THE UNIVERSITY OF CHICAGO

DISTRIBUTIONALLY ROBUST CONTROL WITH STATISTICAL METHODS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
ZHAN LIN

CHICAGO, ILLINOIS
DECEMBER 2023

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ALGORITHMS

ACKNOWLEDGMENTS

First, I would like to express my deep gratitude to my advisor, Prof. Mihai Anitescu. He is knowledgeable and patient and can always provide insightful guidance when we explore new domains. His enthusiastic and diligent attitude influenced me profoundly. It is my greatest fortune to complete graduate studies under his supervision. I would also like to thank Prof. Daniel Sanz-Alonso and Prof. Yuehaw Khoo for being my committee members and kindly providing advice. I also had the opportunity to make numerous new friends throughout my time of studying. It is a non-negligible part of my life. I would like to thank Qing Yan, Dake Zhang, Xialiang Dou, Yi Liu, Pinhan Chen, Lizhen Nie, Jiacheng Wang, Ruiyi Yang, Peijun Xiao, Zhisheng Xiao, Xiaodong Chen for my wonderful years with them. Finally, I would like to thank my family. It is their support that enables me to complete this journey.

ABSTRACT

Stochastic control models are a class of mathematical models that have applications in various fields. They are widely used in applications like autopilot, robotics, and financial economics due to their ability to handle uncertainty in decision-making processes. The naive stochastic control models estimate noise distributions directly from the dataset, which makes it vulnerable to both erroneous data points and over-fitting. While distributionally robust optimization techniques have been developed in recent years to tackle one-stage problems, due to the complexities involved in multiple stages, it is not sufficiently developed in control theory. The focus of this thesis is to develop distributionally robust control with statistical methods.

In this thesis, we will cover four topics relating to distributionally robust control. We will begin by discussing the framework of distributionally robust control, bridging the gap between classical risk-averse control and Wasserstein control. Then, we dive into risk-averse control and generalize the classical risk-averse linear quadratic Gaussian control to the mixture of Gaussian scenarios. We will discuss scenarios with and without uncertainty on components' probability. While the latter can be solved with a closed-form solution, the former is more complex due to the curse of dimensionality. We propose a relaxation of the former and prove a minimax theorem for it. Following that, we introduce group lasso into distributionally robust control as a tool for outlier robustness. We provide a controller that is selectively robust on high-influential points. It can also handle scenarios where erroneous data points are present in the dataset. We also discuss the most general nonlinear non-Gaussian risk-averse control, which may not be solvable but tractable with a sequential approximation.

CHAPTER 1

INTRODUCTION

A well-known aphorism in Statistics states, “All models are wrong.” (Box, 1976, 1979; McCullagh and Nelder, 1989) The complexity of the physical world from the perspective of a modeling endeavor has been widely appreciated by the different scientific communities in the development of modern methods. The pursuit of ideality and simplicity is the natural desire of humanity and it is the starting point of many mathematical theories. For example, when Newton’s laws are stated, it is typically considered in the ideal scenarios, without errors, and with exact positions, velocities and accelerations (Newton and Chittenden, 1850). But the reality is usually different from the story Occam’s razor tells (Punch, 1639). Heisenberg’s uncertainty principle tells us it is never possible to get exact positions and velocities at the same time (Heisenberg, 1927). And more than that, the practical failure of Maxwell’s demon indicates that perturbation and imperfections broadly exist in the real world (Maxwell, 1872). In the community of control theory, classical controls are like Newton’s laws proposed in the ideal scenarios. They are, of course, imperfect models from the perspective of nature. And so appeared the task of robust control to make these models “more useful” even when they are wrong.

1.1 Robust Control and Robust Statistics

Robustness, at least in control problems, may be instantiated in multiple ways. It can be defined technically can be from a few important features. If one aims to avoid extreme cases, this can be assessed for example with a prescribed quantiles, e.g 0.95-quantile. This approach is called Value-at-Risk in the control community (Duffie and Pan, 1997). Or one can use the expectation conditional on the realization of tail values (Rockafellar and Uryasev, 2002). If the target is not avoiding extreme cases but reducing variance, exponential criteria are appropriate here since they penalize tail behaviors (Jacobson, 1973). But if one is very risk averse, worst-case control will be the choice (Korn and Menkens, 2005). Though there can be many kinds of criteria, and new ones are proposed, regularly, we believe that there are only two primary problems in the robust control. The first one is, how is the optimal policy structured? And the other one is, what perturbations or errors on the inputs and their distributions one expects the control policy can tolerate?

From a theoretical perspective, distributionally robust control can be unified by a minimax framework. Such a unified point of view we believe has eluded the control community so far. Though not explicitly expressed, this principle can be identified in several major approaches, which inspired us to use it for a framework. For example, while (Jacobson, 1973) deals with exponential criteria control it also exhibits a deterministic minimax equivalence. We will later show in Chapter 2that it is in fact a minimax problem with Kullback-Leibler divergence which connects with the relative entropy method developed later (Petersen et al., 2000; Kullback and Leibler, 1951). Moreover, when the Wasserstein distance is introduced into the control literature, the minimax framework was used implicitly again (Yang, 2020; Kim and Yang, 2020; Vaserstein, 1969). Inspired by these works and other recent advances in the adversarial learning and computer vision communities, we realized that the key to unifying all these criteria in robust control is to define Wasserstein distance "improperly" (Jiang et al., 2021; Arjovsky et al., 2017; Croitoru et al., 2022) . The recent theoretical advances of Wasserstein distance in optimization problems from different scientific communities stems from its strong duality properties (Gao and Kleywegt, 2016). We know that the traditional definition of Wasserstein distance requires the intrinsic metric to be a true metric, but interestingly, the strong duality properties do not require it (Gao and Kleywegt, 2016). As we show in §2.4, the soft-constraint strong duality properties with generalized Wasserstein distance still hold true in distributionally robust control settings. This suggested us a way to unify many issues in robust control: the choice of the intrinsic pseudo-metric of Wasserstein pseudo-distance decides what scenarios your models can work on.

There is a close connection between robust statistics and robust control. The idea from robust statistics community can help us give an answer to the warning of "All models are wrong" that needs to be addressed by any modeling-centric endeavor as control often is. Robust statistics offers many valuable ideas in helping us to push the envelope of robust control. For example, least absolute shrinkage and selection operator (lasso), elastic net and group lasso suggest an approach to formalize robustness problems with respect to outliers (Tibshirani, 1996; Efron et al., 2004; Zou and Hastie, 2005; Yuan and Lin, 2006). In a related vein, stochastic gradient descent ascent methods provide us a way to optimize with complex models (Beznosikov et al., 2022), but they also suffer from possible misspecification problems that can be addressed by robust approaches. *One objective of this dissertation is to bridge the gap between the control and statistics communities and apply these great ideas to the real world.*

The desire for robustness in decision processes when presented with sensor data streams occurs naturally in the real world. Recent well-known events concern the crashes of the Boeing 737-MAX. The essential reason for that tragedy was the failure of one the angle of attack sensors, which produced unexpected data (Demirci, 2021), as well as the decision of the manufacturer to use only one of the two angle of the attach sensors in the Maneuvering Characteristics Augmentation System (MCAS) control procedure that corrected the airplane pitch. If the controller in would have been designed to identify and handle outliers, this situation could have been averted (When MCAS was corrected it was decided to implement the correcting action only when the two sensors agreed). Moreover, in the motion planning of robotics and unmanned aerial vehicles, quite often the realized motion is different from the planned ones (Latombe, 2012). This difference can be due to Gaussian or non-Gaussian errors or discrepancies, like the sudden changing of winds. It is necessary to consider possible errors to maintain the reliability of plans. As one can thus see, there are in fact two types of robustness problems. The first is how to design a policy that it is robust to possible errors, the second is how to deal with cases where the input data is in fact incorrect.

1.2 Existing Work

As we will later show, we can unify both types of problems with the framework, but the computation tractability is not trivial for most models. This was observed early on in statistical physics where significant computational difficulties were encountered early on. Like models in particle physics and electrodynamics, though we have "Maxwell's equations" as a unifying framework, it is too hard to get analytic or even numerical solutions due to the curse of dimensionality when there is randomness in the problem (Maxwell, 1861; Bellman, 1966). We have a few results for linear quadratic models. But for general models, the answer to the questions we raise here are not easy. In statistical parlance, we are playing a conceptual trade-off between models' generalization errors and tractability. Simplification of the real model is sometimes necessary, and we will see the power of statistical methods in robust control.

Linear Quadratic Models with Gaussian Uncertainty The most basic models in statistics are Gaussian. Risk-free linear quadratic Gaussian problems were proposed in (Kushner, 1971), and then risk-averse models were introduced in (Jacobson, 1973). It was followed by (Schildbach et al., 2013) which changed the risk description towards chance constraints and (Kishida and Cetinkaya, 2022) to conditional Value-at-Risk. An interesting observation is that the latter recovers Jacobson's (Jacobson, 1973) results when it is considered in the stationary setup. (Gattami, 2009) considers the problem with second-order moment constraints. And (Petersen et al., 2000) rephrases it in minimax form with relative entropy penalizer as we mentioned in §1.1. The references (Yang, 2020; Kim and Yang, 2020) instead use a standard Wasserstein distance penalizer. In turn, the works (Scokaert and Rawlings, 1998; Bemporad et al., 2002) discuss the deterministic constrained version problem. And (Xu and Anitescu, 2018) studies the long horizon deterministic constrained convex quadratic problem and shows the exponential decay in the decomposition of the original problem. The references (Cohen et al., 2018) deals with an online version of the problem under the unknown coefficients linear dynamics. In turn, (Mania et al., 2019) shows that online least squares estimation are sufficient for the near-optimal regrets. The references (Alt and Schneider, 2015) discusses the problem with L^1 control cost. Finally, (Ting et al., 2007) studies the problem of our second question, "wrong data", using Bayesian methods to include cases with erroneous data.

Nonlinear Models Nonlinear models are generally far more difficult than linear quadratic models. The straightforward deterministic iterative linear quadratic models are proposed in (Li and Todorov, 2004) from a pure application aspect with little theory. Later, the similar idea is developed into nonlinear stochastic control under Gaussian settings (Todorov and Li, 2005; Farshidian and Buchli, 2015; Nishimura et al., 2021). While some scenario-tree based approaches have been used for nonlinear stochastic control, they are limited to very small problems. As a result, the reinforcement learning point of view, using policy gradient, is the preferred choice in the optimization of complex models (Sutton et al., 1999; Silver et al., 2014). Alpha Go is a famous successful example of deterministic control with complex models (Silver et al., 2016). The development of *robust* reinforcement learning is quite similar with the development of robust control. To that end, (Tessler et al., 2019) introduces robustness concepts into policy gradient methods. The references (Noorani et al., 2022) studies reinforcement learning problems with exponential criteria, whereas(Urpí et al., 2021) optimizes with conditional value-at-risk. The reference (Ren et al., 2020) uses log-likelihood penalizers, which is equivalent to exponential criteria in the sense of exchanging distributions' orders in Kullback-Leibler divergence.

Using Wasserstein Metrics in Robust Formulations It is worth noticing the recent advances in adversarial learning and applications of Wasserstein distance. For a number of reasons, Kullback-Leibler divergence is unfriendly to empirical measures. For example, it can only describe deviations in distributions happened in the same support, while when empirical measures are used this will be too limited. And the computational tractability for empirical measures is far worse than for parametric models. Moreover, an important feature of the Wasserstein distance that makes it so popular in the robust context consists of the strong duality results around it (Gao and Kleywegt, 2016). The Wasserstein distance is thus intensively used in recent research topics in conjunction with robust stochastic control. The reference (Abdullah et al., 2019) uses Wasserstein distance penalizers in the context of a nonlinear formulation, but its methods are based on the expansion approaches and are not exact. An important application of Wasserstein distance of major interested is within generative adversarial networks (Arjovsky et al., 2017). It is pretty intuitive that robust control has many similarities to adversarial learning. In particular, the recent topic of diffusion models are exactly a robust control problem in the context of deep learning models and applications in computer vision (Croitoru et al., 2022), and (Sinha et al., 2017) uses a Wasserstein formulationss for fast adversarial learning. The reference (Jiang et al., 2021) develops this further with complexly modeled using intrinsic distance, whereas(Dou and Anitescu, 2019) deals with robust decision problems under vector autoregressive processes setting with Wasserstein metric. Whereas, (Shafieezadeh Abadeh et al., 2018) develops Wasserstein distributionally robust Kalman filter.

Distributionally robust optimization (DRO) DRO is a topic closely related to robust control and can be interpreted as robust control with only one stage. In the context of control and decision processes in general, robustness to randomness distribution mis-specification is an important endeavor particularly in the context of high-risk installations. Driven by improving the sensitivity of inference to small perturbations many machine learning algorithms have established distributionally robust behaviors and can thus be used to achieve similar objectives in control (Rahimian and Mehrotra, 2019; Staib and Jegelka, 2019; Faccini et al., 2022; Chen and Paschalidis, 2018). An important part of DRO reusable by robust control is the regularization term (Rahimian and Mehrotra, 2019). It can be classified into Wasserstein and ϕ-divergence types. Wasserstein-type DRO is closely related to empirical measures and is thus very popular in data science. Such DRO formulations for machine learning tasks was discussed for lasso (Blanchet and Kang, 2017), logistic regression (Shafieezadeh Abadeh et al., 2015), inference in general (Blanchet et al., 2019) as well as Wasserstein-DRO extensions for many modern machine learning algorithms (Blanchet et al., 2019) . In a related thread, (Sinha et al., 2017) studies stochastic gradient methods in the same context. ϕ-divergence type DRO optimizes the trade-off between expectations and variances, directly relating to (Jacobson, 1973). Reference (Namkoong and Duchi, 2016) studies stochastic gradient methods in DRO with ϕ-divergence whereas (Duchi et al., 2016) provides the equivalence between variance regularization problems and ϕ-divergence DRO. Moreover, (Namkoong and Duchi, 2017) discusses it in parameter learning problems under convex settings, (Shapiro et al., 2021) develops in under a Bayesian setup, whereas (Shapiro and Pichler, 2022) studies it in conditional settings.

Distributionally robust control (DRC) DRC is the natural instantiation of DRO to dynamic optimization. To this end, Wasserstein distributional robust control was discussed in (Yang, 2020; Taşkesen et al., 2023), which aims at providing a DRC tool for empirical measures. (Dixit et al., 2022) presents DRC with total variation distance to optimize a relaxation of conditional Value-at-Risk, whereas (Van Parys et al., 2015) discusses DRC with conditional Value-at-Risk by restricting policies to linear decision rules. The references (Schuurmans and Patrinos, 2021) and (Hakobyan and Yang, 2022) study partially observed linear DRC problems and (Coulson et al., 2019) studies DRC with unknown dynamics. The reference (Guigues et al., 2021) studies upper bounds of risk-averse DRC with stochastic dual dynamic programming methods (Pereira and Pinto, 1991). Finally, (Shapiro, 2021) presents connections between DRO and DRC.

DRC is a relatively new area. While inspired by the literature we described and related works, it is the aim of this thesis to advance the frontier of knowledge of DRC particularly by using statistical techniques.

1.3 Contributions of this Dissertation

The dissertation includes the work of the candidate on four topics of robust control. They are bridge classical risk-averse control and Wasserstein control, risk-averse linear quadratic mixture of Gaussian control, outlier robust control with group lasso, and nonlinear non-Gaussian risk-averse control.

Chapter 2 is of kernel methods for risk-averse linear quadratic controls. It is in fact a work connecting classical approaches (Jacobson, 1973) and Wasserstein robust control as in (Kim and Yang, 2020). An interesting observation of the latter work is that it gives a control algorithm without access to noises' covariance matrices. Exchanging the order of distributions in Kullback-Leibler is in fact giving a log-likelihood estimation and have to be related to the covariance matrices. Thus it is believed that there is something missing in the Wasserstein robust controls. *Our contributions in this chapter include the proposal of a kernel method which can fill the gap between these two directions, the proposal of a unified DRC framework which can unify the usual concepts in the DRC with generalized Wasserstein distance, and the proposal of a risk measure method working for arbitrarily given robustness concepts.* Specifically, if a radial basis function is used in the model, it will be able to recover (Jacobson, 1973)'s and (Kim and Yang, 2020)'s results with different parameters (Buhmann, 2000; Schölkopf et al., 2004). It can be applied in more general settings with different kinds of kernel choices. This is also results in a proposal for a unified framework of robust controls. It in fact indicates that regularizations using Wasserstein distance will have little information about probabilities even if the intrinsic distance needs to take it into consideration.

Chapter 3 is the risk-averse linear quadratic mixture of Gaussian controls and is mainly based on the classical view of robust control. When (Jacobson, 1973) proposed risk-averse linear quadratic Gaussian controls, it is based on the assumption that noises follow Gaussian distribution. This assumption is valid for applications where noises come from the central limit theorem. But there are also plenty of applications that have non-Gaussian noise, and the sample size is not large enough to model it as Gaussian distributions with central limit theorem. For example, in network communication problems (Shih and Hero, 2003), the delay between different nodes most of the time follows a Gaussian distribution. But there is a certain probability that the traffic is congested, and it will be described by another Gaussian distribution that has higher expectations in that regime. So in the designing of stochastic control of applications involving servers like (Raeis et al., 2019; Ali et al., 2018), it is necessary to use an alternative model to handle the Gaussian noises. The aim of our second chapter is to deal with robustness problems under mixture of Gaussian noises. Different from Gaussian noise, applying of mixture of Gaussian noises suffers from the curse of dimensionality and cannot have a closed form. It is easy to derive the Bellman equations but it is hard to solve them either analytically or numerically. *Our contribution in this chapter is the provision of a closed-form solution for risk-averse linear quadratic mixture of Gaussian control when there is no uncertainty on components' probability, and a tractable concave relaxation for risk-averse linear quadratic mixture of Gaussian control when there is uncertainty on components' probability.* In the case that there is no uncertainty on components' probability, we show that dynamic programming of DRC can be solved with induction. And when there is uncertainty on components' probability, we relax the original DRC with an independence condition. And we prove a minimax theorem for it, showing that the relaxation is tractable and concave.

Chapter 4 focuses on the robustness problems to outliers . In real-world applications it is highly likely that outliers, if they exist, vary in a large domain, and only appear a few times, resulting in the difficulties of modeling them, and less likely the majority shifts even a little. For example, in the driving of unmanned aerial vehicles, most of the time winds can be modeled following Gaussian distributions. But the sudden changing of either winds' direction or intensity is possible. The main reliability problem of unmanned aerial vehicles will be how to be robust to such sudden changes. Wasserstein distance was recently introduced into robust control (Kim and Yang, 2020). With ideas from lasso, elastic net and group lasso, the compound intrinsic norm is usedy for outliers robustness problems (Tibshirani, 1996; Efron et al., 2004; Zou and Hastie, 2005; Yuan and Lin, 2006). *The contribution of this chapter is a outlier robust controller with generalized Wasserstein distance composing of squared 2-norm and 2-norm, which can be thought of as a version of group lasso in DRC.* This optimization formulation replaces the Kullback-Leibler divergence that appeared in the classical control (Jacobson, 1973) with the new composite Wasserstein distance. Outliers can possibly have a high impact on the cost-to-go functions. So in the equivalent form of the minimax problem, we can see that the robust controller is, in fact, a controller selecting out high-influence data points and optimizing the worst case that they are shifting in a certain range. Though this is a multistage problem of DRO lasso problems, we still can get efficient computational tractability under the linear quadratic settings.

Chapter 5 is nonlinear non-Gaussian risk-averse control. Nonlinear control problems are important and useful but hard. It is still unclear how to find efficiently a global minimum for a nonlinear optimization problem with nonlinear dynamics. For example, the previous work on the deterministic nonlinear control is purely based on the numerical experimental results (Li and Todorov, 2004). So this chapter is not aiming to solve general nonlinear control problems but to provide a tractable algorithm for nonlinear non-Gaussian risk-averse control. *Our contribution in this chapter is the proposal of a sequentially approximated distributionally robust control algorithm for nonlinear non-Gaussian risk-averse control.* The approach to deal with risk-averse nonlinear controls is to separate deterministic parts from the stochastic parts. The separation we propose is not unique. It is just like finding a trajectory of a deterministic system first and then describing the stochastic component as a perturbation to this trajectory. We assume that the deterministic parts can be solved efficiently, whereas the stochastic parts are approached with a sequence of quadratic programming problems with different deterministic trajectories. And we prove a minimax theorem for the stochastic part. Combining these two controllers gives a risk-averse nonlinear controller.

CHAPTER 2

BRIDGE CLASSICAL RISK-AVERSE CONTROL AND WASSERSTEIN CONTROL

2.1 Introduction

Linear quadratic Gaussian (LQG) control is one of the most basic models in the control theory. It assumes linear dynamics and quadratic objectives and optimizes the expectation of the objective. In some scenarios, the tail behavior of the objective function needs to be taken into consideration. And thus, risk-averse linear quadratic Gaussian control is developed in the (Jacobson, 1973). It deals with the trade-off problem between expectation and variance by optimizing an exponential agent of the original objective since the exponential function has heavier weights on tail values, which means that for any real random variable f, the agent objective is

$$\frac{1}{\gamma}\log\mathbb{E}\left[\exp\left\{\gamma f\right\}\right] = \mathbb{E}\left[f\right] + \frac{\gamma}{2}\mathbb{V}[f] + O(\gamma^2) \tag{2.1}$$

Later, it is followed by (Petersen et al., 2000) for relative entropy penalty. Instead of optimizing the expectation of the objective, it optimizes a minimax problem with Kullback-Leibler divergence. If we denote the original distribution of f to be P, then it optimizes

$$\max_{\tilde{P}, KL(\tilde{P}\|P)\leq\omega} \mathbb{E}_{f\sim\tilde{P}}\left[f\right] \tag{2.2}$$

Recently (Kim and Yang, 2020) uses Wasserstein penalty in the LQG, which optimizes

$$\max_{\tilde{P}} \mathbb{E}_{f\sim\tilde{P}}\left[f\right] - \frac{1}{\gamma}W_2^2(\tilde{P}, P) \tag{2.3}$$

where W_2 is Wasserstein-2 distance and here, P is the empirical distribution of the data. An interesting observation for the Wasserstein controller in (Kim and Yang, 2020) is that it will not depend on the covariance matrices of noises, but the classical LQG did. This paper aims to deal with this gap between the two types of controllers. We find a way to unify all these three types of controllers as the same type of controller with different parameters.

The main points of this paper include,

1. We provide a kernelized method to connect classical risk-averse LQG controllers and Wasserstein robust controllers. We show Wasserstein robust controllers are, in fact, an over-robust choice in the linear quadratic control problems if kernelized methods are not used. We argue from probabilities aspects, geometries aspects, and physical aspects that the kernel method is necessary.

2. We introduce general sense Wasserstein discrepancy for control problems here. We provide a framework to unify the usual concepts in robust controls and provide proof for soft-constraint strong duality properties in general control problems.

3. We claim Wasserstein distance has problems of redundant robustness in all control problems that noises have underlying geometric properties. We provide a locally linear embedding (LLE) algorithm to deal with it. We show that this algorithm is, in fact, in the framework of the kernelized method.

4. We provide a highly tractable method to evaluate risk with arbitrarily given robustness concepts and for general nonlinear control problems.

2.2 Problem Formulation and Assumptions

This paper will consider linear quadratic control problems (LQP) with Wasserstein penalty terms. We will use kernelized distance in the Wasserstein distance, which is

$$\begin{aligned} V_t(x_t) = \min_{u_t} \max_{\tilde{P}_t, \varepsilon \sim \tilde{P}_t} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] - \frac{1}{\gamma} W_t(\tilde{P}_t, P_t) \\ s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \end{aligned} \tag{2.4}$$

where $\gamma > 0, V_T(x_T) = 0, x_t \in \mathbb{R}^{d_1}, u_t \in \mathbb{R}^{d_2}, \varepsilon_t \in R^{d_3}$, ε_t are independent, Q_t, R_t, A_t, B_t, C_t are matrices in proper dimension. P_t are empirical measures on n data points $\varepsilon_{t,i}$. We define the *kernelized Wasserstein distance* as

$$W_t(\tilde{P}_t, P_t) = \min_{\Gamma=\Pi(\tilde{P}_t,P_t)} \int c_t(x,y) d\Gamma(x,y) \tag{2.5}$$

This concept is previously discussed in (Oh et al., 2019) from application aspects. However, its role in the control theory and its connection with classical risk-averse control are not discussed. In this paper, we will use $c_t(x,y)$ constructed by kernels,

$$c_t(x,y) = K_{t,y}(x,x) - 2K_{t,y}(x,y) + K_{t,y}(y,y) \tag{2.6}$$

where $K_{t,y}(x,y)$ are any positive definite kernel functions defined on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_1}$; which in turn implies that $c_t(x,y) \geq 0$. Notice that we do not use Wasserstein distance in the usual sense but generalize it to the discrepancy sense. We will claim its validity in §2.4.

2.2.1 Assumptions

We will use common assumptuions in the risk-averse control,

Assumption 2.1. $Q_t \geq 0$.

Assumption 2.2. $R_t \geq 0$.

Assumption 2.3. $rank(B_t) = d_2$.

Assumption 2.4. *The minimum of* (2.4) *exists.*

For notation simplicity, if not specified, $\|\cdot\|$ are always assumed to be $\|\cdot\|_2$ in this paper. Moreover, if we mention x_{t+1} without dynamics at any place, then the dynamics are assumed to be the one in (2.4).

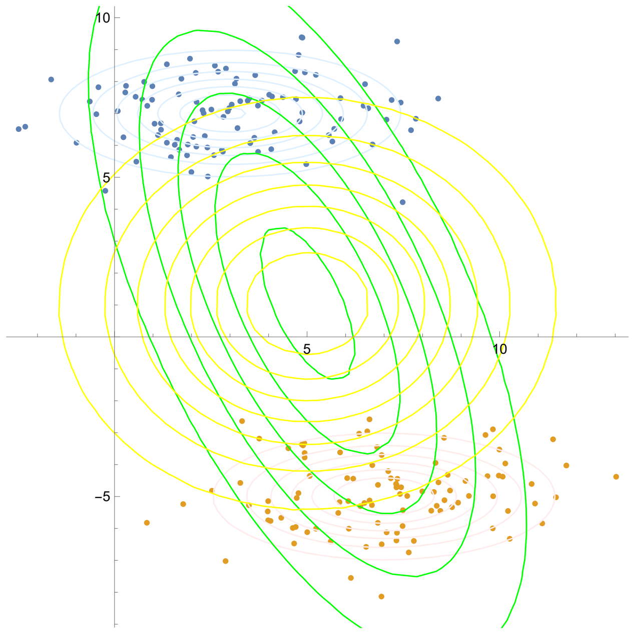


Figure 2.1: This is contour fitting for the case of a two-component Gaussian mixture. Risk-averse LQG will model them as one Gaussian, as the green curve shows. Standard Wasserstein robust controllers model it as the yellow curve shows. We can independently recover two components' contours with a radial basis function kernel.

2.3 Local Density Estimation

We will use the kernel

$$K_t(x, y) = x^T \Sigma_{t,y}^{-1} y \tag{2.7}$$

and radial basis function (RBF) inside it (Broomhead and Lowe, 1988; Scholkopf et al., 1997),

$$\begin{aligned}\Sigma_{t,y} = \frac{1}{\sum_{i=1}^{n} w_{t,i}(y)} \sum_{i=1}^{n} w_{t,i}(y) \left(\varepsilon_{t,i} - \sum_{j=1}^{n} \frac{w_{t,j}(y)}{\sum_{i=1}^{n} w_{t,i}(y)} \varepsilon_{t,j} \right)^T \Bigg(\varepsilon_{t,i} \\ - \sum_{j=1}^{n} \frac{w_{t,j}(y)}{\sum_{i=1}^{n} w_{t,i}(y)} \varepsilon_{t,j} \Bigg) + e^{-\frac{\sigma^2}{2}} \mathbf{I}_{d_3},\end{aligned} \tag{2.8}$$

$$w_{t,i}(y) = e^{-\frac{\|y - \varepsilon_{t,i}\|^2}{2\sigma^2}}. \tag{2.9}$$

σ is a hyper-parameter to tune. In two limit regimes it connects classical risk-averse linear quadratic Gaussian controllers in (Jacobson, 1973) and Wasserstein robust controllers in (Kim and Yang, 2020). It is the weighted sample convariance (Price, 1972) by putting weights $w_{t,i}(y)$ on $\varepsilon_{t,i}$. Let $\sigma \to 0$, the data point $\varepsilon_{t,i}$ closest to y will have normalized weight $\frac{w_{t,i}(y)}{\sum_{j=1}^{n} w_{t,j}(y)} \to 1$ and thus the first term of the $\Sigma_{t,y}$ is 0. Then $\Sigma_{t,y} \to \mathbf{I}_{d_3}$ and

$$c_t(x, y) \to \|x - y\|^2, \tag{2.10}$$

which is the distance used in (Kim and Yang, 2020). Let $\sigma \to \infty$, we have

$$\Sigma_{t,y} \to \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{t,j} \varepsilon_{t,j}^T. \tag{2.11}$$

Define $\Sigma_t = \frac{1}{n} \sum_{j=1}^{n} \varepsilon_{t,j} \varepsilon_{t,j}^T$, then we have

$$c_t(x, y) \approx (x - y)^T \Sigma_t^{-1} (x - y). \tag{2.12}$$

We will show (2.12) recovers (Jacobson, 1973)'s results at Theorem 2.3 and Lemma 2.5.

Theorem 2.1. (2.4) *is equivalent to*

$$\begin{aligned} V_t(x_t) = \min_{u_t} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E} &\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V(x_{t+1}) \right] \\ &- \frac{1}{\gamma n} \sum_{i=1}^{n} \left(\hat{\varepsilon}_{t,i} - \varepsilon_{t,i} \right)^T \Sigma_{t,i}^{-1} \left(\hat{\varepsilon}_{t,i} - \varepsilon_{t,i} \right) \\ &s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t \end{aligned} \tag{2.13}$$

where $\hat{P}_t$ are empirical measures on n data points $\hat{\varepsilon}_{t,i}$, and $\hat{\varepsilon}_{t,i}$ are any data points in $\mathbb{R}^{d_3}$.

Proof. Observe that maximizing over $\hat{P}_t$ is equivalent to maximizing over $\hat{\varepsilon}_{t,i}$, for $i = 1, 2, \ldots, n$. This is by Theorem 2.8. □

Theorem 2.2. *Define*

$$V_t(x_t) = \min_{u_t} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E} \left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1}) - \frac{1}{\gamma} (\hat{\varepsilon}_t - \varepsilon_t)^T \Sigma_t^{-1} (\hat{\varepsilon}_t - \varepsilon_t) \right] \tag{2.14}$$

where the expectation is over the random selection $1,\ldots,n$*, which is equivalent to* (2.13)*, and*

$$V_T\left(x_T\right)=0 \tag{2.15}$$

for any point $x_T\in\mathbb{R}^{d_1}$*. Then* $\forall t<T$*, we have*

$$V_t\left(x_t\right)=x_t^T\Xi_t x_t+2\sum_{\tau=t}^{T}\sum_{i=1}^{n}x_t^T\Xi_{t,\tau,i}\varepsilon_{\tau,i}+z_t \tag{2.16}$$

where $\Xi_t\in\mathbb{R}^{d_1,d_1},\Xi_{t,\tau,i}\in\mathbb{R}^{d_1,d_3}$*,* $z_t\in\mathbb{R}$*. Here,* z_t *is a constant that depends on* t *but not on* x_t *and thus does not affect solutions of optimization problems for time indices smaller than* t*. The optimal policy of* (2.14) *is*

$$u_t=U_t x_t+\sum_{\tau=t}^{T-1}\sum_{j=1}^{n}U_{t,\tau,j}\varepsilon_{\tau,j}, \tag{2.17}$$

and $U_t\in\mathbb{R}^{d_2,d_1},U_{t,\tau,j}\in\mathbb{R}^{d_2,d_3}$*. Here, for* $t<T$*, we have*

$$\begin{aligned}U_t=-\Bigg(&B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\Bigg(B_t\\&+\gamma C_t\mathbb{E}\left[\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_t^{-1}\right)^{-1}\right]C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t\Bigg)\\&+R_t\Bigg)^{-1}B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\Bigg(A_t\\&+\gamma C_t\mathbb{E}\left[\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_t^{-1}\right)^{-1}\right]C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)A_t\Bigg),\end{aligned} \tag{2.18}$$

$$\begin{aligned}U_{t,t,j}=-\frac{1}{n}\Bigg(&B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\Bigg(B_t\\&+\gamma C_t\mathbb{E}\left[\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_t^{-1}\right)^{-1}\right]C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t\Bigg)\\&+R_t\Bigg)^{-1}B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_{t,j}^{-1}\right)^{-1}\Sigma_{t,j}^{-1},\end{aligned} \tag{2.19}$$

and for $\tau > t$,

$$U_{t,\tau,j} = -\Bigg(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\Bigg(B_t + \gamma C_t \mathbb{E}\left[\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_t^{-1}\right)^{-1}\right]C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t\Bigg) + R_t\Bigg)^{-1}\Bigg(B_t^T\Xi_{t+1,\tau,j} + \gamma B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\mathbb{E}\left[\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_t^{-1}\right)^{-1}\right]C_t^T\Xi_{t+1,\tau,j}\Bigg). \tag{2.20}$$

Moreover, for $t < T$ we have

$$\Xi_t = \mathbb{E}\Bigg[\left(A_t+B_tU_t+C_t\hat{A}_t+C_t\hat{B}_tU_t\right)^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t+B_tU_t+C_t\hat{A}_t+C_t\hat{B}_tU_t\right) + U_t^TR_tU_t - \frac{1}{\gamma}\left(\hat{A}_t+\hat{B}_tU_t\right)^T\Sigma_t^{-1}\left(\hat{A}_t+\hat{B}_tU_t\right)\Bigg], \tag{2.21}$$

$$\Xi_{t,t,j} = \mathbb{E}\left[\left(A_t+B_tU_t+C_t\hat{A}_t+C_t\hat{B}_tU_t\right)^T\right]\left(Q_{t+1}+\Xi_{t+1}\right)B_tU_{t,t,j} + \frac{1}{n}\Bigg(A_t+B_tU_t + C_t\hat{A}_{t,j}+C_t\hat{B}_{t,j}-\frac{1}{\gamma}U_t\Bigg)^T C_t\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_{t,j}^{-1}\right)^{-1}\Sigma_{t,j}^{-1} + U_t^TR_tU_{t,t,j} - \frac{1}{n\gamma}\left(\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_{t,j}^{-1}\right)^{-1}\Sigma_{t,j}^{-1}-\mathbf{I}_{d_3}\right), \tag{2.22}$$

and for $\tau > t$,

$$\Xi_{t,\tau,j} = \mathbb{E}\Bigg[\left(A_t+B_tU_t+C_t\hat{A}_t+C_t\hat{B}_tU_t\right)^T\left(Q_{t+1}+\Xi_{t+1}\right)\Bigg(\left(Q_{t+1}+\Xi_{t+1}\right)B_tU_{t,\tau,j} + \gamma C_t\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_t^{-1}\right)^{-1}C_t^T\Xi_{t+1,\tau,j}+\Xi_{t+1,\tau,j}\Bigg) + U_t^TR_tU_{t,\tau,j} - \left(\hat{A}_t+\hat{B}_tU_t\right)^T\Sigma_t^{-1}\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_t^{-1}\right)^{-1}C_t^T\Xi_{t+1,\tau,j}\Bigg]. \tag{2.23}$$

In equations 2.18–2.23, for $t < T$ we use the notation:

$$\hat{A}_{t,i} = \gamma\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_{t,i}^{-1}\right)^{-1}C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)A_t, \tag{2.24}$$

$$\hat{B}_{t,i} = \gamma\left(-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t+\Sigma_{t,i}^{-1}\right)^{-1}C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t. \tag{2.25}$$

Proof. This can be shown by induction. Firstly, it is true for $t = T$. Since $V_T(x_T) = 0$, we only need to choose relevant matrices to be zero. Assume it holds for $t = m+1$ where $m \in \mathbb{N}$, then for $t = m$, by the dynamic $x_{t+1} = A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t$ and the induction, we have

$$
\begin{aligned}
V_t(x_t) = \mathbb{E}\Bigg[&(A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t) + u_t^T R_t u_t \\
&+ 2 \sum_{\tau=t+1}^{T} \sum_{i=1}^{n} (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T \Xi_{t+1,\tau,i} \varepsilon_{\tau,i} - \frac{1}{\gamma} (\hat{\varepsilon}_t - \varepsilon_t)^T \Sigma_t^{-1} (\hat{\varepsilon}_t - \varepsilon_t) \Bigg] + z_{t+1}.
\end{aligned} \tag{2.26}
$$

By the optimality conditions for the inner problem applied to 2.14, at the maximum point $\hat{\varepsilon}_{t,i}$, for $i = 1, 2 \dots n$, we get

$$
\begin{aligned}
&2C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \hat{\varepsilon}_{t,i} + 2C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t) \\
&- \frac{2}{\gamma} \Sigma_{t,i}^{-1} \left(\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\right) + 2 \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} C_t^T \Xi_{t+1,\tau,j} \varepsilon_{\tau,j} = 0.
\end{aligned} \tag{2.27}
$$

$$
\begin{aligned}
\hat{\varepsilon}_{t,i} &= \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_{t,i}^{-1}\right)^{-1} \Bigg(\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t) + \Sigma_{t,i}^{-1} \varepsilon_{t,i} \\
&\qquad + \gamma \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} C_t^T \Xi_{t+1,\tau,j} \varepsilon_{\tau,j} \Bigg) \\
&= \hat{A}_{t,i} x_t + \hat{B}_{t,i} u_t + \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_{t,i}^{-1}\right)^{-1} \Sigma_{t,i}^{-1} \varepsilon_{t,i} \\
&\quad + \gamma \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_{t,i}^{-1}\right)^{-1} \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} C_t^T \Xi_{t+1,\tau,j} \varepsilon_{\tau,j}.
\end{aligned} \tag{2.28}
$$

Computing now the expectation with respect to the random selection now, we get

$$
\begin{aligned}
\mathbb{E}\left[\hat{\varepsilon}_t\right] = {} & \gamma \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1}\right] C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) A_t x_t \\
& + \gamma \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1}\right] C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t u_t \\
& + \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1} \Sigma_t^{-1} \varepsilon_t\right] \\
& + \gamma \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1}\right] \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} C_t^T \Xi_{t+1,\tau,j} \varepsilon_{\tau,j},
\end{aligned}
\tag{2.29}
$$

where the expectation is over the random selection process.

Using now the optimality condition on u_t, we have by Danskin's theorem (Danskin, 1966) that the derivative of the minimum problem in minimax problem (2.14) equals to the derivative of the minimum problem fixing $\hat{P}_t$ to be maximum point first

$$
\begin{aligned}
& 2B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t u_t + 2B_t^T \left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t x_t + C_t \mathbb{E}\left[\hat{\varepsilon}_t\right]\right) \\
& + 2R_t u_t + 2 \sum_{\tau=t+1}^{T} \sum_{i=1}^{n} B_t^T \Xi_{t+1,\tau,i} \varepsilon_{\tau,i} = 0.
\end{aligned}
\tag{2.30}
$$

Then, solving for u_t, we get that:

$$\begin{aligned}u_t = -\Bigg(& B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \Bigg(B_t \\ & + \gamma C_t \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1}\right] C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \Bigg) \\ & + R_t \Bigg)^{-1} \Bigg(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \Bigg(A_t \\ & + \gamma C_t \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1}\right] C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t \Bigg) x_t \\ & + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1} \Sigma_t^{-1} \varepsilon_t\right] \\ & + \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} \Bigg(B_t^T \Xi_{t+1,\tau,j} \\ & + \gamma B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \mathbb{E}\left[\left(-\gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t + \Sigma_t^{-1}\right)^{-1}\right] C_t^T \Xi_{t+1,\tau,j} \Bigg) \varepsilon_{\tau,j} \Bigg) \\ = & U_t x_t + \sum_{\tau=t}^{T} \sum_{j=1}^{n} U_{t,\tau,j} \varepsilon_{\tau,j}\end{aligned} \tag{2.31}$$

We can now compute the expression of the value function at t in (2.14), by substituting x_{t+1} from the dynamics recursion, and then using the optimal expressions of u_t (2.31) and $\hat{\varepsilon}_{t,i}$, $i = 1, 2, \ldots n$, from (2.28), while using the expectation with respect to the random selection. We get

$$
\begin{aligned}
V_t(x_t) = \mathbb{E} \Bigg[& (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t) + u_t^T R_t u_t \\
& + 2 \sum_{\tau=t+1}^{T} \sum_{i=1}^{n} (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T \Xi_{t+1,\tau,i} \varepsilon_{\tau,i} - \frac{1}{\gamma} (\hat{\varepsilon}_t - \varepsilon_t)^T \Sigma_t^{-1} (\hat{\varepsilon}_t - \varepsilon_t) \Bigg] + z_{t+1} \\
= \mathbb{E} \Bigg[& x_t^T \left(A_t + B_t U_t + C_t \hat{A}_t + C_t \hat{B}_t U_t \right)^T (Q_{t+1} + \Xi_{t+1}) \Big(A_t + B_t U_t + C_t \hat{A}_t \\
& + C_t \hat{B}_t U_t \Big) x_t + x_t^T U_t^T R_t U_t x_t - \frac{1}{\gamma} x_t^T \left(\hat{A}_t + \hat{B}_t U_t \right)^T \Sigma_t^{-1} \left(\hat{A}_t + \hat{B}_t U_t \right) x_t \\
& + 2 x_t^T \left(A_t + B_t U_t + C_t \hat{A}_t + C_t \hat{B}_t U_t \right)^T \Bigg((Q_{t+1} + \Xi_{t+1}) B_t \sum_{\tau=t}^{T} \sum_{j=1}^{n} U_{t,\tau,j} \varepsilon_{\tau,j} \\
& + C_t \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_t^{-1} \right)^{-1} \Sigma_t^{-1} \varepsilon_t \\
& + \gamma C_t \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_t^{-1} \right)^{-1} \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} C_t^T \Xi_{t+1,\tau,j} \varepsilon_{\tau,j} \\
& + \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} \Xi_{t+1,\tau,j} \varepsilon_{\tau,j} \Bigg) + 2 x_t^T U_t^T R_t \sum_{\tau=t}^{T} \sum_{j=1}^{n} U_{t,\tau,j} \varepsilon_{\tau,j} \\
& - \frac{2}{\gamma} x_t^T \left(\hat{A}_t + \hat{B}_t U_t \right)^T \Sigma_t^{-1} \Bigg(\left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_t^{-1} \right)^{-1} \Sigma_t^{-1} \varepsilon_t \\
& + \gamma \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_t^{-1} \right)^{-1} \sum_{\tau=t+1}^{T} \sum_{j=1}^{n} C_t^T \Xi_{t+1,\tau,j} \varepsilon_{\tau,j} - \varepsilon_t \Bigg) \Bigg] + z_t
\end{aligned} \tag{2.32}
$$

Thus, by identifying the relevant matrices in (2.32), we obtain that

$$
\begin{aligned}
\Xi_t = \mathbb{E} \Big[& \left(A_t + B_t U_t + C_t \hat{A}_t + C_t \hat{B}_t U_t \right)^T (Q_{t+1} + \Xi_{t+1}) \left(A_t + B_t U_t + C_t \hat{A}_t + C_t \hat{B}_t U_t \right) \\
& + U_t^T R_t U_t - \frac{1}{\gamma} \left(\hat{A}_t + \hat{B}_t U_t \right)^T \Sigma_t^{-1} \left(\hat{A}_t + \hat{B}_t U_t \right) \Big]
\end{aligned} \tag{2.33}
$$

and that, $\forall t, \tau > t, j$,

$$\Xi_{t,\tau,j} = \mathbb{E}\Bigg[\left(A_t + B_tU_t + C_t\hat{A}_t + C_t\hat{B}_tU_t\right)^T (Q_{t+1} + \Xi_{t+1}) \Bigg((Q_{t+1} + \Xi_{t+1}) B_tU_{t,\tau,j}$$
$$+ \gamma C_t \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_t^{-1}\right)^{-1} C_t^T \Xi_{t+1,\tau,j} + \Xi_{t+1,\tau,j}\Bigg) + U_t^T R_t U_{t,\tau,j}$$
$$- \left(\hat{A}_t + \hat{B}_tU_t\right)^T \Sigma_t^{-1} \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_t^{-1}\right)^{-1} C_t^T \Xi_{t+1,\tau,j}\Bigg], \tag{2.34}$$

and that $\forall t, j$,

$$\Xi_{t,t,j} = \mathbb{E}\left[\left(A_t + B_tU_t + C_t\hat{A}_t + C_t\hat{B}_tU_t\right)^T\right] (Q_{t+1} + \Xi_{t+1}) B_tU_{t,t,j} + \frac{1}{n}\Bigg(A_t + B_tU_t$$
$$+ C_t\hat{A}_{t,j} + C_t\hat{B}_{t,j} - \frac{1}{\gamma}U_t\Bigg)^T C_t \left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_{t,j}^{-1}\right)^{-1} \Sigma_{t,j}^{-1}$$
$$+ U_t^T R_t U_{t,t,j} - \frac{1}{n\gamma}\left(\left(-\gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t + \Sigma_{t,j}^{-1}\right)^{-1} \Sigma_{t,j}^{-1} - \mathbf{I}_{d_3}\right). \tag{2.35}$$

This completes the induction hypothesis and the proof.

□

To see how this framework connects (Jacobson, 1973) and (Kim and Yang, 2020), we consider two distributionally robust control problems: one with Wasserstein distance from (2.3), (2.36), and one with KL-divergence, (2.38). The Wasserstein distributionally robust control problem is

$$V_t(x_t) = \min_{u_t} \max_{\tilde{P}_t, \varepsilon \sim \tilde{P}_t} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] - \frac{1}{\gamma} W_t(\tilde{P}_t, P_t^W), \tag{2.36}$$

where P_t^W is an empirical measure with mean $\tilde{\mu}_t$ and covariance $\tilde{\Sigma}_t$, and where we take the terminal condition to be $V_T(x_T) = 0$. The "generalized" Wasserstein distance we will use is

$$W_t(\tilde{P}_t, P_t) = \min_{\Gamma = \Pi(\tilde{P}_t, P_t)} \int (x-y)^T \tilde{\Sigma}_t (x-y) d\Gamma(x,y). \tag{2.37}$$

Note that, when $\tilde{\Sigma}$ is taken to be the identity matrix, the problem (2.36)–(2.37) uses the classical Wasserstein W_2^2 distance and is thus equivalent to the setup in (Kim and Yang, 2020). On the other hand, the KL-divergence distributionally robust control problem from (Jacobson, 1973) is stated as:

$$V_t(x_t) = \min_{u_t} \max_{\tilde{P}_t, \varepsilon \sim \tilde{P}_t} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] - \frac{2}{\gamma} KL(\tilde{P}_t, P_t^K). \quad (2.38)$$

Here P_t^K is a $N(\tilde{\mu}_t, \tilde{\Sigma}_t)$ and we take again as terminal condition $V_T(x_T) = 0$. We will later show that the KL controller (2.38) is equivalent to the risk-averse controller from Theorem 2.10. Our framework allows us to consider the relationship between the following two workflows: get the data for the noise, $\varepsilon_{t,i}$, $i = 1, 2, \ldots, m$ compute the empirical covariance and then (a) compute the Wasserstein risk-averse controller, relative to the empirical distribution, by solving (2.36) with metric (2.37) or (b) compute the KL risk-averse controller by solving (2.38) relative to the normal distribution with covariance matrix the empirical covariance itself. We then claim that (2.36) and (2.38) have *the same optimal control policy*, which illuminates the subtle relationship between the two distributionally robust control paradigms.

We show this in Theorem 2.3 and Lemma 2.5,

Theorem 2.3. *Observe that, by Theorem 2.1, the problem* (2.36)–(2.37) *is equivalent to*

$$V_t(x_t) = \min_{u_t} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \left\{ \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t - \frac{1}{\gamma}(\hat{\varepsilon}_t - \varepsilon_t)^T \tilde{\Sigma}_t^{-1} (\hat{\varepsilon}_t - \varepsilon_t) + V(x_{t+1})\right]\right\}, \quad (2.39)$$

where $x_{t+1} = A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t$, *and*

$$V_T(x_T) = 0 \quad (2.40)$$

for any point $x_T \in \mathbb{R}^{d_1}$. *Then* $\forall t < T$, *we have the optimal control for problem* (2.39)–(2.40), *and, consequently, for problem* (2.36)–(2.37) *satisfies* $u_t = U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_\tau$ *where* $U_t \in \mathbb{R}^{d_2,d_1}$ *and*

$$V_t(x_t) = x_t^T \Xi_t x_t + 2 \sum_{\tau=t}^{T-1} x_t^T \Xi_{t,\tau} \tilde{\mu}_\tau + z_t, \quad (2.41)$$

where $\Xi_t \in \mathbb{R}^{d_1,d_1}$, $z_t \in \mathbb{R}$. *Here,* z_t *is a constant that depends on* t *but not on to* x_t *and thus does not affect solutions of optimization problems for time indices smaller than* t. *Here, for* $t < T$, U_t *and* $U_{t,t}$ *are defined as*

$$U_t = -\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t + R_t + B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_t\right)^{-1}B_t^T\left(Q_{t+1}\right. \\ \left.+\,\Xi_{t+1}\right)\left(A_t + C_t\tilde{A}_t\right), \tag{2.42}$$

$$U_{t,t} = -\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t + R_t + B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_t\right)^{-1}B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{C}_t, \tag{2.43}$$

and for $t > T$,

$$U_{t,\tau} = -\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t + R_t + B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_t\right)^{-1}B_t^T\left(Q_{t+1}\right. \\ \left.+\,\Xi_{t+1}\right)\left(\gamma C_t\tilde{C}_t\tilde{\Sigma}_t C_t\Xi_{t+1,\tau} + B_t^T\Xi_{t+1,\tau}\right). \tag{2.44}$$

Moreover, Ξ_t *and* $\Xi_{t,t}$ *are defined as*

$$\Xi_t = \left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right)^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right) \\ + U_t^TR_tU_t - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_tU_t\right)^T\tilde{\Sigma}_t^{-1}\left(\tilde{A}_t + \tilde{B}_tU_t\right), \tag{2.45}$$

$$\Xi_{t,t} = \left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right)\left(Q_{t+1}+\Xi_{t+1}\right)\left(B_tU_{t,t} + C_t\tilde{B}_tU_{t,t} + C_t\tilde{C}_t\right) \\ + U_t^TR_tU_{t,t} - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_tU_t\right)\tilde{\Sigma}_t^{-1}\tilde{C}_t \tag{2.46}$$

and for $\tau > t$,

$$\Xi_{t,\tau} = \left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right)^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(B_tU_{t,\tau} + C_t\tilde{B}_tU_{t,\tau}\right. \\ \left.+\,\gamma C_t\tilde{C}_t\tilde{\Sigma}_tC_t^T\Xi_{t+1,\tau}\right) \\ + U_t^TR_tU_{t,\tau} - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_tU_t\right)^T\tilde{\Sigma}_t^{-1}\left(\tilde{B}_tU_{t,\tau} + \gamma\tilde{\Sigma}_t^{-1}\tilde{C}_t\tilde{\Sigma}_tC_t^T\Xi_{t+1,\tau}\right), \tag{2.47}$$

where

$$\tilde{A}_t = \gamma \left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t, \tag{2.48}$$

$$\tilde{B}_t = \gamma \left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t, \tag{2.49}$$

$$\tilde{C}_t = \left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \tilde{\Sigma}_t^{-1}. \tag{2.50}$$

Proof. This can be shown by induction. Firstly it is true for $t = T$. Since $V_T(x_T) = 0$, we only need to choose relevant matrices to be zero. Assume it holds for $t = m+1$ where $m \in \mathbb{N}$, then for $t = m$, by the dynamic $x_{t+1} = A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t$ and the induction, we have

$$\begin{aligned} V_t\left(x_t\right) = \min_{u_t} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \Bigg\{ & x_{t+1}^T \left(Q_{t+1} + \Xi_{t+1}\right) x_{t+1} + u_t^T R_t u_t + 2 \sum_{\tau=t+1}^{T-1} x_{t+1}^T \Xi_{t+1,\tau} \tilde{\mu}_\tau \\ & - \frac{1}{\gamma} \left(\hat{\varepsilon}_t - \varepsilon_t\right)^T \tilde{\Sigma}_t^{-1} \left(\hat{\varepsilon}_t - \varepsilon_t\right) \Bigg\} + z_{t+1}. \end{aligned} \tag{2.51}$$

Applying the optimality conditions on $\hat{\varepsilon}_{t,i}$, i=1,2,...,n in the inner problem, same as we did in Theorem 2.2, results in

$$\begin{aligned} & C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \hat{\varepsilon}_{t,i} + C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) \\ & + \sum_{\tau=t+1}^{T-1} C_t^T \Xi_{t+1,\tau} \tilde{\mu}_\tau - \frac{1}{\gamma} \tilde{\Sigma}_t^{-1} \left(\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\right) = 0, \end{aligned} \tag{2.52}$$

and therefore, by solving for $\hat{\varepsilon}_{t,i}$, for $i = 1, 2, \ldots, n$, and then, forcing γ out, also in

$$\begin{aligned}
\hat{\varepsilon}_{t,i} &= -\left(C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t - \frac{1}{\gamma}\tilde{\Sigma}_t^{-1}\right)^{-1}\Bigg(C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t x_t + B_t u_t\right) \\
&\quad + \sum_{\tau=t+1}^{T-1} C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau\Bigg) - \left(C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t - \frac{1}{\gamma}\tilde{\Sigma}_t^{-1}\right)^{-1}\frac{1}{\gamma}\tilde{\Sigma}_t^{-1}\varepsilon_{t,i} \\
&= \gamma\left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}\Bigg(C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t x_t + B_t u_t\right) \\
&\quad + \sum_{\tau=t+1}^{T-1} C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau\Bigg) + \left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}\tilde{\Sigma}_t^{-1}\varepsilon_{t,i} \\
&= \tilde{A}_t x_t + \tilde{B}_t u_t + \gamma\sum_{\tau=t+1}^{T-1}\tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau + \tilde{C}_t\varepsilon_{t,i}
\end{aligned} \tag{2.53}$$

where $\tilde{A}_t$ is defined in (2.48), $\tilde{B}_t$ is defined in (2.49) and $\tilde{C}_t$ is defined in (2.50).

Applying the expectation operator, over the random selection, to (2.52), implies that

$$\begin{aligned}
&C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\mathbb{E}\left[\hat{\varepsilon}_t\right] + C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t x_t + B_t u_t\right) \\
&+ \sum_{\tau=t+1}^{T-1} C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau - \frac{1}{\gamma}\tilde{\Sigma}_t^{-1}\mathbb{E}\left[\hat{\varepsilon}_t - \varepsilon_t\right] = 0.
\end{aligned} \tag{2.54}$$

By $\mathbb{E}\left[\varepsilon_t\right] = \tilde{\mu}_t$, and solving for $\mathbb{E}\left[\hat{\varepsilon}_t\right]$ in (2.54), we get that

$$\mathbb{E}\left[\hat{\varepsilon}_t\right] = \tilde{A}_t x_t + \tilde{B}_t u_t + \gamma\sum_{\tau=t+1}^{T-1}\tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau + \tilde{C}_t\tilde{\mu}_t. \tag{2.55}$$

Applying now the optimality condition on u_t in the outer problem, we have by Danskin's theorem (Danskin, 1966) that the derivative of the minimum problem in minimax problem (2.14) equals to the derivative of the minimum problem fixing $\hat{P}_t$ to be maximum point first

$$\begin{aligned}
&B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t u_t + B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t x_t + C_t\mathbb{E}\left[\hat{\varepsilon}_t\right]\right) \\
&+ R_t u_t + \sum_{\tau=t+1}^{T-1} B_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau = 0,
\end{aligned} \tag{2.56}$$

and, consequently, by solving for u_t and using (2.42), that

$$
\begin{aligned}
u_t = & -\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t+R_t+B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_t\right)^{-1}B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\Big(A_t \\
& +C_t\tilde{A}_t\Big)x_t-\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t+R_t+B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_t\right)^{-1}\Bigg(B_t^T\big(Q_{t+1} \\
& +\Xi_{t+1}\big)C_t\left(\gamma\sum_{\tau=t+1}^{T-1}\tilde{C}_t\tilde{\Sigma}_tC_t\Xi_{t+1,\tau}\tilde{\mu}_\tau+\tilde{C}_t\tilde{\mu}_t\right)+\sum_{\tau=t+1}^{T-1}B_t^T\Xi_{t+1,\tau}\tilde{\mu}_\tau\Bigg) \\
= & \; U_tx_t+\sum_{\tau=t}^{T-1}U_{t,\tau}\tilde{\mu}_\tau.
\end{aligned}
\tag{2.57}
$$

Plugging in u_t and $\hat{\varepsilon}_{t,i}$ from (2.53) back into $V_t(x_t)$, we have

$$
\begin{aligned}
&V_t(x_t) \\
&= \mathbb{E}\Bigg[\Bigg(A_t x_t + B_t\left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right) \\
&\quad + C_t\left(\tilde{A}_t x_t + \tilde{B}_t\left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right) + \gamma \sum_{\tau=t+1}^{T-1} \tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau + \tilde{C}_t\varepsilon_t\right)\Bigg)^T (Q_{t+1} \\
&\quad + \Xi_{t+1})\Bigg(A_t x_t + B_t\left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right) \\
&\quad + C_t\left(\tilde{A}_t x_t + \tilde{B}_t\left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right) + \gamma \sum_{\tau=t+1}^{T-1} \tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau + \tilde{C}_t\varepsilon_t\right)\Bigg) \\
&\quad + \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right)^T R_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right) \\
&\quad - \frac{1}{\gamma}\left(\tilde{A}_t x_t + \tilde{B}_t\left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right) + \gamma \sum_{\tau=t+1}^{T-1} \tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau + \tilde{C}_t\varepsilon_t\right)^T \tilde{\Sigma}_t^{-1}\Bigg(\tilde{A}_t x_t \\
&\quad + \tilde{B}_t\left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\tilde{\mu}_t\right) + \gamma \sum_{\tau=t+1}^{T-1} \tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\tilde{\mu}_\tau + \tilde{C}_t\varepsilon_t\Bigg)\Bigg] + z_{t+1} \\
&= x_t^T\Big(\left(A_t + B_t U_t + C_t\tilde{A}_t + C_t\tilde{B}_t U_t\right)^T (Q_{t+1} + \Xi_{t+1})\left(A_t + B_t U_t + C_t\tilde{A}_t + C_t\tilde{B}_t U_t\right) \\
&\quad + U_t^T R_t U_t - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_t U_t\right)^T \tilde{\Sigma}_t^{-1}\left(\tilde{A}_t + \tilde{B}_t U_t\right)\Big) x_t \\
&\quad + 2x_t^T\Big(\left(A_t + B_t U_t + C_t\tilde{A}_t + C_t\tilde{B}_t U_t\right)(Q_{t+1} + \Xi_{t+1})\left(B_t U_{t,t} + C_t\tilde{B}_t U_{t,t} + C_t\tilde{C}_t\right) \\
&\quad + U_t^T R_t U_{t,t} - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_t U_t\right)\tilde{\Sigma}_t^{-1}\tilde{C}_t\Big)\tilde{\mu}_t + 2\sum_{\tau=t+1}^{T-1} x_t^T\Big(\Big(A_t + B_t U_t + C_t\tilde{A}_t \\
&\quad + C_t\tilde{B}_t U_t\Big)^T (Q_{t+1} + \Xi_{t+1})\left(B_t U_{t,\tau} + C_t\tilde{B}_t U_{t,\tau} + \gamma C_t\tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\right) + U_t^T R_t U_{t,\tau} \\
&\quad - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_t U_t\right)^T \tilde{\Sigma}_t^{-1}\left(\tilde{B}_t U_{t,\tau} + \gamma\tilde{\Sigma}_t^{-1}\tilde{C}_t\tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau}\right)\Big)\tilde{\mu}_\tau + z_t \\
&= x_t^T \Xi_t x_t + 2\sum_{\tau=t}^{T-1} x_t^T \Xi_{t,\tau}\tilde{\mu}_\tau + z_t.
\end{aligned}
\tag{2.58}
$$

This completes the proof.

∎

Lemma 2.4. *For any random variable X following a distribution P which has density $dP(x)$, the maximum point $\tilde{P}$ of the problem*

$$\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}\left[X\right] - \frac{1}{\gamma} KL(\tilde{P}\|P) \tag{2.59}$$

also has same support with P if the maximum in (2.59) *exists, and its density satisfies the condition*

$$d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{2.60}$$

Consequently, we have the identity

$$\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}\left[X\right] - \frac{1}{\gamma} KL(\tilde{P}\|P) = \int x d\tilde{P}(x) - \frac{1}{\gamma} \log\left(\frac{d\tilde{P}(x)}{dP(x)}\right) d\tilde{P}(x) = \frac{1}{\gamma}\log \mathbb{E}_P\left[e^{\gamma X}\right], \tag{2.61}$$

where $\mathbb{E}_P\left[e^{\gamma X}\right]$ is the normalized constant for the $d\tilde{P}(x)$ in (2.60)*.*

Proof. Consider the function

$$\log \mathbb{E}_P\left[e^{\gamma X}\right] = \log \int e^{\gamma x} dP(x). \tag{2.62}$$

For any distributions $\tilde{P}(x)$ that has density and same support with $P(x)$,

$$\begin{aligned} \log \int e^{\gamma x} dP(x) &= \log \int e^{\gamma x}\left(\frac{dP(x)}{d\tilde{P}(x)}\right) d\tilde{P}(x) \\ &\geq \int \left(\gamma x d\tilde{P}(x) + \log\left(\frac{dP(x)}{d\tilde{P}(x)}\right) d\tilde{P}(x)\right), \end{aligned} \tag{2.63}$$

where the last inequality comes from Jensen's inequality in probabilistic setting (Durrett, 2019). Divide γ on the both hand side of (2.63), we have

$$\frac{1}{\gamma}\log \mathbb{E}_P\left[e^{\gamma X}\right] \geq \mathbb{E}_{\tilde{P}}\left[x\right] - \frac{1}{\gamma} KL(\tilde{P}\|P). \tag{2.64}$$

Equality can be attained from Jensen's inequality with the condition

$$d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{2.65}$$

Note that the distribution constructed from (2.65) always exists if (2.62) is not infinity. Thus if (2.62) is not infinite, the maximum of (2.59) exists.

Asssume now (2.62) is infinite. By constructing $\tilde{P}(x)$, we get

$$d\tilde{P}_R(x) \sim dP(x) \times e^{\gamma x} \times I(\|x\| \leq R), \tag{2.66}$$

where $R \in \mathbb{R}$, we have $\tilde{P}_R$ always exists and feasible in (2.59). Denote C_R to be the normalized constant such that

$$d\tilde{P}_R(x) = \frac{1}{C_R} dP(x) \times e^{\gamma x} \times I(\|x\| \leq R). \tag{2.67}$$

Since (2.62) is infinite, we must have that

$$\lim_{R \to \infty} C_R = \lim_{R \to \infty} \int e^{\gamma x} \times I(\|x\| \leq R) dP(x) = \infty$$

Let $R \to \infty$, we have

$$\lim_{R \to \infty} \mathbb{E}_{\tilde{P}_R}[X] - \frac{1}{\gamma} KL(\tilde{P}_R \| P) = \lim_{R \to \infty} \int x d\tilde{P}_R(x) - \frac{1}{\gamma} \log\left(\frac{d\tilde{P}_R(x)}{dP(x)}\right) d\tilde{P}_R(x) = \lim_{R \to \infty} \int -\frac{1}{\gamma} \log\left(\frac{d\tilde{P}_R(x)}{e^{\gamma x} dP(x)}\right) d\tilde{P}_R(x) = \lim_{R \to \infty} \int_{\|x\| \leq R} \frac{1}{\gamma} \frac{\log C_R}{C_R} e^{\gamma x} dP(x) = \lim_{R \to \infty} \frac{1}{\gamma} \log C_R = \infty. \tag{2.68}$$

From (2.68) and (2.64) we have that the existence of the maximum of (2.59) is equivalent to the finite property of (2.62). When finiteness of either occurs, we can invoke (2.64) and (2.65) to claim our conclusion. This completes the proof.

□

Lemma 2.5. *Consider the KL-divergence distributionally robust control problem with Gaussian noise,* $P_t = N(\tilde{\mu}_t, \tilde{\Sigma}_t)$,

$$V_t(x_t) = \min_{u_t} \max_{\tilde{P}_t, \tilde{\varepsilon}_t \sim \tilde{P}_t} \left\{ \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t - \frac{2}{\gamma} KL(\tilde{P}_t \| P_t) + V(x_{t+1}) \right] \right\}, \tag{2.69}$$

where $x_{t+1} = A_t x_t + B_t u_t + C_t \tilde{\varepsilon}_t$, *and*

$$V_T(x_T) = 0, \tag{2.70}$$

for any point $x_T \in \mathbb{R}^{d_1}$. *Then* $\forall t < T$, *we that the optimal control solution of* (2.69) *satisfies* $u_t = U_t x_t + \sum_{\tau=t}^{T_1} U_{t,\tau}\tilde{\mu}_t$ *where* $U_t \in \mathbb{R}^{d_2,d_1}, U_{t,\tau} \in \mathbb{R}^{d_2,d_3}$ *and*

$$V_t(x_t) = x_t^T \Xi_t x_t + 2\sum_{\tau=t}^{T_1} x_t^T \Xi_{t,\tau}\tilde{\mu}_t + z_t, \tag{2.71}$$

where $\Xi_t \in \mathbb{R}^{d_1,d_1}, \Xi_{t,\tau} \in \mathbb{R}^{d_1,d_3}, z_t \in \mathbb{R}$. *Here,* z_t *is a constant that depends on* t *but not on to* x_t *and thus does not affect solutions of optimization problems for time indices smaller than* t. *For* $t < T$, U_t *and* $U_{t,t}$ *are defined as*

$$\begin{aligned} U_t = -\left(B_t^T(Q_{t+1}+\Xi_{t+1})B_t + R_t + B_t^T(Q_{t+1}+\Xi_{t+1})C_t\tilde{B}_t\right)^{-1} B_t^T (Q_{t+1} \\ + \Xi_{t+1})\left(A_t + C_t\tilde{A}_t\right), \end{aligned} \tag{2.72}$$

$$U_{t,t} = -\left(B_t^T(Q_{t+1}+\Xi_{t+1})B_t + R_t + B_t^T(Q_{t+1}+\Xi_{t+1})C_t\tilde{B}_t\right)^{-1} B_t^T(Q_{t+1}+\Xi_{t+1})C_t\tilde{C}_t, \tag{2.73}$$

and for $t > T$,

$$\begin{aligned} U_{t,\tau} = -\left(B_t^T(Q_{t+1}+\Xi_{t+1})B_t + R_t + B_t^T(Q_{t+1}+\Xi_{t+1})C_t\tilde{B}_t\right)^{-1} B_t^T (Q_{t+1} \\ + \Xi_{t+1})\left(\gamma C_t\tilde{C}_t\tilde{\Sigma}_t C_t\Xi_{t+1,\tau} + B_t^T\Xi_{t+1,\tau}\right). \end{aligned} \tag{2.74}$$

Moreover, Ξ_t *and* $\Xi_{t,t}$ *are defined as*

$$\begin{aligned} \Xi_t = &\left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right)^T (Q_{t+1}+\Xi_{t+1})\left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right) \\ &+ U_t^T R_t U_t - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_tU_t\right)^T \tilde{\Sigma}_t^{-1}\left(\tilde{A}_t + \tilde{B}_tU_t\right), \end{aligned} \tag{2.75}$$

$$\begin{aligned} \Xi_{t,t} = &\left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right)(Q_{t+1}+\Xi_{t+1})\left(B_tU_{t,t} + C_t\tilde{B}_tU_{t,t} + C_t\tilde{C}_t\right) \\ &+ U_t^T R_t U_{t,t} - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_tU_t\right)\tilde{\Sigma}_t^{-1}\tilde{C}_t \end{aligned} \tag{2.76}$$

and for $\tau > t$,

$$\Xi_{t,\tau} = \left(A_t + B_tU_t + C_t\tilde{A}_t + C_t\tilde{B}_tU_t\right)^T (Q_{t+1} + \Xi_{t+1}) \left(B_tU_{t,\tau} + C_t\tilde{B}_tU_{t,\tau} + \gamma C_t\tilde{C}_t\tilde{\Sigma}_tC_t^T\Xi_{t+1,\tau}\right) + U_t^TR_tU_{t,\tau} - \frac{1}{\gamma}\left(\tilde{A}_t + \tilde{B}_tU_t\right)^T \tilde{\Sigma}_t^{-1}\left(\tilde{B}_tU_{t,\tau} + \gamma\tilde{\Sigma}_t^{-1}\tilde{C}_t\tilde{\Sigma}_tC_t^T\Xi_{t+1,\tau}\right), \tag{2.77}$$

where

$$\tilde{A}_t = \gamma\left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} C_t^T (Q_{t+1} + \Xi_{t+1}) A_t, \tag{2.78}$$

$$\tilde{B}_t = \gamma\left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} C_t^T (Q_{t+1} + \Xi_{t+1}) B_t, \tag{2.79}$$

$$\tilde{C}_t = \left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} \tilde{\Sigma}_t^{-1}. \tag{2.80}$$

Proof. This can be shown by induction. Firstly it is true for $t = T$. Since $V_T(x_T) = 0$, we only need to choose relevant matrices to be zero. Assume it holds for $t = m+1$ where $m \in \mathbb{N}$, then for $t = m$, by the dynamic $x_{t+1} = A_tx_t + B_tu_t + C_t\hat{\varepsilon}_t$ and the induction, we have

$$V_t(x_t) = \min_{u_t} \max_{\tilde{P}_t, \tilde{\varepsilon}_t \sim \tilde{P}_t} \left\{ \mathbb{E}\left[x_{t+1}^T (Q_{t+1} + \Xi_{t+1}) x_{t+1} + u_t^TR_tu_t + 2\sum_{\tau=t+1}^{T-1} x_{t+1}^T\Xi_{t+1,\tau}\tilde{\mu}_\tau - \frac{2}{\gamma}KL\left(\tilde{P}_t \| P_t\right)\right]\right\} + z_{t+1}. \tag{2.81}$$

Fixing u_t first, then by Lemma 2.4, the maximum point of the inner problem $\tilde{P}_t$ still belongs to Gaussian family and

$$d\tilde{P}_t \sim dP_t \times \exp\left\{\frac{\gamma}{2}\left((A_tx_t + B_tu_t + C_t\varepsilon_t)^T (Q_{t+1} + \Xi_{t+1}) (A_tx_t + B_tu_t + C_t\varepsilon_t) + 2\sum_{\tau=t+1}^{T-1} (A_tx_t + B_tu_t + C_t\varepsilon_t)^T \Xi_{t+1,\tau}\tilde{\mu}_\tau\right)\right\}. \tag{2.82}$$

Since $P_t = N(\tilde{\mu}_t, \tilde{\Sigma}_t)$, we have

$$\tilde{P}_t = N\left(\left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \left(\tilde{\Sigma}_t^{-1} \tilde{\mu}_\tau + \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right)\left(A_t x_t + B_t u_t\right) + \gamma \sum_{\tau=t+1}^{T-1} C_t^T \Xi_{t+1,\tau} \tilde{\mu}_\tau\right), \left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1}\right). \tag{2.83}$$

We now apply the optimality condition to the outer minimization problem of (2.81), and by Danskin's theorem the gradient w.r.t. u_t of the outer problem is the gradient w.r.t u_t of the original problem when $\tilde{P}_t$ fixed to the maximum point. We then have

$$\begin{aligned} &B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t u_t + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right)\left(A_t x_t + C_t \mathbb{E}\left[\tilde{\varepsilon}_t\right]\right) \\ &+ R_t u_t + \sum_{\tau=t+1}^{T-1} B_t^T \Xi_{t+1,\tau} \tilde{\mu}_\tau = 0. \end{aligned} \tag{2.84}$$

Solving for u_t from this equation we get

$$\begin{aligned} u_t = &-\left(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t + R_t + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_t\right)^{-1} B_t^T \left(Q_{t+1} + \Xi_{t+1}\right)\left(A_t + C_t \tilde{A}_t\right) x_t \\ &- \left(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t + R_t + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_t\right)^{-1} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\gamma \sum_{\tau=t+1}^{T-1} \tilde{C}_t \tilde{\Sigma}_t C_t \Xi_{t+1,\tau} \tilde{\mu}_\tau + \tilde{C}_t \tilde{\mu}_t\right) + \sum_{\tau=t+1}^{T-1} B_t^T \Xi_{t+1,\tau} \tilde{\mu}_\tau\right) \\ = &\, U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_\tau, \end{aligned} \tag{2.85}$$

where we used the definitions of U_t, (2.73), and $U_{t,\tau}$, (2.73)–(2.74).

The KL-divergence between two Gaussian $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ in $\mathbb{R}^{d_3}$ (Duchi, 2007) is

$$KL\left(N(\mu_1, \Sigma_1) \| N(\mu_2, \Sigma_2)\right) = \frac{1}{2}\left(\log \frac{|\Sigma_2|}{|\Sigma_1|} - d_3 + \left(\mu_1 - \mu_2\right)^T \Sigma_2^{-1} \left(\mu_1 - \mu_2\right) + tr\{\Sigma_2^{-1} \Sigma_1\}\right). \tag{2.86}$$

Thus, using the optimal control, (2.85), and the expression of the KL divergence, (2.86), in the objective function (2.81), we get

$$
\begin{aligned}
V_t(x_t) = \mathbb{E} \Bigg[& \left(A_t x_t + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_t \right) + C_t \tilde{\varepsilon}_t \right)^T (Q_{t+1} + \Xi_{t+1}) \Bigg(A_t x_t \\
& + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_t \right) + C_t \tilde{\varepsilon}_t \Bigg) \\
& + \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_t \right)^T R_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_t \right) \\
& + 2 \sum_{\tau=t+1}^{T-1} \left(A_t x_t + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_t \right) + C_t \tilde{\varepsilon}_t \right)^T \Xi_{t+1,\tau} \tilde{\mu}_\tau \\
& - \frac{1}{\gamma} \Bigg(\left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \right)^{-1} \Bigg(\tilde{\Sigma}_t^{-1} \tilde{\mu}_t + \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) \\
& \left(A_t + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_t \right) \right) + \gamma \sum_{\tau=t+1}^{T-1} C_t^T \Xi_{t+1,\tau} \tilde{\mu}_\tau \Bigg) - \tilde{\mu}_t \Bigg)^T \tilde{\Sigma}_t^{-1} \\
& \Bigg(\left(\tilde{\Sigma}_t^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \right)^{-1} \Bigg(\tilde{\Sigma}_t^{-1} \tilde{\mu}_t + \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) \Bigg(A_t + B_t \Bigg(U_t x_t \\
& + \sum_{\tau=t}^{T-1} U_{t,\tau} \tilde{\mu}_t \Bigg) \Bigg) + \gamma \sum_{\tau=t+1}^{T-1} C_t^T \Xi_{t+1,\tau} \tilde{\mu}_\tau \Bigg) - \tilde{\mu}_t \Bigg) \Bigg] + Constant + z_{t+1}
\end{aligned}
\tag{2.87}
$$

$$
\begin{aligned}
&= x_t^T \Bigg(\left(A_t + B_t U_t + C_t \tilde{A}_t + C_t \tilde{B}_t U_t \right)^T (Q_{t+1} + \Xi_{t+1}) \left(A_t + B_t U_t + C_t \tilde{A}_t + C_t \tilde{B}_t U_t \right) \\
&\qquad + U_t^T R_t U_t - \frac{1}{\gamma} \left(\tilde{A}_t + \tilde{B}_t U_t \right)^T \tilde{\Sigma}_t^{-1} \left(\tilde{A}_t + \tilde{B}_t U_t \right) \Bigg) x_t \\
&\quad + 2 x_t^T \Bigg(\left(A_t + B_t U_t + C_t \tilde{A}_t + C_t \tilde{B}_t U_t \right) (Q_{t+1} + \Xi_{t+1}) \left(B_t U_{t,t} + C_t \tilde{B}_t U_{t,t} + C_t \tilde{C}_t \right) \\
&\qquad + U_t^T R_t U_{t,t} - \frac{1}{\gamma} \left(\tilde{A}_t + \tilde{B}_t U_t \right) \tilde{\Sigma}_t^{-1} \tilde{C}_t \Bigg) \tilde{\mu}_t + 2 \sum_{\tau=t+1}^{T-1} x_t^T \Bigg(\Big(A_t + B_t U_t + C_t \tilde{A}_t \\
&\qquad + C_t \tilde{B}_t U_t \Big)^T (Q_{t+1} + \Xi_{t+1}) \left(B_t U_{t,\tau} + C_t \tilde{B}_t U_{t,\tau} + \gamma C_t \tilde{C}_t \tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau} \right) + U_t^T R_t U_{t,\tau} \\
&\qquad - \frac{1}{\gamma} \left(\tilde{A}_t + \tilde{B}_t U_t \right)^T \tilde{\Sigma}_t^{-1} \left(\tilde{B}_t U_{t,\tau} + \gamma \tilde{\Sigma}_t^{-1} \tilde{C}_t \tilde{\Sigma}_t C_t^T \Xi_{t+1,\tau} \right) \Bigg) \tilde{\mu}_\tau + z_t \\
&= x_t^T \Xi_t x_t + 2 \sum_{\tau=t}^{T-1} x_t^T \Xi_{t,\tau} \tilde{\mu}_\tau + z_t,
\end{aligned}
$$

where Ξ_t is defined in (2.75), $\Xi_{t,\tau}$ is defined in (2.76)–(2.77), $\tilde{A}_t$ is defined in (2.78), $\tilde{C}_t$ is defined in (2.80) .

This completes the proof.

□

Remark 2.5.1. *By Theorem 2.3 and Lemma 2.5, with any given dataset, if we use the (Jacobson, 1973)'s controller with fitted mean and covariance, then we get exactly the same controllers as Wasserstein DRC problem.*

Figure 2.1 explains difference between those controllers. To see the yellow curve is the contour fitting for standard Wasserstein distance, just plug $\tilde{\Sigma}_t = \mathbf{I}_{d_3}$ into Theorem 2.3 and compare it with Lemma 2.5.

2.4 Unified Understanding of Distributional Robustness

2.4.1 Distributionally Robust Control

We want to give a unified framework of distributionally robust control here to have a better understanding of the essence of robust control. For any functions f_t and g_t, consider problems

$$\begin{aligned} &\min_{u_t(x_t)} \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_{t+1}\left(x_{t+1}, u_t\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} h_t(\tilde{P}_t, P_t) \\ &s.t. x_{t+1} = g_t\left(x_t, u_t, \varepsilon_t\right), \end{aligned} \tag{2.88}$$

where h_t are any similarity measures between two distributions. It can be Kullback-Leibler divergences or Wasserstein distance. We generalize Wasserstein distance to a more general setting with non-negative functions $c_t(\xi, \zeta)$

$$W_t(\tilde{P}_t, P_t) = \min_{\Gamma = \Pi(\tilde{P}_t, P_t)} \int c_t(\xi, \zeta) d\Gamma(\xi, \zeta) \tag{2.89}$$

i.e. $\forall \zeta \in supp\left(P_t\right), c_t\left(\zeta, \zeta\right) = 0$. The choice of c_t decides what sort of robustness we will have in the problem. We want to point out that, as we indicated in the previous section, §2.3, several paradigms of robust and/or risk-averse control, e.g. Wasserstein and Kullback Leibler, can be unified by (2.88) and (2.89) with properly designed c_t in the setting of control theory. Moroever, since other robust control approaches (e.g. H_∞) can be shown to be equivalent to Wasserstein they can also be incorporated in this framework (Kim and Yang, 2020).

We allow γ to be negative to accommodate circumstances with some fraction of erroneous data.When γ is negative, the minimax problem becomes a minimum problem with two stages, that is,

$$\begin{aligned} &\min_{u_t(x_t)} \min_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_{t+1}\left(x_{t+1}, u_t\right)\right] + \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t) \\ &s.t. x_{t+1} = g_t\left(x_t, u_t, \varepsilon_t\right). \end{aligned} \tag{2.90}$$

It has been used in many applications outside the control theory. For example, in image denoising, the objective to repair an image y is

$$\min_x \|y - x\|_2^2 + \eta R(x), \tag{2.91}$$

where $R(x)$ is the negative log-likelihood of a guessed image x under the image generative model, which is in fact an agent for Kullback-Leibler divergence (Fan et al., 2019).

With this insight, we give a unified framework for distributionally robust control using a generalized Wasserstein distance, which encompasses risk-averse KL, H_∞ formulations, and standard Wasserstein settings:

$$\begin{aligned} \min_{u_t(x_t)} \frac{1}{\gamma} \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \gamma \mathbb{E} \left[\sum_{t=0}^{T-1} f_{t+1}(x_{t+1}, u_t) \right] - \sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t) \\ s.t. x_{t+1} = g_t(x_t, u_t, \varepsilon_t). \end{aligned} \tag{2.92}$$

This in fact defined a monotonically non-decreasing function of γ whose domain is $\mathbb{R}$ by providing its definition at $\gamma = 0$, as follows

$$F_\gamma(u_t) = \begin{cases} \mathbb{E} \left[\sum_{t=0}^{T-1} f_{t+1}(x_{t+1}, u_t) \right] & , \gamma = 0 \\ \frac{1}{\gamma} \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \gamma \mathbb{E} \left[\sum_{t=0}^{T-1} f_{t+1}(x_{t+1}, u_t) \right] - \sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t) & , \gamma \neq 0 \end{cases} \tag{2.93}$$

Though we do not pursue this in this paper, this gives a way to evaluate risk under arbitrary robustness concepts. Since $F_\gamma(u_t)$ is a non-decreasing function, we can use a uniform upper bound for a set of candidate policies $u_t(x_t)$ and bisect to get maximized γ as their risk. Even for general nonlinear problems, which are hard to solve optimal policy, we can easily verify risks, since it is a concave/convex function of $\tilde{P}_t$ and in many cases strongly concave/convex.

For example, if we have two candidate policies $u_t^1(x_t)$ and $u_t^2(x_t)$, we will wonder whether they are reliable to make certain amounts of profits. Assume we want to evaluate their risk to make 1,000,000 profits, we can perform bisect to get γ_1 and γ_2 for $u_t^1(x_t)$ and $u_t^2(x_t)$ correspondingly such that

$$F_{\gamma_1}(u_t^1(x_t)) = F_{\gamma_2}(u_t^2(x_t)) = 1000000. \tag{2.94}$$

If we get $\gamma_1 > 0 > \gamma_2$, then it tells us that even if there are some erroneous data points in the dataset, $u_t^1(x_t)$ still can make this amount of profit. And if we want $u_t^2(x_t)$ also make this amount of profit, then it would be possible only if there are some erroneous data points in the dataset. The order between γ_1 and γ_2 tells which policy is more reliable at this profit level. We can then define the risk of a policy $u_t(x_t)$ at a certain profit level L by

$$\begin{aligned} &\max_{\gamma} \gamma \\ &s.t. F_{\gamma}(u_t(x_t)) \leq L. \end{aligned} \tag{2.95}$$

(2.95) doesn't require the continuity of $F_{\gamma}(u_t(x_t))$ on γ and thus can be applied with general bisection algorithms. In the case $F_{\gamma}(u_t(x_t))$ lacks of continuity on γ, an alternative problem may need to be considered at the same time

$$\begin{aligned} &\min_{\gamma} \gamma \\ &s.t. F_{\gamma}(u_t(x_t)) \geq L. \end{aligned} \tag{2.96}$$

Notice that this quantification of risk is also close to the discussion of convex measures of risk (Frittelli and Gianin, 2005) in Economics. We in fact uses $\frac{1}{\gamma}W(\tilde{P}, P)$ as the penalty function in their dual representation. What's different is that we define γ and generalized Wasserstein distance separately here and thus can quantify risk with maximized γ.

A metaphor of the difference between Wasserstein distance and KL-divergence is that Wasserstein distance describes the microstates discrepancy of thermodynamics systems, while KL-divergence describes the macrostates discrepancy of thermodynamics systems (Dechant and Sakurai, 2019). It is natural for particles in microstates to get anisotropic forces, and thus we define $c_t(x, y)$ asymmetrically, in a physical sense, with positions y and the cost of energy to move y to x. The minimization inside Wasserstein distance is then simply thing like the principle of least action in analytical mechanics (Lagrange, 1853). And thus, philosophically, we can, of course, describe the macrostates with the microstates.

The following result invokes strong duality for an extension of a case presented in (Gao and Kleywegt, 2016). That is,

Lemma 2.6. *Let Ξ be a set and $\mathcal{P}(\Xi)$ be the set of Borel probability measures on Ξ. Given $\nu \in \mathcal{P}(\Xi)$ and $\Psi \in L^1(\nu)$, for any $\theta > 0$ and $p \in [1, \infty)$, define W_p to be standard Wasserstein-p distance, the strong duality holds for the primal*

$$\sup_{\mu \in \mathcal{P}(\Xi)} \left\{ \int_{\Xi} \Psi(\xi) : W_p(\mu, \nu) \le \theta \right\} \tag{2.97}$$

and the dual

$$\inf_{\lambda \ge 0} \left\{ \lambda \theta^p - \int_{\Xi} \inf_{\xi \in \Xi} [\lambda d^p(\xi, \zeta) - \Psi(\xi)] \nu(d\zeta) \right\} \tag{2.98}$$

The strong duality still holds if W_p is replaced by generalized Wasserstein distance defined in (2.89)*, that is, the primal*

$$\sup_{\mu \in \mathcal{P}(\Xi)} \left\{ \int_{\Xi} \Psi(\xi) : W(\mu, \nu) \le \theta \right\} \tag{2.99}$$

and the dual

$$\inf_{\lambda \ge 0} \left\{ \lambda \theta - \int_{\Xi} \inf_{\xi \in \Xi} [\lambda c(\xi, \zeta) - \Psi(\xi)] \nu(d\zeta) \right\} \tag{2.100}$$

Proof. The strong duality between (2.97) and (2.98) is by the Theorem 1 and Proposition 2 of (Gao and Kleywegt, 2016). The strong duality between (2.97) and (2.98) is by applying Remark 2 of (Gao and Kleywegt, 2016) to generalized Wasserstein distance. □

Theorem 2.7. *Choose h_t to be W_t from* (2.89) *and P_t to be empirical measures on n data points, and assume that $\gamma > 0$. Then, the distributionally robust control problem* (2.88) *is equivalent to*

$$\begin{aligned} \min_{u_t(x_t)} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t\left(x_{t+1}, u_t\right)\right] - \frac{1}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} c_t\left(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}\right) \\ s.t. x_{t+1} = g_t\left(x_t, u_t, \hat{\varepsilon}_t\right) \end{aligned} \tag{2.101}$$

where $\hat{P}_t$ are empirical measures on n data points $\hat{\varepsilon}_{t,i}$ and $\hat{\varepsilon}_{t,i}$ can be arbitrary points in $\mathbb{R}^{d_3}$.

Proof. For a given optimal control function sequence, $u_t(x_t)$, $t = 0, 1, \dots T-1$, assume that the distributions that are solutions of the inner minimization problem, $\left\{\tilde{P}_t^0\right\}$ of (2.101), satisfy $W_t(\tilde{P}_t^1, P_t) = \omega_t$, for $t = 0, 1 \dots, T-1$. We claim that

$$\left\{\tilde{P}_t^0\right\}_{t=0,1,\dots,T-1} \in arg \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \leq \omega_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right]. \tag{2.102}$$

We will prove this claim by contrapositive. If (2.102) does not hold, then another distribution sequence $\left\{\tilde{P}_t^1\right\}_{t=0,1,\dots,T-1}$ is a maximum point of (2.102), and in particular it is a feasible point with a strictly better objective.

That is, it satisfies $W_t(\tilde{P}_t^1, P_t) \leq \omega_t$, for $t = 0, 1, T-1$, and

$$\mathbb{E}_{\varepsilon_t \sim \tilde{P}_t^1}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] > \mathbb{E}_{\varepsilon_t \sim \tilde{P}_t^0}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t).\right] \tag{2.103}$$

Since, due to the feasibility, we must have that

$$-\frac{1}{\gamma}\sum_{t=0}^{T-1} W_t(\tilde{P}_t^1, P_t) \geq -\frac{1}{\gamma}\sum_{t=0}^{T-1} \omega_t = -\frac{1}{\gamma}\sum_{t=0}^{T-1} W_t(\tilde{P}_t^1, P_t), \tag{2.104}$$

adding (2.103) and (2.104), we get

$$\begin{aligned} &\mathbb{E}_{\varepsilon_t \sim \tilde{P}_t^1} \left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\tilde{P}_t^1, P_t) \\ &> \mathbb{E}_{\varepsilon_t \sim \tilde{P}_t^0} \left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\tilde{P}_t^0, P_t) \end{aligned} \tag{2.105}$$

which contradicts the assumption that $\left\{\tilde{P}_t^0\right\}$, for $t = 0, 1, \ldots, T-1$, is a solution of the inner maximization problem at given $u_t(x_t)$, $t = 0, 1, \ldots T-1$. We thus conclude that (2.102) holds.

Then we have that

$$\begin{aligned} &\max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E} \left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t) \\ &= \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \leq \omega_t} \mathbb{E} \left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t) \\ &= \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \leq \omega_t} \mathbb{E} \left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} \omega_t. \end{aligned} \tag{2.106}$$

By Lemma 2.6,

$$
\begin{aligned}
&\max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \le \omega_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] \\
&= \min_{\lambda_0 \ge 0} \Bigg\{ \lambda_0 \omega_0 \\
&\quad - \int \min_{\hat{\varepsilon}_0} \Bigg\{ \lambda_0 c_0\left(\hat{\varepsilon}_0, \tilde{\varepsilon}_0\right) - \max_{\varepsilon_0 = \hat{\varepsilon}_0, \forall t \ge 1, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \le \omega_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] \Bigg\} dP_0\left(\tilde{\varepsilon}_0\right) \Bigg\} \\
&= \min_{\lambda_0 \ge 0} \Bigg\{ \lambda_0 \omega_0 \\
&\quad - \frac{1}{n} \sum_{i=1}^{n} \min_{\hat{\varepsilon}_{0,i}} \Bigg\{ \lambda_0 c_0\left(\hat{\varepsilon}_{0,i}, \varepsilon_{0,i}\right) - \max_{\varepsilon_0 = \hat{\varepsilon}_{0,i}, \forall t \ge 1, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \le \omega_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] \Bigg\} \Bigg\} \\
&= \min_{\lambda_0 \ge 0} \max_{\hat{\varepsilon}_{0,i}} \Bigg\{ \lambda_0 \omega_0 \\
&\quad - \frac{1}{n} \sum_{i=1}^{n} \Bigg(\lambda_0 c_0\left(\hat{\varepsilon}_{0,i}, \varepsilon_{0,i}\right) - \max_{\tilde{P}_t, \varepsilon_0 = \hat{\varepsilon}_{0,i}, \forall t \ge 1, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \le \omega_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] \Bigg) \Bigg\} \\
&= \min_{\lambda_0 \ge 0} \max_{\hat{\varepsilon}_{0,i}} \Bigg\{ \frac{1}{n} \sum_{i=1}^{n} \Bigg(\max_{\tilde{P}_t, \varepsilon_0 = \hat{\varepsilon}_{0,i}, \forall t \ge 1, \varepsilon_t \sim \tilde{P}_t, W_t(\tilde{P}_t, P_t) \le \omega_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] \Bigg) \\
&\qquad\qquad - \frac{\lambda_0}{n} \sum_{i=1}^{n} c_0\left(\hat{\varepsilon}_{0,i}, \varepsilon_{0,i}\right) + \lambda_0 \omega_0 \Bigg\} \\
&= \dots \\
&= \min_{\lambda_0 \ge 0} \max_{\hat{\varepsilon}_{0,i}} \min_{\lambda_1 \ge 0} \max_{\hat{\varepsilon}_{1,i}} \dots \min_{\lambda_{T-1} \ge 0} \max_{\hat{\varepsilon}_{T-1,i}} \Bigg\{ \sum_{i_0=1}^{n} \sum_{i_1=1}^{n} \dots \sum_{i_{T-1}=1}^{n} \frac{1}{n^T} \sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \Bigg|_{\varepsilon_t = \hat{\varepsilon}_{t, i_t}} \\
&\qquad\qquad - \sum_{t=0}^{T-1} \frac{\lambda_t}{n} \sum_{i=1}^{n} c_t\left(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}\right) + \sum_{t=0}^{T-1} \lambda_t \omega_t \Bigg\}
\end{aligned}
\tag{2.107}
$$

Since $\lambda_{T-1} = \frac{1}{\gamma}$ is feasible in the inner problem of (2.107), we have

$$
\begin{aligned}
&\min_{\lambda_{T-1}\geq 0}\max_{\hat{\varepsilon}_{T-1,i}}\left\{\sum_{i_0=1}^{n}\sum_{i_1=1}^{n}\cdots\sum_{i_{T-1}=1}^{n}\frac{1}{n^T}\sum_{t=0}^{T-1}f_t(x_{t+1},u_t)\bigg|_{\varepsilon_t=\hat{\varepsilon}_{t,i_t}}-\sum_{t=0}^{T-1}\frac{\lambda_t}{n}\sum_{i=1}^{n}c_t\left(\hat{\varepsilon}_{t,i},\varepsilon_{t,i}\right)\right.\\
&\qquad\left.+\sum_{t=0}^{T-1}\lambda_t\omega_t\right\}\leq\max_{\hat{\varepsilon}_{T-1,i}}\left\{\sum_{i_0=1}^{n}\sum_{i_1=1}^{n}\cdots\sum_{i_{T-1}=1}^{n}\frac{1}{n^T}\sum_{t=0}^{T-1}f_t(x_{t+1},u_t)\bigg|_{\varepsilon_t=\hat{\varepsilon}_{t,i_t}}\right.\\
&\qquad\left.-\sum_{t=0}^{T-2}\frac{\lambda_t}{n}\sum_{i=1}^{n}c_t\left(\hat{\varepsilon}_{t,i},\varepsilon_{t,i}\right)-\frac{1}{n\gamma}\sum_{i=1}^{n}c_{T-1}\left(\hat{\varepsilon}_{T-1,i},\varepsilon_{T-1,i}\right)+\sum_{t=0}^{T-2}\lambda_t\omega_t+\frac{1}{\gamma}\omega_{T-1}\right\}
\end{aligned}
\tag{2.108}
$$

Similarly, we can do that for λ_{T-2}, and we have

$$
\begin{aligned}
&\min_{\lambda_{T-2}\geq 0}\max_{\hat{\varepsilon}_{T-2,i}}\min_{\lambda_{T-1}\geq 0}\max_{\hat{\varepsilon}_{T-1,i}}\left\{\sum_{i_0=1}^{n}\sum_{i_1=1}^{n}\cdots\sum_{i_{T-1}=1}^{n}\frac{1}{n^T}\sum_{t=0}^{T-1}f_t(x_{t+1},u_t)\bigg|_{\varepsilon_t=\hat{\varepsilon}_{t,i_t}}\right.\\
&\qquad\qquad-\sum_{t=0}^{T-1}\frac{\lambda_t}{n}\sum_{i=1}^{n}c_t\left(\hat{\varepsilon}_{t,i},\varepsilon_{t,i}\right)\\
&\qquad\left.+\sum_{t=0}^{T-1}\lambda_t\omega_t\right\}\leq\min_{\lambda_{T-2}\geq 0}\max_{\hat{\varepsilon}_{T-2,i}}\max_{\hat{\varepsilon}_{T-1,i}}\left\{\sum_{i_0=1}^{n}\sum_{i_1=1}^{n}\cdots\sum_{i_{T-1}=1}^{n}\frac{1}{n^T}\sum_{t=0}^{T-1}f_t(x_{t+1},u_t)\bigg|_{\varepsilon_t=\hat{\varepsilon}_{t,i_t}}\right.\\
&\qquad\left.-\sum_{t=0}^{T-2}\frac{\lambda_t}{n}\sum_{i=1}^{n}c_t\left(\hat{\varepsilon}_{t,i},\varepsilon_{t,i}\right)-\frac{1}{n\gamma}\sum_{i=1}^{n}c_{T-1}\left(\hat{\varepsilon}_{T-1,i},\varepsilon_{T-1,i}\right)+\sum_{t=0}^{T-2}\lambda_t\omega_t+\frac{1}{\gamma}\omega_{T-1}\right\}\\
&\leq\max_{\hat{\varepsilon}_{T-2,i}}\max_{\hat{\varepsilon}_{T-1,i}}\left\{\sum_{i_0=1}^{n}\sum_{i_1=1}^{n}\cdots\sum_{i_{T-1}=1}^{n}\frac{1}{n^T}\sum_{t=0}^{T-1}f_t(x_{t+1},u_t)\bigg|_{\varepsilon_t=\hat{\varepsilon}_{t,i_t}}-\sum_{t=0}^{T-3}\frac{\lambda_t}{n}\sum_{i=1}^{n}c_t\left(\hat{\varepsilon}_{t,i},\varepsilon_{t,i}\right)\right.\\
&\qquad\left.-\frac{1}{n\gamma}\sum_{t=T-2}^{T-1}\sum_{i=1}^{n}c_t\left(\hat{\varepsilon}_{t,i},\varepsilon_{t,i}\right)+\sum_{t=0}^{T-3}\lambda_t\omega_t+\frac{1}{\gamma}\sum_{t=T-2}^{T-1}\omega_t\right\}
\end{aligned}
\tag{2.109}
$$

Apply this from $t = T-1$ to $t = 0$, we have

$$\min_{\lambda_0 \geq 0} \max_{\hat{\varepsilon}_{0,i}} \min_{\lambda_1 \geq 0} \max_{\hat{\varepsilon}_{1,i}} \dots \min_{\lambda_{T-1} \geq 0} \max_{\hat{\varepsilon}_{T-1,i}} \left\{ \sum_{i_0=1}^{n} \sum_{i_1=1}^{n} \dots \sum_{i_{T-1}=1}^{n} \frac{1}{n^T} \sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \Bigg|_{\varepsilon_t = \hat{\varepsilon}_{t,i_t}} \right.$$
$$\left. - \sum_{t=0}^{T-1} \frac{\lambda_t}{n} \sum_{i=1}^{n} c_t\left(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}\right) + \sum_{t=0}^{T-1} \lambda_t \omega_t \right\} \leq \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \left\{ \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] \right.$$
$$\left. - \frac{1}{\gamma n} \sum_{t=0}^{T-1} \sum_{i=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}) \right\} + \frac{1}{\gamma} \sum_{t=0}^{T-1} \omega_t \tag{2.110}$$

Therefore, from (2.110) and (2.107), we obtain that

$$\max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] \leq \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \left\{ \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] \right.$$
$$\left. - \frac{1}{\gamma n} \sum_{t=0}^{T-1} \sum_{i=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}) \right\} + \frac{1}{\gamma} \sum_{t=0}^{T-1} \omega_t \tag{2.111}$$

Now by subtracting the last term from both sides and using (2.106), we get

$$\max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t)$$
$$\leq \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \left\{ \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma n} \sum_{t=0}^{T-1} \sum_{i=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}) \right\}. \tag{2.112}$$

Also notice that for any empirical measures $\hat{P}_t$ and P_t, we have that

$$\frac{1}{n} \sum_{t=0}^{T-1} \sum_{i=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}) \geq \sum_{t=0}^{T-1} W_t\left(\hat{P}_t, P_t\right). \tag{2.113}$$

To justify this, observe that W_t is the optimal transport cost.

$$\sum_{t=0}^{T-1} W_t(\hat{P}_t, P_t) = \sum_{t=0}^{T-1} \min_{\Gamma = \Pi(\hat{P}_t, P_t)} \int c_t(x, y) d\Gamma(x, y). \tag{2.114}$$

We have that $\hat{P}_t$ is a measure over Ξ_t with mass only at $\hat{\varepsilon}_{t,i}$, for $i = 1, 2, \ldots, n$, and P_t is measure over Ξ_t with mass only at $\varepsilon_{t,i}$. Therefore a measure $\Gamma(x, y)$ over $\Xi_t \times \Xi_t$ whose marginal in x is $\hat{P}_t$, and in y is P_t can have mass only at points $(\hat{\varepsilon}_{t,i}, \varepsilon_{t,j})$. We denote that mass by $\Gamma_{t,i,j} \geq 0$, for $i, j = 1, 2, \ldots N$.

In that case, the right-hand side of (2.114) is simply the linear program

$$\begin{aligned}
&\min_{\Gamma_{t,i,j}} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \sum_{j=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,j}) \Gamma_{t,i,j} \\
&s.t. 0 \leq \Gamma_{t,i,j} \leq 1, \quad t = 0, \ldots, T-1, \quad i = 1, \ldots, n, j = 1, \ldots, n \\
&\quad \sum_{j=1}^{n} \Gamma_{t,i,j} = \frac{1}{n}, t = 0, \ldots, T-1, \quad i = 1, \ldots, n \\
&\quad \sum_{i=1}^{n} \Gamma_{t,i,j} = \frac{1}{n}, t = 0, \ldots, T-1, \quad j = 1, \ldots, n.
\end{aligned} \tag{2.115}$$

By choosing $\Gamma_{t,i,i} = \frac{1}{n}$ for $t = 0, \ldots, T-1, i = 1, \ldots, n$, we get a feasible solution of (2.115), which is the left-hand side of (2.113) and thus it is also an upper bound for the right-hand side of (2.114). Therefore, (2.113) holds.

Therefore

$$\begin{aligned}
&\max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \left\{ \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] - \frac{1}{\gamma n} \sum_{t=0}^{T-1} \sum_{i=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}) \right\} \\
&\overset{(2.113)}{\leq} \max_{\hat{P}_t, \varepsilon_t \sim \hat{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\hat{P}_t, P_t) \\
&\leq \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t),
\end{aligned} \tag{2.116}$$

since the last maximum is taken over a larger set.

Finally, since (2.112) and (2.116) are reverse inequalities between the same quantities, we obtain

$$
\begin{aligned}
&\max_{\tilde{P}_t,\varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] - \frac{1}{\gamma}\sum_{t\ =0}^{T-1} W_t(\tilde{P}_t, P_t) \\
&= \max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} \left\{ \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] - \frac{1}{\gamma n}\sum_{t=0}^{T-1}\sum_{i=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}) \right\},
\end{aligned}
\tag{2.117}
$$

which completes our proof. □

Similar to the result in Theorem 2.7, we have equivalence between the dynamic programming recursion formulated with Wasserstein distance and a finite-dimensional version.

Theorem 2.8.

$$
\begin{aligned}
V_t(x_t) = \min_{u_t} \max_{\tilde{P}_t,\varepsilon_t \sim \tilde{P}_t} &\mathbb{E}\left[f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right] - \frac{1}{\gamma} W_t(\tilde{P}_t, P_t) \\
&s.t. x_{t+1} = g_t\left(x_t, u_t, \varepsilon_t\right)
\end{aligned}
\tag{2.118}
$$

is equivalent to

$$
\begin{aligned}
V_t(x_t) = \min_{u_t} \max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} &\mathbb{E}\left[f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right] - \frac{1}{\gamma n}\sum_{i=1}^{n} c_t(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}) \\
&s.t. x_{t+1} = g_t\left(x_t, u_t, \hat{\varepsilon}_t\right),
\end{aligned}
\tag{2.119}
$$

where $\hat{P}_t$ are empirical measures on n data points $\hat{\varepsilon}_{t,i}$ and $\hat{\varepsilon}_{t,i}$ can be arbitrary points in $\mathbb{R}^{d_3}$.

Proof. This is the same with the proof of Theorem 2.7 but only one stage. □

2.4.2 Worst case

Theorem 2.9. *Consider worst-case optimization problems with $\varepsilon_t \in \Xi_t$,*

$$
\begin{aligned}
V_t(x_t) = \min_{u_t} \max_{\varepsilon_t \in \Xi_t} &\left\{f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right\} \\
&s.t. x_{t+1} = g_t\left(x_t, u_t, \varepsilon_t\right)
\end{aligned}
\tag{2.120}
$$

The robust formulation (2.120) *is equivalent to the generalized Wassersein version* (2.118) *for* P_t *empirical measures on* n *data points and the choice of cost function*

$$c_t(x,y) = \begin{cases} 0, x \in \Xi_t \\ \infty, x \notin \Xi_t \end{cases}. \tag{2.121}$$

Proof. Consider

$$V_t(x_t) = \min_{u_t} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \left\{ \mathbb{E}\left[f_t\left(g_t\left(x_t, u_t, \hat{\varepsilon}_t\right), u_t\right) + V_{t+1}\left(g_t\left(x_t, u_t, \hat{\varepsilon}_t\right)\right)\right] - \frac{1}{n\gamma}\sum_{i=1}^{n} c_t\left(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}\right) \right\} \tag{2.122}$$

and

$$V_T(x_T) = 0 \tag{2.123}$$

Since the choice $\hat{\varepsilon}_{t,i} = \varepsilon_{t,i} \in \Xi_t$, $i = 1, 2, \ldots, N$ results in a finite value of the inner problem objective, it must be that any solution of (2.122) satisfies $\hat{\varepsilon}_{t,i} \in \Xi_t$, $i = 1, 2, \ldots n$. With the choice of transport cost, we get that $\frac{1}{n\gamma}\sum_{i=1}^{n} c_t\left(\hat{\varepsilon}_{t,i}, \varepsilon_{t,i}\right) = 0$.

Therefore, each $\hat{\varepsilon}_{t,i}$ in the support of $\hat{P}_t$ needs to satisfy

$$\hat{\varepsilon}_{t,i} \in \arg\max_{\hat{\varepsilon}_t \in \Xi_t} \left\{f_t\left(g_t\left(x_t, u_t, \hat{\varepsilon}_t\right), u_t\right) + V_{t+1}\left(g_t\left(x_t, u_t, \hat{\varepsilon}_t\right)\right)\right\}. \tag{2.124}$$

As a result, at the inner maximum, each component under the expectation will have the same value, the one of the optimal objective in (2.124). Therefore the expectation is over a constant, and results in the recursion for the worst-case robust optimization recursion.

Thus, for any t, (2.122) is equivalent to

$$V_t(x_t) = \min_{u_t} \max_{\hat{\varepsilon}_t \in \Xi_t} \left\{f_t\left(g_t\left(x_t, u_t, \hat{\varepsilon}_t\right), u_t\right) + V_{t+1}\left(g_t\left(x_t, u_t, \hat{\varepsilon}_t\right)\right)\right\} \tag{2.125}$$

Apply this backward from $t = T - 1$, and we get V_t to be cost-to-go functions of (2.120).

□

2.4.3 H_∞

Consider the H_∞ control problem where noises $\varepsilon_t \in \Xi_t, t = 0 \ldots, T-1$,

$$\min_{u_t(x_t)} \sup_{\varepsilon_t \in \Xi_t} \mathbb{E}\left[\frac{\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)}{\sum_{t=0}^{T-1}\|\varepsilon_t\|^2}\right] \quad (2.126)$$
$$s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t.$$

The expectation is on a single sample ε_t every stage and thus unnecessary, but we keep it to clarify the equivalence with our generalized Wasserstein formulation. (2.126) is equivalent to the problem

$$\min_{\lambda} \min_{u_t(x_t)} \max_{\varepsilon_t \in \Xi_t} \lambda$$
$$s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \quad (2.127)$$
$$\mathbb{E}\left[\frac{\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)}{\sum_{t=0}^{T-1}\|\varepsilon_t\|^2}\right] \le \lambda.$$

The innermost problem can be understood as a feasibility problem whose value is the parameter defining the constraint, λ, if feasible, and ∞ otherwise (Kim and Yang, 2020). Note that the last inequality in (2.127) can be reformulated as

$$\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) - \lambda \sum_{t=0}^{T-1}\|\varepsilon_t\|^2\right] \le 0. \quad (2.128)$$

For any fixed λ, we only need to solve the problem

$$\min_{u_t(x_t)} \max_{\varepsilon_t \in \Xi_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) - \lambda \sum_{t=0}^{T-1}\|\varepsilon_t\|^2\right] \quad (2.129)$$
$$s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t$$

.

to check whether the innermost problem in (2.127) is feasible.

Consequently, the smallest λ for which the optimal value of (2.129) is non-positive is the optimal objective of (2.126).

Observe now that (2.129) is a Wasserstein DRC problem (2.101) with $c_t(x, y) = \|\cdot\|_2^2$, $n = 1$, and $P_t = \delta(0)$, i.e the reference density has all mass at 0. Similarly to out discussion about risk values, it can be shown that the optimal objective of (2.129) is increasing with decreasing λ. Therefore, the optimal value of (2.126), and the optimal policy that goes with that corresponds to the smallest λ for which the optimal value of (2.129) is non-positive.

That value of λ can be obtained by bisection as we did in the discussion of risk in (2.95). The optimal policy of the H_∞ problem is the same as the Wasserstein DRC policy of (2.129) at that value of λ. It is the optimal control policy which is most reliable among all policies at profit level 0 in the setting of (2.95).

2.4.4 KL divergence

Theorem 2.10. *Consider exponential criteria optimization problems in (Jacobson, 1973),*

$$
\begin{aligned}
&\min_{u_t(x_t)} \frac{1}{\gamma} \log \mathbb{E}_P \left[\exp \left\{ \gamma \left(\sum_{t=0}^{T-1} f_{t+1}\left(x_{t+1}, u_t\right) \right) \right\} \right] \\
&s.t. x_{t+1} = g_t\left(x_t, u_t, \varepsilon_t\right)
\end{aligned}
\tag{2.130}
$$

(2.88) *has same optimal policy with* (2.130)*, when replacing* $\sum_{t=0}^{T-1} h_t(\tilde{P}_t, P_t)$ *with* $h(\tilde{P}, P) = KL(\tilde{P}\|P)$*, where* $\tilde{P}$ *has same support with* P *and* $P = \prod_{t=0}^{T-1} P_t$*.*

Proof. By Lemma 2.4, we have

$$
\frac{1}{\gamma} \log \mathbb{E}_P \left[\exp \left\{ \gamma \left(\sum_{t=0}^{T-1} f_t\left(x_{t+1}, u_t\right) \right) \right\} \right] \geq \mathbb{E}_{\tilde{P}} \left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t) \right] - \frac{1}{\gamma} KL(\tilde{P}\|P). \tag{2.131}
$$

Equality can be attained with

$$d\tilde{P} \sim dP \times \exp\left\{\gamma\left(\sum_{t=0}^{T-1} f_t\left(x_{t+1}, u_t\right)\right)\right\}. \tag{2.132}$$

□

Lemma 2.11. (2.130) *equals*

$$\begin{aligned} &\min_{u_t(x_t)} \mathbb{E}\left[\sum_{t=0}^{T-1} f_{t+1}(x_{t+1}, u_t)\right] + \frac{\gamma}{2}\mathbb{V}\left[\sum_{t=0}^{T-1} f_{t+1}(x_{t+1}, u_t)\right] + O(\gamma^2) \\ &s.t. x_{t+1} = g(x_t, u_t, \varepsilon_t) \end{aligned} \tag{2.133}$$

Proof. This is by the expansion of the exponential and logarithm functions, which gives

$$\begin{aligned} \frac{1}{\gamma}\log \mathbb{E}\left[\exp\{\gamma X\}\right] &= \mathbb{E}\left[X\right] + \frac{\gamma}{2}\left(\mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2\right) + O(\gamma^2) \\ &= \mathbb{E}\left[X\right] + \frac{\gamma}{2}\mathbb{V}\left[X\right] + O(\gamma^2). \end{aligned} \tag{2.134}$$

□

From (2.132), we learn that the maximum point $\tilde{P}$ may have cross terms from

$$\exp\left\{\gamma\left(\sum_{t=0}^{T-1} f_t\left(x_{t+1}, u_t\right)\right)\right\}. \tag{2.135}$$

. In a simple case like linear quadratic control, it tells us that the maximum $\tilde{P}$ will no longer have a cartesian production structure.

Theorem 2.12. *Consider the dynamic programming of KL-divergence DRC,*

$$\begin{aligned} V_t(x_t) &= \min_{u_t} \max_{\tilde{P}_t} \mathbb{E}\left[f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right] - \frac{1}{\gamma} KL(\tilde{P}_t \| P_t) \\ &s.t. x_{t+1} = g(x_t, u_t, \varepsilon_t), \end{aligned} \tag{2.136}$$

where $\tilde{P}_t$ has same support with P_t. We have

$$V_t(x_t) = \min_{u_t} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma\left(f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right)\right\}\right]$$
$$= \min_{u_t(x_t),\ldots,u_{T-1}(x_{T-1})} \mathbb{E}\left[\sum_{\tau=t}^{T-1} f_{\tau+1}(x_{\tau+1}, u_\tau)\right] + \frac{\gamma}{2}\mathbb{V}\left[\sum_{\tau=t}^{T-1} f_{\tau+1}(x_{\tau+1}, u_\tau)\right] + O(\gamma^2). \tag{2.137}$$

Proof. First by Lemma 2.4,

$$V_t(x_t) = \min_{u_t} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma\left(f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right)\right\}\right]. \tag{2.138}$$

Then by expanding $V_{t+1}(x_{t+1})$ with the recursion in (2.138) and using the dynamics we obtain that

$$V_t(x_t) = \min_{u_t} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma\left(f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right)\right\}\right]$$
$$= \min_{u_t} \frac{1}{\gamma} \log \mathbb{E}\Bigg[\exp\Bigg\{\gamma f_{t+1}(x_{t+1}, u_t) \tag{2.139}$$
$$+ \min_{u_{t+1}(x_{t+1})} \left\{\log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1}) + V_{t+2}(x_{t+2})\right\} \ |x_{t+1}\right]\right\}\Bigg\}\Bigg].$$

Note that

$$V_t(x_t) = \min_{u_t} \frac{1}{\gamma} \log \int \exp\left\{\gamma f_{t+1}(x_{t+1}, u_t)\right\} \times$$
$$\min_{u_{t+1}(x_{t+1})} \left\{\exp\left\{\log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1}) + V_{t+2}(x_{t+2})\right\} \ |x_{t+1}\right]\right\}\right\} dP(\varepsilon_t). \tag{2.140}$$

We also have that

$$\exp\left\{\gamma f_{t+1}(x_{t+1}, u_t)\right\} > 0 \tag{2.141}$$

and

$$\exp\left\{\log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1}) + V_{t+2}(x_{t+2})\right\} \ |x_{t+1}\right]\right\} > 0. \tag{2.142}$$

Thus, the minimum inside the integral in (2.140) can be moved out of the integral. Then we have

$$\begin{aligned} V_t(x_t) = &\min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \int \exp\left\{\gamma f_{t+1}(x_{t+1}, u_t)\right\} \times \\ &\exp\left\{\log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1}) + V_{t+2}(x_{t+2})\right\} \ |x_{t+1}\right]\right\} dP(\varepsilon_t) \\ = &\min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma f_{t+1}(x_{t+1}, u_t)\right.\right. \\ &\qquad \left.\left. + \log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1}) + V_{t+2}(x_{t+2})\right\} \ |x_{t+1}\right]\right\}\right]. \end{aligned} \tag{2.143}$$

By a similar argument applied from $t+2$ to $T-1$, we have

$$\begin{aligned} &V_t(x_t) \\ &= \min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma f_{t+1}(x_{t+1}, u_t)\right.\right. \\ &\qquad\qquad \left.\left. + \log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1}) + V_{t+2}(x_{t+2})\right\} \ |x_{t+1}\right]\right\}\right] \\ &= \min_{u_t, u_{t+1}(x_{t+1}), u_{t+2}(x_{t+2})} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma f_{t+1}(x_{t+1}, u_t) + \log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1})\right.\right.\right.\right. \\ &\left.\left.\left.\left. + \log \mathbb{E}\left[\exp\left\{\gamma f_{t+3}(x_{t+3}, u_{t+2}) + V_{t+3}(x_{t+3})\right\} \ |x_{t+2}\right]\right\} \ |x_{t+1}\right]\right\}\right] \\ &= \cdots \\ &= \min_{u_t, u_{t+1}(x_{t+1}), u_{t+2}(x_{t+2}), \ldots, u_{T-1}(x_{T-1})} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma f_{t+1}(x_{t+1}, u_t)\right.\right. \\ &\quad \left.\left. + \log \mathbb{E}\left[\exp\left\{\gamma f_{t+2}(x_{t+2}, u_{t+1}) + \log \mathbb{E}\left[\exp\left\{\gamma f_{t+3}(x_{t+3}, u_{t+2}) + \ldots\right\} \ |x_{t+2}\right]\right\} \ |x_{t+1}\right]\right\}\right]. \end{aligned} \tag{2.144}$$

Note that by the expansion of the exponential function, same as in Lemma 2.11, we get for the stage $T-1$, since $V_T(x_T) = 0$, that

$$\begin{aligned} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma f_T(x_T, u_{T-1}) \ |x_{T-1}\right\}\right] = &\ \mathbb{E}\left[f_T(x_T, u_{T-1}) \ |x_{T-1}\right] \\ &+ \frac{\gamma}{2} \mathbb{V}\left[f_T(x_T, u_{T-1}) \ |x_{T-1}\right] + O(\gamma^2). \end{aligned} \tag{2.145}$$

For stage $T-2$, we have

$$\frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma f_{T-1}(x_{T-1}, u_{T-2}) + \log \mathbb{E}\left[\exp\left\{\gamma f_T(x_T, u_{T-1}) \ |x_{T-1}\right\}\right]\right\} \ |x_{T-2}\right]. \tag{2.146}$$

Then we plug the right-hand side of (2.145) for the second term, which gives

$$\frac{1}{\gamma}\log\mathbb{E}\left[\exp\left\{\gamma f_{T-1}(x_{T-1},u_{T-2})+\gamma\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\right.\right.$$
$$\left.\left.+\frac{\gamma^2}{2}\mathbb{V}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]+O(\gamma^3)\right\}\ |x_{T-2}\right]. \tag{2.147}$$

Again by the expansion of the exponential and logarithm function, same as in Lemma 2.11 and combining the $O(\gamma^2)$ terms, the expansion of (2.147) gives

$$\mathbb{E}\left[f_{T-1}(x_{T-1},u_{T-2})+\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]+\frac{\gamma}{2}\mathbb{V}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]+O(\gamma^2)\ |x_{T-2}\right]$$
$$+\frac{\gamma}{2}\mathbb{V}\left[f_{T-1}(x_{T-1},u_{T-2})+\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\right.$$
$$\left.+\frac{\gamma}{2}\mathbb{V}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]+O(\gamma^2)\ |x_{T-2}\right]$$
$$=\mathbb{E}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})+\frac{\gamma}{2}\mathbb{V}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]$$
$$+\frac{\gamma}{2}\mathbb{V}\left[f_{T-1}(x_{T-1},u_{T-2})+\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]+O(\gamma^2). \tag{2.148}$$

Note that conditional on x_{T-1}, $f_{T-1}(x_{T-1},u_{T-2})$ is a constant (not a random variable), we get that

$$\mathbb{E}\left[\mathbb{V}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]=\mathbb{E}\left[\mathbb{V}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right] \tag{2.149}$$

and

$$\mathbb{V}\left[f_{T-1}(x_{T-1},u_{T-2})+\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]$$
$$=\mathbb{V}\left[\mathbb{E}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]. \tag{2.150}$$

Also

$$\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-2}\right]=\mathbb{E}\left[\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]. \tag{2.151}$$

Using (2.151), we get

$$
\begin{aligned}
&\mathbb{E}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-2}\right]+\frac{\gamma}{2}\mathbb{E}\left[\mathbb{V}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]\\
&\quad+\frac{\gamma}{2}\mathbb{V}\left[f_{T-1}(x_{T-1},u_{T-2})+\mathbb{E}\left[f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]+O(\gamma^2)\\
&\overset{(2.149),\ (2.150)}{=}\mathbb{E}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-2}\right]\\
&\quad+\frac{\gamma}{2}\mathbb{E}\left[\mathbb{V}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]\\
&\quad+\frac{\gamma}{2}\mathbb{V}\left[\mathbb{E}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-1}\right]\ |x_{T-2}\right]+O(\gamma^2)\\
&=\mathbb{E}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-2}\right]\\
&\quad+\frac{\gamma}{2}\mathbb{V}\left[f_{T-1}(x_{T-1},u_{T-2})+f_T(x_T,u_{T-1})\ |x_{T-2}\right]+O(\gamma^2),
\end{aligned}
\tag{2.152}
$$

where the last identity follows from the variance decomposition formula. By applying a similar argument from stage $T-3$ to t we complete the proof.

□

(2.130) is well known as an agent for optimizaing a trade-off between expectations and variances. An alternative viewpoint for it is penalizing log-likelihood.

Theorem 2.13.

$$
\begin{aligned}
V_t(x_t)&=\min_{u_t}\max_{\tilde{P}_t}\mathbb{E}\left[f_{t+1}(x_{t+1},u_t)+V_{t+1}(x_{t+1})\right]-\frac{1}{\gamma}KL(\tilde{P}_t\|P_t)\\
s.t.x_{t+1}&=g(x_t,u_t,\varepsilon_t)
\end{aligned}
\tag{2.153}
$$

is equivalent to

$$
\begin{aligned}
&V_t(x_t)=\min_{u_t}\max_{\tilde{P}_t}\mathbb{E}\left[f_{t+1}(x_{t+1},u_t)+V_{t+1}(x_{t+1})\right]+\frac{1}{\gamma}\left(\mathbb{E}_{\tilde{P}_t}\left[\log dP_t(\varepsilon)\right]-\mathbb{E}_{\tilde{P}_t}\left[\log d\tilde{P}_t(\varepsilon)\right]\right)\\
&\quad s.t.x_{t+1}=g(x_t,u_t,\varepsilon_t).
\end{aligned}
\tag{2.154}
$$

Proof. This is by the definition of KL-divergence

$$
KL(\tilde{P}_t\|P_t)=\int\log\left(\frac{d\tilde{P}_t(\varepsilon_t)}{dP_t(\varepsilon_t)}\right)d\tilde{P}_t(\varepsilon_t). \tag{2.155}
$$

$\square$

Remark 2.13.1. *This can be understood as the process of selecting an alternative distribution hypothesis $\tilde{P}_t$. From Gibbs' inequality, we have that*

$$0 \leq KL(\tilde{P}_t \| P_t) = \mathbb{E}_{\tilde{P}_t}\left[\log dP_t(\varepsilon)\right] - \mathbb{E}_{\tilde{P}_t}\left[\log d\tilde{P}_t(\varepsilon)\right],$$

which can be interpreted as a difference between log-likelihoods. Therefore the penalty on the KL divergence in (2.155) *can be interpreted as a penalization on the log-likelihood difference (the second term in* (2.155)*) that comes from the observed distribution P_t under the hypothesis $\tilde{P}_t$.*

We now show that generalized Wasserstein distance DRC can approximate the KL-divergence DRC in the same sense of penalizing likelihood. We first prove the following lemma

Lemma 2.14. *For $c_t(\xi, \zeta) = \|\xi - \zeta\|^2$, define $K_h(\eta)$ to be Gaussian kernel (Li and Racine, 2007) with $h > 0$, then*

$$c_t(\xi, \eta) = 2h^2 \left(\int \log \hat{P}_{t,\xi}(\eta) d\hat{P}_{t,\xi}(\eta) - \int \log \hat{P}_{t,\zeta}(\eta) d\hat{P}_{t,\xi}(\eta) \right), \tag{2.156}$$

where $\hat{P}_{t,\xi}(\eta)$ is the kernel estimated probability measure with only 1 data point ξ

$$\begin{aligned} \hat{P}_{t,\xi}(\eta) &= K_h(\eta - \xi) \\ &= \frac{1}{\left(2\pi h^2\right)^{\frac{d_3}{2}}} e^{-\frac{\|\eta - \xi\|^2}{2h^2}}. \end{aligned} \tag{2.157}$$

Proof.

$$
\begin{aligned}
\int \log \hat{P}_{t,\zeta}(\eta) d\hat{P}_{t,\xi}(\eta) &\overset{(2.157)}{=} \int -\frac{1}{2h^2}\|\zeta - \eta\|^2 d\hat{P}_{t,\xi}(\eta) + \log \frac{1}{(2\pi h^2)^{\frac{d_3}{2}}} \\
&= -\frac{1}{2h^2}\mathbb{E}_{X\sim \hat{P}_{t,\xi}}\left[\|X-\zeta\|^2\right] + \log \frac{1}{(2\pi h^2)^{\frac{d_3}{2}}} \\
&= -\frac{1}{2h^2}\left(\mathbb{E}_{X\sim \hat{P}_{t,\xi}}\left[\|X - \mathbb{E}X + \mathbb{E}X - \zeta\|^2\right]\right) + \log \frac{1}{(2\pi h^2)^{\frac{d_3}{2}}} \\
&= -\frac{1}{2h^2}\|\xi - \zeta\|^2 - \frac{d_3}{2} + \log \frac{1}{(2\pi h^2)^{\frac{d_3}{2}}}. \qquad (2.158)
\end{aligned}
$$

The last identity occurs since the entries of $X - \mathbb{E}X$ are independent Gaussian random variables with mean zero and variance h^2. For the first term in (2.157), we get

$$
\int \log \hat{P}_{t,\xi}(\eta) d\hat{P}_{t,\xi}(\eta) = -\frac{d_3}{2} + \log \frac{1}{(2\pi h^2)^{\frac{d_3}{2}}}, \qquad (2.159)
$$

by taking $\zeta \to \xi$ in (2.158).

Thus, by subtracting (2.159) from (2.158), we get

$$
\int \log \hat{P}_{t,\xi}(\eta) d\hat{P}_{t,\xi}(\eta) - \int \log \hat{P}_{t,\zeta}(\eta) d\hat{P}_{t,\xi}(\eta) = \frac{1}{2h^2}\|\xi - \zeta\|^2. \qquad (2.160)
$$

□

Remark 2.14.1. *Consider the dynamic programming of Wasserstein DRC,*

$$
\begin{aligned}
V_t(x_t) &= \min_{u_t} \max_{\tilde{P}_t, \tilde{\varepsilon}_t \sim \tilde{P}_t} \mathbb{E}\left[f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right] - \frac{1}{\gamma} W^2(\tilde{P}_t, P_t) \\
s.t. x_{t+1} &= g_t(x_t, u_t, \tilde{\varepsilon}_t),
\end{aligned} \qquad (2.161)
$$

where P_t is empirical measure on n data points $\{\varepsilon_t\}_n$, $W^2(\tilde{P}_t, P_t)$ is standard Wasserstein distance with $c_t(\xi, \zeta) = \|\xi - \zeta\|^2$.

Notice that the strong duality of (2.161) *gives*

$$
\begin{aligned}
V_t(x_t) = \min_{u_t} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E}\left[f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right] - \frac{1}{\gamma}\sum_{i=1}^{n} \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 \\
s.t. x_{t+1} = g_t(x_t, u_t, \hat{\varepsilon}_t)
\end{aligned}
\tag{2.162}
$$

and (2.156) *from Lemma 2.14 gives*

$$
\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 = 2h^2 \left(\mathbb{E}_{\hat{P}_{t,\hat{\varepsilon}_{t,i}}} \left[\log d\hat{P}_{t,\varepsilon_{t,i}}\right] - \mathbb{E}_{\hat{P}_{t,\hat{\varepsilon}_{t,i}}} \left[\log d\hat{P}_{t,\hat{\varepsilon}_{t,i}}\right] \right). \tag{2.163}
$$

We can now understand the standard Wasserstein DRC as modeling local density for each data points by $\hat{P}_{t,\varepsilon_{t,i}}$ *first, and then looking for alternative distribution hypothesis from the Gaussian distribution family by penalizing the log-likelihood difference. It will then use* $\hat{\varepsilon}_{t,i}$ *from the maximized distribution to the risk-averse control.*

However, estimating the local density with only 1 data point would be inaccurate, and we can group them up to increase the number of data points to k to have a better estimation of the local density.

Theorem 2.15. *Consider the dynamic programming of Wasserstein DRC,*

$$
\begin{aligned}
V_t(x_t) = \min_{u_t} \max_{\tilde{P}_t^k, (\tilde{\varepsilon}_{t,1}, \ldots, \tilde{\varepsilon}_{t,k}) \sim \tilde{P}_t^k, \tilde{\varepsilon}_t \sim Rand(\tilde{\varepsilon}_{t,1}, \ldots, \tilde{\varepsilon}_{t,k})} & \mathbb{E}\left[f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right] \\
& - \frac{1}{\gamma} W(\tilde{P}_t^k, P_t^k) \\
s.t. x_{t+1} = g(x_t, u_t, \tilde{\varepsilon}_t),
\end{aligned}
\tag{2.164}
$$

where P_t^k *is empirical measure on* m *groups of* k *joint data points* $(\varepsilon_{t,1}, \ldots, \varepsilon_{t,k})$. $W(\tilde{P}_t, P_t)$ *is the generalized Wasserstein distance with*

$$
c_t((\xi_1, \ldots, \xi_k), (\zeta_1, \ldots, \zeta_k)) = \int \log \hat{P}_{t,\xi}^k(\eta) d\hat{P}_{t,\xi}^k(\eta) - \int \log \hat{P}_{t,\zeta}^k(\eta) d\hat{P}_{t,\xi}^k(\eta). \tag{2.165}
$$

$\hat{P}^k_{t,\zeta}(\xi_1,\dots,\xi_k)$ is the kernel estimated probability measure with k data points $\zeta = (\zeta_1,\dots,\zeta_k)$ whose density function is

$$d\hat{P}^k_{t,\xi}(\eta) = \frac{1}{k}\sum_{i=1}^{k} K_h(\eta - \xi_i), \tag{2.166}$$

where $K_h(\eta)$ is Gaussian kernel with $h > 0$

$$K_h(\eta) = \frac{1}{(2\pi h^2)^{\frac{d_3}{2}}} e^{-\frac{\|\eta\|^2}{2h^2}} \tag{2.167}$$

and h may depend on k.

Then the generalized Wasserstein distance with cost function (2.165) *is well-defined. Assume the dataset contains m groups of data points $\left\{\varepsilon^i_t := (\varepsilon^i_{t,1},\dots,\varepsilon^i_{t,k})\right\}_{i=1,2,\dots,m}$,* (2.164) *is equivalent to*

$$\begin{aligned}
V_t(x_t) = \min_{u_t} & \max_{\hat{P}_t = \left\{\hat{\varepsilon}^i_t := (\hat{\varepsilon}^i_{t,1},\dots,\hat{\varepsilon}^i_{t,k})\right\}_m} \mathbb{E}\left[f_{t+1}(x_{t+1}, u_t) + V_{t+1}(x_{t+1})\right] \\
&+ \frac{1}{\gamma m}\sum_{i=1}^{m}\left(\mathbb{E}_{\hat{P}^k_{t,\hat{\varepsilon}^i_t}}\left[\log d\hat{P}^k_{t,\varepsilon^i_t}(\varepsilon)\right] - \mathbb{E}_{\hat{P}^k_{t,\hat{\varepsilon}^i_t}}\left[\log d\hat{P}^k_{t,\hat{\varepsilon}^i_t}(\varepsilon)\right]\right) \\
& s.t. x_{t+1} = g(x_t, u_t, \hat{\varepsilon}_t),
\end{aligned} \tag{2.168}$$

where the expectation is over the random selection process that selects $\hat{\varepsilon}^i_t$ from

$$\left\{\hat{\varepsilon}^i_t := (\hat{\varepsilon}^i_{t,1},\dots,\hat{\varepsilon}^i_{t,k})\right\}_{i=1,2,\dots,m} \tag{2.169}$$

first, and then randomly selects $\hat{\varepsilon}_t$ from $(\hat{\varepsilon}^i_{t,1},\dots,\hat{\varepsilon}^i_{t,k})$.

Remark 2.15.1. *In this formulation, the maximization step searches for $\tilde{P}^k_t$. And then the random process draw $(\tilde{\varepsilon}_{t,1},\dots,\tilde{\varepsilon}_{t,k})$ from $\tilde{P}^k_t$, following by drawing $\tilde{\varepsilon}_t$ uniformly from $(\tilde{\varepsilon}_{t,1},\dots,\tilde{\varepsilon}_{t,k})$.*

Alternatively, we can view the original random process in the expectation as randomly draw from m group first and then draw data points from the group, which doesn't change the expectation as long as each data points appear same times. Then (2.164) *just performs risk-averse operations in the first step.*

Proof. We only need to verify this definition of c_t is proper. Note that

$$\begin{aligned} c_t((\xi_1,\dots,\xi_k),(\zeta_1,\dots,\zeta_k)) &= \int \log \hat{P}^k_{t,\xi}(x) d\hat{P}^k_{t,\xi}(x) - \int \log \hat{P}^k_{t,\zeta}(x) d\hat{P}^k_{t,\xi}(x) \\ &= KL(\hat{P}^k_{t,\xi} \| \hat{P}^k_{t,\zeta}), \end{aligned} \tag{2.170}$$

we have

$$c_t((\zeta_1,\dots,\zeta_k),(\zeta_1,\dots,\zeta_k)) = 0 \tag{2.171}$$

and

$$c_t((\xi_1,\dots,\xi_k),(\zeta_1,\dots,\zeta_k)) \geq 0. \tag{2.172}$$

Then (2.164) is equivalent to

$$\begin{aligned} V_t(x_t) = \min_{u_t} & \max_{\hat{P}_t = \left\{\hat{\varepsilon}^i_t := (\hat{\varepsilon}^i_{t,1},\dots,\hat{\varepsilon}^i_{t,k})\right\}_m} \mathbb{E}\left[f_{t+1}(x_{t+1},u_t) + V_{t+1}(x_{t+1})\right] \\ & + \frac{1}{\gamma m} \sum_{i=1}^m \left(\int \log d\hat{P}^k_{t,\varepsilon^i_t}(\varepsilon) d\hat{P}^k_{t,\hat{\varepsilon}_t}(\varepsilon) - \int d \log \hat{P}^k_{t,\hat{\varepsilon}_t}(\varepsilon) d\hat{P}^k_{t,\hat{\varepsilon}_t}(\varepsilon) \right) \\ & s.t. x_{t+1} = g(x_t,u_t,\hat{\varepsilon}_t). \end{aligned} \tag{2.173}$$

Note that

$$\int \log d\hat{P}^k_{t,\varepsilon^i_t}(\varepsilon) d\hat{P}^k_{t,\hat{\varepsilon}_t}(\varepsilon) = \mathbb{E}_{\hat{P}^k_{t,\hat{\varepsilon}_t}} \left[\log d\hat{P}^k_{t,\varepsilon^i_t}(\varepsilon) \right] \tag{2.174}$$

and

$$\int \log d\hat{P}^k_{t,\hat{\varepsilon}_t}(\varepsilon) d\hat{P}^k_{t,\hat{\varepsilon}_t}(\varepsilon) = \mathbb{E}_{\hat{P}^k_{t,\hat{\varepsilon}_t}} \left[\log d\hat{P}^k_{t,\hat{\varepsilon}_t}(\varepsilon) \right], \tag{2.175}$$

this completes the proof.

□

Remark 2.15.2. *This explains the intuition of our framework in §2.3. Notice that when $m = 1$ and $k = n$, it recovers KL-divergence DRC controller approximately, which restricts the alternative distribution hypothesis to a parametric family. When $m = n$ and $k = 1$, this is standard Wasserstein DRC controller. With decreasing m from n to 1 and grouping nearby data points, we perform a sample selection process, which is why we define* (2.8)*. Under* *certain conditions, DRC controller for $m = 1$ and $k = n$ will be exact KL-divergence DRC as we have shown.*

Remark 2.15.3. *From the non-parametric statistics (Devroye, 1985), when n and k go to infinity, the error that comes from kernel density estimation will go to 0. Thus, when n and k go to infinity, though the above alternative distribution is restricted to certain distribution families, it will not differ significantly from the whole distribution space.*

2.4.5 Standard Wasserstein Distance

Theorem 2.16. *Consider the Wasserstein DRC problem,*

$$\begin{aligned} \min_{u_t(x_t)} \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t\left(x_{t+1}, u_t\right)\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1} W_2^2(\tilde{P}_t, P_t) \\ s.t. x_{t+1} = g_t\left(x_t, u_t, \varepsilon_t\right), \end{aligned} \tag{2.176}$$

where

$$W_2(\tilde{P}_t, P_t) = \min_{\Gamma = \Pi(\tilde{P}_t, P_t)} \left(\int \|\xi - \zeta\|^2 d\Gamma(\xi, \zeta)\right)^{\frac{1}{2}}. \tag{2.177}$$

(2.176) *is equivalent to*

$$\begin{aligned} \min_{u_t(x_t)} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t\left(x_{t+1}, u_t\right)\right] - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n} \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 \\ s.t. x_{t+1} = g_t\left(x_t, u_t, \hat{\varepsilon}_t\right). \end{aligned} \tag{2.178}$$

Standard Wasserstein distance is one of the most robust distances in the setting of distributionally robust control. It ignores any underlying geometry and probabilities of problems, figure 2.2 is an explanation of why it can be too robust in many scenarios.

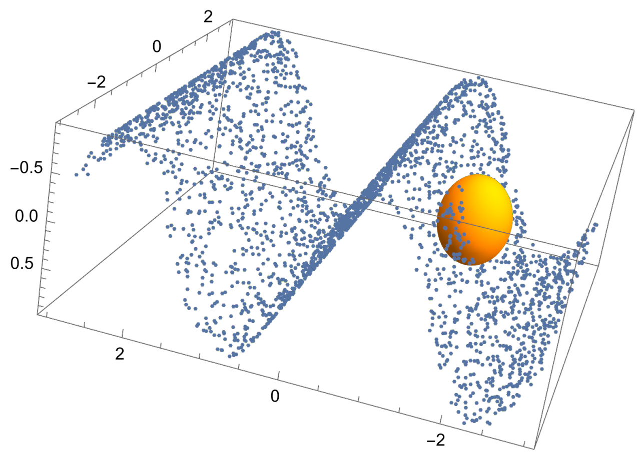


Figure 2.2: Blue points are data points. The orange ball is the range of possible shifting from one point that will be considered in the robust problem. Assuming that noises are distributed on specific manifolds, standard Wasserstein distance ignores any underlying geometry of the problem. It ensures it is robust in all directions, including every possible shifting in the ball, even outside the manifold. This explains why it will bring redundant robustness. In this sense, it is the most robust distance in the distributionally robust control.

2.5 Distribution on Manifold

From the above argument, we see the direct applications of standard Wasserstein distance will have the problem of redundant robustness. One way to avoid this problem is to use the kernel trick used in (2.8), which is shown in figure 2.3, explaining why it can remedy the redundant robustness. We provide an alternative way to deal with this problem with locally linear embedding (Roweis and Saul, 2000).

Consider LLE Wasserstein robust problems,

$$\begin{aligned} &\min_{u_t(x_t)} \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1} W_t(\tilde{P}_t, P_t) \\ &\qquad s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \end{aligned} \tag{2.179}$$

Define $s_t(y) = \left\{z_{t,k}\right\}_K$ to be K nearest neighbors of y in the set $\left\{\varepsilon_{t,i}\right\}_n$, where K is a parameter to tune, then we define

$$W_t(\tilde{P}_t, P_t) = \min_{\Gamma=\Pi(\tilde{P}_t,P_t)} \int c_t\,(x,y)\, d\Gamma(x,y) \tag{2.180}$$

where

$$c_t\,(x,y) = \|Proj_{t,y}(x) - Proj_{t,y}(y)\|^2 + \alpha\|\big(x - Proj_{t,y}(x)\big) - \big(y - Proj_{t,y}(y)\big)\,\|^2, \tag{2.181}$$

and α is another parameter to tune. Here, $Proj_{t,y}(x)$ is the projection on the regression hyperplane of the nearest K neighbors of y, that is

$$Proj_{t,y}(x) = \arg \min_{e\in Span\{s_t(y)\}} \|x-e\|^2, \tag{2.182}$$

where

$$Span\{s_t(y)\} = \{x : a_{t,y}^T x + b_{t,y} = 0, a_{t,y}, b_{t,y} \in \arg\min_{a,b} \sum_{e\in s_t(y)} \Big(a^T e + b\Big)^2\}. \tag{2.183}$$

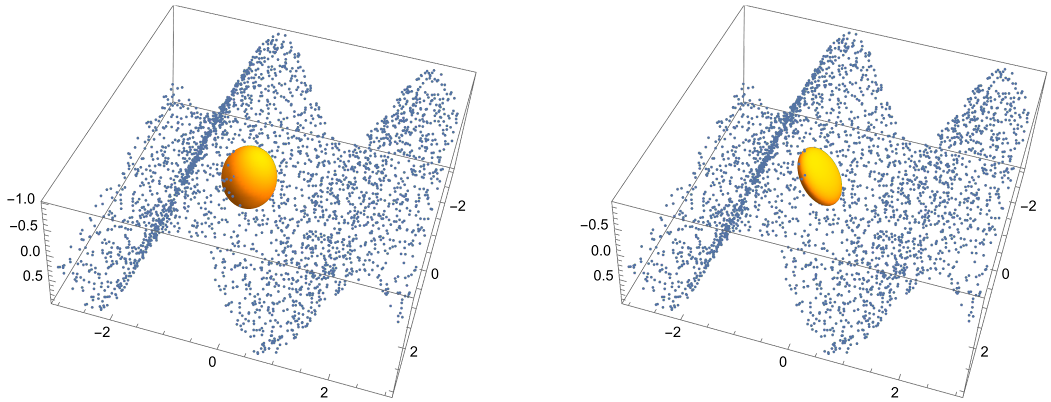
It's not hard to verify that c_t is non-negative, $c_t(y,y) = 0$, and $\forall y$, convex in x. Then we have

Theorem 2.17. (2.179) *is equivalent to*

$$\begin{aligned}
\min_{u_t(x_t)} \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t} \mathbb{E}&\left[\sum_{t=0}^{T-1}\Big(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\Big)\right] \\
&- \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\Big(\|Proj_{t,\varepsilon_{t,i}}\left(\hat{\varepsilon}_{t,i}\right) - Proj_{t,\varepsilon_{t,i}}\left(\varepsilon_{t,i}\right)\|^2 \\
&+\alpha\|\Big(\hat{\varepsilon}_{t,i} - Proj_{t,\varepsilon_{t,i}}\left(\hat{\varepsilon}_{t,i}\right)\Big) - \Big(\varepsilon_{t,i} - Proj_{t,\varepsilon_{t,i}}\left(\varepsilon_{t,i}\right)\Big)\|^2\Big) \\
s.t. x_{t+1} &= A_t x_t + B_t u_t + C_t\hat{\varepsilon}_t
\end{aligned} \tag{2.184}$$

Proof. This is by Theorem 2.7. □

We can compute the basis of the hyperplane with simple algebra, but it will involve matrix computations and increase the time complexity. Instead, we use a soft-constraint version here,



(a) Standard Wasserstein Distance (b) Kernelized Wasserstein Distance

Figure 2.3: Comparison between standard Wasserstein distance and kernelized Wasserstein distance. By putting higher weights on the nearby points, the local geometry structure is recovered and thus remedies the redundant robustness.

Theorem 2.18. *The soft-constraint version of* (2.184) *is*

$$
\begin{aligned}
&\min_{u_t(x_t)} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \max_{\hat{a}_{t,i}, \hat{b}_{t,i}, \hat{e}_{t,i}, e_{t,i}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\Bigg(\|e_{t,i} - \hat{e}_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \hat{e}_{t,i} - \varepsilon_{t,i} + e_{t,i}\|^2 + \rho_1 \sum_{e \in s_t(\varepsilon_{t,i})}\left(\hat{a}_{t,i}^T e + \hat{b}_{t,i}\right)^2 \\
&+ \rho_2\left(\left(\hat{a}_{t,i}^T \hat{e}_{t,i} + \hat{b}_{t,i}\right)^2 + \left(\hat{a}_{t,i}^T e_{t,i} + \hat{b}_{t,i}\right)^2\right) + \rho_3\left(\|e_{t,i} - \varepsilon_{t,i}\|^2 + \|\hat{e}_{t,i} - \hat{\varepsilon}_{t,i}\|^2\right)\Bigg) \\
&s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t,
\end{aligned}
\tag{2.185}
$$

which is a special case of the kernel method with kernel constructed by the inner maximum point $\hat{a}_{t,i}, \hat{b}_{t,i}, z_{t,i}, e_{t,i}$ *of* (2.185)*,* $\rho_1, \rho_2, \rho_3 > 0$*,* $\hat{a}_{t,i} \in \mathbb{R}^{d_3}, \hat{b}_{t,i} \in \mathbb{R}, z_{t,i}, e_{t,i} \in \mathbb{R}^{d_3}$*.*

Proof. This is by relaxing the equality and minimum condition in (2.182) and (2.183). □

CHAPTER 3

RISK-AVERSE MIXTURE OF GAUSSIAN CONTROL

3.1 Introduction

In stochastic programming, the optimization objective is selected to be the expectation. Given the model f, we optimize

$$\mathbb{E}\left[f\right]. \tag{3.1}$$

(3.1), in fact, assumes two impossible things. The one is the ground truth distribution is known, and the other one is in the application of the stochastic models, the realized data points will strictly follow the same distribution.

The first one is not hard to understand. If anyone claims a group of data points following Gaussian, they do not mean they know the parameter of $N(\mu, \Sigma)$ that generates the data points, but more closely mean that the cumulative distribution function of the data points is close to a Gaussian with estimated parameters $N(\hat{\mu}, \hat{\Sigma})$. The most information one can know is $\hat{\mu}$ and $\hat{\Sigma}$, but the ground truth parameters μ and Σ can never be learned. Pessimistically, the ground truth distribution for general cases cannot even be described with a finite number of parameters.

The other one is about the limit-sense definition of distributions. Even if you know the ground truth parameters for $N(\mu, \Sigma)$ in (3.1), in the application, if you run 1000 times on it, it is a question of whether the distribution of data points is closer to $N(\mu, \Sigma)$ or $N(\hat{\mu}, \hat{\Sigma})$, where $\hat{\mu}$ and $\hat{\Sigma}$ are parameters estimated from the data points directly? In the first 1000 runs, it may look like $N(\hat{\mu}_1, \hat{\Sigma}_1)$. But in another 1000 runs, it maybe look like the other Gaussian $N(\hat{\mu}_2, \hat{\Sigma}_2)$. It just means if we use $N(\hat{\mu}_1, \hat{\Sigma}_1)$ in (3.1) for the first 1000 runs and use $N(\hat{\mu}_2, \hat{\Sigma}_2)$ for the second 1000 runs, we should be able to have better results. This occurs because distributions are defined in the limit long-run sense. If you only sample limited times from a distribution, the samples will never follow the same distribution strictly.

Instead, control theory has two frameworks to deal with these problems under Gaussian settings. The first one is risk-averse control (Jacobson, 1973), which optimizes a trade-off between expectations and variances,

$$\mathbb{E}\left[f\right] + \frac{\gamma}{2}\mathbb{V}\left[f\right] \tag{3.2}$$

with agent

$$\frac{1}{\gamma}\log\mathbb{E}\left[\exp\{\gamma f\}\right] = \mathbb{E}[f] + \frac{\gamma}{2}\mathbb{V}\left[f\right] + O(\gamma^2). \tag{3.3}$$

The second one is minimax control with relative entropy (Petersen et al., 2000), which put constraints on relative entropy,

$$\max_{\tilde{P}, KL(\tilde{P}\|P)\leq\gamma} \mathbb{E}_{\tilde{P}}\left[f\right]. \tag{3.4}$$

(3.2) can be understood as penalizing statistical fluctuations by penalizing variances, while (3.4) describes the difference between the ground truth distribution P and the realized distributions $\tilde{P}$ with Kullback-Leibler divergence. (3.2) is called risk-averse control since it penalizes variance while optimizing the expectation. (3.2) and (3.4), in fact, can be proven equivalent and work for Gaussian distribution.

In this paper, we will discuss the risk-averse control problem under mixture of Gaussian distribution. In many applications, noise will deviate from Gaussian heavily. Instead, they may be modeled with mixture of Gaussian family. And thus, our method is designed for mixture of Gaussian distribution but can work for more general distribution families. It brings significant convenience to model uncertainty coming from multiple different sources. We will discuss two types of RLQMG problems. The one is with uncertainty on competent probability, and the other one is without that uncertainty. We show that the first one is a highly nonlinear optimization problem but can be approached by putting an independence assumption on the components' risk-averse probability. And luckily, the second one can be solved exactly.

This paper is organized in the following order. The problem and assumptions are stated in the second section. In the third section, we show the equivalence between (3.2) and (3.4).

And we will use the equivalence form to discuss RLQMG. In the fourth section, we discuss the RLQMG without uncertainty on components' probability. In the fifth section, we discuss RLQMG with uncertainty on components' probability. In the sixth section, we give an independence relaxation for the RLQMG with uncertainty on components' probability. In the last section, we provide experiment details for them.

3.2 Problem Formulation and Assumptions

The problem we consider in this paper is

$$
\begin{aligned}
&\min_{u_t(x_t)} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma \sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right\}\right] \\
&s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \\
&\qquad \varepsilon_t \sim \sum_{i=1}^{n} \pi_{t,i} N\left(\mu_{t,i}, \Sigma_{t,i}\right), \varepsilon_t \text{ are independent.}
\end{aligned}
\tag{3.5}
$$

3.2.1 Assumptions

We will use common assumptions in the risk-averse control,

Assumption 3.1. $Q_t \geq 0$.

Assumption 3.2. $R_t \geq 0$.

Assumption 3.3. $rank(B_t) = d_2$.

Assumption 3.4. $\forall t, i, 0 < \pi_{t,i} \leq 1, \sum_{i=1}^{n} \pi_{t,i} = 1$.

For notation simplicity, if not specified, $\|\cdot\|$ are always assumed to be $\|\cdot\|_2$ in this paper. And if at any place we mention x_{t+1} without dynamics, then the dynamics are assumed to be the one in (3.5).

3.3 Equivalence

Theorem 3.1. (3.5) *differs from*

$$\min_{u_t(x_t)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] + \frac{\gamma}{2}\mathbb{V}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right]$$

$$s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \tag{3.6}$$

$$\varepsilon_t \sim \sum_{i=1}^{n} \pi_{t,i} N\left(\mu_{t,i}, \Sigma_{t,i}\right), \varepsilon_t \text{ are independent}$$

in $O(\gamma^2)$.

Proof. This is by the expansion of the logarithm and exponential function. □

Lemma 3.2. *For any random variable X following a distribution P which has density or mass $dP(x)$, the maximum point $\tilde{P}$ of the problem*

$$\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}[X] - \frac{1}{\gamma} KL(\tilde{P}\|P) \tag{3.7}$$

also has same support with P if the maximum in (3.7) *exists, and its density satisfies the condition*

$$d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{3.8}$$

Consequently, we have the identity

$$\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}[X] - \frac{1}{\gamma} KL(\tilde{P}\|P) = \int x d\tilde{P}(x) - \frac{1}{\gamma}\log\left(\frac{d\tilde{P}(x)}{dP(x)}\right) d\tilde{P}(x) = \frac{1}{\gamma}\log \mathbb{E}_P\left[e^{\gamma X}\right], \tag{3.9}$$

where $\mathbb{E}_P\left[e^{\gamma X}\right]$ is the normalized constant for the $d\tilde{P}(x)$ in (3.8).

Proof. Consider the function

$$\log \mathbb{E}_P\left[e^{\gamma X}\right] = \log \int e^{\gamma x} dP(x). \tag{3.10}$$

For any distributions $\tilde{P}(x)$ that has density and same support with $P(x)$,

$$\begin{aligned}\log \int e^{\gamma x} dP(x) &= \log \int e^{\gamma x} \left(\frac{dP(x)}{d\tilde{P}(x)}\right) d\tilde{P}(x) \\ &\geq \int \left(\gamma x d\tilde{P}(x) + \log\left(\frac{dP(x)}{d\tilde{P}(x)}\right) d\tilde{P}(x)\right),\end{aligned} \tag{3.11}$$

where the last inequality comes from Jensen's inequality in probabilistic setting (Durrett, 2019).

Divide γ on the both hand side of (3.11), we have

$$\frac{1}{\gamma} \log \mathbb{E}_P \left[e^{\gamma X}\right] \geq \mathbb{E}_{\tilde{P}}\left[x\right] - \frac{1}{\gamma} KL(\tilde{P} \| P). \tag{3.12}$$

Equality can be attained from Jensen's inequality with the condition

$$d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{3.13}$$

Note that the distribution constructed from (3.13) always exists if (3.10) is not infinity. Thus if (3.10) is not infinite, the maximum of (3.7) exists.

Assume now (3.10) is infinite. By constructing $\tilde{P}(x)$, we get

$$d\tilde{P}_R(x) \sim dP(x) \times e^{\gamma x} \times I(\|x\| \leq R), \tag{3.14}$$

where $R \in \mathbb{R}$, we have $\tilde{P}_R$ always exists and feasible in (3.7). Denote C_R to be the normalized constant such that

$$d\tilde{P}_R(x) = \frac{1}{C_R} dP(x) \times e^{\gamma x} \times I(\|x\| \leq R). \tag{3.15}$$

Since (3.10) is infinite, we must have that

$$\begin{aligned}\lim_{R \to \infty} C_R &= \lim_{R \to \infty} \int e^{\gamma x} \times I(\|x\| \leq R) dP(x) \\ &= \infty\end{aligned} \tag{3.16}$$

Let $R \to \infty$, we have

$$\lim_{R \to \infty} \mathbb{E}_{\tilde{P}_R}\left[X\right] - \frac{1}{\gamma} KL(\tilde{P}_R \| P) = \lim_{R \to \infty} \int x d\tilde{P}_R(x) - \frac{1}{\gamma} \log\left(\frac{d\tilde{P}_R(x)}{dP(x)}\right) d\tilde{P}_R(x) = \lim_{R \to \infty} \int - \frac{1}{\gamma} \log\left(\frac{d\tilde{P}_R(x)}{e^{\gamma x} dP(x)}\right) d\tilde{P}_R(x) = \lim_{R \to \infty} \int_{\|x\| \leq R} \frac{1}{\gamma} \frac{\log C_R}{C_R} e^{\gamma x} dP(x) = \lim_{R \to \infty} \frac{1}{\gamma} \log C_R = \infty. \tag{3.17}$$

From (3.17) and (3.12) we have that the existence of the maximum of (3.7) is equivalent to the finite property of (3.10). When finiteness of either occurs, we can invoke (3.12) and (3.13) to claim our conclusion. This completes the proof.

□

Theorem 3.3. (3.5) *is equivalent to*

$$\begin{aligned} &\min_{u_t(x_t)} \max_{\tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma} KL(\tilde{P} \| P) \\ &s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \\ &(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1}) \sim \tilde{P}, \end{aligned} \tag{3.18}$$

where $P = \prod_{t=0}^{T-1}\left(\sum_{i=1}^{n} \pi_{t,i} N(\mu_{t,i}, \Sigma_{t,i})\right)$.

Proof. Fixing $u_t(x_t)$ first, we can apply Lemma 3.2 to the objective of (3.5). Then, (3.9) gives the inner maximization problem. □

Theorem 3.4. *Consider the dynamics programming*

$$\begin{aligned} &V_t(x_t) = \min_{u_t} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right)\right\}\right] \\ &s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \\ &\varepsilon_t \sim \sum_{i=1}^{n} \pi_{t,i} N\left(\mu_{t,i}, \Sigma_{t,i}\right), \varepsilon_t \text{ are independent,} \end{aligned} \tag{3.19}$$

whose terminal condition is

$$V_T(X_T) = 0. \tag{3.20}$$

$V_0(x_0)$ *solves* (3.5).

Proof. First by Lemma 3.2,

$$V_t(x_t) = \min_{u_t} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right)\right\}\right]. \tag{3.21}$$

Expand $V_{t+1}(x_{t+1})$, we have

$$\begin{aligned} V_t(x_t) = \min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \mathbb{E} \Bigg[\exp \Bigg\{ \gamma \Bigg(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \\ + \frac{1}{\gamma} \log \mathbb{E} \left[\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \right] \Bigg) \Bigg\} \Bigg] . \end{aligned} \tag{3.22}$$

As

$$\begin{aligned} &\exp \Bigg\{ \gamma \Bigg(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \\ &+ \frac{1}{\gamma} \log \mathbb{E} \left[\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \right] \Bigg) \Bigg\} \\ &= \exp \Bigg\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \\ &\quad + \gamma \left(\frac{1}{\gamma} \log \mathbb{E} \left[\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \right] \right) \Bigg\} \\ &= \exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \right\} \\ &\exp \left\{ \gamma \left(\frac{1}{\gamma} \log \mathbb{E} \left[\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \right] \right) \right\} , \end{aligned} \tag{3.23}$$

we have

$$\begin{aligned} V_t(x_t) &= \min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \mathbb{E} \Bigg[\exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \right\} \\ &\exp \left\{ \gamma \frac{1}{\gamma} \log \mathbb{E} \left[\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \right] \right\} \Bigg] \\ &= \min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \mathbb{E} \Bigg[\exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \right\} \\ &\mathbb{E} \left[\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \right] \Bigg] . \end{aligned} \tag{3.24}$$

When conditioned on x_{t+1}, we have

$$\begin{aligned} &\exp \Bigg\{ \gamma \Bigg(x_{t+1}^T Q_{t+1} x_{t+1} \\ &+ u_t^T R_t u_t \Bigg) \Bigg\} \mathbb{E} \left[\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \right] \\ &= \mathbb{E} \Bigg[\exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \right\} \\ &\exp \left\{ \gamma \left(x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} + V_{t+2}(x_{t+2}) \right) \right\} \ |x_{t+1} \Bigg] . \end{aligned} \tag{3.25}$$

Thus

$$
\begin{aligned}
&V_t(x_t) \\
&= \min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \mathbb{E}\Big[\mathbb{E}\Big[\exp\Big\{\gamma\Big(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} \\
&+ V_{t+2}(x_{t+2})\Big)\Big\} \ |x_{t+1}\Big]\Big] \\
&= \min_{u_t, u_{t+1}(x_{t+1})} \frac{1}{\gamma} \log \mathbb{E}\Big[\exp\Big\{\gamma\Big(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + x_{t+2}^T Q_{t+2} x_{t+2} + u_{t+1}^T R_{t+1} u_{t+1} \\
&+ V_{t+2}(x_{t+2})\Big)\Big\}\Big].
\end{aligned} \tag{3.26}
$$

Continue this to the stage T we complete the proof. □

Theorem 3.5. (3.19) *is equivalent to*

$$
\begin{aligned}
V_t(x_t) = &\min_{u_t} \max_{\tilde{\pi}_t, \tilde{P}_{t,i}} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] \\
&- \frac{1}{\gamma} \sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i} \| P_{t,i}) - \frac{1}{\gamma} KL(\tilde{\pi}_t \| \pi_t) \\
s.t. x_{t+1} &= A_t x_t + B_t u_t + C_t \varepsilon_t \\
\varepsilon_t &\sim \sum_{i=1}^{n} \tilde{\pi}_{t,i} \tilde{P}_{t,i},
\end{aligned} \tag{3.27}
$$

whose terminal condition is

$$V_T(x_T) = 0. \tag{3.28}$$

Remark 3.5.1. *Observe that, if x_t is a random variable, then the solution distribution will depend on x_t. While for the case of one cluster, this does not pose major difficulties as shown in (Jacobson, 1973), for multiple clusters this will make for a very difficult backward recursion that we will later address.*

Proof. (3.19) can be formulated as

$$\begin{aligned}
V_t(x_t) &= \min_{u_t} \frac{1}{\gamma} \log \left(\sum_{i=1}^{n} \pi_{t,i} \mathbb{E}_{\varepsilon_t \sim P_{t,i}} \left[\exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1}) \right) \right\} \right] \right) \\
&= \min_{u_t} \frac{1}{\gamma} \log \left(\sum_{i=1}^{n} \pi_{t,i} \exp \left\{ \gamma \right.\right. \\
&\qquad \left.\left. \times \frac{1}{\gamma} \log \mathbb{E}_{\varepsilon_t \sim P_{t,i}} \left[\exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1}) \right) \right\} \right] \right\} \right).
\end{aligned} \tag{3.29}$$

By applying Lemma 3.2 with the mass function $\pi_{t,i}$, (3.29) is equivalent to

$$\begin{aligned}
V_t(x_t) = \min_{u_t} \max_{\tilde{\pi}_{t,i}} \sum_{i=1}^{n} \tilde{\pi}_{t,i} \frac{1}{\gamma} \log \mathbb{E}_{\varepsilon_t \sim P_{t,i}} \left[\exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1}) \right) \right\} \right] \\
- \frac{1}{\gamma} KL(\tilde{\pi}_t \| \pi_t).
\end{aligned} \tag{3.30}$$

By applying Lemma 3.2 again on the

$$\frac{1}{\gamma} \log \mathbb{E}_{\varepsilon_t \sim P_{t,i}} \left[\exp \left\{ \gamma \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1}) \right) \right\} \right], \tag{3.31}$$

we have (3.29) equivalent to

$$\begin{aligned}
V_t(x_t) = \min_{u_t} \max_{\tilde{\pi}_{t,i}, \tilde{P}_{t,i}} \sum_{i=1}^{n} \tilde{\pi}_{t,i} \mathbb{E}_{\varepsilon_t \sim \tilde{P}_{t,i}} \left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1}) \right] \\
- \frac{1}{\gamma} \sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i} \| P_{t,i}) - \frac{1}{\gamma} KL(\tilde{\pi}_t \| \pi_t).
\end{aligned} \tag{3.32}$$

Combine $\tilde{\pi}_{t,i}$ into the expectation part; now we get the distribution to be mixture of Gaussian $\sum_{i=1}^{n} \tilde{\pi}_{t,i} \tilde{P}_{t,i}$.

□

We can see from (3.27), the original problem (3.5) can be written as

$$\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{\pi}_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1}),\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_t)} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\Bigg]$$

$$s.t. x_{t+1} = A_t x_t + B_t u_t + C_t\varepsilon_t \tag{3.33}$$

where the density of the joint distribution of noises $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1})$ is

$$\sum_{i_0,i_1,\ldots,i_{T-1}} \prod_{t=0}^{T-1} \tilde{\pi}_{t,i_t}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1}) d\tilde{P}_{t,i_t}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1}). \tag{3.34}$$

This is because from Theorem 3.4 and Theorem 3.5, we see the risk-averse distribution

$$\sum_{i=1}^{n} \tilde{\pi}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1}) \tag{3.35}$$

is, in fact, a robust alternative to the condition distribution

$$P(\varepsilon_t|(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})). \tag{3.36}$$

Multiplying all stages together gives the density.

3.4 Without Uncertainty on the Component Probability

There are certain applications that don't have uncertainty on the component probability. For example, if we have 100 students in 3 classes, where they are divided by 40/30/30. And we want to evaluate the teaching effectiveness from quizzes given to all students monthly. Then the risk-averse problem based on students' scores can be formulated as

$$\begin{aligned}V_t(x_t) =& \min_{u_t} \max_{\tilde{P}_{t,1},\tilde{P}_{t,2},\tilde{P}_{t,3},\varepsilon_t \sim \frac{40}{100}\tilde{P}_{t,1}+\frac{30}{100}\tilde{P}_{t,2}+\frac{30}{100}\tilde{P}_{t,3}} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] \\ &- \frac{1}{\gamma}\left(\frac{40}{100}KL(\tilde{P}_{t,1}\|P_{t,1}) + \frac{30}{100}KL(\tilde{P}_{t,2}\|P_{t,2}) + \frac{30}{100}KL(\tilde{P}_{t,3}\|P_{t,3})\right) \\ s.t.\, x_{t+1} &= A_t x_t + B_t u_t + C_t \varepsilon_t \\ \varepsilon_t &\sim P_t.\end{aligned} \tag{3.37}$$

The difference is that, in the above example, we don't have uncertainty in the π; the number of students in each class is always determined. (3.27) is replaced by

$$\begin{aligned}V_t(x_t) =& \min_{u_t} \max_{\tilde{P}_{t,i},\varepsilon_t \sim \sum_{i=1}^n \pi_t \tilde{P}_{t,i}} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] \\ &- \frac{1}{\gamma}\sum_{i=1}^n \pi_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) \\ s.t.\, x_{t+1} &= A_t x_t + B_t u_t + C_t \varepsilon_t \\ \varepsilon_t &\sim \sum_{i=1}^n \pi_{t,i}\tilde{P}_{t,i}.\end{aligned} \tag{3.38}$$

For the convenience that will be shown in the following argument, we use $\frac{2}{\gamma}$ instead of $\frac{1}{\gamma}$ in the (3.38). That is, we will consider

$$\begin{aligned}V_t(x_t) =& \min_{u_t} \max_{\tilde{P}_{t,i},\varepsilon_t \sim \sum_{i=1}^n \pi_t \tilde{P}_{t,i}} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] \\ &- \frac{2}{\gamma}\sum_{i=1}^n \pi_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) \\ s.t.\, x_{t+1} &= A_t x_t + B_t u_t + C_t \varepsilon_t \\ \varepsilon_t &\sim \sum_{i=1}^n \pi_{t,i}\tilde{P}_{t,i}.\end{aligned} \tag{3.39}$$

Theorem 3.6. (3.39) *has solution*

$$V_t(x_t) = x_t^T \Xi_t x_t + 2\xi_t^T x_t + z_t \tag{3.40}$$

and

$$u_t = U_t x_t + \bar{u}_t, \tag{3.41}$$

where $\Xi_t \in \mathbb{R}^{d_1,d_1}, U_t \in \mathbb{R}^{d_2,d_1}, \xi_t \in \mathbb{R}^{d_1}, \bar{u}_t \in \mathbb{R}^{d_2}, z_t \in \mathbb{R}$.

The closed form of the solution for $t < T$ *is*

$$\begin{aligned} U_t = -\Bigg(\sum_{i=1}^{n} \pi_{t,i} \Bigg(& B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \\ & + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\ & + R_t\Bigg)\Bigg)^{-1} \Bigg(\sum_{i=1}^{n} \pi_{t,i} \Bigg(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} \\ & + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \Bigg)\Bigg), \end{aligned} \tag{3.42}$$

$$\begin{aligned} \bar{u}_t = -\Bigg(\sum_{i=1}^{n} \pi_{t,i} \Bigg(& B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \\ & + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\ & + R_t\Bigg)\Bigg)^{-1} \Bigg(\sum_{i=1}^{n} \pi_{t,i} \Bigg(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) \\ & + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\ & + \left(B_t + C_t \tilde{B}_{t,i}\right)^T \xi_{t+1} \Bigg)\Bigg), \end{aligned} \tag{3.43}$$

$$
\begin{aligned}
\Xi_t = \sum_{i=1}^{n} \pi_{t,i} \Bigg(& (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t) \\
& + (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
& - \frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) + U_t^T R_t U_t \Bigg),
\end{aligned} \tag{3.44}
$$

$$
\begin{aligned}
\xi_t = \sum_{i=1}^{n} \pi_{t,i} \Bigg(& (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t \\
& + (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) + U_t^T R_t \bar{u}_t \\
& - \frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\
& + \left(A_t + C_t \tilde{A}_{t,i} + (B_t + C_t \tilde{B}_{t,i}) U_t\right)^T \xi_{t+1} \Bigg),
\end{aligned} \tag{3.45}
$$

$$
\begin{aligned}
z_t = \sum_{i=1}^{n} \pi_{t,i} \Bigg(& \bar{u}_t^T B_t^T (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t + 2 \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right)^T C_t^T \\
& (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t + \\
& \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) + \bar{u}_t^T R_t \bar{u}_t \\
& - \frac{1}{\gamma} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right)^T \Sigma_{t,i}^{-1} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\
& + 2 \xi_{t+1}^T \left(C_t \tilde{\mu}_{t,i} + \left(B_t + C_t \tilde{B}_{t,i}\right) \bar{u}_t\right) \\
& - \frac{1}{\gamma} \Bigg(\log|\Sigma_{t,i}| - \log\left|\left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1}\right| - d_3 \\
& + tr\left\{\Sigma_{t,i}^{-1} \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1}\right\} \Bigg) \\
& + C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} \Bigg) + z_{t+1},
\end{aligned} \tag{3.46}
$$

where

$$\tilde{A}_{t,i} = \gamma\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1} C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)A_t, \tag{3.47}$$

$$\tilde{B}_{t,i} = \gamma\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1} C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t, \tag{3.48}$$

$$\tilde{\mu}_{t,i} = \left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1} \Sigma_{t,i}^{-1}\mu_{t,i}, \tag{3.49}$$

$$\tilde{\xi}_{t+1,i} = \gamma\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1} C_t^T\xi_{t+1}. \tag{3.50}$$

For $t = T$, all these quantities are 0.

Proof. This can be proved by induction. Firstly (3.40) is true for $t = T$ with all relevant matrices and coefficients to be 0. Assume it holds for $t = m+1$, then for $t = m$,

$$\begin{aligned}
&V_t(x_t) \\
&= \min_{u_t} \max_{\tilde{P}_{t,i},\varepsilon_t\sim\sum_{i=1}^n \pi_t\tilde{P}_{t,i}} \mathbb{E}\left[x_{t+1}^T Q_{t+1}x_{t+1} + u_t^T R_t u_t + x_{t+1}^T\Xi_{t+1}x_{t+1} + 2\xi_{t+1}^T x_{t+1} + z_{t+1}\right] \\
&\quad - \frac{2}{\gamma}\sum_{i=1}^n \pi_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) \\
&= \min_{u_t}\sum_{i=1}^n \pi_{t,i}\Bigg(\max_{\tilde{P}_{t,i}}\mathbb{E}_{\varepsilon_t\sim\tilde{P}_{t,i}}\left[x_{t+1}^T Q_{t+1}x_{t+1} + u_t^T R_t u_t + x_{t+1}^T\Xi_{t+1}x_{t+1} + 2\xi_{t+1}^T x_{t+1} + z_{t+1}\right] \\
&\qquad - \frac{2}{\gamma}KL(\tilde{P}_{t,i}\|P_{t,i})\Bigg) + z_{t+1}.
\end{aligned} \tag{3.51}$$

We have

$$\begin{aligned}
&\max_{\tilde{P}_{t,i}}\mathbb{E}_{\varepsilon_t\sim\tilde{P}_{t,i}}\left[x_{t+1}^T Q_{t+1}x_{t+1} + u_t^T R_t u_t + x_{t+1}^T\Xi_{t+1}x_{t+1} + 2\xi_{t+1}^T x_{t+1}\right] - \frac{2}{\gamma}KL(\tilde{P}_{t,i}\|P_{t,i}) \\
&= \max_{\tilde{P}_{t,i}}\mathbb{E}_{\varepsilon_t\sim\tilde{P}_{t,i}}\Big[\left(A_t x_t + B_t u_t + C_t\varepsilon_t\right)^T\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t x_t + B_t u_t + C_t\varepsilon_t\right) + u_t^T R_t u_t \\
&\qquad + 2\xi_{t+1}^T\left(A_t x_t + B_t u_t + C_t\varepsilon_t\right)\Big] - \frac{2}{\gamma}KL(\tilde{P}_{t,i}\|P_{t,i}).
\end{aligned} \tag{3.52}$$

By Lemma 3.2, the maximum point $\tilde{P}_{t,i}$ satisfies

$$d\tilde{P}_{t,i} \sim \exp\left\{\frac{\gamma}{2}\Big((A_t x_t + B_t u_t + C_t \varepsilon_t)^T (Q_{t+1} + \Xi_{t+1})(A_t x_t + B_t u_t + C_t \varepsilon_t) \right. \\ \left. + u_t^T R_t u_t + 2\xi_{t+1}^T (A_t x_t + B_t u_t + C_t \varepsilon_t)\Big)\right\} \times dP_{t,i}. \tag{3.53}$$

Since $P_{t,i}$ is Gaussian, we get that $\tilde{P}_{t,i}$ is still Gaussian,

$$\tilde{P}_{t,i} \sim N\Bigg(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} \Big(\Sigma_{t,i}^{-1}\mu_{t,i} \\ + \gamma C_t^T ((Q_{t+1} + \Xi_{t+1})(A_t x_t + B_t u_t) + \xi_{t+1})\Big), \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1}\Bigg). \tag{3.54}$$

The KL-divergence between two Gaussian $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$ in $\mathbb{R}^{d_3}$ (Duchi, 2007) is

$$KL\left(N(\mu_1, \Sigma_1) \| N(\mu_2, \Sigma_2)\right) = \frac{1}{2}\left(\log\frac{|\Sigma_2|}{|\Sigma_1|} - d_3 + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) + tr\{\Sigma_2^{-1}\Sigma_1\}\right). \tag{3.55}$$

Plug (3.54) and (3.55) into (3.52), we have

$$
\begin{aligned}
&\max_{\tilde{P}_{t,i}} \mathbb{E}_{\varepsilon_t \sim \tilde{P}_{t,i}} \Big[\left(A_t x_t + B_t u_t + C_t \varepsilon_t\right)^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t + C_t \varepsilon_t\right) + u_t^T R_t u_t \\
&\qquad + 2\xi_{t+1}^T \left(A_t x_t + B_t u_t + C_t \varepsilon_t\right) \Big] - \frac{2}{\gamma} KL(\tilde{P}_{t,i} \| P_{t,i}) \\
&= \left(A_t x_t + B_t u_t\right)^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + 2 \bigg(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \\
&\left(\Sigma_{t,i}^{-1} \mu_{t,i} + \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + \gamma C_t^T \xi_{t+1}\right) \bigg)^T \\
&C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + \bigg(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \\
&\left(\Sigma_{t,i}^{-1} \mu_{t,i} + \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + \gamma C_t^T \xi_{t+1}\right) \bigg)^T \\
&C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \bigg(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \\
&\Big(\Sigma_{t,i}^{-1} \mu_{t,i} \\
&+ \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + \gamma C_t^T \xi_{t+1}\Big) \bigg) + u_t^T R_t u_t \\
&\qquad - \frac{1}{\gamma} \Bigg(\bigg(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \\
&\left(\Sigma_{t,i}^{-1} \mu_{t,i} + \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + \gamma C_t^T \xi_{t+1}\right) \bigg) - \mu_{t,i} \Bigg)^T \Sigma_{t,i}^{-1} \\
&\Bigg(\bigg(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \\
&\left(\Sigma_{t,i}^{-1} \mu_{t,i} + \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + \gamma C_t^T \xi_{t+1}\right) \bigg) - \mu_{t,i} \Bigg) \\
&\qquad + 2\xi_{t+1}^T \Bigg(A_t x_t + B_t u_t + C_t \bigg(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \\
&\left(\Sigma_{t,i}^{-1} \mu_{t,i} + \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t x_t + B_t u_t\right) + \gamma C_t^T \xi_{t+1}\right) \bigg) \Bigg) \\
&\qquad - \frac{1}{\gamma} \Bigg(\log|\Sigma_{t,i}| - \log \left| \left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \right| - d_3 \\
&+ tr\{ \Sigma_{t,i}^{-1} \left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \} \Bigg) + C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \Big(\Sigma_{t,i}^{-1} \\
&- \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\Big)^{-1}
\end{aligned}
\tag{3.56}
$$

$$
\begin{aligned}
&= (A_t x_t + B_t u_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t) \\
&\quad + 2 \left(\tilde{A}_{t,i} x_t + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t) \\
&\quad + \left(\tilde{A}_{t,i} x_t + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} x_t \right. \\
&\quad \left. + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right) + u_t^T R_t u_t - \frac{1}{\gamma} \left(\tilde{A}_{t,i} x_t + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i} \right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} x_t \right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \left. + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i} \right) \\
&\quad + 2\xi_{t+1}^T \left(A_t x_t + B_t u_t + C_t \tilde{\mu}_{t,i} + C_t \tilde{A}_{t,i} x_t + C_t \tilde{B}_{t,i} u_t \right) + Constant,
\end{aligned}
$$

where the constant includes everything that doesn't depend on x_t and u_t, such as the trace term in the KL distance between normal densities. (3.51) then becomes

$$
\begin{aligned}
V_t(x_t) = \min_{u_t} \sum_{i=1}^{n} \pi_{t,i} \Bigg(& (A_t x_t + B_t u_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t) \\
& + 2 \left(\tilde{A}_{t,i} x_t + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t x_t + B_t u_t) + \left(\tilde{A}_{t,i} x_t \right. \\
& \left. + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} x_t + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right) \\
& + u_t^T R_t u_t \\
& - \frac{1}{\gamma} \left(\tilde{A}_{t,i} x_t + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i} \right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} x_t + \tilde{B}_{t,i} u_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i} \right) \\
& + 2\xi_{t+1}^T \left(A_t x_t + B_t u_t + C_t \tilde{\mu}_{t,i} + C_t \tilde{A}_{t,i} x_t + C_t \tilde{B}_{t,i} u_t \right) \\
& - \frac{1}{\gamma} \left(\log|\Sigma_{t,i}^{-1}| - \log \left| \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \right)^{-1} \right| - d_3 \right. \\
& \left. + tr\{ \Sigma_{t,i} \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \right)^{-1} \} \right) \Bigg) + z_{t+1}.
\end{aligned}
\tag{3.57}
$$

The optimality condition on u_t gives

$$\begin{aligned}
&\left(\sum_{i=1}^{n}\pi_{t,i}\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t+\tilde{B}_{t,i}^TC_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t\right.\right.\\
&+B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_{t,i}+\tilde{B}_{t,i}^TC_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_{t,i}-\frac{1}{\gamma}\tilde{B}_{t,i}^T\Sigma_{t,i}^{-1}\tilde{B}_{t,i}\\
&\left.\left.+R_t\right)\right)u_t+\left(\sum_{i=1}^{n}\pi_{t,i}\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)A_t+B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{A}_{t,i}\right.\right.\\
&+\tilde{B}_{t,i}^TC_t^T\left(Q_{t+1}+\Xi_{t+1}\right)A_t+\tilde{B}_{t,i}^TC_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{A}_{t,i}\\
&\left.\left.-\frac{1}{\gamma}\tilde{B}_{t,i}\Sigma_{t,i}^{-1}\tilde{A}_{t,i}\right)\right)x_t+\left(\sum_{i=1}^{n}\pi_{t,i}\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\left(\tilde{\mu}_{t,i}+\tilde{\xi}_{t+1,i}\right)\right.\right.\\
&+\tilde{B}_{t,i}^TC_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\left(\tilde{\mu}_{t,i}+\tilde{\xi}_{t+1,i}\right)\\
&\left.\left.-\frac{1}{\gamma}\tilde{B}_{t,i}^T\Sigma_{t,i}^{-1}\left(\tilde{\mu}_{t,i}+\tilde{\xi}_{t+1,i}-\mu_{t,i}\right)+\left(B_t+C_t\tilde{B}_{t,i}\right)^T\xi_{t+1}\right)\right)=0.
\end{aligned} \tag{3.58}$$

Thus we have

$$
\begin{aligned}
u_t = &- \left(\sum_{i=1}^{n} \pi_{t,i} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t \right.\right. \\
&\quad + B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\
&\quad \left.\left. + R_t \right) \right)^{-1} \left(\sum_{i=1}^{n} \pi_{t,i} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) A_t + B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{A}_{t,i} \right.\right. \\
&\quad \left.\left. + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) A_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{A}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \right) \right) x_t \\
&- \left(\sum_{i=1}^{n} \pi_{t,i} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t \right.\right. \\
&\quad + B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\
&\quad \left.\left. + R_t \right) \right)^{-1} \left(\sum_{i=1}^{n} \pi_{t,i} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right) \right.\right. \\
&\quad + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} \right) - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i} \right) \\
&\quad \left.\left. + \left(B_t + C_t \tilde{B}_{t,i} \right)^T \xi_{t+1} \right) \right) \\
= &\; U_t x_t + \bar{u}_t.
\end{aligned} \tag{3.59}
$$

Plug (3.59) into (3.57) we have

$$
\begin{aligned}
V_t = x_t^T \Bigg(\sum_{i=1}^{n} \pi_{t,i} \Bigg(& (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t) \\
& + (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
& - \frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) + U_t^T R_t U_t \Bigg) \Bigg) x_t \\
+ 2 \Bigg(\sum_{i=1}^{n} \pi_{t,i} \Bigg(& (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t \\
& + (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) + U_t^T R_t \bar{u}_t \\
& - \frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\
& + \left(A_t + C_t \tilde{A}_{t,i} + (B_t + C_t \tilde{B}_{t,i}) U_t\right)^T \xi_{t+1} \Bigg) \Bigg)^T x_t \\
+ \sum_{i=1}^{n} \pi_{t,i} \Bigg(& \bar{u}_t^T B_t^T (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t + 2 \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right)^T C_t^T \\
& (Q_{t+1} + \Xi_{t+1}) B_t \bar{u}_t + \\
& \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) + \bar{u}_t^T R_t \bar{u}_t \\
& - \frac{1}{\gamma} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right)^T \Sigma_{t,i}^{-1} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\
& + 2 \xi_{t+1}^T \left(C_t \tilde{\mu}_{t,i} + \left(B_t + C_t \tilde{B}_{t,i}\right) \bar{u}_t \right) \\
& - \frac{1}{\gamma} \Bigg(\log|\Sigma_{t,i}| - \log\left| \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} \right| - d_3 \\
& + tr\left\{ \Sigma_{t,i}^{-1} \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} \right\} \Bigg) \\
& + C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} \Bigg) + z_{t+1} \\
= x_t^T \Xi_t x_t + 2 \xi_t^T x_t + z_t &
\end{aligned}
\tag{3.60}
$$

This completes the proof.

□

And in many settings, we will need to consider unequally uncertainties among distributions, which can be modeled as

$$\begin{aligned} V_t(x_t) = &\min_{u_t} \max_{\tilde{P}_{t,i}, \varepsilon_t \sim \sum_{i=1}^n \pi_t \tilde{P}_{t,i}} \mathbb{E}\left[x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t + V_{t+1}(x_{t+1})\right] \\ &- \sum_{i=1}^n \frac{2}{\gamma_i} \pi_{t,i} KL(\tilde{P}_{t,i} \| P_{t,i}) \\ &s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \\ &\varepsilon_t \sim \sum_{i=1}^n \pi_{t,i} \tilde{P}_{t,i}. \end{aligned} \tag{3.61}$$

Theorem 3.7. (3.61) *has solution*

$$V_t(x_t) = x_t^T \Xi_t x_t + 2\xi_t^T x_t + z_t \tag{3.62}$$

and

$$u_t = U_t x_t + \bar{u}_t, \tag{3.63}$$

where $\Xi_t \in \mathbb{R}^{d_1,d_1}, U_t \in \mathbb{R}^{d_2,d_1}, \xi_t \in \mathbb{R}^{d_1}, \bar{u}_t \in \mathbb{R}^{d_2}, z_t \in \mathbb{R}$*. For* $t = T$ *, all these matrices and coefficients are* 0*.*

The closed form of the solution for $t < T$ *is*

$$\begin{aligned} U_t = -\Bigg(\sum_{i=1}^n \pi_{t,i} \Big(&B_t^T (Q_{t+1} + \Xi_{t+1}) B_t + \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) B_t \\ &+ B_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{B}_{t,i} - \frac{1}{\gamma_i} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\ &+ R_t\Big)\Bigg)^{-1} \Bigg(\sum_{i=1}^n \pi_{t,i} \Big(B_t^T (Q_{t+1} + \Xi_{t+1}) A_t + B_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{A}_{t,i} \\ &+ \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) A_t + \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{A}_{t,i} - \frac{1}{\gamma_i} \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \Big)\Bigg), \end{aligned} \tag{3.64}$$

$$
\begin{aligned}
\bar{u}_t = -\Bigg(\sum_{i=1}^{n} \pi_{t,i} \Big(& B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \\
& + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma_i} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\
& + R_t \Big)\Bigg)^{-1} \Bigg(\sum_{i=1}^{n} \pi_{t,i} \Big(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) \\
& + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) - \frac{1}{\gamma_i} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\
& + \left(B_t + C_t \tilde{B}_{t,i}\right)^T \xi_{t+1} \Big)\Bigg),
\end{aligned} \tag{3.65}
$$

$$
\begin{aligned}
\Xi_t = \sum_{i=1}^{n} \pi_{t,i} \Big(& \left(A_t + B_t U_t\right)^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t + B_t U_t\right) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) \left(A_t + B_t U_t\right) \\
& + \left(A_t + B_t U_t\right)^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
& - \frac{1}{\gamma_i} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) + U_t^T R_t U_t \Big),
\end{aligned} \tag{3.66}
$$

$$
\begin{aligned}
\xi_t = \sum_{i=1}^{n} \pi_{t,i} \Big(& \left(A_t + B_t U_t\right)^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \bar{u}_t \\
& + \left(A_t + B_t U_t\right)^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \bar{u}_t \\
& + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) + U_t^T R_t \bar{u}_t \\
& - \frac{1}{\gamma_i} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\
& + \left(A_t + C_t \tilde{A}_{t,i} + (B_t + C_t \tilde{B}_{t,i}) U_t\right)^T \xi_{t+1} \Big),
\end{aligned} \tag{3.67}
$$

$$
\begin{aligned}
z_t = \sum_{i=1}^{n} \pi_{t,i} \Bigg(& \bar{u}_t^T B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \bar{u}_t + 2 \left(\tilde{B}_{t,i}\bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right)^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t \bar{u}_t + \\
& \left(\tilde{B}_{t,i}\bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right)^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\tilde{B}_{t,i}\bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) + \bar{u}_t^T R_t \bar{u}_t \\
& - \frac{1}{\gamma_i} \left(\tilde{B}_{t,i}\bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right)^T \Sigma_{t,i}^{-1} \left(\tilde{B}_{t,i}\bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\
& + 2\xi_{t+1}^T \left(C_t \tilde{\mu}_{t,i} + \left(B_t + C_t \tilde{B}_{t,i}\right) \bar{u}_t\right) \\
& - \frac{1}{\gamma_i} \Bigg(\log|\Sigma_{t,i}| - \log\left|\left(\Sigma_{t,i}^{-1} - \gamma_i C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1}\right| - d_3 \\
& + tr\left\{\Sigma_{t,i}^{-1} \left(\Sigma_{t,i}^{-1} - \gamma_i C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1}\right\}\Bigg) \\
& + C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \left(\Sigma_{t,i}^{-1} - \gamma_i C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1}\Bigg) + z_{t+1},
\end{aligned}
\tag{3.68}
$$

where

$$
\tilde{A}_{t,i} = \gamma_i \left(\Sigma_{t,i}^{-1} - \gamma_i C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t, \tag{3.69}
$$

$$
\tilde{B}_{t,i} = \gamma_i \left(\Sigma_{t,i}^{-1} - \gamma_i C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) B_t, \tag{3.70}
$$

$$
\tilde{\mu}_{t,i} = \left(\Sigma_{t,i}^{-1} - \gamma_i C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \Sigma_{t,i}^{-1} \mu_{t,i}, \tag{3.71}
$$

$$
\tilde{\xi}_{t+1,i} = \gamma_i \left(\Sigma_{t,i}^{-1} - \gamma_i C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} C_t^T \xi_{t+1}. \tag{3.72}
$$

3.5 With Uncertainty on Component Probability

The more general setting of RLQMG is with the uncertainty on π_t. But this problem is much more difficult than the case without uncertainty. To see this is a nonlinear programming problem, we can come back to (3.27). For the last stage $t = T-1$,

$$
\begin{aligned}
V_{T-1}(x_{T-1}) = \min_{u_{T-1}} \max_{\tilde{\pi}_{T-1}} \max_{\tilde{P}_{t,i}} \sum_{i=1}^{n} \tilde{\pi}_{t,i} \Bigg(& \mathbb{E}_{\varepsilon_{T-1} \sim \tilde{P}_{t,i}} \left[x_T^T Q_T x_T + u_{T-1}^T R_{T-1} u_{T-1}\right] \\
& - \frac{1}{\gamma} KL(\tilde{P}_{T-1,i} \| P_{T-1,i})\Bigg) - \frac{1}{\gamma} KL(\tilde{\pi}_{T-1} \| \pi_{T-1}).
\end{aligned}
\tag{3.73}
$$

By Lemma 3.2 and (3.40), fixing u_{T-1} and $\tilde{\pi}_{T-1}$ first, we have the inner maximum

$$\begin{aligned} &\max_{\tilde{P}_{t,i}} \mathbb{E}_{\varepsilon_{T-1} \sim \tilde{P}_{t,i}} \left[x_T^T Q_T x_T + u_{T-1}^T R_{T-1} u_{T-1} \right] \\ &\quad - \frac{1}{\gamma} KL(\tilde{P}_{T-1,i} \| P_{T-1,i}) = u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i}, \end{aligned} \tag{3.74}$$

where we use $\hat{\Xi}_{T-1,i} \in \mathbb{R}^{d_2,d_2}, \hat{\xi}_{T-1,i} \in \mathbb{R}^{d_2}, \hat{z}_{T-1,i} \in \mathbb{R}$ from Theorem 3.6 for fixed π_t, these matrices and coefficients are constant functions of u_{T-1}.

Then by Lemma 3.2, the maximum point $\tilde{\pi}$ in the outer maximum problem satisfies

$$\tilde{\pi}_{T-1,i} \sim \pi_{T-1,i} \times \exp\left\{ \gamma \left(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \right) \right\}, \tag{3.75}$$

which means

$$\tilde{\pi}_{T-1,i} = \frac{\pi_{T-1,i} \times \exp\left\{ \gamma \left(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \right) \right\}}{\sum_{i=1}^n \pi_{T-1,i} \times \exp\left\{ \gamma \left(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \right) \right\}}. \tag{3.76}$$

Plug (3.76) back into (3.73) we have

$$\begin{aligned} &V_{T-1}(x_{T-1}) \\ &= \min_{u_{T-1}} \sum_{i=1}^n \frac{\pi_{T-1,i} \times \exp\left\{ \gamma \left(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \right) \right\}}{\sum_{i=1}^n \pi_{T-1,i} \times \exp\left\{ \gamma \left(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \right) \right\}} \\ &\Bigg(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \\ &\quad - \frac{1}{\gamma} \Bigg(\log \left(\frac{\pi_{T-1,i} \times \exp\left\{ \gamma \left(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \right) \right\}}{\sum_{i=1}^n \pi_{T-1,i} \times \exp\left\{ \gamma \left(u_{T-1}^T \hat{\Xi}_{T-1,i} u_{T-1} + 2\hat{\xi}_{T-1,i}^T u_{T-1} + \hat{z}_{T-1,i} \right) \right\}} \right) \\ &\qquad - \log\left(\pi_{t,i}\right) \Bigg) \Bigg). \end{aligned} \tag{3.77}$$

Thus, we can see even for one stage, it is a nonlinear optimization problem since u_T appears at the top and the bottom of the fraction. Therefore (a) we no longer have the closed-form solution for the risk-averse problem, and thus (b) since the function is certainly not quadratic, we cannot easily write the $T-2$ step algebra. While we do not carry the complete calculation here, the main difficulty is that, since we cannot solve the problem recursively in closed form, the resulting expression of $V(\cdot)$ has a combinatorial dependence on $\tilde{\pi}_{t,i}$ on x_t which is intractable when numbers of stage increase. Instead, we come back to (3.33) and consider an relaxation,

$$\begin{aligned} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{\pi}_t,\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_t)} \mathbb{E}\Bigg[\sum_{t=0}^{T-1} \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \\ - \frac{1}{\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i} \| P_{t,i}) - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{\pi}_t \| \pi_t) \Bigg] \\ s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t, \end{aligned} \tag{3.78}$$

where the density of the joint distribution of noises $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1})$ is

$$\sum_{i_0,i_1,\ldots,i_{T-1}} \prod_{t\,=0}^{T-1} \tilde{\pi}_{t,i_t} d\tilde{P}_{t,i_t}(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1}). \tag{3.79}$$

That is, we consider $\tilde{\pi}_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_t)$ to be a constant instead of a function. We will call this semi-independence relaxation, as it essentially says the previous state doesn't influence the distribution of π_t, or mathematically

$$P(I_t = i | (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})) = P(I_t = i), \tag{3.80}$$

where I_t is the selected component of t-th stage. As our original distribution has an independence assumption among π_t and $P_{t,i}$, this relaxation just imposes the original assumption to the $\tilde{\pi}_t$ in risk-averse distribution. As independence is usually solid in models like Brownian motion, the consideration of correlated risk-averse distribution is instead redundant and unnecessary. This relaxation can also be understood from a corner case where Gaussians shrinkage to points, then (3.78) is approximately

$$\begin{aligned} &\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{\pi}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t \| \pi_t) \\ &\qquad s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t. \end{aligned} \tag{3.81}$$

Compared to Theorem 3.3, we see that (3.81) is still doing distributionally robust control but with an independent structure. If we removed the independent relaxation, then it becomes

$$\begin{aligned} &\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma} KL(\pi \| \prod_{t=0}^{T-1} \pi_t) \\ &\qquad s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t, \end{aligned} \tag{3.82}$$

where π will have n^T parameters and becomes intractable. What our relaxation does is to move from **questioning the structure of the ground truth distribution** to **questioning the accuracy of parameter estimations**.

We are going to show in this section that

Theorem 3.8. $\exists \gamma_0, \forall \gamma < \gamma_0$,

$$\begin{aligned} &\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{\pi}_t, \tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right. \\ &\quad \left. - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i} \| P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t \| \pi_t)\right] \end{aligned} \tag{3.83a}$$

$$\begin{aligned} &= \max_{\tilde{\pi}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right. \\ &\quad \left. - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i} \| P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t \| \pi_t)\right]. \end{aligned} \tag{3.83b}$$

The proof of Theorem 3.8 will require a few auxiliary theorems, and we will leave it at the end of this section. We provide a sketch of the proof here first, *Proof Sketch.* We start from the maximin on the right hand side, (3.83b)

$$\begin{aligned}&\max_{\tilde{\pi}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right. \\ &\left. - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\right].\end{aligned} \tag{3.84}$$

The proof will proceed as follows

- As we showed in Theorem 3.6, at fixed $\tilde{\pi}_t$, the optimal control policy u_t will be a linear function of x_t. As x_t is linear in $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})$, it results that u_t itself is a linear function of $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})$. We can thus parameterize u_t as a linear function; which we prove in **Theorem 3.9**.

- We will prove in **Lemma 3.10** and **Lemma 3.11** that this function is continuous on γ, which, for γ small enough gives us a bound on the resulting parameters defining u_t. We will then show, also in **Lemma 3.11**, that this bound is also independent of $\tilde{\pi}_t$. Subsequently, using the bounds as constraints on the inner minimization problem will not affect either the optimal solution of the inner and outer problems.

- Subsequently, in **Theorem 3.12**, we show that the inner problem in (3.84) is equivalent to one stated over a bounded domain. Note that the bounds will not themselves be attained, this serves to emphasize that the solution of the inner problem will always be in a compact set for the assumptions stated.

- Subsequently, in **Theorem 3.13** we show that the outer function in (3.84) (which we will then maximize with respect to $\tilde{\pi}_t$) is in effect concave over the compact set defined at the previous point by computing and bounding the corresponding Hessian. The bound on the parameters defining u_t makes the Hessian of the first two terms in (3.84) also bounded. But the Hessian of the last term is $-\frac{1}{\gamma} diag\{\frac{1}{\tilde{\pi}_{0,0}}, \frac{1}{\tilde{\pi}_{0,1}}, \ldots, \frac{1}{\tilde{\pi}_{T-1,n}}\}$, which is smaller than $-\frac{1}{\gamma}\mathbf{I}$. Thus, when $\gamma \to 0$, we must have a negative-definite Hessian.

- Finally, we have convexity on u_t and concavity on $\tilde{\pi}_t$. Applying Sion's minimax theorem (Sion, 1958), as the constraint from the bound doesn't change optimality, we complete the proof of **Theorem 3.8**.

□

Note that the inner minimax problem of (3.83b) and (3.84) is the solution to the problem with prescribed cluster weights, $\pi_t = \tilde{\pi}_t$, which we solved in (3.39) with u_t as a function of $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})$. Applying the results from (3.39), we get the following.

Theorem 3.9. *The optimal control of*

$$\begin{aligned} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\Bigg[&\sum_{t=0}^{T-1} \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \\ &- \frac{1}{\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i} \| P_{t,i}) \Bigg] \end{aligned} \tag{3.85}$$

is a linear decision rule

$$u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1}) = H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau} \varepsilon_\tau + h_t, \tag{3.86}$$

where

$$H_t = U_t \prod_{s=0}^{t-1} \left(A_s + B_s U_s \right), \tag{3.87}$$

$$H_{t,\tau} = \prod_{s=\tau+1}^{t} \left(A_s + B_s U_s \right) C_\tau \tag{3.88}$$

and

$$h_t = U_t \sum_{w=0}^{t-1} \prod_{s=w+1}^{t-1} \left(A_s + B_s U_s \right) B_w \bar{u}_w + \bar{u}_t. \tag{3.89}$$

U_t and $\bar{u}_t$ are defined in Theorem 3.6 with replacing π_t by $\tilde{\pi}_t$. The definition of the product on matrices multiplication is

$$\prod_{s=i}^{j} \left(A_s + B_s U_s \right) = \begin{cases} \left(A_j + B_j U_j \right) \left(A_{j-1} + B_{j-1} U_{j-1} \right) \cdots \left(A_i + B_i U_i \right) & , j \geq i \\ \mathbf{I} & , j = i - 1. \end{cases} \tag{3.90}$$

Proof. The optimal control of (3.39) is a linear decision rule which depends on x_t,

$$u_t(x_t) = U_t x_t + \bar{u}_t. \tag{3.91}$$

As

$$x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t, \tag{3.92}$$

we have

$$\begin{aligned} x_{t+1} &= A_t x_t + B_t(U_t x_t + \bar{u}_t) + C_t \varepsilon_t \\ &= (A_t + B_t U_t) x_t + B_t \bar{u}_t + C_t \varepsilon_t \\ &= \prod_{s=0}^{t} (A_s + B_s U_s)\, x_0 + \sum_{w=0}^{t} \prod_{s=w+1}^{t} (A_s + B_s U_s)\, B_w \bar{u}_w + \sum_{w=0}^{t} \prod_{s=w+1}^{t} (A_s + B_s U_s)\, C_w \varepsilon_w \end{aligned} \tag{3.93}$$

Plug (3.93) into (3.91), we have

$$\begin{aligned} u_t &= U_t \prod_{s=0}^{t-1} (A_s + B_s U_s)\, x_0 + U_t \sum_{w=0}^{t-1} \prod_{s=w+1}^{t-1} (A_s + B_s U_s)\, B_w \bar{u}_w \\ &\quad + U_t \sum_{w=0}^{t-1} \prod_{s=w+1}^{t-1} (A_s + B_s U_s)\, C_w \varepsilon_w + \bar{u}_t \\ &= H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau} \varepsilon_\tau + h_t. \end{aligned} \tag{3.94}$$

□

We now show that the parameters defining the solution in Theorem 3.6 are uniformly bounded with the values of the cluster weights, for all γ small enough.

Lemma 3.10. *$\exists \gamma_0 > 0, L, S > 0$, such that, $\forall \gamma < \gamma_0$, for all values of $\pi_0, \pi_1, \ldots, \pi_{T-1}$ with $0 \le \pi_{t,i} \le 1$ and $\sum_{i=1}^{n} \pi_{t,i} = 1$, we have*

$$\|U_t\| \le L, \forall t \tag{3.95}$$

and

$$\|\Xi_t\| \le S, \forall t. \tag{3.96}$$

Here, U_t and Ξ_t are the quantities defined in Theorem 3.6.

Proof. We are going to prove this recursively and show that $\exists L_t > 0, S_t > 0, t = 0, 1, \ldots, T-1$, under the same condition of the lemma,

$$\|U_t\| \leq L_t, \forall t, \tag{3.97}$$

$$\|\Xi_t\| \leq S_t, \forall t, \tag{3.98}$$

where Ξ_t is defined in (3.44). Without loss of generality, we define $\Xi_T = 0$ and $S_T = 0$ as $V_T(x_T) = 0$. Then by (3.42),

$$\begin{aligned} U_t = -\Bigg(\sum_{i=1}^{n} \pi_{t,i} \Big(& B_t^T (Q_{t+1} + \Xi_{t+1}) B_t + \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) B_t \\ & + B_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\ & + R_t \Big)\Bigg)^{-1} \Bigg(\sum_{i=1}^{n} \pi_{t,i} \Big(B_t^T (Q_{t+1} + \Xi_{t+1}) A_t + B_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{A}_{t,i} \\ & + \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) A_t + \tilde{B}_{t,i}^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \tilde{A}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \Big)\Bigg), \end{aligned} \tag{3.99}$$

where $\tilde{A}_{t,i}$ and $\tilde{B}_{t,i}$ are defined in (3.47) and (3.48),

$$\tilde{A}_{t,i} = \gamma \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} C_t^T (Q_{t+1} + \Xi_{t+1}) A_t, \tag{3.100}$$

$$\tilde{B}_{t,i} = \gamma \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1}) C_t\right)^{-1} C_t^T (Q_{t+1} + \Xi_{t+1}) B_t. \tag{3.101}$$

(3.100) gives an upper bound for $\|\tilde{A}_{t,i}\|$,

$$\|\tilde{A}_{t,i}\| \leq \gamma \left\| \left(\Sigma_{t,i}^{-1} - \gamma C_t^T (Q_{t+1} + \Xi_{t+1})\right)^{-1} \right\| \left\| C_t^T (Q_{t+1} + \Xi_{t+1}) A_t \right\|. \tag{3.102}$$

For any stage t, if we have a γ_0^{t+1} and S_{t+1} such that $\forall \gamma < \gamma_0^{t+1}$, $\|\Xi_{t+1}\| \leq S_{t+1}$, then $\forall \gamma < \gamma_0^{t+1}, \exists m_t$, s.t.

$$\left\| C_t^T (Q_{t+1} + \Xi_{t+1}) A_t \right\| \leq m_t. \tag{3.103}$$

Also, as

$$\begin{aligned}\left\|\left(\Sigma_{t,i}^{-1}-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\right)^{-1}\right\| &= \left\|\Sigma_{t,i}^{\frac{1}{2}}\left(\mathbf{I}-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\right)^{-1}\Sigma_{t,i}^{\frac{1}{2}}\right\| \\ &\leq \left\|\Sigma_{t,i}^{\frac{1}{2}}\right\|\left\|\left(\mathbf{I}-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\right)^{-1}\right\|\left\|\Sigma_{t,i}^{\frac{1}{2}}\right\| \\ &= \left\|\Sigma_{t,i}\right\|\left\|\mathbf{I}-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\right\|^{-1},\end{aligned} \tag{3.104}$$

and $\|\Xi_{t+1}\|\leq S_{t+1}$, we can choose a $\hat{\gamma}_t^a < \gamma_0^{t+1}$, which we can find an lower bound n_t, $\forall\gamma < \hat{\gamma}_t^a$,

$$\begin{aligned}\left\|\mathbf{I}-\gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\right\| &\geq \left\|\mathbf{I}\right\|-\gamma\left\|C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)\right\| \\ &\geq n_t.\end{aligned} \tag{3.105}$$

Then by (3.102), $\forall\gamma < \hat{\gamma}_t^a$, we have

$$\|\tilde{A}_{t,i}\|\leq \gamma m_t n_t = \gamma a_t, \tag{3.106}$$

where $a_t = m_t n_t$ is a constant not depending on π. Similarly, we can find $\hat{\gamma}_t^b < \gamma_0^{t+1}$ and b_t, $\forall\gamma < \hat{\gamma}_t^b$, we have

$$\|\tilde{B}_{t,i}\|\leq \gamma b_t, \tag{3.107}$$

where b_t is a constant not depending on π. Now we come back to (3.99),

$$\begin{aligned}&\|U_t\| \\ &\leq \left\|\left(\sum_{i=1}^n \pi_{t,i}\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t+\tilde{B}_{t,i}^T C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)B_t+B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_{t,i}\right.\right.\right. \\ &\left.\left.\left.+\tilde{B}_{t,i}^T C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{B}_{t,i}-\frac{1}{\gamma}\tilde{B}_{t,i}^T\Sigma_{t,i}^{-1}\tilde{B}_{t,i}+R_t\right)\right)^{-1}\right\|\left\|\sum_{i=1}^n \pi_{t,i}\left(B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)A_t\right.\right. \\ &+B_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{A}_{t,i}+\tilde{B}_{t,i}^T C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)A_t+\tilde{B}_{t,i}^T C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\tilde{A}_{t,i} \\ &\left.\left.-\frac{1}{\gamma}\tilde{B}_{t,i}\Sigma_{t,i}^{-1}\tilde{A}_{t,i}\right)\right\|.\end{aligned} \tag{3.108}$$

If we choose $\gamma < \min\{\hat{\gamma}_t^a, \hat{\gamma}_t^b\}$, for the second term, as $0 \le \pi_{t,i} \le 1, \sum_{i=1}^n \pi_{t,i} = 1$, we have

$$\begin{aligned}
&\left\| \sum_{i=1}^n \pi_{t,i} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} \right.\right. \\
&\quad \left.\left. + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \right) \right\| \\
&\quad \le \max_i \left\| B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t \right. \\
&\qquad\qquad \left. + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \right\|.
\end{aligned} \tag{3.109}$$

For any i, by triangle inequality,

$$\begin{aligned}
&\left\| \left(B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t + B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t \right.\right. \\
&\quad \left.\left. + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \right) \right\| \\
&\quad \le \left\| B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t \right\| + \left\| B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} \right\| \\
&\qquad + \left\| \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t \right\| + \left\| \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \tilde{A}_{t,i} \right\| + \frac{1}{\gamma} \left\| \tilde{B}_{t,i} \Sigma_{t,i}^{-1} \tilde{A}_{t,i} \right\|.
\end{aligned} \tag{3.110}$$

Plug in (3.106) and (3.107), we have

$$\begin{aligned}
(3.110) \le &\left\| B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t \right\| + \gamma a_t \left\| B_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \right\| \\
&+ \gamma b_t \left\| C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) A_t \right\| + \gamma^2 a_t b_t \left\| C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t \right\| + \gamma a_t b_t \left\| \Sigma_{t,i}^{-1} \right\|.
\end{aligned} \tag{3.111}$$

As $\|\Xi_{t+1}\| \le S_{t+1}$, and the RHS of (3.111) depends continuously on γ, we can find a $c_t > 0$, such that $\forall \gamma < \min\{\hat{\gamma}_t^a, \hat{\gamma}_t^b\}$,

$$(3.111) \le c_t, \tag{3.112}$$

where c_t is a constant not depending on π. For the first term,

$$\left\| \left(\sum_{i=1}^{n} \pi_{t,i} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} \right.\right.\right.$$
$$\left.\left.\left. + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} + R_t \right) \right)^{-1} \right\| \tag{3.113}$$

will be bounded by the inverse of the minimum eigenvalue of

$$\sum_{i=1}^{n} \pi_{t,i} \left(B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} \right.$$
$$\left. + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} + R_t \right). \tag{3.114}$$

Consider the $i-th$ term of (3.114), as

$$\begin{aligned} &\tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + B_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \tilde{B}_{t,i} \\ &\quad = \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t + \left(\tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t \right)^T, \end{aligned} \tag{3.115}$$

and

$$\begin{aligned} \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t &= \left(\gamma \left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \right)^{-1} C_t^T \left(Q_{t+1} \right.\right. \\ &\qquad \left.\left. + \Xi_{t+1} \right) B_t \right)^T C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t \\ &= \gamma B_t^T \left(Q_{t+1} + \Xi_{t+1} \right)^T C_t \left(\Sigma_{t,i}^{-1} \right. \\ &\qquad \left. - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \right)^{-1} C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) B_t, \end{aligned} \tag{3.116}$$

we can choose $\hat{\gamma}_t^c < \min\{\hat{\gamma}_t^a, \hat{\gamma}_t^b\}$ such that for all i,

$$\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1} \right) C_t \geq 0. \tag{3.117}$$

Then we have

$$(3.115) \geq 0. \tag{3.118}$$

For any x,

$$\begin{aligned} &\pi_{t,i}x^T \Big(B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t \\ &+ B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma}\tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\ &+ R_t \Big) x \ge \pi_{t,i}x^T \left(B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t - \frac{1}{\gamma}\tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \right) x. \end{aligned} \tag{3.119}$$

As $Q_{t+1} > 0$ and $\tilde{B}_{t,i} \le \gamma b_t$, we can choose $\hat{\gamma}_t^d < \hat{\gamma}_t^c$, such that for all $\gamma < \hat{\gamma}_t^d$ and all i,

$$\left\| \frac{1}{\gamma}\tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \right\| \le \frac{1}{2} \min_x x^T \left(Q_{t+1}+\Xi_{t+1}\right) x. \tag{3.120}$$

Then we have

$$\begin{aligned} (3.119) &\ge \pi_{t,i} \frac{1}{2} \min_x x^T \left(Q_{t+1}+\Xi_{t+1}\right) x \\ &\ge \frac{1}{2}\lambda_{\min}(Q_{t+1}) \\ &> 0. \end{aligned} \tag{3.121}$$

As $0 \le \pi_{t,i} \le 1, \sum_{i=1}^n \pi_{t,i} = 1$, we have

$$\begin{aligned} &\sum_{i=1}^n \pi_{t,i}x^T \Big(B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t + B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} \\ &\qquad + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma}\tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} + R_t \Big) x \\ &\ge \min_i x^T \Big(B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t + B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} \\ &\qquad + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma}\tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} + R_t \Big) x \\ &\ge \frac{1}{2}\lambda_{\min}(Q_{t+1}). \end{aligned} \tag{3.122}$$

Thus,

$$(3.113) \le 2\frac{1}{\lambda_{\min}(Q_{t+1})}. \tag{3.123}$$

Combine this with (3.112), we have $\forall \gamma < \hat{\gamma}_t^d$,

$$\|U_t\| \le 2\frac{1}{\lambda_{\min}(Q_{t+1})} c_t, \tag{3.124}$$

where c_t is a constant not depending on π. Thus we pick

$$L_t = 2\frac{1}{\lambda_{\min}(Q_{t+1})}c_t. \tag{3.125}$$

Now we come to find S_t for Ξ_t. The triangle inequality gives

$$\begin{aligned}
\|\Xi_t\| = \Bigg\|\sum_{i=1}^{n} \pi_{t,i} \Bigg(&(A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t) \\
&+ \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t) \\
&+ (A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
&+ \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) \\
&- \frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right) + U_t^T R_t U_t \Bigg)\Bigg\| \\
\leq \sum_{i=1}^{n} \pi_{t,i} \Bigg(&\left\|(A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t)\right\| \\
&+ \left\|\left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t)\right\| \\
&+ \left\|(A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| \\
&+ \left\|\left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| \\
&+ \left\|\frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| + \left\|U_t^T R_t U_t\right\|\Bigg).
\end{aligned} \tag{3.126}$$

As $0 \leq \pi_{t,i} \leq 1, \sum_{i=1}^{n} \pi_{t,i} = 1$, we have

$$\begin{aligned}
\|\Xi_t\| \leq \max_i &\left\|(A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t)\right\| \\
&+ \left\|\left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t)\right\| \\
&+ \left\|(A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| \\
&+ \left\|\left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| \\
&+ \left\|\frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| + \left\|U_t^T R_t U_t\right\|.
\end{aligned} \tag{3.127}$$

As $\left\|\tilde{A}_{t,i}\right\|$ and $\left\|\tilde{B}_{t,i}\right\|$ are in $O(\gamma)$, when $\gamma \to 0$, RHS of (3.127) will converge to a constant. Note that RHS is a continuous function of γ in $(0, \infty)$. Thus, we can pick $\gamma_0^t = \hat{\gamma}_t^d$ and

$$\begin{aligned} S_t = \max_{0<\gamma<\gamma_0^t} \max_i & \left\|(A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t)\right\| \\ & + \left\|\left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) (A_t + B_t U_t)\right\| \\ & + \left\|(A_t + B_t U_t)^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| \\ & + \left\|\left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T (Q_{t+1} + \Xi_{t+1}) C_t \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| \\ & + \left\|\frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)\right\| + \left\|U_t^T R_t U_t\right\|. \end{aligned} \tag{3.128}$$

So far, we have gotten L_t and S_t for the stage t. By induction, we can complete the proof for (3.97) and (3.98) for all t. By choosing $\gamma_0 = \gamma_0^0, L = \max_t L_t, S = \max_t S_t$, we complete the proof. □

We now prove that the parameters defining the optimal policy in Theorem 3.9 are uniformly bounded with respect to the cluster weights for all γ small enough.

Lemma 3.11. *$\exists \gamma_0 > 0, M_1, M_2, M_3 > 0$, such that, $\forall \gamma < \gamma_0$, for all values of $\pi_0, \pi_1, \ldots, \pi_{T-1}$ with $0 \le \pi_{t,i} \le 1$ and $\sum_{i=1}^n \pi_{t,i} = 1$, we have*

$$\|H_{t,\tau}\| \le M_1, \forall t, \tau, \tag{3.129}$$

$$\|H_t\| \le M_2, \forall t, \tag{3.130}$$

$$\|h_t\| \le M_3, \forall t. \tag{3.131}$$

Here, the quantities $H_{t,\tau}$, H_t, and h_t parameterize the linear policy defined in Theorem 3.9.

Proof. With γ_0 from Lemma 3.10, we have

$$\begin{aligned} \|H_{t,\tau}\| &= \left\| \prod_{s=\tau+1}^{t} (A_s + B_s U_s) C_\tau \right\| \\ &\le \max\{\|C_t\|, \prod_{s=\tau+1}^{t} (\|A_s\| + \|B_s\| \|U_s\|) \|C_\tau\|\}. \end{aligned} \tag{3.132}$$

By (3.95),

$$\left\|H_{t,\tau}\right\| \leq \max\{\|C_t\|, \prod_{s=\tau+1}^{t} \left(\|A_s\|+\|B_s\|L\right)\|C_\tau\|\}. \tag{3.133}$$

Choosing

$$M_1 = \max_{t,\tau} \max\{\|C_t\|, \prod_{s=\tau+1}^{t} \left(\|A_s\|+\|B_s\|L\right)\|C_\tau\|\} \tag{3.134}$$

gives (3.129). The argument is similar for H_t. And for h_t, note

$$\begin{aligned}\bar{u}_t = -\Bigg(\sum_{i=1}^{n} \pi_{t,i} \bigg(& B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t \\ & + B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \tilde{B}_{t,i} - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \tilde{B}_{t,i} \\ & + R_t\bigg)\Bigg)^{-1} \Bigg(\sum_{i=1}^{n} \pi_{t,i} \bigg(B_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) \\ & + \tilde{B}_{t,i}^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) - \frac{1}{\gamma} \tilde{B}_{t,i}^T \Sigma_{t,i}^{-1} \left(\tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\ & + \left(B_t + C_t \tilde{B}_{t,i}\right)^T \xi_{t+1}\bigg)\Bigg),\end{aligned} \tag{3.135}$$

where

$$\tilde{\mu}_{t,i} = \left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t\right)^{-1} \Sigma_{t,i}^{-1} \mu_{t,i}, \tag{3.136}$$

$$\tilde{\xi}_{t+1,i} = \gamma \left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t\right)^{-1} C_t^T \xi_{t+1}, \tag{3.137}$$

$$\begin{aligned}\xi_t = \sum_{i=1}^{n} \pi_{t,i} \bigg(& \left(A_t + B_t U_t\right)^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t \bar{u}_t \\ & + \left(A_t + B_t U_t\right)^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) \\ & + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) B_t \bar{u}_t \\ & + \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T C_t^T \left(Q_{t+1}+\Xi_{t+1}\right) C_t \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i}\right) + U_t^T R_t \bar{u}_t \\ & - \frac{1}{\gamma} \left(\tilde{A}_{t,i} + \tilde{B}_{t,i} U_t\right)^T \Sigma_{t,i}^{-1} \left(\tilde{B}_{t,i} \bar{u}_t + \tilde{\mu}_{t,i} + \tilde{\xi}_{t+1,i} - \mu_{t,i}\right) \\ & + \left(A_t + C_t \tilde{A}_{t,i} + (B_t + C_t \tilde{B}_{t,i}) U_t\right)^T \xi_{t+1}\bigg).\end{aligned} \tag{3.138}$$

The inverse part in (3.135) has been discussed in (3.113). The norm of the second part can be bounded simply by triangle inequality, providing that

$$\|\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\|^{-1} \tag{3.139}$$

is bounded, which has also been discussed in (3.104). □

We now show that the minimax problem we described in Theorem 3.9 can equivalently be solved on a compact set.

Theorem 3.12. $\exists \gamma_0, M_1, M_2, M_3, N$, *such that,* $\forall \gamma < \gamma_0$, (3.85) *is equivalent to*

$$\begin{aligned}
&\min_{(H_t, H_{t,\tau}, h_t)\in\mathcal{H}} \max_{K_{t,i}\in\mathcal{K}, \tilde{P}_{t,i}=N(\hat{\mu}_{t,i},\hat{\Sigma}_{t,i})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right.\\
&\qquad\qquad \left. - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i})\right]\\
&\qquad s.t. u_t(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{t-1}) = H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau}\varepsilon_\tau + h_t, \forall t,\\
&\qquad \hat{\mu}_{t,i} = \left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1} \Sigma_{t,i}^{-1}\mu_{t,i} + \gamma K_{t,i}\begin{pmatrix} 1 \\ \varepsilon_0 \\ \varepsilon_1 \\ \dots \\ \varepsilon_{t-1}\end{pmatrix}, \forall t, i,\\
&\qquad \hat{\Sigma}_{t,i} = \left(\Sigma_{t,i}^{-1} - \gamma C_t^T \left(Q_{t+1} + \Xi_{t+1}\right) C_t\right)^{-1}, \forall t, i,
\end{aligned} \tag{3.140}$$

where

$$\mathcal{H} = \left\{(H_t, H_{t,\tau}, h_t) | \|H_{t,\tau}\| \leq M_1, \|H_t\| \leq M_2, \|h_t\| \leq M_3\right\}, \tag{3.141}$$

$$\mathcal{K} = \left\{K_{t,i} | \|K_{t,i}\| \leq N\right\}. \tag{3.142}$$

M_1, M_2, M_3, N *are quantities that don't depend on the value of* π.

Proof. We can choose γ_0 to be γ_0 in Lemma 3.11. Then, by Theorem 3.9 and Lemma 3.11, we have (3.85) equal to

$$\begin{aligned} \min_{(H_t,H_{t,\tau},h_t)\in\mathcal{H}} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{T-1})} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) \\ - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i})\Bigg] \\ s.t. u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1}) = H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau}\varepsilon_\tau + h_t, \forall t. \end{aligned} \tag{3.143}$$

By (3.54), the maximum point $\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})$ is a Gaussian with mean

$$\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}\left(\Sigma_{t,i}^{-1}\mu_{t,i} + \gamma C_t^T\left(\left(Q_{t+1}+\Xi_{t+1}\right)\left(A_t x_t + B_t u_t\right) + \xi_{t+1}\right)\right) \tag{3.144}$$

and covariance

$$\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}, \tag{3.145}$$

where ξ_t is defined in (3.45). Expand x_t with dynamics and plug (3.86) into (3.144), we can see the mean is a linear function of $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1}$, which gives $K_{t,i}$. As $H_t, H_{t,\tau}, h_t$ are bounded, $K_{t,i}$ are naturally bounded by a function of M_1, M_2, M_3. □

After all the bounds we obtained before, we are now ready to prove the concavity result, which will allow us to invoke duality results and thus reverse the order of the min max.

Theorem 3.13. *With* $\gamma_0, \mathcal{H}, \mathcal{K}$ *defined in Theorem 3.12,* $\exists\gamma_1 < \gamma_0$*, such that,* $\forall\gamma < \gamma_1$ *and* $(H_t, H_{t,\tau}, h_t) \in \mathcal{H}$*, we have that*

$$\begin{aligned} \max_{K_{t,i}\in\mathcal{K}, \tilde{P}_{t,i}=N(\hat{\mu}_{t,i},\hat{\Sigma}_{t,i})} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) \\ - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\Bigg] \end{aligned} \tag{3.146}$$

is concave in $\tilde{\pi}_{t,i}$.

Proof. By plugging dynamics and (3.86) into (3.146), (3.146) can be written as

$$\begin{aligned} \max_{K_{t,i} \in \mathcal{K}, \tilde{P}_{t,i}=N(\hat{\mu}_{t,i},\hat{\Sigma}_{t,i})} & x_0^T \Lambda_0 x_0 + \sum_{t=0}^{T-1} x_0^T \Lambda_{1,t} \mathbb{E}\left[\varepsilon_t\right] + \sum_{t=0}^{T-1}\sum_{\tau=t}^{T-1} \mathbb{E}\left[\varepsilon_t\right]^T \Lambda_{2,t,\tau} \mathbb{E}\left[\varepsilon_\tau\right] \\ + \sum_{t=0}^{T-1} \mathbb{E}\left[\varepsilon_t^T \Lambda_{3,t} \varepsilon_t\right] & - \mathbb{E}\left[\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\right], \end{aligned} \tag{3.147}$$

where $\Lambda_0, \Lambda_{1,t}, \Lambda_{2,t,\tau}, \Lambda_{3,t}$ are matrices as a function of $(H_t, H_{t,\tau}, h_t)$ and not depending on $\tilde{\pi}_{t,i}, K_{t,i}, \hat{\Sigma}_{t,i}$. As we will not use their values, we do not give them here. Note that $\mathcal{K}$ is a compact set, by Danskin's theorem (Danskin, 1966), at any given point $(H_t, H_{t,\tau}, h_t) \in \mathcal{H}$, the derivative of (3.147) w.r.t. $\tilde{\pi}_{t,i}$ is

$$\begin{aligned} & \sum_{t=0}^{T-1} x_0^T \Lambda_{1,t} \nabla_\pi \mathbb{E}\left[\varepsilon_t\right] + \sum_{t=0}^{T-1}\sum_{\tau=t}^{T-1} \nabla_\pi \left(\mathbb{E}\left[\varepsilon_t\right]^T \Lambda_{2,t,\tau}\mathbb{E}\left[\varepsilon_\tau\right]\right) + \sum_{t=0}^{T-1} \nabla_\pi \mathbb{E}\left[\varepsilon_t^T \Lambda_{3,t}\varepsilon_t\right] \\ & - \nabla_\pi \mathbb{E}\left[\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i})\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\log\tilde{\pi}_{t,i} + 1 - \log \pi_{t,i}\right), \end{aligned} \tag{3.148}$$

where $\tilde{P}_{t,i}$ in (3.148) is taken to be the maximum point. With (3.148), we can compute Hessian and it is

$$\begin{aligned} & \sum_{t=0}^{T-1} x_0^T \Lambda_{1,t} \nabla_\pi^2 \mathbb{E}\left[\varepsilon_t\right] + \sum_{t=0}^{T-1}\sum_{\tau=t}^{T-1} \nabla_\pi^2 \left(\mathbb{E}\left[\varepsilon_t\right]^T \Lambda_{2,t,\tau}\mathbb{E}\left[\varepsilon_\tau\right]\right) + \sum_{t=0}^{T-1} \nabla_\pi^2 \mathbb{E}\left[\varepsilon_t^T \Lambda_{3,t}\varepsilon_t\right] \\ & - \nabla_\pi^2 \mathbb{E}\left[\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n} \tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i})\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\frac{1}{\tilde{\pi}_{t,i}}. \end{aligned} \tag{3.149}$$

Note that $\mathbb{E}\left[\varepsilon_t\right]$ is the expectation under the density

$$\sum_{i_0,i_1,\ldots,i_{T-1}} \prod_{t=0}^{T-1} \tilde{\pi}_{t,i_t} d\tilde{P}_{t,i_t}(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1}) \tag{3.150}$$

and can be quite complicated; we don't expand it there. However, as $0 \le \pi_{t,i} \le 1$ and $K_{t,i}$ are bounded, the Hessian of the first three terms should be bounded when $\gamma < \gamma_0$. For the fourth term, as $\tilde{P}_{t,i}$ is Gaussian, and KL-divergence between two Gaussian $N(\mu_1, \Sigma_1), N(\mu_2, \Sigma_2)$ is given by

$$KL\left(N(\mu_1,\Sigma_1)\|N(\mu_2,\Sigma_2)\right) = \frac{1}{2}\left(\log\frac{|\Sigma_2|}{|\Sigma_1|} - d_3 + (\mu_1-\mu_2)^T\Sigma_2^{-1}(\mu_1-\mu_2) + tr\{\Sigma_2^{-1}\Sigma_1\}\right). \tag{3.151}$$

Thus,

$$KL\left(\tilde{P}_{t,i}\|P_{t,i}\right) = \frac{1}{2}\left(\log\frac{\left|\Sigma_{t,i}\right|}{\left|\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}\right|} - d_3\right. \tag{3.152}$$

$$+\left(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}\Sigma_{t,i}^{-1}\mu_{t,i} + \gamma K_{t,i}\begin{pmatrix}1\\ \varepsilon_0\\ \varepsilon_1\\ \dots\\ \varepsilon_{t-1}\end{pmatrix} - \mu_{t,i}\right)^T\Sigma_{t,i}^{-1}$$

$$\left(\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}\Sigma_{t,i}^{-1}\mu_{t,i} + \gamma K_{t,i}\begin{pmatrix}1\\ \varepsilon_0\\ \varepsilon_1\\ \dots\\ \varepsilon_{t-1}\end{pmatrix} - \mu_{t,i}\right)$$

$$\left. + tr\{\Sigma_{t,i}^{-1}\left(\Sigma_{t,i}^{-1} - \gamma C_t^T\left(Q_{t+1}+\Xi_{t+1}\right)C_t\right)^{-1}\}\right).$$

Essentially, we see the Hessian of the fourth term

$$\mathbb{E}\left[\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i}KL(\tilde{P}_{t,i}\|P_{t,i})\right] \tag{3.153}$$

is also bounded. While $\frac{1}{\tilde{\pi}_{t,i}} \geq 1$, the last term adds at least $-\frac{1}{\gamma}$ to the diagonal of the Hessian. When $\gamma \to 0$, as the diagonal goes to negative infinity, the Hessian goes to negative definite, and we have (3.146) is concave in $\tilde{\pi}_{t,i}$. □

Now we can prove Theorem 3.8,

Proof for Theorem 3.8. With γ_0 to be γ_1 in Theorem 3.13 and $\mathcal{H},\mathcal{K}$ from Theorem 3.12,

$$\begin{aligned}
&\max_{\tilde{\pi}_t}\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})}\max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right.\\
&\quad\left.-\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i}KL(\tilde{P}_{t,i}\|P_{t,i})-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{\pi}_t\|\pi_t)\right]\\
&=\max_{\tilde{\pi}_t}\min_{(H_t,H_{t,\tau},h_t)\in\mathcal{H}}\max_{K_{t,i}\in\mathcal{K},\tilde{P}_{t,i}=N(\hat{\mu}_{t,i},\hat{\Sigma}_{t,i})}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right.\\
&\quad\left.-\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i}KL(\tilde{P}_{t,i}\|P_{t,i})-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{\pi}_t\|\pi_t)\right].
\end{aligned} \tag{3.154}$$

As the objective is convex in $(H_t, H_{t,\tau}, h_t)$, and by Theorem 3.13, concave in $\tilde{\pi}_{t,i}$, by Sion's minimax theorem, we have

$$\begin{aligned}
&\max_{\tilde{\pi}_t}\min_{(H_t,H_{t,\tau},h_t)\in\mathcal{H}}\max_{K_{t,i}\in\mathcal{K},\tilde{P}_{t,i}=N(\hat{\mu}_{t,i},\hat{\Sigma}_{t,i})}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right.\\
&\qquad\left.-\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i}KL(\tilde{P}_{t,i}\|P_{t,i})-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{\pi}_t\|\pi_t)\right]\\
&=\min_{(H_t,H_{t,\tau},h_t)\in\mathcal{H}}\max_{\tilde{\pi}_t}\max_{K_{t,i}\in\mathcal{K},\tilde{P}_{t,i}=N(\hat{\mu}_{t,i},\hat{\Sigma}_{t,i})}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right.\\
&\quad\left.-\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i}KL(\tilde{P}_{t,i}\|P_{t,i})-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{\pi}_t\|\pi_t)\right].
\end{aligned} \tag{3.155}$$

As changing maximin to minimax doesn't influence the innermost problem, we have

$$\begin{aligned}
&\min_{(H_t,H_{t,\tau},h_t)\,\in\mathcal{H}} \max_{\tilde{\pi}_t} \max_{K_{t,i}\,\in\mathcal{K},\tilde{P}_{t,i}=N(\hat{\mu}_{t,i},\hat{\Sigma}_{t,i})} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) \\
&\qquad - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\Bigg] \\
&= \min_{(H_t,H_{t,\tau},h_t)\in\mathcal{H}} \max_{\tilde{\pi}_t} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) \\
&- \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\Bigg].
\end{aligned} \tag{3.156}$$

$(H_t, H_{t,\tau}, h_t) \in \mathcal{H}$ is a constraint put on u_t, which will make the objective non-decreasing,

$$\begin{aligned}
&\min_{(H_t,H_{t,\tau},h_t)\,\in\mathcal{H}} \max_{\tilde{\pi}_t} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) \\
&\qquad - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\Bigg] \\
&\geq \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{\pi}_t} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) \\
&- \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\Bigg].
\end{aligned} \tag{3.157}$$

Note the LHS of (3.157) equal

$$\begin{aligned}
&\max_{\tilde{\pi}_t} \min_{(H_t,H_{t,\tau},h_t)\,\in\mathcal{H}} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\Bigg[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right) \\
&\qquad - \frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\Bigg],
\end{aligned} \tag{3.158}$$

and apply max-min inequality to the RHS of (3.157), the RHS is greater or equal to

$$\begin{aligned}\max_{\tilde{\pi}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}&\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right. \\ -\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) &\left. - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\right],\end{aligned} \tag{3.159}$$

which is, by (3.154),

$$\begin{aligned}\max_{\tilde{\pi}_t} \min_{(H_t,H_{t,\tau},h_t)\in\mathcal{H}} \max_{\tilde{P}_{t,i}(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}&\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right. \\ &\left. -\frac{1}{\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\tilde{\pi}_{t,i} KL(\tilde{P}_{t,i}\|P_{t,i}) - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{\pi}_t\|\pi_t)\right].\end{aligned} \tag{3.160}$$

We see (3.158) and (3.160) are same. Thus, all inequalities here are equalities. We complete the proof. □

Note that the model discussed in this section limits γ to be the same for any of the mixture components. However, this assumption can be relaxed and can be replaced by γ_i for each of the Gaussian components and one parameter, γ for the vector of cluster weights. In particular, this will allow us to have different risk aversion for different portions of the uncertainty space. For example, we may choose to be more risk-averse to outlying components of the distribution. The theorems in this section still apply as long as $\max\{\gamma_1, \gamma_2, \ldots, \gamma_n, \gamma\}$ is sufficiently small. We now discuss our algorithm allowing for different γ and describe its pseudocode in Algorithm 1 and Algorithm 2.

Algorithm The objective of the algorithm is to compute a control policy that solves (3.5); that is the linear quadratic risk-averse control with Gaussian mixture noise in the dynamics. As we discussed in §3.5, the solution of this problem results in a complex backward recursion for the optimal weights (3.76) and, consequently, for the objective function (3.77), which we cannot reduce from it sum form. Consequently, when carrying out the backward recursion, the resulting expressions contain as many combinations as the number of stages resulting in a combinatorial explosion.

To address this issue, we consider a relaxation whereby the weights of the adversarial density do not depend on the prior realizations of the noise, (3.79). For this relaxation, Theorem 3.8 allows us to use Sion's theorem for γ sufficiently small, and solve an inner problem with the fixed weights, which has an explicit solution by §3.4, and then solve an outer optimization problem in the weights by computing the gradient of the fixed weight objective relative to these weights. It should be noted that in §3.4 the risk parameter is $\frac{\gamma}{2}$ instead γ. This should be changed back to γ in the implementation. We describe that algorithm in Algorithm 1 and the computation of the gradients in Algorithm 2.

In Algorithm 1, we perform a gradient ascent with backtracking and Wolfe conditions. It will intensively use Theorem 3.6 to compute objective values and gradients. Note that the optimal control policy in Theorem 3.6 is computed recursively, we can apply chain rule to lower time complexity significantly when computing gradients. We provide a back-propagation algorithm in Algorithm 2 for the computation of gradients. It first performs a backward run to get quantities. And then it computes one-step gradients for recursions in Theorem 3.6, which is the first loop in Algorithm 2. In the second loop of Algorithm 2, the previous one-step gradients are summed up by chain rule. It finally returns $\frac{\partial V_0}{\pi_{t,i}}$ for all t, i to Algorithm 1.

3.6 Experiment

3.6.1 Without Uncertainty on Components' Probability

We consider a planning problem with 30 stages. Each stage will have a demand $\varepsilon_{t,i}$, which comes from two independent sources and has 50/50 chances for each one occurring. At the beginning of each stage, we take the decision u_t and pay a cost $c_1 u_t^2$ to fulfill u_t unit

Algorithm 1: Risk-Averse Gradient Ascent with Uncertainty on Components' Probability

Data: $\pi_{t,i}, \mu_{t,i}, \Sigma_{t,i}, A_t, B_t, C_t, R_t, Q_t$, risk parameters γ_i, initial step size ϵ_0, step size shrinkage rate β, minimal step size $\tilde{\epsilon}$, maximum steps N, c_1, c_2 for Wolfe conditions, $t = 0, 1, \ldots, T-1, i = 1, 2, \ldots, n$.

Result: $U_t, \bar{u}_t$ and control policy $u_t(x_t) = U_t x_t + \bar{u}_t, t = 0, 1, \ldots, T-1$.

```
for t = 0, ..., T − 1 do
    for i = 1. ..., n do
        π̂_{t,i} = π_{t,i};
k = 0;
while k < N and ϵ > ϵ̃ do
    ϵ = ϵ_0;
    //Compute the objective value and gradient
    Compute V_0 at π̂_{t,i} by (3.40);
    Compute gradient ∇π at π̂_{t,i} with Algorithm 2;
    while ϵ > ϵ̃ do
        //Project to feasible region
        for t = 0, ..., T − 1 do
            for i = 1, ..., n do
                π̃_{t.i} = π̂_{t,i} − ϵ∇π_{t,i};
                π̃_{t.i} = min{1, max{0, π̃_{t.i}}};
            π̃_t = Σ_{i=1}^n π̃_{t,i};
            for i = 1, ..., n do
                π̃_{t,i} = π̃_{t,i} / π̃_t;
        //Check Wolfe Condition
        Compute Ṽ_0 at π̃_{t,i} by (3.40);
        Compute gradient ∇̃π at π̃_{t,i} with Algorithm 2;
        if Ṽ_0 ≤ c_1 (V_0 − ϵ‖∇π‖^2) and −∇̃π^T ∇π ≤ −c_2‖∇π‖^2 then
            //Satisfies Wolfe Condition and exits
            Break;
        else
            //Backtracking
            ϵ = βϵ;
    k+=1;
Get U_t and ū_t by (3.42) and (3.43);
```

Algorithm 2: Gradient Computation in Risk-Averse Mixture of Gaussian Control

Data: risk parameter γ_i, current position $\hat{\pi}_{t,i}$, system parameters $\mu_{t,i}, \Sigma_{t,i}, A_t, B_t, C_t, R_t, Q_t, t = 0, 1, \ldots, T-1, i = 1, \ldots, n.$

Result: $\nabla \pi_{t,i} = \frac{\partial V_0(x_0)}{\partial \pi_{t,i}}(\hat{\pi}_{t,i}), t = 0, 1, \ldots, T-1, i = 1, 2, \ldots, n.$

```
//Backward run to compute all quantities
```

Solve all matrices in Theorem 3.6 from $T-1$ to 0 at $\hat{\pi}_{t,i}$;

```
//Compute one-step gradients based on dependency
```

for $t = 0, \ldots, T-1$ **do**

- Compute $\frac{\partial U_t}{\partial \Xi_{t+1}}$ at $\hat{\pi}_{t,i}$ from (3.64) ;
- Compute $\frac{\partial \bar{u}_t}{\partial \Xi_{t+1}}, \frac{\partial \bar{u}_t}{\partial \xi_{t+1}}$ at $\hat{\pi}_{t,i}$ from (3.65);
- Compute $\frac{\partial \Xi_t}{\partial \Xi_{t+1}}, \frac{\partial \Xi_t}{\partial U_t}$ at $\hat{\pi}_{t,i}$ from (3.66);
- Compute $\frac{\partial \xi_t}{\partial \Xi_{t+1}}, \frac{\partial \xi_t}{\partial U_t}, \frac{\partial \xi_t}{\partial \xi_{t+1}}, \frac{\partial \xi_t}{\partial \bar{u}_t}$ at $\hat{\pi}_{t,i}$ from (3.67);
- Compute $\frac{\partial z_t}{\partial \Xi_{t+1}}, \frac{\partial z_t}{\partial \xi_{t+1}}, \frac{\partial z_t}{\partial \bar{u}_t}$ at $\hat{\pi}_{t,i}$ from (3.68);
- **for** $i = 1. \ldots, n$ **do**
  - Compute $\frac{\partial U_t}{\partial \tilde{A}_{t,i}}, \frac{\partial U_t}{\partial \tilde{B}_{t,i}}, \frac{\partial U_t}{\partial \pi_{t,i}}$ at $\hat{\pi}_{t,i}$ from (3.64);
  - Compute $\frac{\partial \bar{u}_t}{\partial \tilde{B}_{t,i}}, \frac{\partial \bar{u}_t}{\partial \tilde{\xi}_{t+1,i}}, \frac{\partial \bar{u}_t}{\partial \tilde{\mu}_{t,i}}, \frac{\partial \bar{u}_t}{\partial \pi_{t,i}}$ at $\hat{\pi}_{t,i}$ from (3.65);
  - Compute $\frac{\partial \Xi_t}{\partial \tilde{A}_{t,i}}, \frac{\partial \Xi_t}{\partial \tilde{B}_{t,i}}, \frac{\partial \Xi_t}{\partial \pi_{t,i}}$ at $\pi_{t,i}$ from (3.66);
  - Compute $\frac{\partial \xi_t}{\partial \tilde{A}_{t,i}}, \frac{\partial \xi_t}{\partial \tilde{B}_{t,i}}, \frac{\partial \xi_t}{\partial \tilde{\xi}_{t+1,i}}, \frac{\partial \xi_t}{\partial \tilde{\mu}_{t,i}}, \frac{\partial \xi_t}{\partial \pi_{t,i}}$ at $\hat{\pi}_{t,i}$ from (3.67);
  - Compute $\frac{\partial z_t}{\partial \tilde{B}_{t,i}}, \frac{\partial z_t}{\partial \tilde{\xi}_{t+1,i}}, \frac{\partial z_t}{\partial \tilde{\mu}_{t,i}}, \frac{\partial z_t}{\partial \pi_{t,i}}$ at $\hat{\pi}_{t,i}$ from (3.68);
  - Compute $\frac{\partial \tilde{A}_{t,i}}{\partial \Xi_{t+1}}$ at $\hat{\pi}_{t,i}$ from (3.69);
  - Compute $\frac{\partial \tilde{B}_{t,i}}{\partial \Xi_{t+1}}$ at $\hat{\pi}_{t,i}$ from (3.70);
  - Compute $\frac{\partial \tilde{\mu}_{t,i}}{\partial \Xi_{t+1}}$ at $\hat{\pi}_{t,i}$ from (3.71);
  - Compute $\frac{\partial \tilde{\xi}_{t+1,i}}{\partial \Xi_{t+1}}, \frac{\partial \tilde{\xi}_{t+1,i}}{\partial \xi_{t+1}}$ at $\hat{\pi}_{t+1,i}$ from (3.72);

//Backpropagation

for $t = 0, \dots, T-1$ **do**

  for $i = 1, \dots, n$ **do**

    for $\tau = t, t-1, \dots, 0$ **do**

$\frac{\partial U_\tau}{\partial \pi_{t,i}} = \frac{\partial U_\tau}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial U_\tau}{\partial \pi_{t,i}}$ by applying chain rule to (3.64);

$\frac{\partial \bar{u}_\tau}{\partial \pi_{t,i}} = \frac{\partial \bar{u}_\tau}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial \bar{u}_\tau}{\partial \xi_{\tau+1}} \frac{\partial \xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial \bar{u}_\tau}{\partial \pi_{t,i}}$ by applying chain rule to (3.65);

$\frac{\partial \Xi_\tau}{\partial \pi_{t,i}} = \frac{\partial \Xi_\tau}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial \Xi_\tau}{\partial U_\tau} \frac{\partial U_\tau}{\partial \pi_{t,i}} + \frac{\partial \Xi_\tau}{\partial \pi_{t,i}}$ by applying chain rule to (3.66);

$\frac{\partial \xi_\tau}{\partial \pi_{t,i}} = \frac{\partial \xi_\tau}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial \xi_\tau}{\partial U_\tau} \frac{\partial U_\tau}{\partial \pi_{t,i}} + \frac{\partial \xi_\tau}{\partial \xi_{\tau+1}} \frac{\partial \xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial \xi_\tau}{\partial \pi_{t,i}}$ by applying chain rule to (3.67);

$\frac{\partial z_\tau}{\partial \pi_{t,i}} = \frac{\partial z_\tau}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial z_\tau}{\partial \xi_{\tau+1}} \frac{\partial \xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial z_\tau}{\partial \bar{u}_t} \frac{\partial \bar{u}_t}{\partial \pi_{t,i}} + \frac{\partial z_{\tau+1}}{\partial \pi_{t,i}}$ by applying chain rule to (3.68);

      for $j = 1, \dots, n$ **do**

$\frac{\partial \tilde{A}_{\tau,j}}{\partial \pi_{t,i}} = \frac{\partial \tilde{A}_{\tau,j}}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}}$ by applying chain rule to (3.69);

$\frac{\partial \tilde{B}_{\tau,j}}{\partial \pi_{t,i}} = \frac{\partial \tilde{B}_{\tau,j}}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}}$ by applying chain rule to (3.70);

$\frac{\partial \bar{\mu}_{\tau,j}}{\partial \pi_{t,i}} = \frac{\partial \bar{\mu}_{\tau,j}}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}}$ by applying chain rule to (3.71);

$\frac{\partial \tilde{\xi}_{\tau+1,j}}{\partial \pi_{t,i}} = \frac{\partial \tilde{\xi}_{\tau+1,j}}{\partial \Xi_{\tau+1}} \frac{\partial \Xi_{\tau+1}}{\partial \pi_{t,i}} + \frac{\partial \tilde{\xi}_{\tau+1,j}}{\partial \xi_{\tau+1}} \frac{\partial \xi_{\tau+1}}{\partial \pi_{t,i}}$ by applying chain rule to (3.72);

      for $j = 1, \dots, n$ **do**

$\frac{\partial U_\tau}{\partial \pi_{t,i}} = \frac{\partial U_\tau}{\partial \pi_{t,i}} + \frac{\partial U_\tau}{\partial \tilde{A}_{\tau,j}} \frac{\partial \tilde{A}_{\tau,j}}{\partial \pi_{t,i}} + \frac{\partial U_\tau}{\partial \tilde{B}_{\tau,j}} \frac{\partial \tilde{B}_{\tau,j}}{\partial \pi_{t,i}}$ by applying chain rule to (3.64);

$\frac{\partial \bar{u}_\tau}{\partial \pi_{t,i}} = \frac{\partial \bar{u}_\tau}{\partial \pi_{t,i}} + \frac{\partial \bar{u}_\tau}{\partial \tilde{B}_{\tau,j}} \frac{\partial \tilde{B}_{\tau,j}}{\partial \pi_{t,i}} + \frac{\partial \bar{u}_\tau}{\partial \tilde{\xi}_{\tau+1,j}} \frac{\partial \tilde{\xi}_{\tau+1,j}}{\partial \pi_{t,i}} + \frac{\partial \bar{u}_\tau}{\partial \bar{\mu}_{\tau,j}} \frac{\partial \bar{\mu}_{\tau,j}}{\partial \pi_{t,i}}$ by applying chain rule to (3.65);

$\frac{\partial \Xi_\tau}{\partial \pi_{t,i}} = \frac{\partial \Xi_\tau}{\partial \pi_{t,i}} + \frac{\partial \Xi_\tau}{\partial \tilde{A}_{\tau,j}} \frac{\partial \tilde{A}_{\tau,j}}{\partial \pi_{t,i}} + \frac{\partial \Xi_\tau}{\partial \tilde{B}_{\tau,j}} \frac{\partial \tilde{B}_{\tau,j}}{\partial \pi_{t,i}}$ by applying chain rule to (3.66);

$\frac{\partial \xi_\tau}{\partial \pi_{t,i}} = \frac{\partial \xi_\tau}{\partial \pi_{t,i}} + \frac{\partial \xi_\tau}{\partial \tilde{A}_{\tau,j}} \frac{\partial \tilde{A}_{\tau,j}}{\partial \pi_{t,i}} + \frac{\partial \xi_\tau}{\partial \tilde{B}_{\tau,j}} \frac{\partial \tilde{B}_{\tau,j}}{\partial \pi_{t,i}} + \frac{\partial \xi_\tau}{\partial \tilde{\xi}_{\tau+1,j}} \frac{\partial \tilde{\xi}_{\tau+1,j}}{\partial \pi_{t,i}} + \frac{\partial \xi_\tau}{\partial \bar{\mu}_{\tau,j}} \frac{\partial \bar{\mu}_{\tau,j}}{\partial \pi_{t,i}}$ by applying chain rule to (3.67);

$\frac{\partial z_\tau}{\partial \pi_{t,i}} = \frac{\partial z_\tau}{\partial \tilde{B}_{\tau,j}} \frac{\partial \tilde{B}_{\tau,j}}{\partial \pi_{t,i}} + \frac{\partial z_t}{\partial \tilde{\xi}_{\tau+1,j}} \frac{\partial \tilde{\xi}_{\tau+1,j}}{\partial \pi_{t,i}} + \frac{\partial z_\tau}{\partial \bar{\mu}_{\tau,j}} \frac{\partial \bar{\mu}_{\tau,j}}{\partial \pi_{t,i}} + \frac{\partial z_\tau}{\partial \pi_{t,i}}$ by applying chain rule to (3.68);

$\frac{\partial z_\tau}{\partial \pi_{t,i}} = \frac{\partial z_\tau}{\partial \pi_{t,i}} + \frac{\partial z_{\tau+1}}{\partial \pi_{t,i}}$;

$\nabla \pi_{t,i} = x_0^T \frac{\partial \Xi_0}{\pi_{t,i}} x_0 + 2x_0^T \frac{\partial \xi_0}{\partial \pi_{t,i}} + \frac{\partial z_0}{\partial \pi_{t,i}}$;

	.05 quantile	.25 quantile	.50 quantile	.75 quantile	.95 quantile
$\gamma_1 = 0.005, \gamma_2 = 0.005$	75891.7	79523.6	82163.7	84886.8	88906.7
$\gamma_1 = 0., \gamma_2 = 0.$	77578.8	81546.	84383.3	87304.3	91604.6

Table 3.1: Quantiles of objective values for risk-free and risk-averse controllers.

demand. Denote the difference between realized demands and $x_{t-1} + u_t$ to be x_t, that is, $x_t = x_{t-1} + u_t - \varepsilon_t$. This difference will be penalized by $c_2 x_t^2$. The demand $\varepsilon_{t,i}$ is modeled as $N(\mu_{t,i}, \Sigma_{t,i})$, $t = 0, 1, 2, \ldots, 29, i = 1, 2$, where

$$\mu_{t,i} = \begin{cases} d_0 + \mu_i, t \equiv 0, 1, 2, 3, 4 \mod 7 \\ d_1 + \mu_i, t \equiv 5, 6 \mod 7 \end{cases}. \tag{3.161}$$

We consider different uncertainty levels for demands from different sources with the model in (3.11). Here, we assume that the weight of the two clusters is known and it is exactly 0.5 for each.

In our experiment, we first input the correct parameters $\mu_{t,i}$ to the model to check the tail behavior. Then, we choose $\mu_{t,i}$ different from the one used in the design of controllers to simulate the error in the estimation of $\mu_{t,i}$, with an error parameter $e_{t,i}$ (which represents the difference between the true parameter and the one used in the computation of the risk-averse policy). We choose the other parameter values in the model as $c_1 = 10, c_2 = 5, d_0 = 7, d_1 = 3, \mu_1 = 9, \mu_2 = 6, \Sigma_{t,i} = \mathbf{I}_1$.

The tail behavior is shown in figure 3.1, with $\gamma_1 = \gamma_2 = 0.005$. As we can see, the risk-averse controller produces more mass in the green area and has a lighter tail as expected. We now choose the misspecified $\mu_{t,i}$ by having an error of 3 at each stage. We now compare risk-averse controllers and risk-free controllers under this scenario. As risk-aversion parameters, we choose $\gamma_1 = \gamma_2 = 0.005$. The results are displayed in figure 3.2 and table 3.1. And we can see the risk-averse controller will perform much better comparing to risk-free controller in this scenario.

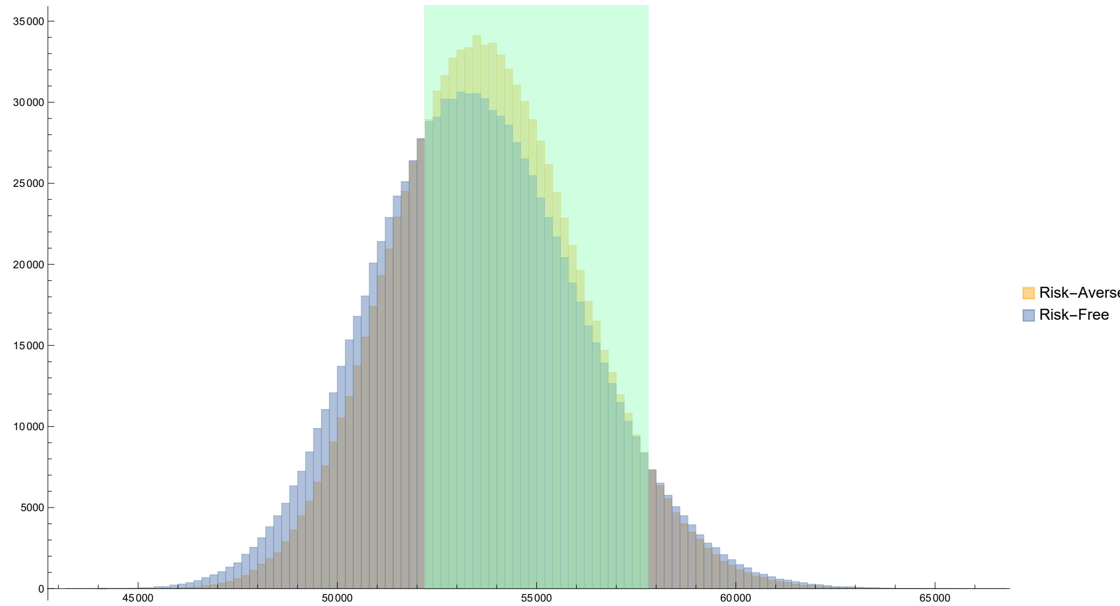


Figure 3.1: Tail behavior under exact parameters.

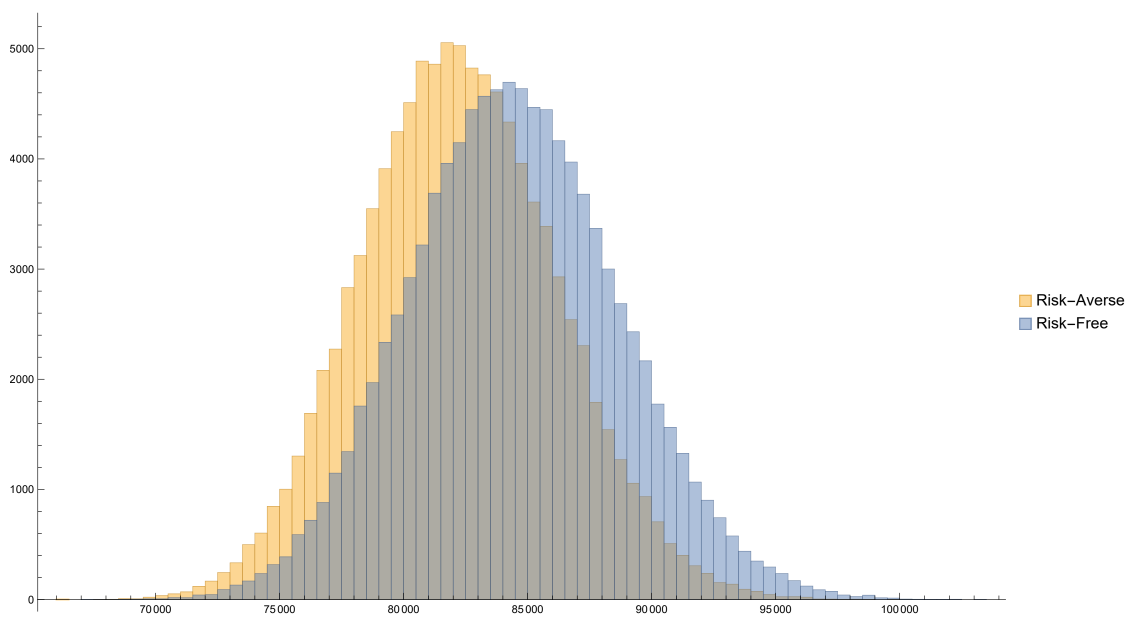


Figure 3.2: Risk-averse control under inexact parameters.

	.05 quantile	.25 quantile	.50 quantile	.75 quantile	.95 quantile
$\gamma_1 = 0.007, \gamma_2 = 0., e_{t,1} = 6., e_{t,2} = 0.$	71466.8	80730.8	87723.	94680.7	105296.
$\gamma_1 = 0., \gamma_2 = 0., e_{t,1} = 6., e_{t,2} = 0.$	72561.4	82028.4	89156.3	97073.	108517.
$\gamma_1 = 0.007, \gamma_2 = 0.002, e_{t,1} = 6., e_{t,2} = 2.$	83153.5	91502.7	97647.6	104135.	113629.
$\gamma_1 = 0., \gamma_2 = 0., e_{t,1} = 6., e_{t,2} = 2.$	85428.5	94282.9	100728.	107425.	117491.

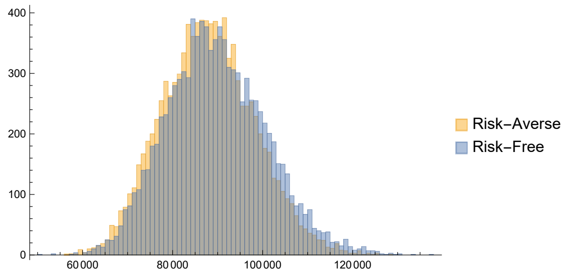
Table 3.2: Quantiles for risk-averse and risk-free controllers when the main uncertainty is on the $\mu_{t,1}$.

	.05 quantile	.25 quantile	.50 quantile	.75 quantile	.95 quantile
$\gamma_1 = 0., \gamma_2 = 0.007, e_{t,1} = 0., e_{t,2} = 6.$	76409.4	80086.3	82924.8	85763.5	89674.7
$\gamma_1 = 0., \gamma_2 = 0., e_{t,1} = 0., e_{t,2} = 6.$	77541.3	81489.3	84357.8	87216.5	91424.6
$\gamma_1 = 0.002, \gamma_2 = 0.007, e_{t,1} = 2., e_{t,2} = 6.$	90144.4	92863.5	94743.6	96661.7	99502.8
$\gamma_1 = 0., \gamma_2 = 0., e_{t,1} = 2., e_{t,2} = 6.$	92607.2	95556.4	97557.7	99638.	102632.

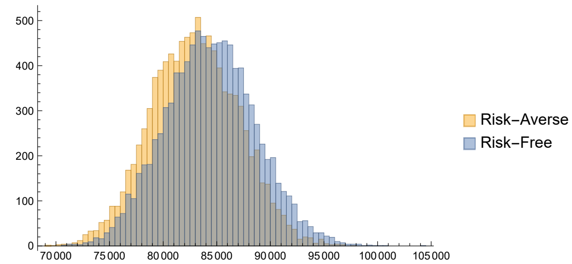
Table 3.3: Quantiles for risk-averse and risk-free controllers when the main uncertainty is on the $\mu_{t,2}$.

We also consider the problem of having different uncertainty levels. We set the error to be $e_{t,1} = 6, e_{t,2} = 0$ or $e_{t,1} = 0, e_{t,2} = 6$. And for each case we set $\gamma_1 = 0.007, \gamma_2 = 0$ or $\gamma_1 = 0, \gamma_2 = 0.007$ (with the choice of zero risk parameter being the one of zero error in the respective case). The simulation results are shown in figure 3.3a and figure 3.3b, which is same with the above conclusion. We then increase the smaller value between $e_{t,1}$ and $e_{t,2}$ to be 2, and set the smaller value between γ_1 and γ_2 to 0.002, then we have figure 3.3c and figure 3.3d, which are more obvious to see the advantage of our risk-averse controllers. The detailed comparison on quantiles of these controllers are listed in tables 3.2 and 3.3.

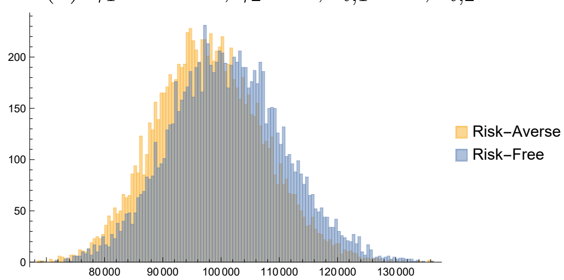
Tuning γ_i when the errors are different is difficult and we do not currently have a crisp principle by which to do it, though our choices in these experiments seem to behave properly. An appropriate value of γ_i should include the ground truth distribution into consideration but not be exceedingly conservative. As data points sampled from the true distribution may come from different sources, we can sort them in order of confidence, with a smaller γ_i for distributions with reliable sources.



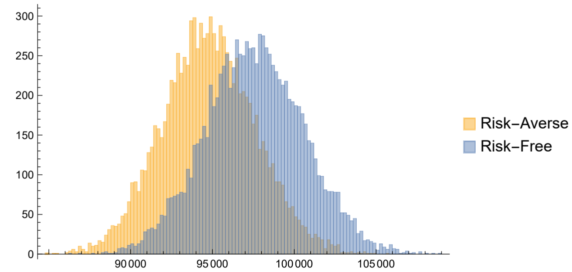
(a) $\gamma_1 = 0.007, \gamma_2 = 0, e_{t,1} = 6, e_{t,2} = 0$



(b) $\gamma_1 = 0, \gamma_2 = 0.007, e_{t,1} = 0, e_{t,2} = 6$



(c) $\gamma_1 = 0.007, \gamma_2 = 0.002, e_{t,1} = 6, e_{t,2} = 2$



(d) $\gamma_1 = 0.002, \gamma_2 = 0.007, e_{t,1} = 2, e_{t,2} = 6$

Figure 3.3: Risk-averse control with different uncertainty levels.

3.6.2 With Uncertainty on Components' Probability

We now use Algorithm 1 to solve the same model used in §3.6.1. γ is chosen to be the same for all distributions with a value 0.0002. The tail behavior under the exact parameter is shown in figure 3.4, where the risk-averse controller has more mass in the green area. We also compare this risk-averse controller with the risk-free controller and the risk-averse controller in §3.6.1 under inexact parameters. The ground truth value of $\pi_{t,i}$ is chosen to be $(0.75, 0.25)$ instead of the model input value $(0.5, 0.5)$. The histogram figure 3.5 shows that the controller with weight robustness outperforms others significantly.

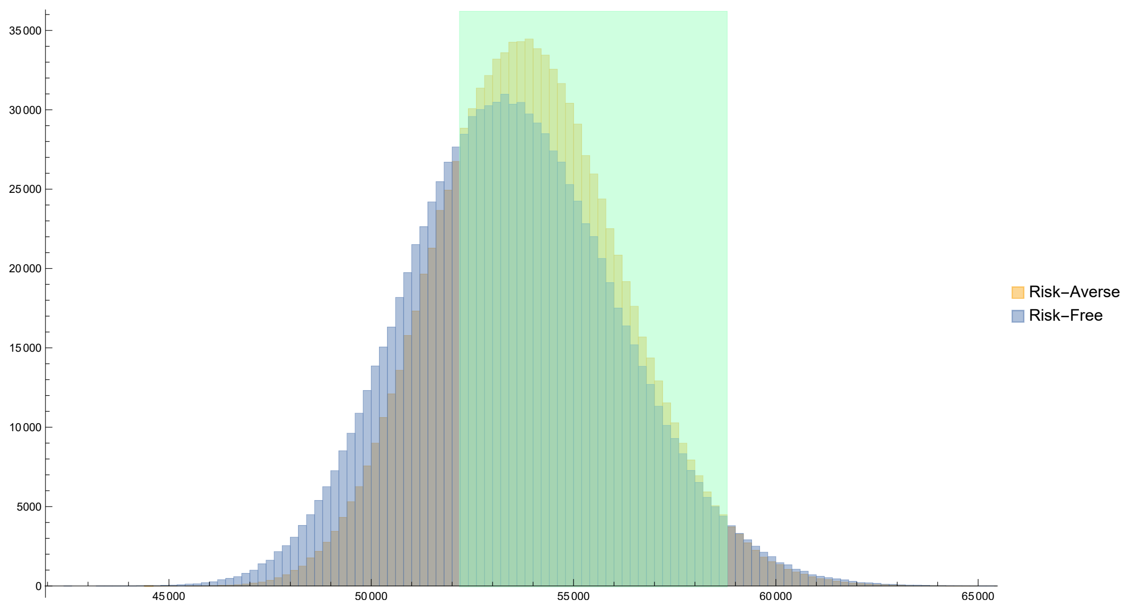


Figure 3.4: Tail behavior under exact parameters.

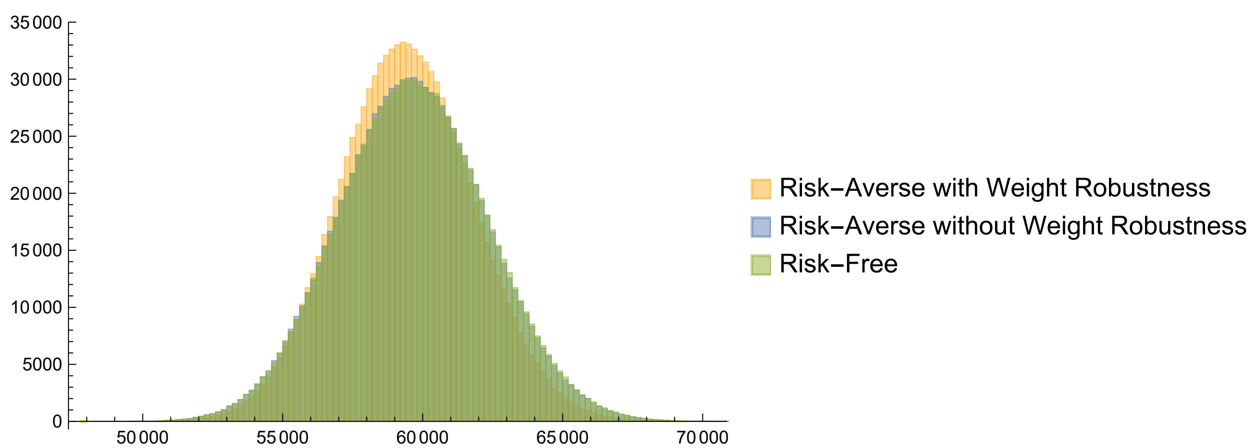


Figure 3.5: Risk-averse control under inexact parameters.

CHAPTER 4

OUTLIER ROBUST CONTROL WITH GROUP LASSO

4.1 Introduction

Robustness, at least in control problems, is a term that has multiple definitions. Some describe it with the portion of data falling into tails like value-at-risk (Philippe, 2001) and conditional value-at-risk (Uryasev, 2000), while some define it as a trade-off between expectation and variance, optimized over which using exponential criteria (Jacobson, 1973), and some consider extreme cases having worst-case optimizations (Tadmor, 1990). It is not hard to understand why there are so many definitions for a single term in the control problems, since we have different desires when data is not generated in the expected way, and we have different ways to characterize the deviations that happened in the dataset. There is a viewpoint understanding robustness as the concept of optimizing under statistical fluctuation, which is previously used in (Petersen et al., 2000; Tzortzis et al., 2015; Yang, 2020). Their methods are successful when the dataset is well-sampled. However, when outliers exist in the dataset, people usually lack knowledge about them. And the target of this paper is to deal with this problem in the robust linear quadratic control model.

Robust control problems are, in fact, quite close to the discussion of distributionally robust optimization(DRO) (Rahimian and Mehrotra, 2019). While they have problems in solving things like moments constraints, with the convenience of discrete data points, empirical measures are more widely used in the recent discussion, like (Namkoong and Duchi, 2016; Kim and Yang, 2020; Sinha et al., 2017). (Sinha et al., 2017) is the first paper we know applying Wasserstein distance in the discussion of robust optimization, while (Yang, 2020; Kim and Yang, 2020) shift from the general optimization and focus on applying W_2 into linear quadratic controls. The difference between control problems and general optimizations is that we usually can solve Bellman equations, thus possibly having lower time complexity. All of their results are based on the strong duality result of Wasserstein constraints optimization (Gao and Kleywegt, 2023). We will keep the setting of discrete data points in this paper, and also use the strong duality result.

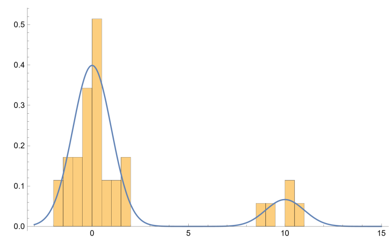
In this paper, we will mainly study the problem of robustness to the outliers. The term outlier, in fact, has two types of meanings. The one is, you have a few unusual data points outside the majority, on which you have little knowledge and it may vary in a large domain. Another one is, your dataset is contaminated and has erroneous data points. The latter one is previously studied by (Ting et al., 2007) with Bayesian methods. In this paper, we propose to use generalized Wasserstein distance with $\|x\|_2^2+\alpha\|x\|$ term to deal with the robustness problems of outliers. And we will show that, though it has two definitions for outliers, our model can handle both cases by simply changing the sign of the penalty. We summarize the points of this paper here,

1. We show that using $\|x\|_2^2+\alpha\|x\|$ in the generalized Wasserstein distance is a natural generalization of exponential criteria linear quadratic Gaussian (LQG) problems. And thus with this penalty term, they are in fact risk-averse LQG problems.

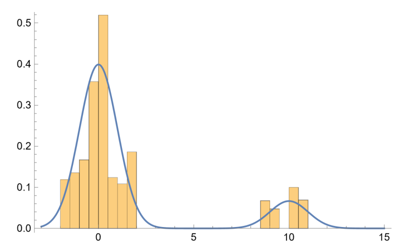
2. We provide equivalent forms of (4.1) and (4.2) which have intuitive interpretations. We explain why they are robust to outliers and what's the difference between those two equations. In fact, they correspond to two scenarios where outliers appear.

3. We provide tractable algorithms to solve (4.1) and (4.2), which transfers (4.1) and (4.2) into the form of group lasso (Tibshirani, 1996; Yuan and Lin, 2006) and makes use of ADMM iterations (Boyd et al., 2011) to solve it.

We want to use a few plots to illustrate the view of statistical fluctuation and the difference between a few distributions' similarity measures. In the case that we have a dataset, in the stochastic control, we want to optimize over expectation to minimize the designed performance. However, when we sampled again or applied it in reality, we may not see the same dataset again, and this is called statistical fluctuation. When KL-divergence is applied in the optimization problems, it is, in fact, considering perturbation weights on each data point, as shown in the figure 4.1, to minimize the worst criteria in the local domain. Different from KL-divergence, W_2 considers the shifting that happened in each data point, as shown in the figure 4.2. Our proposed penalty describes a domain for outliers only, as shown in figure 4.3, keeping the majority fixed. Most of the risk-averse control papers assume the risk parameters γ to be positive. However, we point out there that $\gamma < 0$ also has important meanings. We will clearly see the role of $\gamma < 0$ and its important applications later.

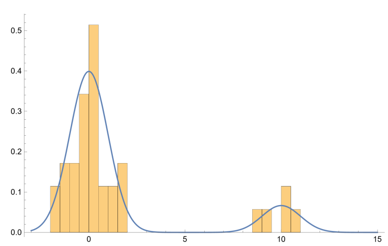


(a) Input dataset

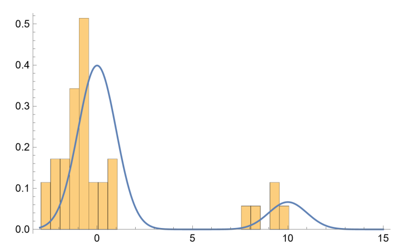


(b) Perturbed dataset

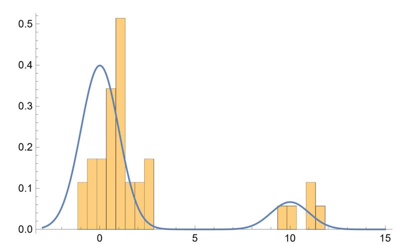
Figure 4.1: Local KL-divergence domain



(a) Input dataset

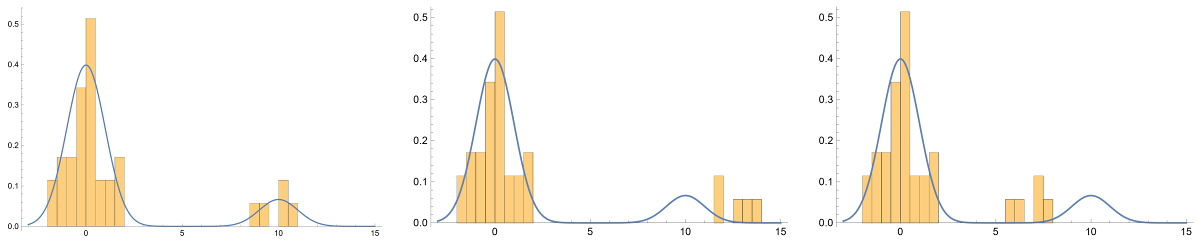


(b) Perturbed dataset 1



(c) Perturbed dataset 2

Figure 4.2: Local W_2 domain



(a) Input dataset (b) Perturbed dataset 1 (c) Perturbed dataset 2

Figure 4.3: Local Outlier Robust W domain

4.2 Problem Formulation and Assumptions

In this paper, we will consider the linear quadratic control problem (LQP) with Wasserstein penalty. We consider two problems,

$$\begin{aligned} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1} W(\tilde{P}_t, P_t) \\ s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t, \end{aligned} \tag{4.1}$$

where $\gamma > 0$ and

$$\begin{aligned} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \min_{\tilde{P}_t, \varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1} W(\tilde{P}_t, P_t) \\ s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t, \end{aligned} \tag{4.2}$$

where $\gamma < 0$. We assume $\alpha > 0$, $x_t \in \mathbb{R}^{d_1}, u_t \in \mathbb{R}^{d_2}, \varepsilon_t \in R^{d_3}$, ε_t are independent noise, Q_t, R_t, A_t, B_t, C_t are matrices in proper dimension. In our settings, P_t are empirical measures on n data points $\varepsilon_{t,i}$ at each stage. $W(\tilde{P}_t, P_t)$ is generalized Wasserstein distance equipped with $d(x,y) = \|x-y\|_2^2 + \alpha\|x-y\|_2$,

$$W(\tilde{P}_t, P_t) = \min_{\Gamma=\Pi(\tilde{P}_t,P_t)} \int \|x-y\|_2^2 + \alpha\|x-y\|_2 d\Gamma(x,y). \tag{4.3}$$

Notice that (4.3) is not a distance in the usual sense, it is instead a discrepancy between two distributions. An unified view of (4.1) and (4.2) is

$$\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \frac{1}{\gamma} \max_{\tilde{P}_t,\varepsilon_t\sim\tilde{P}_t} \gamma\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \sum_{t=0}^{T-1} W(\tilde{P}_t, P_t) \tag{4.4}$$
$$s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t.$$

4.2.1 Assumptions

We will use common assumptions in the risk-averse control,

Assumption 4.1. $Q_t \geq 0$.

Assumption 4.2. $R_t \geq 0$.

Assumption 4.3. $rank(B_t) = d_2$.

Assumption 4.4. *The minimum of* (4.4) *exists.*

For notation simplicity, if not specified, $\|\cdot\|$ are always assumed to be $\|\cdot\|_2$ in this paper. And if at any place we mention x_{t+1} without dynamics, then the dynamics are assumed to be the one in (4.4).

4.3 Generalization of Risk-Averse LQG

It is a long time since the discussion of risk-averse linear quadratic problems. There are many kinds of risk-averse controllers coming from optimizing different sorts of criteria, like value-at-risk (Philippe, 2001), conditional value-at-risk (Uryasev, 2000), and exponential criteria (Jacobson, 1973). We are going to start from the exponential criteria to see that (4.1) and (4.2) are natural generalizations of classical risk-averse LQP.

In the exponential criteria risk-averse LQG, the optimization problem is

$$\begin{aligned}&\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\exp\left\{\gamma\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right\}\right] \\ &s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t\end{aligned} \tag{4.5}$$

where $\gamma > 0$ and ε_t are i.i.d. Gaussian. Notice that $\gamma > 0$ is a general assumption in the risk-averse control. But we want to argue that, $\gamma < 0$ also has a role in risk-averse control. We will explain the reason later and use an extension of (4.5) in the discussion to allow $\gamma < 0$.

$$\begin{aligned}&\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \frac{1}{\gamma}\log\mathbb{E}\left[\exp\left\{\gamma\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right\}\right] \\ &s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t\end{aligned} \tag{4.6}$$

It is well known the exponential criteria gives a risk-averse controller since it is a trade-off between expectations and variances when γ approaches 0.

Lemma 4.1.

$$\begin{aligned}&\frac{1}{\gamma}\log\mathbb{E}\left[\exp\left\{\gamma\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right\}\right] \\ &= \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] + \frac{\gamma}{2}\mathbb{V}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] + o(\gamma)\end{aligned} \tag{4.7}$$

Proof.

$$\exp\{\gamma x\} = 1 + \gamma x + \frac{\gamma^2}{2}x^2 + o(\gamma^2) \tag{4.8}$$

Thus for any Gaussian random variables X, since its moment generating function exists outside 0,

$$
\begin{aligned}
\log \mathbb{E}\left[\exp\{\gamma X\}\right] &= \log\left(1+\gamma\mathbb{E}\left[X\right]+\frac{\gamma^2}{2}\mathbb{E}\left[X\right]^2+o(\gamma^2)\right) \\
&= \gamma\mathbb{E}\left[X\right]+\frac{\gamma^2}{2}\mathbb{E}\left[X^2\right]-\frac{\gamma^2}{2}\mathbb{E}\left[X\right]^2+o(\gamma^2) \\
&= \gamma\mathbb{E}\left[X\right]+\frac{\gamma^2}{2}\mathbb{V}\left[X\right]+o(\gamma^2)
\end{aligned} \tag{4.9}
$$

□

For cases $\gamma > 0$ and $\gamma < 0$, each of them corresponds to one of (4.1) and (4.2).

Lemma 4.2. *For any random variable X following a distribution P which has density $dP(x)$, the maximum point $\tilde{P}$ of the problem*

$$
\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}\left[X\right] - \frac{1}{\gamma}KL(\tilde{P}\|P) \tag{4.10}
$$

also has same support with P if $\gamma > 0$ and the maximum in (4.10) exists, and its density satisfies the condition

$$
d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{4.11}
$$

Consequently, we have the identity

$$
\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}\left[X\right] - \frac{1}{\gamma}KL(\tilde{P}\|P) = \int x d\tilde{P}(x) - \frac{1}{\gamma}\log\left(\frac{d\tilde{P}(x)}{dP(x)}\right)d\tilde{P}(x) = \frac{1}{\gamma}\log \mathbb{E}_P\left[e^{\gamma X}\right], \tag{4.12}
$$

where $\mathbb{E}_P\left[e^{\gamma X}\right]$ is the normalized constant for the $d\tilde{P}(x)$ in (4.11)*.*

Proof. Consider the function

$$
\log \mathbb{E}_P\left[e^{\gamma X}\right] = \log \int e^{\gamma x} dP(x). \tag{4.13}
$$

For any distributions $\tilde{P}(x)$ that has density and same support with $P(x)$,

$$
\begin{aligned}
\log \int e^{\gamma x} dP(x) &= \log \int e^{\gamma x}\left(\frac{dP(x)}{d\tilde{P}(x)}\right)d\tilde{P}(x) \\
&\geq \int \left(\gamma x d\tilde{P}(x) + \log\left(\frac{dP(x)}{d\tilde{P}(x)}\right)d\tilde{P}(x)\right),
\end{aligned} \tag{4.14}
$$

where the last inequality comes from Jensen's inequality in probabilistic setting (Durrett, 2019). Divide γ on the both hand side of (4.14), we have

$$\frac{1}{\gamma}\log \mathbb{E}_P\left[e^{\gamma X}\right] \geq \mathbb{E}_{\tilde{P}}\left[x\right] - \frac{1}{\gamma}KL(\tilde{P}\|P). \tag{4.15}$$

Equality can be attained from Jensen's inequality with the condition

$$d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{4.16}$$

Note that the distribution constructed from (4.16) always exists if (4.13) is not infinity. Thus if (4.13) is not infinite, the maximum of (4.10) exists.

Asssume now (4.13) is infinite. By constructing $\tilde{P}(x)$, we get

$$d\tilde{P}_R(x) \sim dP(x) \times e^{\gamma x} \times I(\|x\| \leq R), \tag{4.17}$$

where $R \in \mathbb{R}$, we have $\tilde{P}_R$ always exists and feasible in (4.10). Denote C_R to be the normalized constant such that

$$d\tilde{P}_R(x) = \frac{1}{C_R} dP(x) \times e^{\gamma x} \times I(\|x\| \leq R). \tag{4.18}$$

Since (4.13) is infinite, we must have that

$$\lim_{R\to\infty} C_R = \lim_{R\to\infty} \int e^{\gamma x} \times I(\|x\| \leq R) dP(x) = \infty$$

Let $R \to \infty$, we have

$$\lim_{R\to\infty} \mathbb{E}_{\tilde{P}_R}\left[X\right] - \frac{1}{\gamma}KL(\tilde{P}_R\|P) = \lim_{R\to\infty}\int x d\tilde{P}_R(x) - \frac{1}{\gamma}\log\left(\frac{d\tilde{P}_R(x)}{dP(x)}\right) d\tilde{P}_R(x) = \lim_{R\to\infty}\int -\frac{1}{\gamma}\log\left(\frac{d\tilde{P}_R(x)}{e^{\gamma x}dP(x)}\right) d\tilde{P}_R(x) = \lim_{R\to\infty}\int_{\|x\|\leq R} \frac{1}{\gamma}\frac{\log C_R}{C_R} e^{\gamma x} dP(x) = \lim_{R\to\infty}\frac{1}{\gamma}\log C_R = \infty. \tag{4.19}$$

From (4.19) and (4.15) we have that the existence of the maximum of (4.10) is equivalent to the finite property of (4.13). When finiteness of either occurs, we can invoke (4.15) and (4.16) to claim our conclusion. This completes the proof.

□

Theorem 4.3. *$\forall \gamma > 0$,*

$$\begin{aligned}
&\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma \sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right\}\right] \\
&s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \\
&\varepsilon_t \sim P_t, \varepsilon_t \text{ are independent.}
\end{aligned} \tag{4.20}$$

is equivalent to

$$\begin{aligned}
&\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\tilde{P},\varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma} KL\left(\tilde{P} \| \prod_{t=0}^{T-1} P_t\right) \\
&s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t.
\end{aligned} \tag{4.21}$$

Proof. This is a direct application of Lemma 4.2. □

Theorem 4.4. *$\forall \gamma < 0$,*

$$\begin{aligned}
&\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \frac{1}{\gamma} \log \mathbb{E}\left[\exp\left\{\gamma \sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right\}\right] \\
&s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t \\
&\varepsilon_t \sim P_t, \varepsilon_t \text{ are independent.}
\end{aligned} \tag{4.22}$$

is equivalent to

$$\begin{aligned}
&\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \min_{\tilde{P}_t,\varepsilon_t \sim \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{\gamma} KL\left(\tilde{P} \| \prod_{t=0}^{T-1} P_t\right) \\
&s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \varepsilon_t.
\end{aligned} \tag{4.23}$$

Proof. This is also an application of Lemma 4.2. But since now $\gamma < 0$, we have to reformulate

(4.22) to be

$$
\begin{aligned}
&\max_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} -\frac{1}{\gamma}\log\mathbb{E}\left[\exp\left\{-\gamma\left(-\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right\}\right)\right]\\
&\qquad s.t. x_{t+1}=A_tx_t+B_tu_t+C_t\varepsilon_t\\
&\qquad \varepsilon_t\sim P_t.
\end{aligned} \tag{4.24}
$$

Then by Lemma 4.2, we have

$$
\begin{aligned}
&-\frac{1}{\gamma}\log\mathbb{E}\left[\exp\left\{-\gamma\left(-\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right\}\right)\right]\\
&=\max_{\tilde{P},\varepsilon_t\sim\tilde{P}_t} -\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right]+\frac{1}{\gamma}KL\left(\tilde{P}\|\prod_{t=0}^{T-1}P_t\right).
\end{aligned} \tag{4.25}
$$

This gives the equivalent form of (4.24)

$$
\begin{aligned}
&\max_{u_t(x_t)}\max_{\tilde{P},\varepsilon_t\sim\tilde{P}_t} -\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^TQ_{t+1}x_{t+1}+u_t^TR_tu_t\right)\right]+\frac{1}{\gamma}KL\left(\tilde{P}\|\prod_{t=0}^{T-1}P_t\right)\\
&s.t. x_{t+1}=A_tx_t+B_tu_t+C_t\varepsilon_t\\
&\qquad \varepsilon_t\sim P_t,
\end{aligned} \tag{4.26}
$$

which is equivalent to (4.23).

□

KL-divergence is a measure comparing the similarity between two distributions and is not necessary. There are many other choices other than KL-divergence to compare the similarity, for example, W_2 or our proposed generalized Wasserstein distance. We claim the equivalent forms of (4.1) and (4.2) here to point out that the two scenarios, $\gamma > 0$ for minimax problems and $\gamma < 0$ for minimum problems, are, in fact, from a unified form.

4.4 Minimax Problem

We will first consider the case $\gamma > 0$ and make use of a strong duality result to reformulate (4.1). With what we have shown in §2.4 for general forms, we have

Theorem 4.5. (4.1) *is equivalent to*

$$\begin{aligned}\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t} \mathbb{E}&\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\ &- \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right) \\ s.t. x_{t+1} &= A_t x_t + B_t u_t + C_t\hat{\varepsilon}_t.\end{aligned} \tag{4.27}$$

Proof. This is a direct application of Theorem 2.7. □

The minimax problem is hard to optimize as it is over function space. Instead, we consider the problem that exchanges the optimization order,

$$\begin{aligned}\max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} &\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\ &- \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right).\end{aligned} \tag{4.28}$$

Theorem 4.6. $\exists\gamma_0, \forall\gamma < \gamma_0,$

$$\begin{aligned}\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t} &\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\ &- \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right)\end{aligned} \tag{4.29a}$$

$$\begin{aligned}= \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} &\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\ &- \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right).\end{aligned} \tag{4.29b}$$

We will need a few lemmas and theorems for the proof of Theorem 4.6. The first one is the closed-form solution of risk-free linear quadratic control when noises have non-zero mean.

Lemma 4.7 (Bellman equation)**.** *Define*

$$V_t(x_t) = \min_{u_t} \mathbb{E}\left[(A_t x_t + B_t u_t + C_t\hat{\varepsilon}_t)^T Q_{t+1}(A_t x_t + B_t u_t + C_t\hat{\varepsilon}_t) + u_t^T R_t u_t \right. \\ \left. + V_{t+1}(A_t x_t + B_t u_t + C_t\hat{\varepsilon}_t)\right], \tag{4.30}$$

with the terminal condition

$$V_T(x_T) = 0, \tag{4.31}$$

then we have

$$V_t(x_t) = x_t^T \Xi_{t,0} x_t + \sum_{\tau=t}^{T-1} x_t^T \Xi_{t,1}^{\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] + \sum_{\tau_1=t}^{T-1}\sum_{\tau_2=\tau_1}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_{\tau_1}\right]^T \Xi_{t,2}^{\tau_1,\tau_2} \mathbb{E}\left[\hat{\varepsilon}_{\tau_2}\right] + \sum_{\tau=t}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_\tau^T \tilde{\Xi}_\tau \hat{\varepsilon}_\tau\right] \tag{4.32}$$

and the optimal control policy of (4.30) *is*

$$u_t(x_t) = U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right], \tag{4.33}$$

where $\Xi_{t,0}^{\tau}, \Xi_{t,1}^{\tau}, \Xi_{t,2}^{\tau_1,\tau_2}, \tilde{\Xi}_\tau, U_t, U_{t,\tau}$ *are matrices in proper dimension, and*

$$U_t = -\left(B_t^T Q_{t+1} B_t + R_t + B_t^T \Xi_{t+1,0} B_t\right)^{-1}\left(B_t^T Q_{t+1} A_t + B_t^T \Xi_{t+1,0} A_t\right), \tag{4.34}$$

$$U_{t,t} = -\left(B_t^T Q_{t+1} B_t + R_t + B_t^T \Xi_{t+1,0} B_t\right)^{-1}\left(B_t^T Q_{t+1} C_t + B_t^T \Xi_{t+1,0} C_t\right), \tag{4.35}$$

for $\tau > t$,

$$U_{t,\tau} = -\frac{1}{2}\left(B_t^T Q_{t+1} B_t + R_t + B_t^T \Xi_{t+1,0} B_t\right)^{-1} B_t^T \Xi_{t+1,1}^{\tau}, \tag{4.36}$$

$$\Xi_{t,0} = (A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0}\right)(A_t + B_t U_t) + U_t^T R_t U_t, \tag{4.37}$$

$$\Xi_{t,1}^{t} = 2\left(A_t + B_t U_t\right)^T \left(Q_{t+1} + \Xi_{t+1,0}\right)\left(C_t + B_t U_{t,t}\right) + 2U_t^T R_t U_{t,t}, \tag{4.38}$$

for $\tau > t$,

$$\Xi_{t,1}^{\tau} = 2\left(A_t + B_t U_t\right)^T \left(Q_{t+1} + \Xi_{t+1,0}\right) B_t U_{t,\tau} + 2U_t^T R_t U_{t,\tau} + \left(A_t + B_t U_t\right)^T \Xi_{t+1,1}^{\tau}, \quad (4.39)$$

$$\Xi_{t,2}^{t,t} = U_{t,t}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0}\right) B_t + R_t\right) U_{t,t} + 2C_t^T \left(Q_{t+1} + \Xi_{t+1,0}\right) B_t U_{t,t}, \quad (4.40)$$

for $\tau > t$,

$$\begin{aligned} \Xi_{t,2}^{t,\tau} = &\, 2U_{t,t}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0}\right) B_t + R_t\right) U_{t,\tau} \\ &+ 2C_t^T \left(Q_{t+1} + \Xi_{t+1,0}\right) B_t U_{t,\tau} + C_t^T \Xi_{t+1,1}^{\tau} + U_{t,t}^T B_t^T \Xi_{t+1,1}^{\tau}, \end{aligned} \quad (4.41)$$

for $\tau_1 > t, \tau_2 = \tau_1$,

$$\Xi_{t,2}^{\tau_1,\tau_2} = \Xi_{t+1,2}^{\tau_1,\tau_2} + U_{t,\tau_1}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0}\right) B_t + R_t\right) U_{t,\tau_2} + U_{t,\tau_2}^T B_t^T \Xi_{t+1,1}^{\tau_1}, \quad (4.42)$$

for $\tau_1 > t, \tau_2 > \tau_1$,

$$\Xi_{t,2}^{\tau_1,\tau_2} = \Xi_{t+1,2}^{\tau_1,\tau_2} + 2U_{t,\tau_1}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0}\right) B_t + R_t\right) U_{t,\tau_2} + U_{t,\tau_1}^T B_t^T \Xi_{t+1,1}^{\tau_2} + {\Xi_{t+1,1}^{\tau_1}}^T B_t U_{t,\tau_2}, \quad (4.43)$$

$$\tilde{\Xi}_t = C_t^T \left(Q_{t+1} + \Xi_{t+1,0}\right) C_t. \quad (4.44)$$

Proof. This can be shown by induction. Firstly it is true for $t = T$ with all matrices being 0. Then assume it is true for $t = m + 1$, for $t = m$,

$$\begin{aligned} V_t(x_t) = &\min_{u_t} \mathbb{E}\Big[(A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T Q_{t+1} (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t) + u_t^T R_t u_t \\ &\qquad + V_{t+1}(A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)\Big] \\ = &\min_{u_t} \mathbb{E}\Bigg[(A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T Q_{t+1} (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t) + u_t^T R_t u_t \\ &+ (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T \Xi_{t+1,0} (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t) \\ &+ \sum_{\tau=t+1}^{T-1} (A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t)^T \Xi_{t+1,1}^{\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] + \sum_{\tau_1=t+1}^{T-1} \sum_{\tau_2=\tau_1}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_{\tau_1}\right] \Xi_{t,2}^{\tau_1,\tau_2} \mathbb{E}\left[\hat{\varepsilon}_{\tau_2}\right] \\ &+ \sum_{\tau=t+1}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_\tau^T \tilde{\Xi}_\tau \hat{\varepsilon}_\tau\right]\Bigg] \end{aligned} \quad (4.45)$$

This is a quadratic form of u_t, by optimality condition, we have

$$\begin{aligned}&2B_t^T Q_{t+1} A_t x_t + 2B_t^T Q_{t+1} B_t u_t + 2B_t^T Q_{t+1} C_t \mathbb{E}\left[\hat{\varepsilon}_t\right] + 2R_t u_t + 2B_t^T \Xi_{t+1,0} B_t u_t \\ &+ 2B_t^T \Xi_{t+1,0} A_t x_t + 2B_t^T \Xi_{t+1,0} C_t \mathbb{E}\left[\hat{\varepsilon}_t\right] + \sum_{\tau=t+1}^{T-1} B_t^T \Xi_{t+1,1}^{\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] = 0.\end{aligned} \tag{4.46}$$

This solves

$$\begin{aligned}u_t &= -\left(B_t^T Q_{t+1} B_t + R_t + B_t^T \Xi_{t+1,0} B_t\right)^{-1} \Bigg(B_t^T Q_{t+1} A_t x_t + B_t^T Q_{t+1} C_t \mathbb{E}\left[\hat{\varepsilon}_t\right] \\ &\qquad + B_t^T \Xi_{t+1,0} A_t x_t + B_t^T \Xi_{t+1,0} C_t \mathbb{E}\left[\hat{\varepsilon}_t\right] + \frac{1}{2} \sum_{\tau=t+1}^{T-1} B_t^T \Xi_{t+1,1}^{\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right]\Bigg) \\ &= U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right],\end{aligned} \tag{4.47}$$

where

$$U_t = -\left(B_t^T Q_{t+1} B_t + R_t + B_t^T \Xi_{t+1,0} B_t\right)^{-1} \left(B_t^T Q_{t+1} A_t + B_t^T \Xi_{t+1,0} A_t\right), \tag{4.48}$$

$$U_{t,t} = -\left(B_t^T Q_{t+1} B_t + R_t + B_t^T \Xi_{t+1,0} B_t\right)^{-1} \left(B_t^T Q_{t+1} C_t + B_t^T \Xi_{t+1,0} C_t\right), \tag{4.49}$$

for $\tau > t$,

$$U_{t,\tau} = -\frac{1}{2} \left(B_t^T Q_{t+1} B_t + R_t + B_t^T \Xi_{t+1,0} B_t\right)^{-1} B_t^T \Xi_{t+1,1}^{\tau}. \tag{4.50}$$

Then we have

$$
\begin{aligned}
V_t(x_t) = \mathbb{E}\Bigg[&\left(A_t x_t + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) + C_t \hat{\varepsilon}_t \right)^T Q_{t+1} \Bigg(A_t x_t \\
&\qquad + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) + C_t \hat{\varepsilon}_t \Bigg) \\
&+ \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right)^T R_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) \\
&+ \left(A_t x_t + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) + C_t \hat{\varepsilon}_t \right)^T \Xi_{t+1,0} \Bigg(A_t x_t \\
&\qquad + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) + C_t \hat{\varepsilon}_t \Bigg) \\
&+ \sum_{\tau_1=t+1}^{T-1} \left(A_t x_t + B_t \left(U_t x_t + \sum_{\tau_2=t}^{T-1} U_{t,\tau_2} \mathbb{E}\left[\hat{\varepsilon}_{\tau_2}\right] \right) + C_t \hat{\varepsilon}_t \right)^T \Xi_{t+1,1}^{\tau_1} \mathbb{E}\left[\hat{\varepsilon}_{\tau_1}\right] \\
&+ \sum_{\tau_1=t+1}^{T-1} \sum_{\tau_2=\tau_1}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_{\tau_1}\right] \Xi_{t,2}^{\tau_1,\tau_2} \mathbb{E}\left[\hat{\varepsilon}_{\tau_2}\right] + \sum_{\tau=t+1}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_\tau^T \tilde{\Xi}_\tau \hat{\varepsilon}_\tau\right] \Bigg]
\end{aligned}
\tag{4.51}
$$

Multiply terms out and combine them, we have

$$
\begin{aligned}
V_t(x_t) = {} & x_t^T \left((A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) (A_t + B_t U_t) + U_t^T R_t U_t \right) x_t \\
& + 2 x_t^T (A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) \left(C_t \mathbb{E}\left[\hat{\varepsilon}_t\right] + \sum_{\tau=t}^{T-1} B_t U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) \\
& + 2 x_t^T U_t^T R_t \left(\sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) + x_t^T (A_t + B_t U_t)^T \left(\sum_{\tau=t+1}^{T-1} \Xi_{t+1,1}^{\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) \\
& + \mathbb{E}\left[\hat{\varepsilon}_t^T C_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) C_t \hat{\varepsilon}_t \right] \\
& + \left(\sum_{\tau=t}^{T-1} B_t U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) \left(\sum_{\tau=t}^{T-1} B_t U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) \\
& + 2 \mathbb{E}\left[\hat{\varepsilon}_t\right]^T C_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) \left(\sum_{\tau=t}^{T-1} B_t U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) \\
& + \left(\sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right)^T R_t \left(\sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) \\
& + \sum_{\tau_1=t+1}^{T-1} \left(C_t \mathbb{E}\left[\hat{\varepsilon}_t\right] + \sum_{\tau_2=t}^{T-1} U_{t,\tau_2} \mathbb{E}\left[\hat{\varepsilon}_{\tau_2}\right] \right)^T \Xi_{t+1,1}^{\tau_1} \mathbb{E}[\hat{\varepsilon}_{\tau_1}] \\
& + \sum_{\tau_1=t+1}^{T-1} \sum_{\tau_2=\tau_1}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_{\tau_1}\right] \Xi_{t+1,2}^{\tau_1,\tau_2} \mathbb{E}\left[\hat{\varepsilon}_{\tau_2}\right] + \sum_{\tau=t+1}^{T-1} \mathbb{E}\left[\hat{\varepsilon}_\tau^T \tilde{\Xi}_\tau \hat{\varepsilon}_\tau \right].
\end{aligned}
\tag{4.52}
$$

Reformulate (4.52) into the form of (4.32), we have

$$\begin{aligned}
V_t(x_t) = {} & x_t^T \left((A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) (A_t + B_t U_t) + U_t^T R_t U_t \right) x_t \\
& + x_t^T \left(2 (A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) \left(C_t + B_t U_{t,t} \right) + 2 U_t^T R_t U_{t,t} \right) \mathbb{E} \left[\hat{\varepsilon}_t \right] \\
& + \sum_{\tau=t+1}^{T-1} x_t^T \left(2 (A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t U_{t,\tau} + 2 U_t^T R_t U_{t,\tau} \right. \\
& \qquad \left. + (A_t + B_t U_t)^T \Xi_{t+1,1}^{\tau} \right) \mathbb{E} \left[\hat{\varepsilon}_\tau \right] \\
& + \sum_{\tau_1=t}^{T-1} \sum_{\tau_2=t}^{T-1} \mathbb{E} \left[\hat{\varepsilon}_{\tau_1} \right]^T U_{t,\tau_1}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t + R_t \right) U_{t,\tau_2} \mathbb{E} \left[\hat{\varepsilon}_{\tau_2} \right] \\
& + \sum_{\tau=t}^{T-1} \mathbb{E} \left[\hat{\varepsilon}_t \right]^T \left(2 C_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t U_{t,\tau} \right) \mathbb{E} \left[\hat{\varepsilon}_\tau \right] + \sum_{\tau=t+1}^{T-1} \mathbb{E} \left[\hat{\varepsilon}_t \right]^T C_t^T \Xi_{t+1,1}^{\tau} \mathbb{E} \left[\hat{\varepsilon}_\tau \right] \\
& + \sum_{\tau_1=t+1}^{T-1} \sum_{\tau_2=t}^{T-1} \mathbb{E} \left[\hat{\varepsilon}_{\tau_2} \right]^T U_{t,\tau_2}^T \Xi_{t+1,1}^{\tau_1} \mathbb{E} \left[\hat{\varepsilon}_{\tau_1} \right] + \sum_{\tau_1=t+1}^{T-1} \sum_{\tau_2=\tau_1}^{T-1} \mathbb{E} \left[\hat{\varepsilon}_{\tau_1} \right] \Xi_{t+1,2}^{\tau_1,\tau_2} \mathbb{E} \left[\hat{\varepsilon}_{\tau_2} \right] \\
& + \mathbb{E} \left[\hat{\varepsilon}_t^T C_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) C_t \hat{\varepsilon}_t \right] + \sum_{\tau=t+1}^{T-1} \mathbb{E} \left[\hat{\varepsilon}_\tau^T \tilde{\Xi}_\tau \hat{\varepsilon}_\tau \right].
\end{aligned} \tag{4.53}$$

Thus,

$$\Xi_{t,0} = (A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) (A_t + B_t U_t) + U_t^T R_t U_t, \tag{4.54}$$

$$\Xi_{t,1}^{t} = 2 (A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) \left(C_t + B_t U_{t,t} \right) + 2 U_t^T R_t U_{t,t}, \tag{4.55}$$

for $\tau > t$,

$$\Xi_{t,1}^{\tau} = 2 (A_t + B_t U_t)^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t U_{t,\tau} + 2 U_t^T R_t U_{t,\tau} + (A_t + B_t U_t)^T \Xi_{t+1,1}^{\tau}, \tag{4.56}$$

$$\Xi_{t,2}^{t,t} = U_{t,t}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t + R_t \right) U_{t,t} + 2 C_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t U_{t,t}, \tag{4.57}$$

for $\tau > t$,

$$\begin{aligned}
\Xi_{t,2}^{t,\tau} = {} & 2 U_{t,t}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t + R_t \right) U_{t,\tau} \\
& + 2 C_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t U_{t,\tau} + C_t^T \Xi_{t+1,1}^{\tau} + U_{t,t}^T B_t^T \Xi_{t+1,1}^{\tau},
\end{aligned} \tag{4.58}$$

for $\tau_1 > t, \tau_2 = \tau_1$,

$$\Xi_{t,2}^{\tau_1,\tau_2} = \Xi_{t+1,2}^{\tau_1,\tau_2} + U_{t,\tau_1}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t + R_t \right) U_{t,\tau_2} + U_{t,\tau_2}^T B_t^T \Xi_{t+1,1}^{\tau_1}, \quad (4.59)$$

for $\tau_1 > t, \tau_2 > \tau_1$,

$$\Xi_{t,2}^{\tau_1,\tau_2} = \Xi_{t+1,2}^{\tau_1,\tau_2} + 2U_{t,\tau_1}^T \left(B_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) B_t + R_t \right) U_{t,\tau_2} + U_{t,\tau_1}^T B_t^T \Xi_{t+1,1}^{\tau_2} + {\Xi_{t+1,1}^{\tau_1}}^T B_t U_{t,\tau_2}, \quad (4.60)$$

$$\tilde{\Xi}_t = C_t^T \left(Q_{t+1} + \Xi_{t+1,0} \right) C_t. \quad (4.61)$$

□

Remark 4.7.1. *Through the recursion we see $V_0(x_0)$ is a quadratic function of*

$$\left(\hat{\varepsilon}_{0,1}^T, \hat{\varepsilon}_{0,2}^T, \ldots, \hat{\varepsilon}_{0,n}^T, \hat{\varepsilon}_{1,1}^T, \hat{\varepsilon}_{1,2}^T, \ldots, \hat{\varepsilon}_{1,n}^T, \ldots, \hat{\varepsilon}_{T-1,1}^T, \hat{\varepsilon}_{T-1,2}^T, \ldots, \hat{\varepsilon}_{T-1,n}^T \right)^T. \quad (4.62)$$

Theorem 4.8. *The optimal control of* (4.30) *is given by*

$$u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1}) = H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau} \varepsilon_\tau + \sum_{\tau=0}^{T-1} h_{t,\tau} \mathbb{E}\left[\varepsilon_\tau\right], \quad (4.63)$$

where

$$H_t = U_t \prod_{\tau=0}^{t-1} \left(A_\tau + B_\tau U_\tau \right), \quad (4.64)$$

$$H_{t,\tau} = \sum_{s=0}^{t-1} U_t \prod_{w=s+1}^{t-1} \left(A_w + B_w U_w \right) C_\tau, \quad (4.65)$$

$$h_{t,\tau} = \begin{cases} \sum_{s=0}^{\tau} U_t \left(\prod_{w=s+1}^{t-1} \left(A_w + B_w U_w \right) B_s U_{s,\tau} \right) & , \tau < t \\ \sum_{s=0}^{t-1} U_t \left(\prod_{w=s+1}^{t-1} \left(A_w + B_w U_w \right) B_s U_{s,\tau} \right) + U_{t,\tau} & , \tau \geq t \end{cases}. \quad (4.66)$$

The definition of the product on matrices multiplication is

$$\prod_{s=i}^{j} \left(A_s + B_s U_s \right) = \begin{cases} \left(A_j + B_j U_j \right) \left(A_{j-1} + B_{j-1} U_{j-1} \right) \cdots \left(A_i + B_i U_i \right) & , j \geq i \\ \mathbf{I} & , j = i-1. \end{cases} \quad (4.67)$$

Proof. By (4.33),

$$u_t(x_t) = U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\mathbb{E}\left[\hat{\varepsilon}_\tau\right]. \tag{4.68}$$

Also, by dynamics,

$$\begin{aligned}
x_{t+1} &= A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t \\
&= A_t x_t + B_t \left(U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau}\mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) + C_t \hat{\varepsilon}_t \\
&= \left(A_t + B_t U_t\right) x_t + C_t \hat{\varepsilon}_t + \sum_{\tau=t}^{T-1} B_t U_{t,\tau}\mathbb{E}\left[\hat{\varepsilon}_\tau\right] \\
&= \left(A_t + B_t U_t\right)\left(A_{t-1} + B_{t-1}U_{t-1}\right) x_{t-1} + \left(C_t \hat{\varepsilon}_t + (A_t + B_t U_t) C_{t-1}\hat{\varepsilon}_{t-1}\right) \\
&\quad + \left(\sum_{\tau=t}^{T-1} B_t U_{t,\tau}\mathbb{E}\left[\hat{\varepsilon}_\tau\right] + \left(A_t + B_t U_t\right) \sum_{\tau=t-1}^{T-1} B_{t-1}U_{t-1,\tau}\mathbb{E}\left[\hat{\varepsilon}_\tau\right] \right) \\
&= \dots \\
&= \prod_{\tau=0}^{t} \left(A_\tau + B_\tau U_\tau\right) x_0 + \sum_{s=0}^{t} \prod_{\tau=s+1}^{t} \left(A_\tau + B_\tau U_\tau\right) C_s \hat{\varepsilon}_s \\
&\quad + \sum_{s=0}^{t} \prod_{\tau=s+1}^{t} \left(A_\tau + B_\tau U_\tau\right) \sum_{w=\tau}^{T-1} B_s U_{s,w}\mathbb{E}\left[\hat{\varepsilon}_w\right] \\
&= \prod_{\tau=0}^{t} \left(A_\tau + B_\tau U_\tau\right) x_0 + \sum_{s=0}^{t} \prod_{\tau=s+1}^{t} \left(A_\tau + B_\tau U_\tau\right) C_s \hat{\varepsilon}_s \\
&\quad + \sum_{w=0}^{T-1} \sum_{s=0}^{\min\{w,t\}} \left(\prod_{\tau=s+1}^{t} \left(A_\tau + B_\tau U_\tau\right) B_s U_{s,w} \right) \mathbb{E}\left[\hat{\varepsilon}_w\right].
\end{aligned} \tag{4.69}$$

Thus,

$$
\begin{aligned}
u_t &= U_t x_t + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \\
&= U_t \left(\prod_{\tau=0}^{t-1} \left(A_\tau + B_\tau U_\tau\right) x_0 + \sum_{s=0}^{t-1} \prod_{\tau=s+1}^{t-1} \left(A_\tau + B_\tau U_\tau\right) C_s \hat{\varepsilon}_s \right. \\
&\quad \left. + \sum_{w=0}^{T-1} \sum_{s=0}^{\min\{w,t-1\}} \left(\prod_{\tau=s+1}^{t-1} \left(A_\tau + B_\tau U_\tau\right) B_s U_{s,w} \right) \mathbb{E}\left[\hat{\varepsilon}_w\right] \right) + \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right] \\
&= U_t \prod_{\tau=0}^{t-1} \left(A_\tau + B_\tau U_\tau\right) x_0 + \sum_{s=0}^{t-1} U_t \prod_{\tau=s+1}^{t-1} \left(A_\tau + B_\tau U_\tau\right) C_s \hat{\varepsilon}_s \\
&\quad + \sum_{w=0}^{t-1} \left(\sum_{s=0}^{w} U_t \left(\prod_{\tau=s+1}^{t-1} \left(A_\tau + B_\tau U_\tau\right) B_s U_{s,w} \right) \right) \mathbb{E}\left[\hat{\varepsilon}_w\right] \\
&\quad + \sum_{w=t}^{T-1} \left(\sum_{s=0}^{t-1} U_t \left(\prod_{\tau=s+1}^{t-1} \left(A_\tau + B_\tau U_\tau\right) B_s U_{s,w} \right) + U_{t,w} \right) \mathbb{E}\left[\hat{\varepsilon}_w\right].
\end{aligned}
\tag{4.70}
$$

□

Theorem 4.9.

$$
\begin{aligned}
&\max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \min_{u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \right] \\
&\quad - \frac{1}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\| \right) \\
&= \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \min_{h_t, u_t = H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau} \varepsilon_t + h_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t \right) \right] \\
&\quad - \frac{1}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\| \right),
\end{aligned}
\tag{4.71}
$$

where $h_t \in \mathbb{R}^{d_2}$, H_t and $H_{t,\tau}$ are defined in (4.64) *and* (4.65)*.*

Proof. This is a direct application of Theorem 4.8. □

Theorem 4.10. *Parameterize* $u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})$ *with*

$$u_t = H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau}\varepsilon_\tau + h_t, \tag{4.72}$$

where $h_t \in \mathbb{R}^{d_2}$, H_t *and* $H_{t,\tau}$ *are defined in* (4.64) *and* (4.65), *then* $\exists\gamma_0, \forall\gamma < \gamma_0$,

$$\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{n\gamma}\sum_{t\ =0}^{T-1}\sum_{i\ =1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right) \tag{4.73}$$

is a concave function of $\hat{\varepsilon}_{t,i}$.

Proof. As $\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|$ is a concave function of $\hat{\varepsilon}_{t,i}$, we only need to prove the concavity of

$$\mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] - \frac{1}{n\gamma}\sum_{t\ =0}^{T-1}\sum_{i\ =1}^{n}\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2. \tag{4.74}$$

We compute the Hessian of (4.74) to check concavity. With the parameterization (4.72), the first term of (4.74) gives a quadratic form of $\hat{\varepsilon}_{t,i}$,

$$\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\left(\sum_{\tau=0}^{t}\left(\prod_{s=\tau+1}^{t} A_s\right)\left(B_\tau\left(H_\tau x_0 + \sum_{w=0}^{s-1} H_{\tau,w}\hat{\varepsilon}_w + h_\tau\right) + C_\tau\hat{\varepsilon}_\tau\right)\right)^T Q_{t+1} \right.\right. \tag{4.75}$$
$$\left(\sum_{\tau=0}^{t}\left(\prod_{s=\tau+1}^{t} A_s\right)\left(B_\tau\left(H_\tau x_0 + \sum_{w=0}^{s-1} H_{\tau,w}\hat{\varepsilon}_w + h_\tau\right) + C_\tau\hat{\varepsilon}_\tau\right)\right)$$
$$\left.\left. + \left(H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau}\varepsilon_\tau + h_t\right)^T R_t \left(H_t x_0 + \sum_{\tau=0}^{t-1} H_{t,\tau}\varepsilon_\tau + h_t\right)\right)\right].$$

Note that h_t doesn't influence the Hessian of the first term, and $H_t, H_{t,\tau}$ are given. It means for arbitrary values of h_t and $\hat{\varepsilon}_{t,i}$, the Hessian of the first term is always fixed, which is decided by the system itself. Note the Hessian of the second term of (4.74) is $-\frac{1}{n\gamma}\mathbf{I}$. Thus, when $\gamma \to 0$, we must have the total Hessian negative-definite. This gives the concavity. □

Now we can prove Theorem 4.6,

Proof of Theorem 4.6. We start from a restricted maximin problem and choose γ_0 to be γ_0 in Theorem 4.10,

$$\begin{aligned}&\max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t,\|\hat{\varepsilon}_{t,i}\|\leq r}\ \min_{u_t(\varepsilon_0,\varepsilon_1,\dots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1}x_{t+1}+u_t^T R_t u_t\right)\right]\\&-\frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right),\end{aligned} \tag{4.76}$$

where r is a constant that uniformly bounds $\hat{\varepsilon}_{t,i}$. Then, by Theorem 4.8 and Theorem 4.9, (4.76) equals

$$\begin{aligned}&\max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t,\|\hat{\varepsilon}_{t,i}\|\leq r}\ \min_{h_t,u_t=H_t x_0+\sum_{\tau=0}^{t-1}H_{t,\tau}\varepsilon_\tau+h_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1}x_{t+1}+u_t^T R_t u_t\right)\right]\\&-\frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right).\end{aligned} \tag{4.77}$$

The objective of (4.77) is a convex function of h_t and, by Theorem 4.10, a concave function of $\hat{\varepsilon}_{t,i}$. Note that the constraint $\|\hat{\varepsilon}_{t,i}\|\leq r$ limit $\hat{\varepsilon}_{t,i}$ to a compact set. Thus, we can apply Sion's minimax theorem and have (4.77) equal

$$\begin{aligned}&\min_{h_t,u_t=H_t x_0+\sum_{\tau=0}^{t-1}H_{t,\tau}\varepsilon_\tau+h_t}\ \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t,\|\hat{\varepsilon}_{t,i}\|\leq r} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1}x_{t+1}+u_t^T R_t u_t\right)\right]\\&-\frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right).\end{aligned} \tag{4.78}$$

As (4.78) is a minimization problem putting constraints on u_t, it is always greater than the unconstraint problem

$$\begin{aligned}&\min_{u_t(\varepsilon_0,\varepsilon_1,\dots,\varepsilon_{t-1})}\ \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t,\|\hat{\varepsilon}_{t,i}\|\leq r} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1}x_{t+1}+u_t^T R_t u_t\right)\right]\\&-\frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right).\end{aligned} \tag{4.79}$$

The max-min inequality gives

$$
\begin{aligned}
(4.79) \geq \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t, \|\hat{\varepsilon}_{t,i}\| \leq r} \min_{u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})} & \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
& - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right).
\end{aligned} \tag{4.80}
$$

As the RHS of (4.80) equals (4.77), this means all inequalities here are, in fact, equalities. So far, we prove for all r that

$$
\begin{aligned}
& \min_{u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t, \|\hat{\varepsilon}_{t,i}\| \leq r} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
& \quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right)
\end{aligned} \tag{4.81a}
$$

$$
\begin{aligned}
= & \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t, \|\hat{\varepsilon}_{t,i}\| \leq r} \min_{u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
& \quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right).
\end{aligned} \tag{4.81b}
$$

For the unconstrained maximin problem

$$
\begin{aligned}
\max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \min_{u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})} & \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
& - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right),
\end{aligned} \tag{4.82}
$$

denote $\hat{\varepsilon}_t^*$ to be its maximum point and $r_1 = \max_{t,i}\|\hat{\varepsilon}_{t,i}^*\|$. At this r_1, denote its minimum point to be u_t^*. Then, $\forall r \geq r_1$, we have

$$
\begin{aligned}
&\max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right)
\end{aligned} \tag{4.83a}
$$

$$
\begin{aligned}
&= \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t,\|\hat{\varepsilon}_{t,i}\|\le r_1} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right)
\end{aligned} \tag{4.83b}
$$

$$
\begin{aligned}
&= \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t,\|\hat{\varepsilon}_{t,i}\|\le r} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right)
\end{aligned} \tag{4.83c}
$$

$$
\begin{aligned}
&= \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t,\|\hat{\varepsilon}_{t,i}\|\le r} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right).
\end{aligned} \tag{4.83d}
$$

As the maximum and minimum point of (4.83c) are fixed when $r \ge r_1$, the minimum point of (4.83d) is also fixed, and the same with the minimum point of (4.83c). Denote this minimum point to be u_t^* and $\hat{\varepsilon}_t^{**}$ to be the maximum point of the problem

$$
\max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^{*T} R_t u_t^*\right)\right] - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right). \tag{4.84}
$$

With $r_2 = \max_{t,i}\|\hat{\varepsilon}_{t,i}^{**}\|$, we have

$$
\begin{aligned}
&\max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^{*T} R_t u_t^*\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right) && (4.85a)\\
&= \max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t, \|\hat{\varepsilon}_{t,i}\| \le \max\{r_1,r_2\}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^{*T} R_t u_t^*\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right) && (4.85b)\\
&= \max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t, \|\hat{\varepsilon}_{t,i}\| \le \max\{r_1,r_2\}} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right) && (4.85c)\\
&= \max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right). && (4.85d)
\end{aligned}
$$

As minimax problem is always greater than or equal to maximin problem,

$$
\begin{aligned}
&\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right) \\
&\ge \max_{\hat{P}_t,\hat{\varepsilon}_t \sim \hat{P}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right),
\end{aligned} \tag{4.86}
$$

and we see the lower bound is attained at u_t^* by (4.85), we have

$$\begin{aligned}
&\max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right) \\
&= \min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\
&\quad - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right).
\end{aligned} \tag{4.87}$$

□

4.5 Group Lasso

Group Lasso is a tool widely used in statistics to select significant coefficients among lots of possibly unrelated coefficients. It has been used intensively in fields like biology and computer science but has not yet been advanced in control theory. The standard form of group lasso can be formulated as

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|_2^2 + \lambda \sum_{g=1}^{G} c_g \|\beta_g\|_2, \tag{4.88}$$

where X is data matrix, y is response, β is coefficient that are concatenated by $\beta_g, g = 1, \ldots, G$, that is

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_G \end{pmatrix}. \tag{4.89}$$

The coefficient β is divided into G groups $\beta_1, \beta_2, \ldots, \beta_G$ and the penalty term $\|\beta_g\|_2$ ensures that when λ is decreased from infinity to 0, β_g will change from zero to non-zero in the order of their importance. That is, only a few high-influential predictors will be selected when λ is properly tuned.

To see the role of group lasso in the robust control, we come back to the previous equivalence and see that

$$\begin{aligned}\max_{\hat{P}_t, \hat{\varepsilon}_t \sim \hat{P}_t} \min_{u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1} x_{t+1} + u_t^T R_t u_t\right)\right] \\ - \frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|^2 + \alpha\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|\right) \\ s.t. x_{t+1} = A_t x_t + B_t u_t + C_t \hat{\varepsilon}_t\end{aligned} \tag{4.90}$$

is, in fact, a group lasso problem, with $\beta_g = \hat{\varepsilon}_{t,i} - \varepsilon_{t,i}$. This is because in the recursion of Lemma 4.7, the objective value of expectation depends quadratically on $\hat{\varepsilon}_{t,i}$. And thus, when we change the sign of (4.90), it becomes group lasso in a non-standard form. The interesting point here is that according to group lasso, $\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}$ will be selected in the order of their influence. That is, with a properly tuned γ in our model, we will only be robust to those points that really need to be considered. And when γ is negative, we will modify those points that can reduce the objective value significantly, which would possibly be classified as erroneous outliers. When γ is positive, this process is reversed and now becomes robust to outliers, which means control policies will be selectively robust to high-influential points.

The standard form group lasso (4.88) has a ADMM iteration to solve (Boyd et al., 2011),

$$\begin{aligned}\beta^k &= (X^T X + \rho I)^{-1}(X^T y + \rho(\alpha^{k-1} - w^{k-1})), \\ \alpha_g^k &= R_{\frac{\lambda c_g}{\rho}}(\beta_g^k + w_g^{k-1}), g = 1, \ldots, G, \\ w^k &= w^{k-1} + \beta^k - \alpha^k,\end{aligned} \tag{4.91}$$

where

$$R_t(x) = (1 - \frac{t}{\|x\|_2})_+ x \tag{4.92}$$

and ρ is a parameter to tune.

To form (4.29b) into the standard form of group lasso, from (4.32), we have

Theorem 4.11.

$$V_0(x_0) = \left(\hat{\varepsilon}_{0,1}^T, \ldots, \hat{\varepsilon}_{T-1,n}^T\right) \Theta \begin{pmatrix} \hat{\varepsilon}_{0,1} \\ \vdots \\ \hat{\varepsilon}_{T-1,n} \end{pmatrix} + x_0^T \tilde{\Theta} \begin{pmatrix} \hat{\varepsilon}_{0,1} \\ \vdots \\ \hat{\varepsilon}_{T-1,n} \end{pmatrix} + x_0^T \Xi_{0,0} x_0, \tag{4.93}$$

where

$$\Theta_{(t,i),(t,i)} = \frac{1}{n}\tilde{\Xi}_t + \frac{1}{n^2}\Xi_{0,2}^{t,t}, \tag{4.94}$$

for $j \neq i$,

$$\begin{aligned} \Theta_{(t,i),(t,j)} &= \Theta_{(t,j),(t,i)} \\ &= \frac{1}{2n^2}\Xi_{0,2}^{t,t}, \end{aligned} \tag{4.95}$$

for $\tau \neq t$,

$$\begin{aligned} \Theta_{(t,i),(\tau,j)} &= \Theta_{(\tau,j),(t,i)} \\ &= \frac{1}{2n^2}\Xi_{0,2}^{\min\{t,\tau\},\max\{t,\tau\}}, \end{aligned} \tag{4.96}$$

$$\tilde{\Theta}_{(t,i)} = \frac{1}{n}\Xi_{0,1}^{t}. \tag{4.97}$$

By denote $\hat{\epsilon} = \left(\hat{\varepsilon}_{0,1}^T, \ldots, \hat{\varepsilon}_{T-1,n}^T\right)^T$ and $\epsilon = \left(\varepsilon_{0,1}^T, \ldots, \varepsilon_{T-1,n}^T\right)^T$, (4.29b) can be reformulated as

$$
\begin{aligned}
&\max_{\hat{\epsilon}} \hat{\epsilon}^T \Theta \hat{\epsilon} + x_0^T \tilde{\Theta} \hat{\epsilon} + x_0^T \Xi_{0,0} x_0 - \frac{1}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2^2 + \alpha \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2 \right) \\
&= \max_{\hat{\epsilon}} (\hat{\epsilon} - \epsilon + \epsilon)^T \Theta (\hat{\epsilon} - \epsilon + \epsilon) + x_0^T \tilde{\Theta} (\hat{\epsilon} - \epsilon + \epsilon) + x_0^T \Xi_{0,0} x_0 \\
&- \frac{1}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2^2 + \alpha \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2 \right) \\
&= \max_{\hat{\epsilon}-\epsilon} (\hat{\epsilon} - \epsilon)^T \Theta (\hat{\epsilon} - \epsilon) + \epsilon^T \Theta (\hat{\epsilon} - \epsilon) + x_0^T \tilde{\Theta} (\hat{\epsilon} - \epsilon) \\
&- \frac{1}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2^2 + \alpha \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2 \right) + Constant \qquad (4.98) \\
&= \min_{\hat{\epsilon}-\epsilon} -(\hat{\epsilon} - \epsilon)^T \Theta (\hat{\epsilon} - \epsilon) - \epsilon^T \Theta (\hat{\epsilon} - \epsilon) - x_0^T \tilde{\Theta} (\hat{\epsilon} - \epsilon) \\
&+ \frac{1}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2^2 + \alpha \|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2 \right) + Constant \\
&= \min_{\hat{\epsilon}-\epsilon} \frac{1}{n\gamma} (\hat{\epsilon} - \epsilon)^T (\hat{\epsilon} - \epsilon) - (\hat{\epsilon} - \epsilon)^T \Theta (\hat{\epsilon} - \epsilon) - \epsilon^T \Theta (\hat{\epsilon} - \epsilon) - x_0^T \tilde{\Theta} (\hat{\epsilon} - \epsilon) \\
&+ \frac{\alpha}{n\gamma} \sum_{t=0}^{T-1} \sum_{i=1}^{n} \left(\|\hat{\varepsilon}_{t,i} - \varepsilon_{t,i}\|_2 \right) + Constant
\end{aligned}
$$

Thus we have

Theorem 4.12. (4.29b) *is equivalent to a group lasso problem in the form of* (4.88) *with*

$$X = (\frac{1}{n\gamma}\mathbf{I} - \Theta)^{\frac{1}{2}}, \tag{4.99}$$

$$y = \frac{1}{2}(\frac{1}{n\gamma}\mathbf{I} - \Theta)^{-\frac{1}{2}} \left(\Theta\varepsilon + \tilde{\Theta}^T x_0 \right), \tag{4.100}$$

$$\lambda = \frac{2\alpha}{n\gamma}, c_g = 1. \tag{4.101}$$

We give the algorithm for the minimax problem in Algorithm 3.

4.6 Minimum Problem

The minimum problem is in the opposite direction of the minimax problem. It can be formed into a group lasso following the same procedure in §4.5. We give the detailed quantities here.

Theorem 4.13. (4.2) *is equivalent to*

$$\begin{aligned}\min_{u_t(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})} \min_{\hat{P}_t,\hat{\varepsilon}_t\sim\hat{P}_t} \mathbb{E}&\left[\sum_{t=0}^{T-1}\left(x_{t+1}^T Q_{t+1}x_{t+1}+u_t^T R_t u_t\right)\right]\\ &+\frac{1}{n\gamma}\sum_{t=0}^{T-1}\sum_{i=1}^{n}\left(\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|^2+\alpha\|\hat{\varepsilon}_{t,i}-\varepsilon_{t,i}\|\right)\\ s.t. x_{t+1}&=A_t x_t+B_t u_t+C_t\hat{\varepsilon}_t.\end{aligned} \tag{4.102}$$

Proof. This is a direct application of Theorem 2.7. □

Follow the same notation in §4.5, we have

Theorem 4.14. (4.102) *is a group lasso problem with*

$$X=(\frac{1}{n\gamma}\mathbf{I}+\Theta)^{\frac{1}{2}}, \tag{4.103}$$

$$y=-\frac{1}{2}(\frac{1}{n\gamma}\mathbf{I}+\Theta)^{-\frac{1}{2}}\left(\Theta\varepsilon+\tilde{\Theta}^T x_0\right), \tag{4.104}$$

$$\lambda=\frac{2\alpha}{n\gamma}, c_g=1. \tag{4.105}$$

We give the algorithm for the minimum problem in Algorithm 4.

Algorithm 3: Outlier Robust Control for $\gamma > 0$

Data: $\varepsilon_{t,i}, A_t, B_t, C_t, R_t, Q_t$, risk parameters γ, relative norm parameter α, ADMM parameter ρ, maximum iteration step K, minimum update distance $\epsilon_{\min}$, $t = 0, 1, \ldots, T-1, i = 1, 2, \ldots, n$.

Result: $U_t, \bar{u}_t$ and control policy $u_t(x_t) = U_t x_t + \bar{u}_t, t = 0, 1, \ldots, T-1$.

Solve (4.30) at $\varepsilon_{t,i}$ and get $\Xi^t_{0,1}, \Xi^{t,\tau}_{0,2}, \tilde{\Xi}_t$ for all t, τ;

for $t = 0, 1, \ldots, T-1$ **do**
- **for** $\tau = 0, 1, \ldots, T-1$ **do**
  - **if** $t == \tau$ **then**
    - **for** $i = 1, 2, \ldots, n$ **do**
      - **for** $j = 1, 2, \ldots, n$ **do**
        - **if** $i == j$ **then**
          - $\Theta_{(t,i),(t,i)} = \frac{1}{n}\tilde{\Xi}_t + \frac{1}{n^2}\Xi^{t,t}_{0,2}$;
        - **else**
          - $\Theta_{(t,i),(t,j)} = \Theta_{(t,j),(t,i)} = \frac{1}{2n^2}\Xi^{t,t}_{0,2}$;
      - $\tilde{\Theta}_{(t,i)} = \frac{1}{n}\Xi^t_{0,1}$;
  - **else**
    - **for** $i = 1, 2, \ldots, n$ **do**
      - **for** $j = 1, 2, \ldots, n$ **do**
        - $\Theta_{(t,i),(\tau,j)} = \Theta_{(\tau,j),(t,i)} = \frac{1}{2n^2}\Xi^{t,\tau}_{0,2}$;

$X = (\frac{1}{n\gamma}\mathbf{I} - \Theta)^{\frac{1}{2}}$;

$y = \frac{1}{2}(\frac{1}{n\gamma}\mathbf{I} - \Theta)^{-\frac{1}{2}}\left(\Theta\varepsilon + \tilde{\Theta}^T x_0\right)$;

$\lambda = \frac{2\alpha}{n\gamma}$;

for $t = 0, 1, \ldots, T-1$ **do**
- **for** $i = 1, 2, \ldots, n$ **do**
  - $\alpha^0_{(t,i)} = 0$;

$\beta^0 = (X^T X)^{-1} X^T y$;

$w^0 = 0$;

for $k = 1, 2, \ldots, K$ **do**
- $\beta^k = (X^T X + \rho I)^{-1}(X^T y + \rho(\alpha^{k-1} - w^{k-1}))$;
- **if** $\|\beta^k - \beta^{k-1}\| \leq \varepsilon_{min}$ **then**
  - Break;
- **for** $t = 0, 1, \ldots, T-1$ **do**
  - **for** $i = 1, 2, \ldots, n$ **do**
    - $\alpha^k_{(t,i)} = R_{\frac{\lambda}{\rho}}(\beta^k_{(t,i)} + w^{k-1}_{(t,i)})$, where $R_t(x)$ is defined in (4.92);
  - $w^k = w^{k-1} + \beta^k - \alpha^k$;

for $t = 0, 1, \ldots, T-1$ **do**
- **for** $i = 1, 2, \ldots, n$ **do**
  - $\hat{\varepsilon}_{t,i} = \beta_{(t,i)} + \varepsilon_{t,i}$;

Compute U_t from (4.34);

Compute $\bar{u}_t = \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right]$ from (4.35) and (4.36);

4.7 Experiment

We consider a five stages model

$$\begin{aligned} &\min_{u_t(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1})} \sum_{t=0}^{4} \left(5x_{t+1}^2 + 10u_t^2\right) \\ &s.t. x_{t+1} = x_t + u_t - \varepsilon_t \\ &\varepsilon_t \sim 0.85N(0,1) + 0.15N(15,3), \end{aligned} \tag{4.106}$$

which mixes $N(15,3)$ into $N(0,1)$ with a low probability and $x_0 = 0$. The $N(15,3)$ can be either erroneous data points or true outliers. We sample from this mixture of Gaussian 100 data points for each stage. The histogram of all data points is shown in figure 4.4. It can be seen that apart from Gaussian around 0, we have a few data points away from the origin.

We consider the case that those extra data points are erroneous data points first, which corresponds to $\gamma < 0$. And we choose $\gamma = -100, \alpha = 300$. And in the simulation, we only generate noises from $N(0,1)$. The numerical result is shown in figure 4.5. It shows that our controller reduces the objective value significantly.

Algorithm 4: Outlier Robust Control for $\gamma < 0$

Data: $\varepsilon_{t,i}, A_t, B_t, C_t, R_t, Q_t$, risk parameters γ, relative norm parameter α, ADMM parameter ρ, maximum iteration step K, minimum update distance $\epsilon_{\min}$, $t = 0, 1, \ldots, T-1, i = 1, 2, \ldots, n$.

Result: $U_t, \bar{u}_t$ and control policy $u_t(x_t) = U_t x_t + \bar{u}_t, t = 0, 1, \ldots, T-1$.

Solve (4.30) at $\varepsilon_{t,i}$ and get $\Xi_{0,1}^{t}, \Xi_{0,2}^{t,\tau}, \tilde{\Xi}_t$ for all t, τ;

```
for t = 0,1,...,T-1 do
    for τ = 0,1,...,T-1 do
        if t == τ then
            for i = 1,2,...,n do
                for j = 1,2,...,n do
                    if i == j then
                        Θ_(t,i),(t,i) = (1/n) Ξ̃_t + (1/n²) Ξ^{t,t}_{0,2};
                    else
                        Θ_(t,i),(t,j) = Θ_(t,j),(t,i) = (1/(2n²)) Ξ^{t,t}_{0,2};
                Θ̃_(t,i) = (1/n) Ξ^t_{0,1};
        else
            for i = 1,2,...,n do
                for j = 1,2,...,n do
                    Θ_(t,i),(τ,j) = Θ_(τ,j),(t,i) = (1/(2n²)) Ξ^{t,τ}_{0,2};
```

$X = (\frac{1}{n|\gamma|}\mathbf{I} + \Theta)^{\frac{1}{2}}$;

$y = -\frac{1}{2}(\frac{1}{n|\gamma|}\mathbf{I} + \Theta)^{-\frac{1}{2}} \left(\Theta\varepsilon + \tilde{\Theta}^T x_0\right)$;

$\lambda = \frac{2\alpha}{n|\gamma|}$;

```
for t = 0,1,...,T-1 do
    for i = 1,2,...,n do
        α^0_(t,i) = 0;
```

$\beta^0 = (X^T X)^{-1} X^T y$;

$w^0 = 0$;

for $k = 1, 2, \ldots, K$ **do**
  $\beta^k = (X^T X + \rho I)^{-1}(X^T y + \rho(\alpha^{k-1} - w^{k-1}))$;
  if $\|\beta^k - \beta^{k-1}\| \leq \varepsilon_{min}$ **then**
    Break;
  for $t = 0, 1, \ldots, T-1$ **do**
    for $i = 1, 2, \ldots, n$ **do**
      $\alpha^k_{(t,i)} = R_{\frac{|\lambda|}{\rho}}(\beta^k_{(t,i)} + w^{k-1}_{(t,i)})$, where $R_t(x)$ is defined in (4.92);
    $w^k = w^{k-1} + \beta^k - \alpha^k$;
for $t = 0, 1, \ldots, T-1$ **do**
  for $i = 1, 2, \ldots, n$ **do**
    $\hat{\varepsilon}_{t,i} = \beta_{(t,i)} + \varepsilon_{t,i}$;
Compute U_t from (4.34);
Compute $\bar{u}_t = \sum_{\tau=t}^{T-1} U_{t,\tau} \mathbb{E}\left[\hat{\varepsilon}_\tau\right]$ from (4.35) and (4.36);

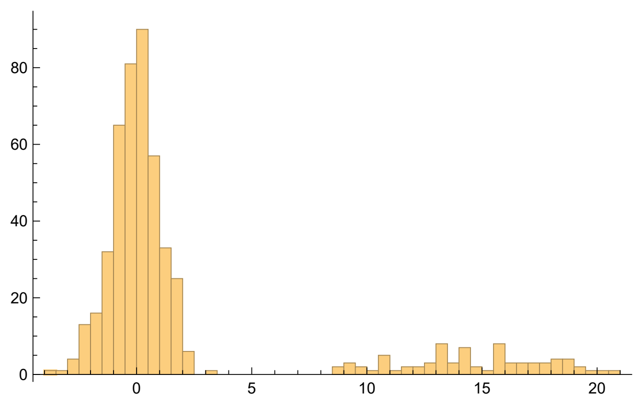


Figure 4.4: Histogram of contaminated data points.

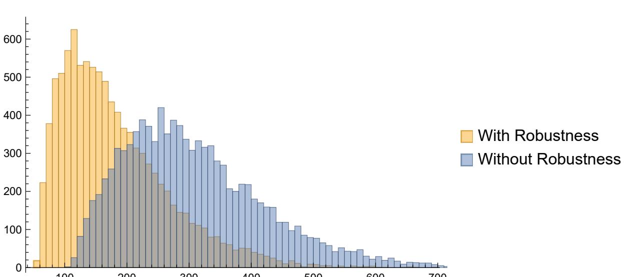
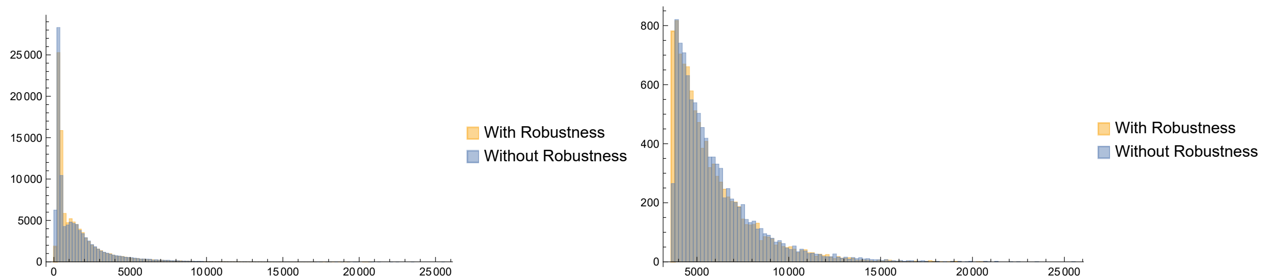


Figure 4.5: Outlier robust control when erroneous data points exist.



(a) With all objective values in 100K runs. (b) With top 10K objective values in 100K runs.

Figure 4.6: Histogram of objective values when outliers exist.

Then, we consider the case that those data points are outliers, which will repeatedly occur but cannot be modeled with $N(0, 1)$. And we will generate noises from the mixture of Gaussian to simulate this behavior. We choose $\gamma = 0.02, \alpha = 0.06$, and the histogram for objective values in 100,000 runs is shown in figure 4.6a. The tail behavior cannot be told from the histogram as the tail falls in a long range. Thus, we take the greatest 10,000 objective values and check its tail behavior carefully in figure 4.6b. It can be seen it has much more mass in the first bin. And the robust controller has less mass in most of the following bins. We also check the quantile plot with robust controller versus risk-free controller. The plot is shown in figure 4.7. The robust controller is, as expected, with a lighter tail.

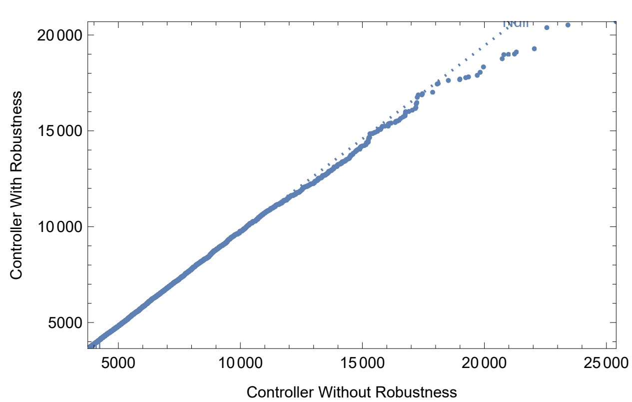


Figure 4.7: Quantile plot when outliers exist.

CHAPTER 5

NONLINEAR NON-GAUSSIAN RISK-AVERSE CONTROL

5.1 Introduction

In general cases, nonlinear optimization is not believed to be solvable due to the limited information from the problems themselves. Especially in control theory, when the number of stages increases, nonlinear control problems become extremely hard. In the deterministic control, a tractable approximation was proposed about two decades ago, which mainly focuses on the experimental aspects with few theoretic guarantees (Li and Todorov, 2004). Later, it is extended to the stochastic case with a similar iterative scheme (Todorov and Li, 2005; Sideris and Bobrow, 2005). After a decade, it is generalized to the risk-averse case (Farshidian and Buchli, 2015). Then, (Nishimura et al., 2021) tries to generalize the problem to the non-Gaussian noise in the distributionally robust control framework. However, their understanding of non-Gaussian risk-averse control is still limited and inexact. (Nishimura et al., 2021)'s work attempts to solve it with ambiguity sets constructed from Gaussian. This doesn't essentially differ from the Gaussian risk-averse control.

While some may think Gaussian is good enough for risk-averse control, we point out here there are crucial differences when noise is non-Gaussian. This point can be explained with a simple example. Assume that noise follows a mixture of Gaussian $0.99N(0,1) + 0.01N(30,1)$. Then we perturb it to be $0.98N(0,1) + 0.02N(30,1)$. This perturbation is subtle, and the KL-divergence between these two distributions is 0.0039 as small as expected. However, if we model the original distribution to be Gaussian $N(0.3, 9.91)$, then the KL-divergence between it and the perturbed distribution would grow insanely to 1.4937. It just means in the risk-averse Gaussian control, the perturbed distribution is far from worth to be considered. When distributions have compact supports, the difference becomes more significant. Thus, in this paper, we will discuss nonlinear risk-averse control problems under non-Gaussian noise.

5.2 Problem Formulation

We consider nonlinear risk-averse control under non-Gaussian noise in this paper,

$$\begin{aligned} &\min_{u_t(x_0,\varepsilon_0,\ldots,\varepsilon_{t-1})} \frac{1}{\gamma}\log \mathbb{E}\left[\exp\left\{\gamma\left(\sum_{t=0}^{T-1} f_t(x_{t+1},u_t)\right)\right\}\right] \\ &s.t. x_{t+1} = g_t(x_t,u_t) + h_t(x_t,u_t)\varepsilon_t, \end{aligned} \tag{5.1}$$

where $\gamma > 0$, $x_t \in \mathbb{R}^{d_1}, u_t(\varepsilon_0,\ldots,\varepsilon_{t-1}) \in \mathbb{R}^{d_2}, \varepsilon_t \in \mathbb{R}^{d_3}$, f_t, g_t, u_t are functions. ε_t are independent noises. The joint distribution of $(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{T-1})$ is denoted as $P\left(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{T-1}\right) = \prod_{t=0}^{T-1} P_t(\varepsilon_t)$. We assume P_t to be empirical measures and they are all independent in this paper. And the optimization is over the function space in search of the optimal control policy $u_t^*(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})$.

For notation simplicity, we will omit $(\varepsilon_0,\ldots,\varepsilon_{t-1})$ when use $u_t(\varepsilon_0,\ldots,\varepsilon_{t-1})$ in this paper. u_t should be always referring to a function $u_t(\varepsilon_0,\ldots,\varepsilon_{t-1})$.

5.3 Distributionally Robust Control Understanding

(5.1) is interesting as an objective approximates a trade-off between expectation and variance,

Theorem 5.1. (5.1) *is equivalent to*

$$\begin{aligned} &\min_{u_t} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1},u_t)\right] + \frac{\gamma}{2}\mathbb{V}\left[\sum_{t=0}^{T-1} f_t(x_{t+1},u_t)\right] + O(\gamma^2) \\ &s.t. x_{t+1} = g_t(x_t,u_t) + h_t(x_t,u_t)\varepsilon_t, \end{aligned} \tag{5.2}$$

Proof. This is by the expansion of the logarithm and exponential function. □

We can also understand this problem from the aspect of distributionally robust control.

Lemma 5.2. *For any random variable X following a distribution P which has density or mass $dP(x)$, the maximum point $\tilde{P}$ of the problem*

$$\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}[X] - \frac{1}{\gamma} KL(\tilde{P} \| P) \tag{5.3}$$

also has same support with P if the maximum in (5.3) exists, and its density satisfies the condition

$$d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{5.4}$$

Consequently, we have the identity

$$\max_{\tilde{P}} \mathbb{E}_{\tilde{P}}[X] - \frac{1}{\gamma} KL(\tilde{P} \| P) = \int x d\tilde{P}(x) - \frac{1}{\gamma} \log\left(\frac{d\tilde{P}(x)}{dP(x)}\right) d\tilde{P}(x) = \frac{1}{\gamma} \log \mathbb{E}_P\left[e^{\gamma X}\right], \tag{5.5}$$

where $\mathbb{E}_P\left[e^{\gamma X}\right]$ is the normalized constant for the $d\tilde{P}(x)$ in (5.4).

Proof. Consider the function

$$\log \mathbb{E}_P\left[e^{\gamma X}\right] = \log \int e^{\gamma x} dP(x). \tag{5.6}$$

For any distributions $\tilde{P}(x)$ that has density and same support with $P(x)$,

$$\begin{aligned} \log \int e^{\gamma x} dP(x) &= \log \int e^{\gamma x} \left(\frac{dP(x)}{d\tilde{P}(x)}\right) d\tilde{P}(x) \\ &\geq \int \left(\gamma x d\tilde{P}(x) + \log\left(\frac{dP(x)}{d\tilde{P}(x)}\right) d\tilde{P}(x)\right), \end{aligned} \tag{5.7}$$

where the last inequality comes from Jensen's inequality in probabilistic setting (Durrett, 2019).

Divide γ on the both hand side of (5.7), we have

$$\frac{1}{\gamma} \log \mathbb{E}_P\left[e^{\gamma X}\right] \geq \mathbb{E}_{\tilde{P}}[x] - \frac{1}{\gamma} KL(\tilde{P} \| P). \tag{5.8}$$

Equality can be attained from Jensen's inequality with the condition

$$d\tilde{P}(x) \sim dP(x) \times e^{\gamma x}. \tag{5.9}$$

Note that the distribution constructed from (5.9) always exists if (5.6) is not infinity. Thus, if (5.6) is not infinite, the maximum of (5.3) exists.

Assume now (5.6) is infinite. By constructing $\tilde{P}(x)$, we get

$$d\tilde{P}_R(x) \sim dP(x) \times e^{\gamma x} \times I(\|x\| \leq R), \tag{5.10}$$

where $R \in \mathbb{R}$, we have $\tilde{P}_R$ always exists and feasible in (5.3). Denote C_R to be the normalized constant such that

$$d\tilde{P}_R(x) = \frac{1}{C_R} dP(x) \times e^{\gamma x} \times I(\|x\| \leq R). \tag{5.11}$$

Since (5.6) is infinite, we must have that

$$\begin{aligned} \lim_{R \to \infty} C_R &= \lim_{R \to \infty} \int e^{\gamma x} \times I(\|x\| \leq R) dP(x) \\ &= \infty \end{aligned} \tag{5.12}$$

Let $R \to \infty$, we have

$$\begin{aligned} &\lim_{R \to \infty} \mathbb{E}_{\tilde{P}_R}[X] - \frac{1}{\gamma} KL(\tilde{P}_R \| P) = \lim_{R \to \infty} \int x d\tilde{P}_R(x) - \frac{1}{\gamma} \log\left(\frac{d\tilde{P}_R(x)}{dP(x)}\right) d\tilde{P}_R(x) = \lim_{R \to \infty} \int \\ &-\frac{1}{\gamma} \log\left(\frac{d\tilde{P}_R(x)}{e^{\gamma x} dP(x)}\right) d\tilde{P}_R(x) = \lim_{R \to \infty} \int_{\|x\| \leq R} \frac{1}{\gamma} \frac{\log C_R}{C_R} e^{\gamma x} dP(x) = \lim_{R \to \infty} \frac{1}{\gamma} \log C_R = \infty. \end{aligned} \tag{5.13}$$

From (5.13) and (5.8) we have that the existence of the maximum of (5.3) is equivalent to the finite property of (5.6). When finiteness of either occurs, we can invoke (5.8) and (5.9) to claim our conclusion. This completes the proof. □

Theorem 5.3. (5.1) *is equivalent to*

$$\begin{aligned} &\min_{u_t} \max_{\tilde{P}, (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{T-1}) \sim \tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)\right] - \frac{1}{\gamma} KL(\tilde{P} \| P) \\ &s.t. x_{t+1} = g_t(x_t, u_t) + h_t(x_t, u_t)\varepsilon_t. \end{aligned} \tag{5.14}$$

Proof. This is a direct application of Lemma 5.2. For every fixed u_t, $\sum_{t=0}^{T-1} f_t(x_{t+1}, u_t)$ is the random variable X in Lemma 5.2. □

5.4 Sequential Approximation of Nonlinear Control

Nonlinear control is among the hardest problems in control theory. It doesn't have clear solutions for general problems, but it can be approximated with tractable algorithms. The approximation we will use in this paper is the sequential trajectory-based linear quadratic approximation. In the iteration of each step, we will set up a trajectory first and then update the linear quadratic approximation of the original problem around the trajectory. That is, we decompose x_t and u_t into

$$x_t(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{t-1}) = \bar{x}_t + \tilde{x}_t(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{t-1}), \tag{5.15}$$

$$u_t(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{t-1}) = \bar{u}_t + \tilde{u}_t(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{t-1}). \tag{5.16}$$

And we iteratively perform linearization at $(\bar{x}_t, \bar{u}_t)$, which gives

$$\begin{aligned}
&\min_{\bar{x}_t, \bar{u}_t} \min_{u_t} \max_{\tilde{P}, (\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{T-1}) \sim \tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1} \Big((x_{t+1} - \bar{x}_{t+1})^T Q_{t+1} (x_{t+1} - \bar{x}_{t+1}) \right. \\
&\quad + (u_t - \bar{u}_t)^T R_t (u_t - \bar{u}_t) + 2 (x_{t+1} - \bar{x}_{t+1})^T S_{t+1} (u_t - \bar{u}_t) + M_{t+1}(x_{t+1} - \bar{x}_{t+1}) \\
&\quad \left. + N_t(u_t - \bar{u}_t) + f_{t+1}(\bar{x}_{t+1}, \bar{u}_t) \Big) \right] - \frac{1}{\gamma} KL(\tilde{P} \| P) \\
&s.t. x_{t+1} = A_t(x_t - \bar{x}_t) + B_t(u_t - \bar{u}_t) + C_t \varepsilon_t + g_t(\bar{x}_t, \bar{u}_t),
\end{aligned} \tag{5.17}$$

where $M_{t+1} = \nabla_x f_{t+1}(\bar{x}_{t+1}, \bar{u}_t), N_t = \nabla_u f_{t+1}(\bar{x}_{t+1}, \bar{u}_t), Q_{t+1} = \nabla_{xx} f_{t+1}(\bar{x}_{t+1}, \bar{u}_t), R_t = \nabla_{uu} f_{t+1}(\bar{x}_{t+1}, \bar{u}_t), S_{t+1} = \nabla_{xu} f_{t+1}(\bar{x}_{t+1}, \bar{u}_t), A_t = \nabla_x g(\bar{x}_t, \bar{u}_t), B_t = \nabla_u g(\bar{x}_t, \bar{u}_t), C_t = h_t(\bar{x}_t, \bar{u}_t)$. In each step of the outermost problem, we fix the trajectory $\bar{x}_t, \bar{u}_t$ and solve the inner problem first. Then, we perform gradient descent on the outer problem to minimize the approximated objective.

Note that when ε_t are all Gaussian, the inner minimization problem of (5.17) is a classical risk-averse linear quadratic Gaussian problem. Thus, by reformulating matrices in (5.17) into standard form, we can solve the inner minimization problem. The standard form we will use is

Theorem 5.4.

$$\min_{\tilde{u}_t} \max_{\tilde{P},(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{T-1})\sim\tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}KL(\tilde{P}\|P) \tag{5.18}$$
$$s.t.\tilde{x}_{t+1}=\tilde{A}_t\tilde{x}_t+\tilde{B}_t\tilde{u}_t+\tilde{C}_t\varepsilon_t,$$

where

$$\tilde{x}_t=\Gamma_{t,1}\begin{pmatrix}x_t\\1\end{pmatrix},\tilde{u}_t(x_0,\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1})=\tilde{u}_t(\Gamma_{t,1}\begin{pmatrix}x_0\\\varepsilon_0\\\varepsilon_1\\\vdots\\\varepsilon_{t-1}\\1\end{pmatrix})=\Gamma_{t,2}u_t(x_0,\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1}), \tag{5.19}$$

$$\tilde{A}_t=\begin{pmatrix}A_t & -A_t\bar{x}_t-B_t\bar{u}_t+g_t(\bar{x}_t,\bar{u}_t)\\0 & 1\end{pmatrix}\Gamma_{t,1}^T,\tilde{B}_t=\begin{pmatrix}B_t\\0\end{pmatrix}\Gamma_{t,2}^T,\tilde{C}_t=\begin{pmatrix}C_t\\0\end{pmatrix}, \tag{5.20}$$

$\tilde{Q}_t,\tilde{R}_t,\tilde{\Gamma}_{t,1},\tilde{\Gamma}_{t,2}$ *are from the orthogonal decomposition*

$$\begin{pmatrix}Q_t & V_1 & S_t\\V_1^T & V_2 & V_3^T\\S_t^T & V_3 & R_t\end{pmatrix}=\Gamma_t^T\begin{pmatrix}\tilde{Q}_t & 0\\0 & \tilde{R}_t\end{pmatrix}\Gamma_t=\begin{pmatrix}\Gamma_{t,1}^T\tilde{Q}_t\Gamma_{t,1} & 0\\0 & \Gamma_{t,2}^T\tilde{R}_t\Gamma_{t,2}\end{pmatrix}, \tag{5.21}$$

with $V_1=-Q_t\bar{x}_t-S_t\bar{u}_t+M_t^T$,$V_2=\bar{x}_t^TQ_t\bar{x}_t+\bar{u}_t^TR_t\bar{u}_t+2\bar{x}_t^TS_t\bar{u}_t-M_t\bar{x}_t-N_t\bar{u}_t+f_t(\bar{x}_t,\bar{u}_t)$,$V_3=-R_t\bar{u}_t-S_t^T\bar{x}_t+N_t^T$.

However, the application of (5.18) is quite limited. As it requires noises to follow Gaussian, this leads to the paper's interesting point, the discussion of nonlinear risk-averse control under non-Gaussian noise.

5.5 Non-Gaussian distributions

(5.18) is a well-defined distributionally robust optimization problem. But it is hard to use for general distributions in distributionally robust control. For example, in distributionally robust control, distributions are usually given stage by stage instead of a unified distribution P. Modeling it as a unified distribution P will bring the curse of dimensionality. Instead, we propose a friendly model for distributionally robust control.

$$\min_{\tilde{u}_t} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{P}_t \| P_t), \tag{5.22}$$

that is, we model the stage-wise risk with $\tilde{P}_t$ independently. It is equivalently requiring the risk-averse distribution $\tilde{P}$ to keep the independent structure as P has. This is not hard to understand why will make sense. In many scenarios, like Brownian motion, we have the solid independence assumption but will have doubt about the accuracy of parameter estimations. (5.22) focus on the robust problem towards errors in the estimation of parameters instead of the model's independent assumption. The difference between (5.18) and (5.22) is stated mathematically in Theorem 5.6 and Theorem 5.7. Before this, we first illustrate how to solve (5.22).

Theorem 5.5. *If $\tilde{Q}_t \geq 0, \tilde{R}_t \geq 0$, then there exists $\gamma_0, \forall \gamma < \gamma_0$,*

$$\min_{\tilde{u}_t} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{P}_t \| P_t) \tag{5.23a}$$

$$= \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{\tilde{u}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{P}_t \| P_t). \tag{5.23b}$$

Proof. Since (5.23b) constraints $\tilde{P}$ to have independent structure, we can apply Lemma 4.7 and have that

$$\max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{\tilde{u}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{P}_t \| P_t) \tag{5.24a}$$

$$= \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{\tilde{u}_t \in \mathcal{U}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{P}_t \| P_t), \tag{5.24b}$$

where

$$\mathcal{U}_t = \left\{\sum_{\tau=0}^{t-1} U_{t,\tau} \varepsilon_\tau + b_t | \forall \tau, U_{t,\tau} \in \mathbb{R}^{d_2, d_3}, b_t \in \mathbb{R}^{d_2}\right\}. \tag{5.25}$$

This is because that $\tilde{u}_t$ is a linear function of $\tilde{x}_t$ in Lemma 4.7. By expanding $\tilde{x}_t$ with $(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{t-1})$, we can see $\tilde{u}_t$ falling into the set $\mathcal{U}_t$. Then, we parameterize $\tilde{u}_t$ and write (5.24b) as

$$\begin{aligned} &\max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{U_{t,\tau}, b_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1}\right.\right. \\ &\left.\left. + \left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)^T \tilde{R}_t \left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{P}_t \| P_t). \end{aligned} \tag{5.26}$$

From (4.34), we see the solution does not depend on the distribution but only on the matrices from the system. That means the minimum point $U_{t,\tau}$ keeps the same for all distributions in (5.26). We denote this minimum point to be $U_{t,\tau}^*$. Then, We have (5.26) equal

$$\begin{aligned} &\max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{U_{t,\tau} = U_{t,\tau}^*, b_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1}\right.\right. \\ &\left.\left. + \left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)^T \tilde{R}_t \left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)\right)\right] - \frac{1}{\gamma} \sum_{t=0}^{T-1} KL(\tilde{P}_t \| P_t). \end{aligned} \tag{5.27}$$

As P_t are empirical measures, $\tilde{P}_t$ are also empirical measures. $\tilde{P}_t$ can be parameterized by a vector π_t. We compute the Hessian w.r.t. $\pi_{t,i}$, note that

$$\nabla_{\pi_t}^2 KL(\tilde{P}_t \| P_t) = diag\{\frac{1}{\tilde{\pi}_{t,1}}, \frac{1}{\tilde{\pi}_{t,2}}, \dots, \frac{1}{\tilde{\pi}_{t,n}}\}, \tag{5.28}$$

we have

$$\frac{1}{\gamma}\nabla^2_\pi \sum_{t=0}^{T-1} KL(\tilde{P}_t\|P_t) = \frac{1}{\gamma} diag\left\{\nabla^2_{\pi_0} KL(\tilde{P}_0\|P_0), \nabla^2_{\pi_1} KL(\tilde{P}_1\|P_1) \dots \nabla^2_{\pi_{T-1}} KL(\tilde{P}_{T-1}\|P_{T-1})\right\} \tag{5.29}$$

$$\geq \frac{1}{\gamma}\mathbf{I}_{n\times T}.$$

Also, note that b_t's value doesn't influence the Hessian of the first part, and $U^*_{t,\tau}$ is fixed for all $\tilde{P}_t$, we have

$$\nabla^2_\pi \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}^T_{t+1}\tilde{Q}_{t+1}\tilde{x}_{t+1} + \left(\sum_{\tau=0}^{t-1} U^*_{t,\tau}\varepsilon_\tau + b_t\right)^T \tilde{R}_t \left(\sum_{\tau=0}^{t-1} U^*_{t,\tau}\varepsilon_\tau + b_t\right)\right)\right] \tag{5.30}$$

is a constant matrix regardless of $\tilde{P}_t$ and b_t. Thus, we can always find γ to make

$$\nabla^2_\pi \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}^T_{t+1}\tilde{Q}_{t+1}\tilde{x}_{t+1} + \left(\sum_{\tau=0}^{t-1} U^*_{t,\tau}\varepsilon_\tau + b_t\right)^T \tilde{R}_t \left(\sum_{\tau=0}^{t-1} U^*_{t,\tau}\varepsilon_\tau + b_t\right)\right)\right] - \frac{1}{\gamma}\mathbf{I}_{n\times T} \leq 0. \tag{5.31}$$

With such γ, the objective of (5.27) is a concave function of $\tilde{P}$ at any given b_t. It is also a convex function of b_t at any given $\tilde{P}_t$. By Sion's minimax theorem, we have (5.27) equal

$$\begin{aligned}\min_{U_{t,\tau} = U^*_{t,\tau}, b_t} \max_{\tilde{P} = \prod_{t=0}^{T-1}\tilde{P}_t} \mathbb{E}&\left[\sum_{t=0}^{T-1}\left(\tilde{x}^T_{t+1}\tilde{Q}_{t+1}\tilde{x}_{t+1}\right.\right. \\ &\left.\left.+\left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)^T \tilde{R}_t \left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)\right)\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{P}_t\|P_t).\end{aligned} \tag{5.32}$$

In the outermost problem of (5.32), we have a constraint on $\tilde{u}_t$ such that $\tilde{u}_t = \sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t$. It is greater than or equal to the unconstrained problem

$$\min_{\tilde{u}_t} \max_{\tilde{P} = \prod_{t=0}^{T-1}\tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}^T_{t+1}\tilde{Q}_{t+1}\tilde{x}_{t+1} + \tilde{u}^T_t\tilde{R}_t\tilde{u}_t\right)\right] - \frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{P}_t\|P_t) \tag{5.33}$$

$$\leq \min_{U_{t,\tau} = U^*_{t,\tau}, b_t} \max_{\tilde{P} = \prod_{t=0}^{T-1}\tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}^T_{t+1}\tilde{Q}_{t+1}\tilde{x}_{t+1} + \left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)^T \tilde{R}_t \left(\sum_{\tau=0}^{t-1} U_{t,\tau}\varepsilon_\tau + b_t\right)\right)\right]$$

$$-\frac{1}{\gamma}\sum_{t=0}^{T-1} KL(\tilde{P}_t\|P_t).$$

(5.26) = (5.27) gives

$$\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\min_{\tilde{u}_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t) \tag{5.34}$$

$$=\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\min_{U_{t,\tau}=U_{t,\tau}^*,b_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\left(\sum_{\tau=0}^{t-1}U_{t,\tau}\varepsilon_\tau+b_t\right)^T\tilde{R}_t\left(\sum_{\tau=0}^{t-1}U_{t,\tau}\varepsilon_\tau+b_t\right)\right)\right]$$
$$-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t).$$

We have

$$\min_{\tilde{u}_t}\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t) \tag{5.35}$$

$$\overset{(5.33)}{\leq}$$

$$\min_{U_{t,\tau}=U_{t,\tau}^*,b_t}\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\left(\sum_{\tau=0}^{t-1}U_{t,\tau}\varepsilon_\tau+b_t\right)^T\tilde{R}_t\left(\sum_{\tau=0}^{t-1}U_{t,\tau}\varepsilon_\tau+b_t\right)\right)\right]$$
$$-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t)$$

$$\overset{(5.27)=(5.32)}{=}$$

$$\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\min_{U_{t,\tau}=U_{t,\tau}^*,b_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\left(\sum_{\tau=0}^{t-1}U_{t,\tau}\varepsilon_\tau+b_t\right)^T\tilde{R}_t\left(\sum_{\tau=0}^{t-1}U_{t,\tau}\varepsilon_\tau+b_t\right)\right)\right]$$
$$-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t)$$

$$\overset{(5.34)}{=}\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\min_{\tilde{u}_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t).$$

As

$$\begin{aligned}&\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\min_{\tilde{u}_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t)\\ &\leq\min_{\tilde{u}_t}\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t)\end{aligned} \tag{5.36}$$

is always true; we complete the proof. □

Remark 5.5.1. *This theorem tells us that the optimal control policy of* (5.22) *is still linear. To solve* (5.22), *we can switch the optimization order and figure out the worst distribution* $\tilde{P}$ *first. Then, the global optimal control policy is the optimal control policy under this worst distribution. The worst distribution can be found with our coupling method in §5.6.*

Theorem 5.6. *If* $\tilde{Q}_t \geq 0, \tilde{R}_t \geq 0$, *then there exists* $\gamma_0, \forall \gamma < \gamma_0$, *the optimal control policy of* (5.18) *is the same as the optimal control policy of*

$$\min_{\tilde{u}_t} \max_{\tilde{P}, KL(\tilde{P}\|P) \leq c(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right], \tag{5.37}$$

where $c(\gamma)$ *is a non-decreasing function of* γ.

Proof. Replace $c(\gamma)$ in (5.37) by a constant c first. And it is equivalent to

$$\min_{\tilde{u}_t} \max_{\tilde{P}} \min_{w \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - w\left(KL(\tilde{P}\|P) - c\right). \tag{5.38}$$

At any given $\tilde{u}_t$, as the objective of (5.37) is a concave function of $\tilde{P}$, and $KL(\tilde{P}\|P)$ is convex in $\tilde{P}$, it satisfies Slater's condition and we have (5.38) equals

$$\min_{\tilde{u}_t} \min_{w \geq 0} \max_{\tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - w\left(KL(\tilde{P}\|P) - c\right), \tag{5.39}$$

which also equals

$$\min_{w \geq 0} \min_{\tilde{u}_t} \max_{\tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - w\left(KL(\tilde{P}\|P) - c\right). \tag{5.40}$$

Now consider the minimum point w as a function of c. Then $w(c)$ is a non-increasing function of c. When we increase c from 0 to infinity, $w(c)$ should converge to some value w_0. If we pick γ such that $\frac{1}{\gamma} > w_0$, the inverse function of $w(c)$ gives $c(\frac{1}{\gamma})$. Thus $c(\gamma)$ is a non-decreasing function. Now consider the inner problem of (5.40),

$$\min_{\tilde{u}_t} \max_{\tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - w\left(KL(\tilde{P}\|P) - c\right). \tag{5.41}$$

As it is convex in $\tilde{u}_t$, concave in $\tilde{P}$, and $\tilde{P}$ is empirical measure which can be parameterized by π_t, we can apply Sion's minimax theorem and have

$$\min_{\tilde{u}_t} \max_{\tilde{P}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-w\left(KL(\tilde{P}\|P)-c\right) \tag{5.42a}$$

$$=\max_{\tilde{P}} \min_{\tilde{u}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-w\left(KL(\tilde{P}\|P)-c\right). \tag{5.42b}$$

Thus, the optimal control policy will only depend on the maximum point $\tilde{P}$. When coming back to (5.18), for the same reason we can apply Sion's minimax theorem and have it equal

$$\max_{\tilde{P},(\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{T-1})\sim\tilde{P}} \min_{\tilde{u}_t} \mathbb{E}\left[\sum_{t=1}^{T}\left(\tilde{x}_t^T\tilde{Q}_t\tilde{x}_t+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}KL(\tilde{P}\|P). \tag{5.43}$$

Comparing (5.40) and (5.43) at $c=c(\gamma)$, they will have the same maximum point $\tilde{P}$; this completes the proof. ☐

Theorem 5.7. *If $\tilde{Q}_t \geq 0, \tilde{R}_t \geq 0$, then there exists $\gamma_0, \forall\gamma < \gamma_0$, the optimal control policy of* (5.22) *is the same as the optimal control policy of*

$$\min_{\tilde{u}_t} \max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t,\forall t,KL(\tilde{P}_t\|P_t)\leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right], \tag{5.44}$$

where $c_t(\gamma)$ are non-decreasing functions of γ.

Proof. Note (5.44) is not a concave optimization problem of $\tilde{P}_t$ as the set of independent distribution is not convex. We first convexify this set and consider the problem

$$\min_{\tilde{u}_t} \max_{\tilde{P}\in\mathcal{P},\forall t,KL(\tilde{P}_t\|P_t)\leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right], \tag{5.45}$$

where $\tilde{P}_t$ in the constraint $KL(\tilde{P}_t\|P_t)\leq c_t(\gamma)$ is the t-th marginal distribution of $\tilde{P}$ and

$$\mathcal{P}=\left\{\sum_{i=1}^{m}\lambda_i\tilde{P}^i|\tilde{P}^i=\prod_{t=0}^{T-1}\tilde{P}_t^i,\sum_{i=1}^{m}\lambda_i=1,0\leq\lambda_i\leq 1,m>0,\forall t,i,KL(\tilde{P}_t^i\|P_t)\leq c_t(\gamma)\right\}. \tag{5.46}$$

Then, the maximum problem in (5.45) is a concave optimization problem of $\tilde{P}$ with Slater's condition held. Thus, we have strong duality and have (5.45) equal

$$\min_{\tilde{u}_t} \min_{w_t \geq 0} \max_{\tilde{P} \in \mathcal{P}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right). \quad (5.47)$$

Note the objective of (5.45) can be written as

$$\mathbb{E}_{\tilde{P}}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] = \sum_{i=1}^{m} \lambda_i \mathbb{E}_{\tilde{P}^i}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right]. \quad (5.48)$$

For any distribution $\tilde{P} = \sum_{i=1}^{m} \lambda_i \tilde{P}^i \in \mathcal{P}$, we can choose $\tilde{P}^i$ having the greatest

$$\mathbb{E}_{\tilde{P}^i}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right]. \quad (5.49)$$

Then this $\tilde{P}^i$ has independent structure and satisfies constraints, and

$$\begin{aligned}\mathbb{E}_{\tilde{P}^i}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] &\geq \sum_{i=1}^{m} \lambda_i \mathbb{E}_{\tilde{P}^i}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\ &= \mathbb{E}_{\tilde{P}}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right].\end{aligned} \quad (5.50)$$

This just means

$$\begin{aligned}&\max_{\tilde{P} \in \mathcal{P}, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\ &= \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right].\end{aligned} \quad (5.51)$$

Note the strong duality for (5.47) in fact says

$$
\begin{aligned}
&\min_{w_t \geq 0} \max_{\tilde{P} \in \mathcal{P}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&= \max_{\tilde{P} \in \mathcal{P}} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&= \max_{\tilde{P} \in \mathcal{P}, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right].
\end{aligned}
\tag{5.52}
$$

Apply (5.51) and write constraints into the objective,

$$
\begin{aligned}
&\max_{\tilde{P} \in \mathcal{P}, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\
&\overset{(5.51)}{=} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\
&= \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right).
\end{aligned}
\tag{5.53}
$$

The max-min inequality gives

$$
\begin{aligned}
&\max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&\leq \min_{w_t \geq 0} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right).
\end{aligned}
\tag{5.54}
$$

As

$$
\begin{aligned}
&\max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&\leq \max_{\tilde{P} \in \mathcal{P}} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right),
\end{aligned}
\tag{5.55}
$$

we have

$$\begin{aligned}
&\min_{w_t \geq 0} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&\leq \min_{w_t \geq 0} \max_{\tilde{P} \in \mathcal{P}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&\overset{(5.52)}{=} \max_{\tilde{P} \in \mathcal{P}} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right)
\end{aligned} \tag{5.56}$$

and

$$\begin{aligned}
&\min_{w_t \geq 0} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&\overset{(5.54)}{\geq} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&= \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\
&\overset{(5.51)}{=} \max_{\tilde{P} \in \mathcal{P}, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\
&= \max_{\tilde{P} \in \mathcal{P}} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right).
\end{aligned} \tag{5.57}$$

(5.56) and (5.57) are inequalities in different directions, combine them and we have

$$\begin{aligned}
&\min_{w_t \geq 0} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&= \max_{\tilde{P} \in \mathcal{P}} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right) \\
&= \max_{\tilde{P} \in \mathcal{P}, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\
&\overset{(5.51)}{=} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t, \forall t, KL(\tilde{P}_t \| P_t) \leq c_t(\gamma)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] \\
&= \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{w_t \geq 0} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right).
\end{aligned} \tag{5.58}$$

(5.44) then is

$$\min_{w_t \geq 0} \min_{\tilde{u}_t} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right). \tag{5.59}$$

Now consider the inner problem of (5.59)

$$\min_{\tilde{u}_t} \max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right). \tag{5.60}$$

By Theorem 5.5, for sufficiently large w_t, we have strong duality and it equals

$$\max_{\tilde{P} = \prod_{t=0}^{T-1} \tilde{P}_t} \min_{\tilde{u}_t} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T \tilde{Q}_{t+1} \tilde{x}_{t+1} + \tilde{u}_t^T \tilde{R}_t \tilde{u}_t\right)\right] - \sum_{t=0}^{T-1} w_t \left(KL(\tilde{P}_t \| P_t) - c_t(\gamma)\right). \tag{5.61}$$

Now, we can follow the same logic of the proof of Theorem 5.6. This completes the proof. □

5.6 Coupling Method

Consider the maximin problem

$$\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\ \min_{\tilde{u}_t(\tilde{x}_t)} \mathbb{E}\left[\sum_{t=0}^{T-1}\left(\tilde{x}_{t+1}^T\tilde{Q}_{t+1}\tilde{x}_{t+1}+\tilde{u}_t^T\tilde{R}_t\tilde{u}_t\right)\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t). \tag{5.62}$$

By Lemma 4.7, we can replace the inner minimum with

$$V_0(x_0)=x_0^T\Xi_{t,0}x_0+\sum_{\tau=0}^{T-1}x_0^T\Xi_{t,1}^{\tau}\mathbb{E}\left[\varepsilon_\tau\right]+\sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1}\mathbb{E}\left[\varepsilon_{\tau_1}\right]^T\Xi_{t,2}^{\tau_1,\tau_2}\mathbb{E}\left[\varepsilon_{\tau_2}\right]+\sum_{\tau=0}^{T-1}\mathbb{E}\left[\varepsilon_\tau^T\tilde{\Xi}_\tau\varepsilon_\tau\right], \tag{5.63}$$

where $\varepsilon_t\sim\tilde{P}_t$. This gives

$$\begin{aligned}\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\ & x_0^T\Xi_{t,0}x_0+\sum_{\tau=0}^{T-1}x_0^T\Xi_{t,1}^{\tau}\mathbb{E}\left[\varepsilon_\tau\right]+\sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1}\mathbb{E}\left[\varepsilon_{\tau_1}\right]^T\Xi_{t,2}^{\tau_1,\tau_2}\mathbb{E}\left[\varepsilon_{\tau_2}\right]\\ &+\sum_{\tau=0}^{T-1}\mathbb{E}\left[\varepsilon_\tau^T\tilde{\Xi}_\tau\varepsilon_\tau\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t).\end{aligned} \tag{5.64}$$

Theorem 5.8. $\exists\gamma_0,\forall\gamma<\gamma_0$,

$$\begin{aligned}\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t}\ & x_0^T\Xi_{t,0}x_0+\sum_{\tau=0}^{T-1}x_0^T\Xi_{t,1}^{\tau}\mathbb{E}\left[\varepsilon_\tau\right]+\sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1}\mathbb{E}\left[\varepsilon_{\tau_1}\right]^T\Xi_{t,2}^{\tau_1,\tau_2}\mathbb{E}\left[\varepsilon_{\tau_2}\right]\\ &+\sum_{\tau=0}^{T-1}\mathbb{E}\left[\varepsilon_\tau^T\tilde{\Xi}_\tau\varepsilon_\tau\right]-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t)\end{aligned} \tag{5.65}$$

is a concave function of $\tilde{P}_t$*.*

Proof. As $\frac{1}{\gamma}KL(\tilde{P}_t\|P_t)$ is $\frac{1}{\gamma}$-strongly convex in $\tilde{P}_t$, and P_t are empirical measures on a fixed support, with sufficiently small γ the Hessian will be diagonally dominant. Thus, it is concave in $\tilde{P}_t$. □

(5.65) will require a gradient ascent. Alternatively, we have a better way to compute the maximum point under the independent assumption. We introduce two new independent random variables $\xi \sim \tilde{P}$ and $\eta \sim \tilde{P}$, we have (5.64) equals to

$$\max_{\tilde{P}=\prod_{t=0}^{T-1}\tilde{P}_t,(\xi,\eta)\sim\tilde{P}\times\tilde{P}} \frac{1}{2}\mathbb{E}\left[x_0^T\Xi_{t,0}x_0+\sum_{\tau=0}^{T-1}x_0^T\Xi_{t,1}^{\tau}\xi_\tau+\sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1}\xi_{\tau_1}^T\Xi_{t,2}^{\tau_1,\tau_2}\eta_{\tau_2}+\sum_{\tau=0}^{T-1}\xi_\tau^T\tilde{\Xi}_\tau\xi_\tau\right]$$
$$+\frac{1}{2}\mathbb{E}\left[x_0^T\Xi_{t,0}x_0+\sum_{\tau=0}^{T-1}x_0^T\Xi_{t,1}^{\tau}\eta_\tau+\sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1}\eta_{\tau_1}^T\Xi_{t,2}^{\tau_1,\tau_2}\xi_{\tau_2}+\sum_{\tau=0}^{T-1}\eta_\tau^T\tilde{\Xi}_\tau\eta_\tau\right]$$
$$-\frac{1}{\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_t\|P_t). \tag{5.66}$$

Now we relax (5.66) to

$$\max_{\tilde{P}_1=\prod_{t=0}^{T-1}\tilde{P}_{1,t},\tilde{P}_2=\prod_{t=0}^{T-1}\tilde{P}_{2,t},\xi\sim\tilde{P}_1,\eta\sim\tilde{P}_2} \frac{1}{2}\mathbb{E}\left[x_0^T\Xi_{t,0}x_0+\sum_{\tau=0}^{T-1}x_0^T\Xi_{t,1}^{\tau}\xi_\tau+\sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1}\xi_{\tau_1}^T\Xi_{t,2}^{\tau_1,\tau_2}\eta_{\tau_2}\right.$$
$$\left.+\sum_{\tau=0}^{T-1}\xi_\tau^T\tilde{\Xi}_\tau\xi_\tau\right]+\frac{1}{2}\mathbb{E}\left[x_0^T\Xi_{t,0}x_0+\sum_{\tau=0}^{T-1}x_0^T\Xi_{t,1}^{\tau}\eta_\tau+\sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1}\eta_{\tau_1}^T\Xi_{t,2}^{\tau_1,\tau_2}\xi_{\tau_2}+\sum_{\tau=0}^{T-1}\eta_\tau^T\tilde{\Xi}_\tau\eta_\tau\right]$$
$$-\frac{1}{2\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_{1,t}\|P_t)-\frac{1}{2\gamma}\sum_{t=0}^{T-1}KL(\tilde{P}_{2,t}\|P_t), \tag{5.67}$$

where $\tilde{P}_1$ and $\tilde{P}_2$ are independent.

Theorem 5.9. $\exists\gamma_0>0,\forall\gamma<\gamma_0$, (5.67) *is jointly concave in* $\tilde{P}_{1,t}$ *and* $\tilde{P}_{2,t}$.

Proof. The proof is similar to the proof of Theorem 5.8. □

Theorem 5.10. $\exists\gamma_0>0,\forall\gamma<\gamma_0$, (5.67) *equals* (5.65).

Proof. We choose γ_0 to be the γ_0 of Theorem 5.9. Denote the objective value of (5.67) to be $F(\tilde{P}_{1,t},\tilde{P}_{2,t})$. And denote its maximum points to be $\tilde{P}_{1,t}^0$ and $\tilde{P}_{2,t}^0$. As it is symmetric in $\tilde{P}_{1,t}$ and $\tilde{P}_{2,t}$, we have $F(\tilde{P}_{1,t}^0,\tilde{P}_{2,t}^0)=F(\tilde{P}_{2,t}^0,\tilde{P}_{1,t}^0)$. If $\tilde{P}_{1,t}^0\neq\tilde{P}_{2,t}^0$, we choose $\tilde{P}_{1,t}^*=\tilde{P}_{2,t}^*=\frac{1}{2}\left(\tilde{P}_{1,t}^0+\tilde{P}_{2,t}^0\right)$. Then, due to the concavity, we have

$$\begin{aligned}F(\tilde{P}_{1,t}^*,\tilde{P}_{2,t}^*)&\geq\frac{1}{2}\left(F(\tilde{P}_{1,t}^0,\tilde{P}_{2,t}^0)+F(\tilde{P}_{2,t}^0,\tilde{P}_{1,t}^0)\right)\\&=F(\tilde{P}_{1,t}^0,\tilde{P}_{2,t}^0),\end{aligned} \tag{5.68}$$

As $(\tilde{P}^0_{1,t}, \tilde{P}^0_{2,t})$ is the maximum point, we have

$$F(\tilde{P}^*_{1,t}, \tilde{P}^*_{2,t}) = F(\tilde{P}^0_{1,t}, \tilde{P}^0_{2,t}). \tag{5.69}$$

Thus, we have (5.67) $\leq$ (5.65), as for each of its maximum points, we can find a point in (5.65) with the same objective value. Also, note that (5.67) is relaxed from (5.65), we have (5.67) $\geq$ (5.65). Finally, we have (5.67) = (5.65). □

Theorem 5.11. *Given $\tilde{P}_2$, the maximum point $\tilde{P}_1$ is*

$$\begin{aligned} d\tilde{P}_{1,t}(\xi_t) \propto dP_{1,t}(\xi_t) \times \exp\Bigg\{\gamma\Bigg(& x_0^T \Xi^t_{t,1}\xi_t + \sum_{\tau=0}^{t-1} \mathbb{E}_{\tilde{P}_2}[\eta_\tau]^T\, \Xi^{\tau,t}_{\tau,2}\xi_t \\ &+ \sum_{\tau=t}^{T-1} \mathbb{E}_{\tilde{P}_2}[\eta_\tau]^T \left(\Xi^{t,\tau}_{t,2}\right)^T \xi_t + \xi_t^T \tilde{\Xi}_t \xi_t \Bigg)\Bigg\}, \end{aligned} \tag{5.70}$$

and given $\tilde{P}_1$, the maximum point $\tilde{P}_2$ is

$$\begin{aligned} d\tilde{P}_{2,t}(\eta_t) \propto dP_{2,t}(\eta_t) \times \exp\Bigg\{\gamma\Bigg(& x_0^T \Xi^t_{t,1}\eta_t + \sum_{\tau=0}^{t-1} \mathbb{E}_{\tilde{P}_1}[\xi_\tau]^T\, \Xi^{\tau,t}_{\tau,2}\eta_t \\ &+ \sum_{\tau=t}^{T-1} \mathbb{E}_{\tilde{P}_1}[\xi_\tau]^T \left(\Xi^{t,\tau}_{t,2}\right)^T \eta_t + \eta_t^T \tilde{\Xi}_t \eta_t \Bigg)\Bigg\}. \end{aligned} \tag{5.71}$$

Proof. By Lemma 5.2, the maximum point

$$\begin{aligned} &d\tilde{P}_1(\xi_0, \xi_1, \ldots, \xi_{T-1}) \propto dP_1(\xi_0, \xi_1, \ldots, \xi_{T-1}) \\ &\quad \times \exp\Bigg\{\gamma\Bigg(x_0^T \Xi_{t,0} x_0 + \sum_{\tau=0}^{T-1} x_0^T \Xi^\tau_{t,1}\xi_\tau + \sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1} \xi_{\tau_1}^T \Xi^{\tau_1,\tau_2}_{t,2} \mathbb{E}_{\tilde{P}_2}[\eta_{\tau_2}] \\ &\qquad + \sum_{\tau_1=0}^{T-1}\sum_{\tau_2=\tau_1}^{T-1} \mathbb{E}[\eta_{\tau_1}]^T\, \Xi^{\tau_1,\tau_2}_{t,2}\xi_{\tau_2} + \sum_{\tau=0}^{T-1} \xi_\tau^T \tilde{\Xi}_\tau \xi_\tau \Bigg)\Bigg\}. \end{aligned} \tag{5.72}$$

As

$$dP_1(\xi_0, \xi_1, \ldots, \xi_{T-1}) = \prod_{t=0}^{T-1} dP_{1,t}(\xi_t), \tag{5.73}$$

we have $d\tilde{P}_1$ still have the independence structure and

$$d\tilde{P}_{1,t}(\xi_t) \propto dP_{1,t}(\xi_t) \times \exp\left\{\gamma\left(x_0^T \Xi_{t,1}^t \xi_t + \sum_{\tau=0}^{t-1} \mathbb{E}_{\tilde{P}_2}\left[\eta_\tau\right]^T \Xi_{\tau,2}^{\tau,t} \xi_t \right.\right.$$
$$\left.\left. + \sum_{\tau=t}^{T-1} \mathbb{E}_{\tilde{P}_2}\left[\eta_\tau\right]^T \left(\Xi_{t,2}^{t,\tau}\right)^T \xi_t + \xi_t^T \tilde{\Xi}_t \xi_t \right)\right\}. \tag{5.74}$$

The proof is similar for $d\tilde{P}_2$. □

Algorithm 5: Nonlinear Non-Gaussian Risk-Averse Control with Coupling Method

Data: initial point $\bar{x}_t^{\text{init}}, \bar{u}_t^{\text{init}}$, initial step size η_{init}, step size shrinkage rate β, minimal step size ϵ_{stop}, c_1, c_2 for Wolfe conditions, data points Data_t, minimal increase of inner step ϵ_{inc}.

Result: $\bar{x}_t, \bar{u}_t, U_t, U_{t,\tau}$ and ξ_t for control policy $u_t(x_t) = U_t x_t + \sum_{t=\tau}^{T-1} U_{t,\tau}\xi_\tau + \bar{u}_t$.

for $t = 0, \ldots, T-1$ **do**
- $\bar{x}_t = \bar{x}_t^{\text{init}}, \bar{u}_t = \bar{u}_t^{\text{init}}$;

$\eta = \eta_{\text{init}}$;

while $\eta > \epsilon_{stop}$ **do**
- $\eta = \eta_{\text{init}}$;
- Compute gradients $D_0, D_1, \ldots, D_{T-1}$ at $\bar{x}_t, \bar{u}_t$ from Algorithm 6;
- **while** $\epsilon > \tilde{\epsilon}$ *and Wolfe conditions not satisfied* **do**
  - $\eta = \beta\eta$.
- **for** $t = 0, \ldots, T-1$ **do**
  - $(\bar{x}_t, \bar{u}_t) = (\bar{x}_t, \bar{u}_t) - \eta D_t$;

Expand (5.14) to (5.17) at $\bar{x}_t, \bar{u}_t$;

Reformulate (5.17) to (5.18);

Perform the backward run of (4.30) and get its matrices;

$\lambda = \epsilon_{\text{inc}} + 1$;

Set $\tilde{P}_{1,t}$ and $\tilde{P}_{2,t}$ to be uniform distribution over Data_t;

Compute V_0 from the objective of (5.67) at $\tilde{P}_{1,0}, \ldots, \tilde{P}_{1,T-1}$ and $\tilde{P}_{2,0}, \ldots, \tilde{P}_{2,T-1}$;

while $\lambda > \epsilon_{inc}$ **do**
- **for** $t = 0, \ldots, T-1$ **do**
  - Compute $\tilde{P}_{1,t}^{\text{new}}$ from (5.70) with given $\tilde{P}_{2,t}$;
  - Compute $\tilde{P}_{2,t}^{\text{new}}$ from (5.71) with given $\tilde{P}_{1,t}^{\text{new}}$;
  - $\tilde{P}_{1,t} = \tilde{P}_{1,t}^{\text{new}}, \tilde{P}_{2,t} = \tilde{P}_{2,t}^{\text{new}}$;
- Compute V_1 from the objective of (5.67) at $\tilde{P}_{1,0}, \ldots, \tilde{P}_{1,T-1}$ and $\tilde{P}_{2,0}, \ldots, \tilde{P}_{2,T-1}$;
- $\lambda = V_1 - V_0$;
- $V_0 = V_1$;

for $t = 0, \ldots, T-1$ **do**
- $\tilde{P}_t = \frac{1}{2}\left(\tilde{P}_{1,t} + \tilde{P}_{2,t}\right)$;
- $\xi_t = \mathbb{E}_{\tilde{P}_t}\left[\varepsilon_t\right]$;

Algorithm 6: Gradient Computation in Nonlinear Non-Gaussian Risk-Averse Control

Data: point $\bar{x}_t, \bar{u}_t$, data points Data_t, minimal increase of inner step ϵ_{inc}.
Result: Gradient D_t
Expand (5.14) to (5.17) at $\bar{x}_t, \bar{u}_t$;
Reformulate (5.17) to (5.18);
Perform the backward run of (4.30) and get its matrices;
$\lambda = \epsilon_{\text{inc}} + 1$;
Set $\tilde{P}_{1,t}$ and $\tilde{P}_{2,t}$ to be uniform distribution over Data_t;
Compute V_0 from the objective of (5.67) at $\tilde{P}_{1,0}, \dots, \tilde{P}_{1,T-1}$ and $\tilde{P}_{2,0}, \dots, \tilde{P}_{2,T-1}$;
while $\lambda > \epsilon_{inc}$ **do**
  for $t = 0, \dots, T-1$ **do**
    Compute $\tilde{P}^{\text{new}}_{1,t}$ from (5.70) with given $\tilde{P}_{2,t}$;
    Compute $\tilde{P}^{\text{new}}_{2,t}$ from (5.71) with given $\tilde{P}^{\text{new}}_{1,t}$;
    $\tilde{P}_{1,t} = \tilde{P}^{\text{new}}_{1,t}, \tilde{P}_{2,t} = \tilde{P}^{\text{new}}_{2,t}$;
  Compute V_1 from the objective of (5.67) at $\tilde{P}_{1,0}, \dots, \tilde{P}_{1,T-1}$ and $\tilde{P}_{2,0}, \dots, \tilde{P}_{2,T-1}$;
  $\lambda = V_1 - V_0$;
  $V_0 = V_1$;
for $t = 0, \dots, T-1$ **do**
  $\tilde{P}_t = \frac{1}{2}\left(\tilde{P}_{1,t} + \tilde{P}_{2,t}\right)$;
for $t = 0, \dots, T-1$ **do**
  Compute the gradient D_t of (5.62) w.r.t. $\bar{x}_t, \bar{u}_t$ at $\tilde{P}_0, \dots, \tilde{P}_{T-1}$;

REFERENCES

Mohammed Amin Abdullah, Hang Ren, Haitham Bou Ammar, Vladimir Milenkovic, Rui Luo, Mingtian Zhang, and Jun Wang. Wasserstein robust reinforcement learning. *arXiv preprint arXiv:1907.13196*, 2019.

Amjad Ali, Ibrar Yaqoob, Ejaz Ahmed, Muhammad Imran, Kyung Sup Kwak, Adnan Ahmad, Syed Asad Hussain, and Zulfiqar Ali. Channel clustering and qos level identification scheme for multi-channel cognitive radio networks. *IEEE Communications Magazine*, 56 (4):164–171, 2018.

Walter Alt and Christopher Schneider. Linear-quadratic control problems with l1-control cost. *Optimal control applications and methods*, 36(4):512–534, 2015.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

Richard Bellman. Dynamic programming. *Science*, 153(3731):34–37, 1966.

Alberto Bemporad, Manfred Morari, Vivek Dua, and Efstratios N Pistikopoulos. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1):3–20, 2002.

Aleksandr Beznosikov, Eduard Gorbunov, Hugo Berard, and Nicolas Loizou. Stochastic gradient descent-ascent: Unified theory and new efficient methods. *arXiv preprint arXiv:2202.07262*, 2022.

Jose Blanchet and Yang Kang. Distributionally robust groupwise regularization estimator. In *Asian Conference on Machine Learning*, pages 97–112. PMLR, 2017.

Jose Blanchet, Yang Kang, and Karthyek Murthy. Robust wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, 56(3):830–857, 2019.

George EP Box. Science and statistics. *Journal of the American Statistical Association*, 71 (356):791–799, 1976.

George EP Box. Robustness in the strategy of scientific model building. In *Robustness in statistics*, pages 201–236. Elsevier, 1979.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

David S Broomhead and David Lowe. Radial basis functions, multi-variable functional interpolation and adaptive networks. Technical report, Royal Signals and Radar Establishment Malvern (United Kingdom), 1988.

Martin D Buhmann. Radial basis functions. *Acta numerica*, 9:1–38, 2000.

Ruidi Chen and Ioannis C Paschalidis. A robust learning approach for regression models based on distributionally robust optimization. *Journal of Machine Learning Research*, 19 (13), 2018.

Alon Cohen, Avinatan Hasidim, Tomer Koren, Nevena Lazic, Yishay Mansour, and Kunal Talwar. Online linear quadratic control. In *International Conference on Machine Learning*, pages 1029–1038. PMLR, 2018.

Jeremy Coulson, John Lygeros, and Florian Dörfler. Regularized and distributionally robust data-enabled predictive control. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 2696–2701. IEEE, 2019.

Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion models in vision: A survey. *arXiv preprint arXiv:2209.04747*, 2022.

John M Danskin. The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–664, 1966.

Andreas Dechant and Yohei Sakurai. Thermodynamic interpretation of wasserstein distance. *arXiv preprint arXiv:1912.08405*, 2019.

Seref Demirci. The requirements for automation systems based on boeing 737 max crashes. *Aircraft Engineering and Aerospace Technology*, 2021.

Luc Devroye. Nonparametric density estimation. *The L_1 View*, 1985.

Anushri Dixit, Mohamadreza Ahmadi, and Joel W Burdick. Distributionally robust model predictive control with total variation distance. *arXiv preprint arXiv:2203.12062*, 2022.

Xialiang Dou and Mihai Anitescu. Distributionally robust optimization with correlated data from vector autoregressive processes. *Operations Research Letters*, 47(4):294–299, 2019.

John Duchi. Derivations for linear algebra and optimization. *Berkeley, California*, 3(1): 2325–5870, 2007.

John Duchi, Peter Glynn, and Hongseok Namkoong. Statistics of robust optimization: A generalized empirical likelihood approach. *arXiv preprint arXiv:1610.03425*, 2016.

Darrell Duffie and Jun Pan. An overview of value at risk. *Journal of derivatives*, 4(3):7–49, 1997.

Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *The Annals of statistics*, 32(2):407–499, 2004.

Daniel Faccini, Francesca Maggioni, and Florian A Potra. Robust and distributionally robust optimization models for linear support vector machine. *Computers & Operations Research*, 147:105930, 2022.

Linwei Fan, Fan Zhang, Hui Fan, and Caiming Zhang. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art*, 2(1):1–12, 2019.

Farbod Farshidian and Jonas Buchli. Risk sensitive, nonlinear optimal control: Iterative linear exponential-quadratic optimal control with gaussian noise. *arXiv preprint arXiv:1512.07173*, 2015.

Marco Frittelli and Emanuela Rosazza Gianin. Law invariant convex risk measures. *Advances in mathematical economics*, pages 33–46, 2005.

Rui Gao and Anton Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655, 2023.

Rui Gao and Anton J Kleywegt. Distributionally robust stochastic optimization with wasserstein distance. *arXiv preprint arXiv:1604.02199*, 2016.

Ather Gattami. Generalized linear quadratic control. *IEEE Transactions on Automatic Control*, 55(1):131–136, 2009.

Vincent Guigues, Alexander Shapiro, and Yi Cheng. Risk-averse stochastic optimal control: an efficiently computable statistical upper bound. *arXiv preprint arXiv:2112.09757*, 2021.

Astghik Hakobyan and Insoon Yang. Wasserstein distributionally robust control of partially observable linear stochastic systems. *arXiv preprint arXiv:2212.04644*, 2022.

W Heisenberg. Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik. *Zeitschrift für Physik*, 43(3):172–198, 1927.

David Jacobson. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic control*, 18(2):124–131, 1973.

Haoming Jiang, Zhehui Chen, Yuyang Shi, Bo Dai, and Tuo Zhao. Learning to defend by learning to attack. In *International Conference on Artificial Intelligence and Statistics*, pages 577–585. PMLR, 2021.

Kihyun Kim and Insoon Yang. Minimax control of ambiguous linear stochastic systems using the wasserstein metric. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 1777–1784. IEEE, 2020.

Masako Kishida and Ahmet Cetinkaya. Risk-aware linear quadratic control using conditional value-at-risk. *IEEE Transactions on Automatic Control*, 2022.

Ralf Korn and Olaf Menkens. Worst-case scenario portfolio optimization: a new stochastic control approach. *Mathematical Methods of Operations Research*, 62(1):123–140, 2005.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Harold Kushner. Introduction to stochastic control. Technical report, BROWN UNIV PROVIDENCE RI DIV OF APPLIED MATHEMATICS, 1971.

Joseph Louis Lagrange. *Mécanique analytique*, volume 1. Mallet-Bachelier, 1853.

Jean-Claude Latombe. *Robot motion planning*, volume 124. Springer Science & Business Media, 2012.

Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics: theory and practice*. Princeton University Press, 2007.

Weiwei Li and Emanuel Todorov. Iterative linear quadratic regulator design for nonlinear biological movement systems. In *ICINCO (1)*, pages 222–229. Citeseer, 2004.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalence is efficient for linear quadratic control. *Advances in Neural Information Processing Systems*, 32, 2019.

J Clerk Maxwell. Theory of heat. *Astronomical register*, 10:107–107, 1872.

James C Maxwell. Li. on physical lines of force. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(141):338–348, 1861.

P McCullagh and JA Nelder. *Generalized linear models*, chapter 1.1.4. Chapman and Hill, 1989.

Hongseok Namkoong and John C Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. *Advances in neural information processing systems*, 29, 2016.

Hongseok Namkoong and John C Duchi. Variance-based regularization with convex objectives. *Advances in neural information processing systems*, 30, 2017.

Isaac Newton and NW Chittenden. *Newton's principia: The mathematical principles of natural philosophy*. Geo. P. Putnam, 1850.

Haruki Nishimura, Negar Mehr, Adrien Gaidon, and Mac Schwager. Rat ilqr: A risk auto-tuning controller to optimally account for stochastic model mismatch. *IEEE Robotics and Automation Letters*, 6(2):763–770, 2021.

Erfaun Noorani, Christos Mavridis, and John Baras. Risk-sensitive reinforcement learning with exponential criteria. *arXiv preprint arXiv:2212.09010*, 2022.

Jung Hun Oh, Maryam Pouryahya, Aditi Iyer, Aditya P Apte, Allen Tannenbaum, and Joseph O Deasy. Kernel wasserstein distance. *arXiv preprint arXiv:1905.09314*, 2019.

Mario VF Pereira and Leontina MVG Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, 52(1):359–375, 1991.

Ian R Petersen, Matthew R James, and Paul Dupuis. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Transactions on Automatic Control*, 45(3):398–412, 2000.

Jorion Philippe. Value at risk: the new benchmark for managing financial risk. *NY: McGraw-Hill Professional*, 2001.

George R Price. Extension of covariance selection mathematics. *Annals of human genetics*, 35(4):485–490, 1972.

J Punch. Johannes poncius's commentary on john duns scotus's opus oxoniense, book iii, dist. 34, q. 1. *John Duns Scotus Opera Omnia*, 15, 1639.

Majid Raeis, Ali Tizghadam, and Alberto Leon-Garcia. Predicting distributions of waiting times in customer service systems using mixture density networks. In *2019 15th International Conference on Network and Service Management (CNSM)*, pages 1–6. IEEE, 2019.

Hamed Rahimian and Sanjay Mehrotra. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659*, 2019.

Yangang Ren, Jingliang Duan, Shengbo Eben Li, Yang Guan, and Qi Sun. Improving generalization of reinforcement learning with minimax distributional soft actor-critic. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2020.

R Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

Georg Schildbach, Paul Goulart, and Manfred Morari. The linear quadratic regulator with chance constraints. In *2013 European Control Conference (ECC)*, pages 2746–2751. IEEE, 2013.

Bernhard Scholkopf, Kah-Kay Sung, Christopher JC Burges, Federico Girosi, Partha Niyogi, Tomaso Poggio, and Vladimir Vapnik. Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, 45(11): 2758–2765, 1997.

Bernhard Schölkopf, Koji Tsuda, and Jean-Philippe Vert. *Kernel methods in computational biology*. MIT press, 2004.

Mathijs Schuurmans and Panagiotis Patrinos. Data-driven distributionally robust control of partially observable jump linear systems. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 4332–4337. IEEE, 2021.

Pierre OM Scokaert and James B Rawlings. Constrained linear quadratic regulation. *IEEE Transactions on automatic control*, 43(8):1163–1169, 1998.

Soroosh Shafieezadeh Abadeh, Peyman M Mohajerin Esfahani, and Daniel Kuhn. Distributionally robust logistic regression. *Advances in Neural Information Processing Systems*, 28, 2015.

Soroosh Shafieezadeh Abadeh, Viet Anh Nguyen, Daniel Kuhn, and Peyman M Mohajerin Esfahani. Wasserstein distributionally robust kalman filtering. *Advances in Neural Information Processing Systems*, 31, 2018.

Alexander Shapiro. Distributionally robust optimal control and mdp modeling. *Operations Research Letters*, 49(5):809–814, 2021.

Alexander Shapiro and Alois Pichler. Conditional distributionally robust functionals. *Available on Optimization Online*, 2022.

Alexander Shapiro, Enlu Zhou, and Yifan Lin. Bayesian distributionally robust optimization. *arXiv preprint arXiv:2112.08625*, 2021.

Meng-Fu Shih and Alfred O Hero. Unicast-based inference of network link delay distributions with finite mixture models. *IEEE Transactions on Signal Processing*, 51(8):2219–2228, 2003.

Athanasios Sideris and James E Bobrow. An efficient sequential linear quadratic algorithm for solving nonlinear optimal control problems. In *Proceedings of the 2005, American Control Conference, 2005.*, pages 2275–2280. IEEE, 2005.

David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *International conference on machine learning*, pages 387–395. PMLR, 2014.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

Aman Sinha, Hongseok Namkoong, Riccardo Volpi, and John Duchi. Certifying some distributional robustness with principled adversarial training. *arXiv preprint arXiv:1710.10571*, 2017.

Maurice Sion. On general minimax theorems. *Pacific Journal of mathematics*, 8(1):171–176, 1958.

Matthew Staib and Stefanie Jegelka. Distributionally robust optimization and generalization in kernel methods. *Advances in Neural Information Processing Systems*, 32, 2019.

Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. *Advances in neural information processing systems*, 12, 1999.

Gilead Tadmor. Worst-case design in the time domain: The maximum principle and the standard h-infinity problem. *Mathematics of Control, Signals and Systems*, 3(4):301–324, 1990.

Bahar Taşkesen, Dan A Iancu, Çağıl Koçyiğit, and Daniel Kuhn. Distributionally robust linear quadratic control. *arXiv preprint arXiv:2305.17037*, 2023.

Chen Tessler, Yonathan Efroni, and Shie Mannor. Action robust reinforcement learning and applications in continuous control. In *International Conference on Machine Learning*, pages 6215–6224. PMLR, 2019.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

Jo-Anne Ting, Evangelos Theodorou, and Stefan Schaal. A kalman filter for robust outlier detection. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1514–1519. IEEE, 2007.

Emanuel Todorov and Weiwei Li. A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *Proceedings of the 2005, American Control Conference, 2005.*, pages 300–306. IEEE, 2005.

Ioannis Tzortzis, Charalambos D Charalambous, and Themistoklis Charalambous. Dynamic programming subject to total variation distance ambiguity. *SIAM Journal on Control and Optimization*, 53(4):2040–2075, 2015.

Núria Armengol Urpí, Sebastian Curi, and Andreas Krause. Risk-averse offline reinforcement learning. *arXiv preprint arXiv:2102.05371*, 2021.

Stanislav Uryasev. Conditional value-at-risk: Optimization algorithms and applications. In *proceedings of the IEEE/IAFE/INFORMS 2000 conference on computational intelligence for financial engineering (CIFEr)(Cat. No. 00TH8520)*, pages 49–57. IEEE, 2000.

Bart PG Van Parys, Daniel Kuhn, Paul J Goulart, and Manfred Morari. Distributionally robust control of constrained stochastic systems. *IEEE Transactions on Automatic Control*, 61(2):430–442, 2015.

Leonid Nisonovich Vaserstein. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–72, 1969.

Wanting Xu and Mihai Anitescu. Exponentially accurate temporal decomposition for long-horizon linear-quadratic dynamic optimization. *SIAM Journal on Optimization*, 28(3): 2541–2573, 2018.

Insoon Yang. Wasserstein distributionally robust stochastic control: A data-driven approach. *IEEE Transactions on Automatic Control*, 66(8):3863–3870, 2020.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.