on Machine Learning* 11.1 (July 2018).

[RY06] Stephen Ross and John Yinger. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*. Cambridge: MIT Press, 2006.

[RZK19] Steffen Rendle, Li Zhang, and Yehuda Koren. "On the difficulty of evaluating baselines: A study on recommender systems". In: *arXiv preprint arXiv:1905.01395* (2019).

[Sam47] Paul A. Samuelson. *Foundations of Economic Analysis*. Harvard University Press, 1947.

[SAPT02] M. Sznaier, T. Amishima, P.A. Parrilo, and J. Tierno. "A convex approach to robust $\mathcal{H}_2$ performance analysis". In: *Automatica* 38.6 (2002).

[SBR19] Max Simchowitz, Ross Boczar, and Benjamin Recht. "Learning linear dynamical systems with semi-parametric least squares". In: *Conference on Learning Theory*. PMLR. 2019.

[Sen73] Amartya Sen. "Behaviour and the Concept of Preference". In: *Economica* 40.159 (1973).

[SF11] Davide Scaramuzza and Friedrich Fraundorfer. "Visual odometry [tutorial]". In: *IEEE robotics & automation magazine* 18.4 (2011).

[SF20] Max Simchowitz and Dylan Foster. "Naive exploration is optimal for online LQR". In: *International Conference on Machine Learning*. PMLR. 2020.

[SHM+16] David Silver et al. "Mastering the game of Go with deep neural networks and tree search". In: *Nature* 529 (Jan. 2016).

[SJ18] Ashudeep Singh and Thorsten Joachims. "Fairness of exposure in rankings". In: *Conference on Knowledge Discovery & Data Mining*. ACM SIGKDD. 2018.

[SK16] Houda Salhi and Samira Kamoun. "Combined parameter and state estimation algorithms for multivariable nonlinear systems using MIMO Wiener models". In: *Journal of Control Science and Engineering* 2016 (2016).

[SL17] Fereshteh Sadeghi and Sergey Levine. "CAD2RL: Real Single-Image Flight Without a Single Real Image". In: *Robotics: Science and Systems XIII, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, July 12-16, 2017*. 2017.

[SLK+20] Matthew J Salganik, Ian Lundberg, Alexander T Kindel, Caitlin E Ahearn, Khaled Al-Ghoneim, Abdullah Almaatouq, Drew M Altschul, Jennie E Brand, Nicole Bohme Carnegie, Ryan James Compton, et al. "Measuring the predictability of life outcomes with a scientific mass collaboration". In: *Proceedings of the National Academy of Sciences* 117.15 (2020).

[SMTJR18] Max Simchowitz, Horia Mania, Stephen Tu, Michael I. Jordan, and Benjamin Recht. "Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification". In: *Conference on Learning Theory*. Vol. 75. PMLR. 2018.

[Sol18] Joan E. Solsman. "YouTube's AI is the puppet master over most of what you watch". In: *CNET* (Jan. 10, 2018). URL: `https://www.cnet.com/news/youtube-ces-2018-neal-mohan/` (visited on 08/18/2019).

[SPGC20] Dougal Shakespeare, Lorenzo Porcaro, Emilia Gómez, and Carlos Castillo. "Exploring Artist Gender Bias in Music Recommendation". In: *arXiv preprint arXiv:2009.01715* (2020).

[SPUP02] Andrew I Schein, Alexandrin Popescul, Lyle H Ungar, and David M Pennock. "Methods and metrics for cold-start recommendations". In: *Conference on Research and Development in Information Retrieval*. ACM SIGIR. 2002.

[SR19] Tuhin Sarkar and Alexander Rakhlin. "Near optimal finite time identification of arbitrary linear dynamical systems". In: *International Conference on Machine Learning*. PMLR. 2019.

[SRD21] Tuhin Sarkar, Alexander Rakhlin, and Munther A Dahleh. "Finite Time LTI System Identification." In: *J. Mach. Learn. Res.* 22 (2021).

[SRSSP20] Sumeet Singh, Spencer M Richards, Vikas Sindhwani, Jean-Jacques E Slotine, and Marco Pavone. "Learning stabilizable nonlinear dynamics with contraction-based regularization". In: *The International Journal of Robotics Research* (2020).

[SSF09] Joseph Stiglitz, Amartya Sen, and Jean-Paul Fitoussi. "The measurement of economic performance and social progress revisited". In: *Reflections and overview. Commission on the measurement of economic performance and social progress, Paris* (2009).

[SSH20] Max Simchowitz, Karan Singh, and Elad Hazan. "Improper learning for non-stochastic control". In: *Conference on Learning Theory*. PMLR. 2020.

[SSSCJ16] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. "Recommendations as treatments: Debiasing learning and evaluation". In: *International Conference on Machine Learning*. PMLR. 2016.

[ST17] Maarten Schoukens and Koen Tiels. "Identification of block-oriented nonlinear systems starting from linear approximations: A survey". In: *Automatica* 85 (2017).

[ST19] Paige Marta Skiba and Jeremy Tobacman. "Do payday loans cause bankruptcy?" In: *The Journal of Law and Economics* 62.3 (2019).

[Ste11] Harald Steck. "Item popularity and recommendation accuracy". In: *Conference on Recommender Systems*. ACM. 2011.

[Ste18] Harald Steck. "Calibrated recommendations". In: *Conference on Recommender Systems*. ACM. 2018.

[Ste19] Harald Steck. "Embarrassingly shallow autoencoders for sparse data". In: *The World Wide Web Conference*. 2019.

[TCMK21] Lenart Treven, Sebastian Curi, Mojmír Mutný, and Andreas Krause. "Learning Stabilizing Controllers for Unstable Linear Quadratic Regulators from a Single Trajectory". In: *Learning for Dynamics and Control*. PMLR. 2021.

[TDD+20] Andrew J Taylor, Victor D Dorobantu, Sarah Dean, Benjamin Recht, Yisong Yue, and Aaron D Ames. "Towards robust data-driven control synthesis for nonlinear systems with actuation uncertainty". In: *arXiv preprint arXiv:2011.10730* (2020).

[TDLYA19] Andrew J Taylor, Victor D Dorobantu, Hoang M Le, Yisong Yue, and Aaron D Ames. "Episodic learning with control lyapunov functions for uncertain robotic systems". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019.

[TFSM14] Marko Tanaskovic, Lorenzo Fagiano, Roy Smith, and Manfred Morari. "Adaptive Receding Horizon Control for Constrained MIMO Systems". In: *Automatica* 50.12 (2014).

[TMP20] Anastasios Tsiamis, Nikolai Matni, and George Pappas. "Sample complexity of kalman filtering for unknown systems". In: *Learning for Dynamics and Control*. PMLR. 2020.

[TR19] Stephen Tu and Benjamin Recht. "The gap between model-based and model-free methods on the linear quadratic regulator: An asymptotic viewpoint". In: *Conference on Learning Theory*. PMLR. 2019.

[TS14] Koen Tiels and Johan Schoukens. "Wiener system identification with generalized orthonormal basis functions". In: *Automatica* 50.12 (2014).

[Tuf18] Zeynep Tufekci. "YouTube, the Great Radicalizer". In: *The New York Times* (Mar. 2018). URL: `https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html`.

[TWK18] Sarah Tang, Valentin Wüest, and Vijay Kumar. "Aggressive flight with suspended payloads using vision-based control". In: *IEEE Robotics and Automation Letters* 3.2 (2018).

[UFSH19] Jack Umenberger, Mina Ferizbegovic, Thomas B Schön, and Håkan Hjalmarsson. "Robust exploration in linear quadratic reinforcement learning". In: vol. 32. 2019.

[US 07] US Federal Reserve. *Report to the congress on credit scoring and its effects on the availability and affordability of credit*. 2007.

[US18] Jack Umenberger and Thomas B Schön. "Learning convex bounds for linear quadratic control policy synthesis". In: *Advances in Neural Information Processing Systems*. 2018.

[US20] Jack Umenberger and Thomas B Schön. "Optimistic robust linear quadratic dual control". In: *Learning for Dynamics and Control*. PMLR. 2020.

[USL19] Berk Ustun, Alexander Spangher, and Yang Liu. "Actionable recourse in linear classification". In: *Conference on Fairness, Accountability, and Transparency*. ACM. 2019.

[VC11] Saúl Vargas and Pablo Castells. "Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems". In: *Conference on Recommender Systems*. ACM. Chicago, Illinois, USA: Association for Computing Machinery, 2011.

[Ver10] Roman Vershynin. "Introduction to the non-asymptotic analysis of random matrices". In: *arXiv preprint arXiv:1011.3027* (2010).

[VK08] M. Vidyasagar and Rajeeva L. Karandikar. "A learning theory approach to system identification and stochastic adaptive control". In: *Journal of Process Control* 18.3 (2008).

[WDGRT18] Grady Williams, Paul Drews, Brian Goldfain, James M Rehg, and Evangelos A Theodorou. "Information-Theoretic Model Predictive Control: Theory and Applications to Autonomous Driving". In: *IEEE Transactions on Robotics* 34.6 (2018).

[Wig94] Torbjörn Wigren. "Convergence analysis of recursive identification algorithms based on the nonlinear Wiener model". In: *IEEE Transactions on Automatic Control* 39.11 (1994).

[WK02] Zhaoyang Wan and Mayuresh V Kothare. "Robust output feedback model predictive control using off-line linear matrix inequalities". In: *Journal of Process Control* 12.7 (2002).

[WMD19] Y. Wang, N. Matni, and J. C. Doyle. "A System Level Approach to Controller Synthesis". In: *IEEE Transactions on Automatic Control* (2019).

[WMR17] Sandra Wachter, Brent Mittelstadt, and Chris Russell. "Counterfactual explanations without opening the black box: Automated decisions and the GDPR". In: *Harvard Journal of Law & Technology* 31.2 (2017).

[WP95] F. Wu and A. Packard. "Optimal LQG performance of linear uncertain systems using state-feedback". In: *American Control Conference*. Vol. 6. IEEE. 1995.

[WQC+20] Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, et al. "Mind: A large-scale dataset for news recommendation". In: *Annual Meeting of the Association for Computational Linguistics*. 2020.

[WSJ21] Andrew Wagenmaker, Max Simchowitz, and Kevin Jamieson. "Task-Optimal Exploration in Linear Dynamical Systems". In: *arXiv preprint arXiv:2102.05214* (2021).

[WXLGC17] Zeng Wei, Jun Xu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. "Reinforcement learning to rank with Markov decision process". In: *Conference on Research and Development in Information Retrieval*. ACM SIGIR. 2017.

[WZ18] Kim P Wabersich and Melanie N Zeilinger. "Linear model predictive safety certification for learning-based control". In: *Conference on Decision and Control*. IEEE. 2018.

[YCHW19] Zhuoran Yang, Yongxin Chen, Mingyi Hong, and Zhaoran Wang. "Provably global convergence of actor-critic: a case for linear quadratic regulator with ergodic cost". In: *Advances in Neural Information Processing Systems*. 2019.

[YCX+18] Longqi Yang, Yin Cui, Yuan Xuan, Chenyang Wang, Serge Belongie, and Deborah Estrin. "Unbiased offline recommender evaluation for missing-not-at-random implicit feedback". In: *Conference on Recommender Systems*. ACM. 2018.

[YG14] Seungil You and Ather Gattami. "H infinity analysis revisited". In: *arXiv preprint arXiv:1412.6160* (2014).

[YHT+21] Sirui Yao, Yoni Halpern, Nithum Thain, Xuezhi Wang, Kang Lee, Flavien Prost, Ed H Chi, Jilin Chen, and Alex Beutel. "Measuring Recommender System Effects with Simulated Users". In: *arXiv preprint arXiv:2101.04526* (2021).

[YJB76] D. Youla, H. Jabr, and J. Bongiorno. "Modern Wiener-Hopf design of optimal controllers–Part II: The multivariable case". In: *IEEE Transactions on Automatic Control* 21.3 (June 1976).

[YND91] P. M. Young, M. P. Newlin, and J. C. Doyle. "μ analysis with real parametric uncertainty". In: *Conference on Decision and Control*. Vol. 2. IEEE. 1991.

[YP18] Sholeh Yasini and Kristiaan Pelckmans. "Worst-Case Prediction Performance Analysis of the Kalman Filter". In: *IEEE Transactions on Automatic Control* 63.6 (2018).

[Yu94] Bin Yu. "Rates of Convergence for Empirical Processes of Stationary Mixing Sequences". In: *The Annals of Probability* 22.1 (1994).

[ZDG96] Kemin Zhou, John C. Doyle, and Keith Glover. *Robust and Optimal Control*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[ZVRG17] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P. Gummadi. "Fairness Constraints: Mechanisms for Fair Classification". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2017.

[ZWSP08] Yunhong Zhou, Dennis Wilkinson, Robert Schreiber, and Rong Pan. "Large-scale parallel collaborative filtering for the netflix prize". In: *International conference on algorithmic applications in management*. Springer. 2008.