Stochastic Analysis of Mean-Field Games, Portfolio Optimization and Low-Rank Matrix Approximation

by
Peiyao Lai

A Dissertation
Submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE
In partial fulfillment of the requirements for the
Degree of Doctor of Philosophy
in
Mathematical Sciences

May 2024

APPROVED:

Dr. Oren Mangoubi, Advisor
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Qingshuo Song
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Stephan Sturm
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Fangfang Wang
Department of Mathematical Sciences
Worcester Polytechnic Institute

Dr. Aaron Smith
Department of Mathematics and Statistics
University of Ottawa

This page is intentionally left blank

Abstract

This thesis employs stochastic analysis tools to address three distinct problems. Firstly, in Hybrid Linear Quadratic Gaussian (LQG) Mean Field Games (MFGs), we investigate the convergence rate of the N-player linear quadratic Gaussian game towards its asymptotic Mean Field Games, using an explicit coupling method. The two main results are as follows. With some assumptions, one is to characterize the Mean-Field game equilibrium path as well as the associated equilibrium measure. The other is to obtain the convergence rate from the N-player game to that from mean-field games in distribution. The second problem involves finding the robust relative performance maximizing portfolio in an incomplete information setting, where the objective is to find the optimal strategy for an investor maximizing her/his robust utility. In the third problem, we obtain tighter right-singular vector perturbation bounds for rectangular matrices perturbed by Gaussian random matrix noise, by analyzing the perturbed matrix as a Dyson-Bessel matrix-valued diffusion. Applications of the perturbation bounds include the subspace recovery problem and the rank-k matrix approximation problem.

Acknowledgments

First and foremost, my heartfelt gratitude goes to my advisor, Professor Oren Mangoubi, for his steadfast support and invaluable guidance during times of adversity. It was in late 2022 that I turned to Oren during the darkest phase of my Ph.D. journey – feeling lost, battling imposter syndrome, and on the brink of leaving the program altogether. Without Oren's unwavering mentorship and assistance, successfully completing my doctoral journey would have been unattainable.

I extend my appreciation to my committee members and research collaborators, Professors Qingshuo Song, and Stephan Sturm. They have allocated time each week to discuss Mean Field Games and Stochastic Analysis during our independent study sessions, providing valuable guidance and insights while reviewing my thesis drafts.

I am grateful to my two other committee members, Professors Fangfang Wang and Aaron Smith, for their generous commitment. Their valuable contributions to the proposal and defense of this thesis, as well as their thorough review of my drafts and insightful feedback, have been indispensable to its completion.

I must also acknowledge and thank Professor Gu Wang, whose kindness and support were instrumental during the initial years of my Ph.D. journey and my formal examinations. Gu provided invaluable guidance in navigating the realm of academic mathematics, and I am forever appreciative of his contributions to my academic development.

I express my deep gratitude to Professor Yuecai Han of Jilin University, who served as my advisor during my master's studies. Additionally, I am grateful to my schoolteachers at Dingnan Erzhong, particularly Qiaoyan Li and Shenglong Xie, for their inspirational guidance in mathematics.

I am also indebted to my friends at WPI who have enriched my experience. Special thanks to Jiamin Jian and Jiaxuan Ye for the productive collaboration on Mean Field Games. Gratitude to Guillermo, Evan, and Riuji for establishing a study workshop and sharing their insights. My heartfelt thanks also go to my friends for their kindness and companionship: Puen, Di, Juan, Yanying, and Fengcheng.

A heartfelt acknowledgment goes to the Mathematics Department at WPI— Sarah Olson, Rhonda Podell, Louisa Kulp, Mike Malone, and Greg Aubin—for their kindness, generosity, and support. Additionally, I am thankful to the summer teaching program for granting me the chance to develop and instruct Matrices and Linear Algebra courses for undergraduate students at WPI.

Last, but most of all, I express deep gratitude to my family - Amǎ Xiaoqing, and Abá Shusen, Ajǐ Peiwen, and my grandparents, Ajiǎ, Agōng, as well as Jiǎpó and Jiǎgōng (may they rest in peace). Their enduring love and support have been a cornerstone of strength and inspiration throughout this long journey. I'd have never gotten here without them.

Contents

Abstract **iii**

Acknowledgments **iv**

List of Notations **vii**

List of Figures **viii**

1 Hybrid Linear Quadratic Gaussian Mean Field Games **1**

1.1 Introduction and Literature Review 1

1.1.1 Outline 4

1.2 Problem Setup 4

1.2.1 Notations 4

1.2.2 Definitions 5

1.2.3 Main result with time-dependent sensitivity $h(y, t)$ 8

1.2.4 Remarks on the main results 10

1.3 Main Result of MFG problem 11

1.3.1 Overview 11

1.3.2 Generic player's control problem given a population measure 13

1.3.3 Proof of Main Theorem 1.2.1 17

1.4 Main Result of Convergence of N-player Game to MFGs 20

1.4.1 Characterization of N-player game by Riccati system 21

1.4.2 Reduced Riccati form for the equilibrium 22

1.4.3 Proof of convergence rate 25

1.5 Numerical Simulations 30

1.5.1 Simulations of a generic player in MFGs 30

1.5.2 Simulations of convergence of the N-player game to MFGs 31

1.6 Conclusion and Future Work 31

1.6.1 Future work 34

2 Robust Relative Performance for Portfolio Optimization **36**

2.1 Introduction and Literature Review 36

2.1.1 Outline 38

2.2 Problem Setup 39

2.3 Main Results of Robust Relative Performance Problem with Incomplete Information 40

2.3.1 Main theorem and remarks . . . 41
2.3.2 Proof of the Main Theorem 2.3.1 . . . 42
2.3.3 Verification for the case $p \neq 0$ . . . 46
2.4 Value of Information . . . 47
2.4.1 Analysis of Nash equilibrium strategy π_i^* with complete information . . . 47
2.4.2 Value of the Information . . . 51
2.5 Numerical Simulations . . . 51
2.6 Conclusion and Future Work . . . 53

3 Random Matrix Perturbation Bounds for Low-Rank Approximation 56
3.1 Introduction . . . 56
3.1.1 Previous works . . . 58
3.1.2 Outline . . . 59
3.2 Problem Formulation . . . 59
3.2.1 Notations . . . 59
3.2.2 Main results of matrix perturbation bounds . . . 60
3.2.3 Two applications to matrix theory . . . 61
3.3 Preliminaries . . . 63
3.3.1 Dyson Bessel process . . . 63
3.3.2 Right singular vector SDE . . . 64
3.3.3 Previous perturbation bounds . . . 64
3.4 Proofs of Main Results . . . 65
3.4.1 Outline of proof of Main Theorem 3.2.1 . . . 65
3.4.2 Proof of Main Theorem 3.2.1 . . . 66
3.4.3 Proof of subspace recovery bound . . . 74
3.4.4 Proof of rank-k matrix approximation bound . . . 75
3.5 Numerical Simulations . . . 77
3.5.1 Simulations of rank-k matrix approximation . . . 77
3.5.2 Simulations of subspace recovery . . . 80
3.6 Conclusion and Future Work . . . 83

A Appendix 85
A.1 Some explicit solutions on LQG-MFGs . . . 85
A.2 Dynkin's formula for a regime-switching diffusion with a quadratic function . . . 87
A.3 Proof of the property of G . . . 89
A.4 Proof of the existence and uniqueness of the ODE system . . . 91
A.5 Multidimensional Problem on LQG-MFGs . . . 95
A.6 Comparison to bound (3.10) in Theorem 18 of S. O'Rourke et al.[35] . . . 96

Bibliography 97

List of Notations

$[m]_k$	k-th moment of the measure flow m
$\alpha, \alpha_i^{(N)}$	Control process
$\alpha^{(N)}$	Control $(\alpha_i^{(N)}, \alpha_{-i}^{(N)}) = (\alpha_1^{(N)}, \alpha_2^{(N)}, \ldots, \alpha_N^{(N)})$
$\Delta_{ij}(t)$	Singular value gap of $\Phi(t) = A + \sqrt{t}G$
γ, Γ	Specified singular value
$\hat{v}, \hat{V}$	Right singular vector of $\hat{A} = A + \sqrt{T}G$
$\hat{\sigma}, \hat{\Sigma}$	Singular value of $\hat{A} = A + \sqrt{T}G$
$\mathbb{P}, \mathbb{P}^{(N)}$	Probability
$\mathcal{A}, \mathcal{A}^{(N)}$	Admissible set for controls
$\mathcal{F}, \mathbb{F}$	Filtration
$\Omega, \Omega^{(N)}$	Probability space
π	Investment strategy process
$\sigma(t)$	Singular value process of $\Phi(t) = A + \sqrt{t}G$
σ, Σ	Singular value of matrix A
m	Measure flow
Q	Generator of the continuous time Markov chain Y
S	Admissible set for controls
$v(t)$	Right singular vector process of $\Phi(t) = A + \sqrt{t}G$
v, V	Right singular vector of matrix A
W_i, B	Brownian motion
$X, X^{(N)}$	Controlled process
Y	Continuous time Markov chain

List of Figures

1.1 MFGs diagram: infinite-dimensional fixed point condition with m_0 7
1.2 Equivalent MFGs diagram: finite dimensional fixed point condition with $\mu_0 = [m_0]_1$ and $\nu_0 = [m_0]_2$. 13
1.3 Simulations for a_y, b_y and c_y, the solution to Riccati system (1.12). 32
1.4 Simulations for value function V, optimal control α, and conditional second moment ν 32
1.5 μ_t: conditional mean of the population density 33
1.6 ν_t: conditional 2nd moment of the population density 33
1.7 Simulation of player 1's optimal value function V. 34

2.1 Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, \theta = 1.2$ given $\nu_i = 0$ 51
2.2 Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, \theta = 0.7$ given $\nu_i = 0$ 52
2.3 Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, p = 0.5$ given $\nu_i = 0$ 52
2.4 Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, p = -0.5$ given $\nu_i = 0$ 53
2.5 Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $\theta = 0.7, p = 0.5$ given $\nu_i = 0$ 54
2.6 Simulations of robust growth rate λ of θ and n, fix p given $\nu_i = 0.2$ 54
2.7 Simulations of Nash equilibrium growth rate ρ of variant θ and n, fix p given $\nu_i = 0.2$, ... 55

3.1 US census 1990 dataset (data source see [49]): the singular values decay *exponentially* fast. The horizontal axis shows the descending order of singular values, and the vertical axis shows a log plot of corresponding singular values 60
3.2 Simulation of the ratio of l.h.s. and r.h.s. of the bound in Corollary 3.2.2, when $k = 15, d = 15$. 78
3.3 Simulation of the ratio of l.h.s. and r.h.s. of the bound in Corollary 3.2.2, when $k = 10, m = 2150$. 79
3.4 Simulation of the ratio of l.h.s. and r.h.s. of the bound in Corollary 3.2.2, when $m = 850, d = 800$. 80
3.5 Simulation of the error of variable m when $k = 9, d = 15$. 81
3.6 Simulation of the error of variable k when $m = 850, d = 800$. 82
3.7 Simulation of the error of variable d when $k = 10, m = 2350$. 83

1

Hybrid Linear Quadratic Gaussian Mean Field Games

The convergence rate of equilibrium measures of N-player Games with Brownian common noise to its asymptotic Mean Field Game system is known as $O(N^{-1/9})$ with respect to 1-Wasserstein distance, obtained by the monograph [9, Cardaliaguet, Delarue, Lasry, Lions (2019)]. In this chapter, we study the convergence rate of the N-player LQG game with a Markov chain common noise towards its asymptotic Mean Field Game.

The main tool relies on an explicit coupling of the optimal trajectory of the N-player game which is driven by N-dimensional Brownian motion and the mean-field game counterpart which is driven by one-dimensional Brownian motion. The two main results are as follows. With some assumptions, one is to characterize the mean-field game equilibrium path $\hat{X}$ as well as associated equilibrium measure $\hat{m}$. The other is to obtain the convergence rate of $(\hat{X}_{1t}^{(N)}, Y^{(N)})$ from the N-player game to $(\hat{X}_t, Y)$ from mean-field games in distribution. One extension of [34, Jian, Lai, Song, and Ye] made in this thesis is that we showed the main results also hold when the sensitivity function $h(y, t)$ is time dependent.

1.1 Introduction and Literature Review

The field of Mean Field Games (MFGs) has emerged as a powerful framework for modeling strategic interactions among a large number of rational agents. Mean Field Game (MFG) theory is intended to describe an asymptotic limit of complex N-player differential game invariant to a reshuffling of the players' indices, and has attracted resurgent attention from numerous researchers in probability after its pioneering works of [39, Lasry and Lions] and [31, Huang, Caines, and Malhame], and we refer to comprehensive descriptions to the book [10, Carmona and Delarue] and the references therein.

In recent years, there has been a growing interest in extending traditional MFG models to incorporate additional complexities, such as stochastic dynamics and hybrid control structures. One such extension is the Hybrid Linear Quadratic Gaussian (LQG) Mean Field Games, which combines elements of stochastic control theory with Gaussian processes to address dynamic decision-making in uncertain environments.

we study the convergence rate of equilibrium measures of N-player differential game in the context of a Linear-Quadratic (LQ) structure with a common noise to its limiting MFG system. Different from the works mentioned above, the common noise in this thesis is a continuous-time Markov chain (CTMC) instead of Brownian motion, which oftentimes models the real-world control problems associated with hybrid systems. Markov chains are widely used to model systems that exhibit randomness and transition between different states. In various real-world scenarios, especially in economics (see [60]), finance ([69]), biology ([70]), and engineering ([68]), the dynamics of systems can be effectively represented as discrete states with probabilistic transitions between them.

LQ control problems have been widely recognized in the stochastic control theory due to their broad applications. More importantly, LQ structure leads to solvability in a closed form, namely the Ricatti system, and this usually sheds light on many fundamental properties of the control theory. For this reason, LQ structure has also been studied in MFGs with or without common noises for its importance. The related literature include major and minor LQG Mean Field Games system ([29, 50, 18]); social optimal in LQG Mean Field Games ([30, 17]); the LQG Mean Field Games with different model settings ([3, 27, 4, 28]); and LQG Graphon Mean Field Games ([20]). Recently, LQ Mean Field Games with a Brownian motion as the common noise have also been studied in ([1, 59]) with restrictions of the dependence of measure on its mean alone. Moreover, some literature considers various topics of Mean Field control and game problems with Markov chain common noise, see [42, 51, 52].

A fundamental question in this regard is the convergence rate of N-player game to the desired MFG system. A well-known result is about the convergence rate of value functions of the generic player, which can be shown $O(N^{-1})$, see for instance [8, 9, 10, 32]. In particular, [32] establishes the convergence rate of value functions in the sense of

$$J_1^N(\hat{\alpha}_1, \hat{\alpha}_{-1}) \leq J_1^N(\alpha_1, \hat{\alpha}_{-1}) + O(N^{-1}),$$

where J_1^N is the value of the first player in N-player game and $\hat{\alpha}$ is the Nash equilibrium decentralized control process for the MFG problem.

On the contrary, another challenging aspect lies in determining the convergence rate of equilibrium measures, which is complicated by the correlation structures among N players. To be more concrete, we examine the behavior of the $\hat{X}_{it}^{(N)}$, which represents the equilibrium state of the i-th player at time t in the N-player game defined within the probability space $\left(\Omega^{(N)}, \mathcal{F}^{(N)}, \mathbb{F}^{(N)}, \mathbb{P}^{(N)}\right)$. Additionally, we denote $\hat{X}_t$ as the equilibrium path at time t derived from the associated MFG defined in the probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. The question pertains to the convergence of $\hat{X}_{1t}^{(N)}$ as follows:

(Q) The $\mathbb{W}_p$-convergence rate of the representative equilibrium path,

$$\mathbb{W}_p\left(\mathcal{L}\left(\hat{X}_{1t}^{(N)}\right), \mathcal{L}\left(\hat{X}_t\right)\right) = O\left(N^{-?}\right).$$

Here, $\mathbb{W}_p$ denotes the p-Wasserstein metric.

The existing literature extensively explores the convergence rate in this context. For (Q), Theorem 2.4.9 of the monograph [9] establishes a convergence rate of $O(N^{-1/2})$ using the $\mathbb{W}_1$ metric. More recently, [33] addresses (Q) by introducing displacement monotonicity and controlled common noise, and Theorem 2.23 applies the maximum principle of forward-backward propagation of chaos to achieve the same convergence rate. It is important to note that these results are not applicable to the Linear Quadratic Gaussian (LQG) framework, primarily due to the assumption concerning the linear growth of the cost functional.

The main result of this chapter establishes that the equilibrium measures exhibit a convergence rate of 1/2 concerning the 2-Wasserstein distance. The precise statement of this result can be found in Theorem 1.2.2. In comparison to the aforementioned literature, two primary distinctions emerge. Firstly, within the framework of Mean Field Games, the common noise is modeled as a Continuous-Time Markov Chain. Secondly, a significant difference lies in the cost function's behavior, as it does not possess linear growth within the context of the Linear Quadratic Gaussian (LQG) framework.

To obtain the desired convergence rate in this chapter, the first building block is the characterization of the equilibrium measure of the limiting MFG by a finite-dimensional ODE system. The key step leading us to a desired finite-dimensional system is that, instead of searching for the infinite-dimensional function directly, we postulate a Markovian structure via auxiliary processes (1.15) governed by its finite-dimensional coefficient functions, which exhibits the distinct feature of Markov chain common noise relatives to the Brownian motion counterpart.

The next stage towards the convergence rate is to compare the limiting MFG system to a N-player game. In contrast to the characterization of the MFG system, it is relatively routine to solve the N-player game due to its LQ structure. Therefore, the convergence rate problem can be recasted to the following question about the coupling of the two following processes: For two equilibrium processes $\hat{X}$ of MFG in Ω and $\hat{X}_1^{(N)}$ of N-player game in $\Omega^{(N)}$, finding a random process Z^N in Ω whose distribution is identical to $\hat{X}_1^{(N)}$ satisfying the estimate in the form of $\mathbb{E}[|\hat{X}_t - Z_t^N|^2] = O\left(N^{-?}\right)$. For this purpose, we first show an N-invariant algebraic structure of the seemingly intractable κN^3 dimensional ODE system (1.27), which originated from [32, Huang and Yang] as a dimensional reduction in the system with Brownian common noise. Thanks to this N-invariant structure, the complex ODE system (1.27) can be reduced to the ODE system (1.31) whose dimension agrees with the ODE (1.12) of MFG system. Moreover, $\hat{X}_1^{(N)}$ can be represented as a stochastic flow driven by two Brownian motions $W_1^{(N)}$ and $W_{-1}^{(N)} := \frac{1}{\sqrt{N-1}}\sum_{i=2}^N W_i^{(N)}$, which enables us to embed the equilibrium process $\hat{X}_1^{(N)}$ to any probability space having only two Brownian motions.

1.1.1 Outline

The rest of this chapter is outlined as follows: Section 1.2 presents a precise formulation of the problem and two main results. Section 1.3 is devoted to the derivation of our first result: the equilibrium of MFGs. In Section 1.4, we show in detail the convergence of the N-player game to MFGs, which yields our second main result. Section 1.5 demonstrates the convergence by some numerical simulations. The conclusion and some potential future works are summarized in Section 1.6. Appendix A.1 - A.5 are the appendixes that collect some related facts to support the main proof.

1.2 Problem Setup

First, we collect common notations used in this chapter in Subsection 1.2.1. Then, we set up problems on MFGs and the N-player game separately in Subsection 1.2.2. We present our main results with time-dependent sensitivity $h(y,t)$ in Subsection 1.2.3. In Subsection 1.2.4 we discuss some remarks on the main results.

1.2.1 Notations

Let $T > 0$ be a fixed terminal time and $(\Omega, \mathcal{F}_T, \mathbb{F} = \{\mathcal{F}_t : 0 \le t \le T\}, \mathbb{P})$ be a completed filtered probability space satisfying the usual conditions, on which W and B are two independent standard Brownian motions, and Y is a continuous time Markov chain (CTMC) independent of (W, B) taking values in a finite state space $\mathcal{Y} = \{1, 2, \dots, \kappa\}$ with a generator

$$Q = (q_{i,j})_{i,j\in\mathcal{Y}} \tag{1.1}$$

satisfying $q_{i,j} \ge 0$ for all $i \ne j \in \mathcal{Y}$ and $\sum_{i\ne j} q_{i,j} + q_{i,i} = 0$ for each $i \in \mathcal{Y}$. In the above, the Brownian motion B does not play any role in MFG problem formulation until the convergence proof of the N-player game to MFGs.

By $L^p := L^p(\Omega, \mathbb{P})$, we denote the space of random variables X on $(\Omega, \mathcal{F}_T, \mathbb{P})$ with finite p-th moment with norm $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$. We also denote by $L^p_{\mathbb{F}} := L^p_{\mathbb{F}}([0,T] \times \Omega)$ the space of all $\mathbb{F}$-progressively measurable random processes $\alpha = (\alpha_t)_{0\le t\le T}$ satisfying

$$\mathbb{E}\left[\int_0^T |\alpha_t|^p dt\right] < \infty.$$

For any polish (complete separable metric) space $(P, \mathcal{B}(P), d)$, we use δ_x to denote the Dirac measure on the point $x \in P$. Then, the collection of all probabilities m on $(P, \mathcal{B}(P), d)$ having finite k-th moment is denoted by $\mathcal{P}_k(P)$, i.e.

$$[m]_k := \int x^k m(dx) < \infty, \quad \forall m \in \mathcal{P}_k(P).$$

The equilibrium of MFGs with the common noise yields the conditional distribution. For real-valued random variables X and Z in $(\Omega, \mathcal{F}_T, \mathbb{P})$, we denote the distribution of X conditional on $\sigma(Z)$ by $\mathcal{L}(X|Z)$, or equivalently

$$\mathcal{L}(X|Z)(A) = \mathbb{E}[I_A(X)|Z], \quad \forall A \in \mathcal{F}_T.$$

Note that $\mathcal{L}(X|Z)(A) : \Omega \mapsto \mathbb{R}$ is a $\sigma(Z)$-measurable random variable, therefore, $\mathcal{L}(X|Z)$ is $\sigma(Z)$-measurable random probability distribution with k-th moment $[\mathcal{L}(X|Z)]_k = \mathbb{E}[X^k|Z]$, if it exists. We refer to more details on the conditional distribution in Volume II of [10].

Next proposition provides an embedding approach to prove the convergence in distribution, which will be used later in the convergence of the N-player game to MFGs.

Proposition 1.2.1 (Convergence in distribution)**.** *Suppose* $(\Omega^{(N)}, \mathcal{F}_T^{(N)}, \mathbb{P}^{(N)})$ *is a complete probability space. Let* $X^{(N)}$ *and* X *be random variables of* $\Omega^{(N)} \mapsto P$ *and* $\Omega \mapsto P$*, respectively. Then,* $X^{(N)}$ *is convergent in distribution to* X*, denoted by* $X^{(N)} \Rightarrow X$*, if there exists* $Z^N : \Omega \mapsto P$ *satisfying* $\mathcal{L}(Z^N) = \mathcal{L}(X^{(N)})$*, such that* $Z^N \to X$ *holds almost surely, i.e.*

$$\lim_{N\to\infty} d(Z^N, X) = 0, \text{ almost surely in } \mathbb{P},$$

where d *represents the metric assigned to the space* P*.*

In this chapter, we formulate the N-player game in the completed filtered probability space

$$(\Omega^{(N)}, \mathcal{F}_T^{(N)}, \mathbb{F}^{(N)} := \{\mathcal{F}_t^{(N)} : 0 \le t \le T\}, \mathbb{P}^{(N)}),$$

and $Y^{(N)}$ is the continuous time Markov chain in $\Omega^{(N)}$ with the same generator given by (1.1) and $W^{(N)} = (W_i^{(N)} : i = 1, \dots, N)$ is an N-dimensional standard Brownian motion. We assume $Y^{(N)}$ and $W^{(N)}$ are independent of each other.

For better clarity, we use the superscript (N) for a random variable to emphasize the probability space $\Omega^{(N)}$ it belongs to. For example, Proposition 1.2.1 denotes a random variable in $\Omega^{(N)}$ by $X^{(N)}$, while its distribution copy in Ω by Z^N, not by $Z^{(N)}$.

1.2.2 Definitions

The equilibrium of MFGs

In this subsection, we define the equilibrium of MFGs associated with a generic player's stochastic control problem in the probability setting Ω, see Section 1.2.1.

Given a random measure flow $m : (0, T] \times \Omega \mapsto \mathcal{P}_2(\mathbb{R})$, consider a generic player who wants to minimize her expected accumulated cost on $[0, T]$:

$$J(y, x, \alpha) = \mathbb{E}\left[\int_0^T \frac{1}{2}\alpha_s^2 + F(Y_s, X_s, s, m_s)ds + G(Y_T, X_T, m_T)\Big|Y_0 = y, X_0 = x\right] \tag{1.2}$$

with some given cost functions $F : \mathcal{Y} \times \mathbb{R} \times [0,T] \times \mathcal{P}_2(\mathbb{R}) \mapsto \mathbb{R}, G : \mathcal{Y} \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \mapsto \mathbb{R}$ and underlying random processes $(Y, X) : [0,T] \times \Omega \mapsto \mathcal{Y} \times \mathbb{R}$. Among three processes (Y, X, m), the generic player can control the process X via α in the form of

$$X_t = X_0 + \int_0^t \left(\tilde{b}_1(Y_s, s)X_s + \tilde{b}_2(Y_s, s)\alpha_s\right) ds + W_t, \quad \forall t \in [0,T], \tag{1.3}$$

where $\tilde{b}_1(\cdot,\cdot)$ and $\tilde{b}_2(\cdot,\cdot)$ are two deterministic functions. We assume that the initial state X_0 is independent of Y. The process Y of (1.1) represents the common noise and $m = (m_t)_{0\le t\le T}$ is a given random density flow normalized up to total mass one.

The objective of the control problem for the generic player is to find its optimal control $\hat{\alpha} \in \mathcal{A} := L^4_{\mathbb{F}}$ to minimize the total cost, i.e.

$$V[m](y,x) = J[m](y,x,\hat{\alpha}) \le J[m](y,x,\alpha), \quad \forall \alpha \in \mathcal{A}. \tag{1.4}$$

Associated to the optimal control $\hat{\alpha}$, we denote the optimal path by $\hat{X} = (\hat{X}_t)_{0\le t\le T}$. To introduce MFG Nash equilibrium, it is often convenient to highlight the dependence of the optimal path and optimal control of the generic player and its associated value on the underlying density flow m, which are respectively denoted by

$$\hat{X}_t[m], \hat{\alpha}_t[m], \text{ and } V[m].$$

Now, we present the definition of the equilibrium below, see also Volume II page 127 of [10] for a general setup with a common noise.

Definition 1.2.1 (MFG equilibrium measure, equilibrium path and equilibrium control). *Given an initial distribution $\mathcal{L}(X_0) = m_0 \in \mathcal{P}_2(\mathbb{R})$, a random measure flow $\hat{m} = \hat{m}(m_0)$ is said to be an MFG equilibrium measure if it satisfies the fixed point condition*

$$\hat{m}_t = \mathcal{L}(\hat{X}_t[\hat{m}]|Y), \ \forall 0 < t \le T, \quad \textit{almost surely in } \mathbb{P}. \tag{1.5}$$

The path $\hat{X}$ and the control $\hat{\alpha}$ associated to $\hat{m}$ is called the MFG equilibrium path and equilibrium control, respectively. The value function of the control problem associated with the equilibrium measure $\hat{m}$ is called as MFG value function, denoted by

$$U(m_0, y, x) = V[\hat{m}](y,x). \tag{1.6}$$

The flowchart of MFGs diagram is given in Figure 1.1. It is noted from the optimality condition (1.4) and the fixed point condition (1.5) that

$$J[\hat{m}](y,x,\hat{\alpha}) \le J[\hat{m}](y,x,\alpha), \quad \forall \alpha$$

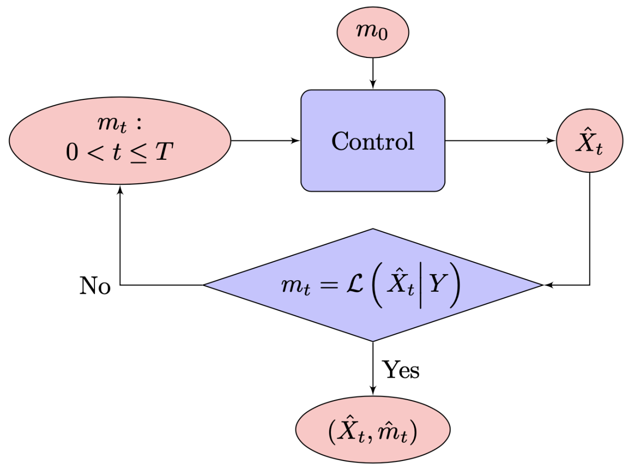


Figure 1.1: MFGs diagram: infinite-dimensional fixed point condition with m_0

holds for the equilibrium measure $\hat{m}$ and its associated equilibrium control $\hat{\alpha}$, while it is not

$$J[\hat{m}](y, x, \hat{\alpha}) \leq J[m](y, x, \alpha), \quad \forall \alpha, m.$$

Otherwise, this problem turns into a McKean-Vlasov control problem discussed in [51].

Equilibrium of the N-player game

The discrete counterpart of MFGs is an N-player game, which is formulated below in the probability space $\Omega^{(N)}$, see Section 1.2.1 for more details on the probability setup.

Recall that, $W_{it}^{(N)}$ and $W_{jt}^{(N)}$ are independent Brownian motions for $j \neq i$ and the common noise $Y^{(N)}$ is the continuous time Markov chain in $\Omega^{(N)}$ with the generator given by (1.1). Let the player i follow the dynamic, for $i = 1, 2, \ldots, N$,

$$dX_{it}^{(N)} = \left(\tilde{b}_1(Y_t^{(N)}, t) X_{it}^{(N)} + \tilde{b}_2(Y_t^{(N)}, t)\alpha_{it}^{(N)}\right) dt + dW_{it}^{(N)}, \quad X_{i0}^{(N)} = x_i^{(N)}. \tag{1.7}$$

The cost function for player i associated to the control $\alpha^{(N)} = (\alpha_i^{(N)} : i = 1, 2, \ldots, N)$ is

$$\begin{aligned} J_i^N(y, x^{(N)}, \alpha^{(N)}) = \mathbb{E}\Bigg[\int_0^T \left(\frac{1}{2}|\alpha_{it}^{(N)}|^2 + F(Y_t^{(N)}, X_{it}^{(N)}, t, \rho(X_t^{(N)}))\right) dt + \\ G(Y_T^{(N)}, X_{iT}^{(N)}, \rho(X_T^{(N)})) \Big| X_0^{(N)} = x^{(N)}, Y_0^{(N)} = y\Bigg], \end{aligned} \tag{1.8}$$

where $x^{(N)} = (x_1^{(N)}, x_2^{(N)}, \ldots, x_N^{(N)})$ is an $\mathbb{R}^N$-valued random vector in $\Omega^{(N)}$ to denote the initial state for N player, $\alpha_i^{(N)} \in \mathcal{A}^{(N)} := L^4_{\mathbb{F}^{(N)}}$, and

$$\rho(x^{(N)}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i^{(N)}}$$

is the empirical measure of a vector $x^{(N)}$ with Dirac measure δ. We use the notation for the control $\alpha^{(N)} = (\alpha_i^{(N)}, \alpha_{-i}^{(N)}) = (\alpha_1^{(N)}, \alpha_2^{(N)}, \dots, \alpha_N^{(N)})$.

Definition 1.2.2 (Equilibrium condition and equilibrium path of N-player game)**.**

1. *The value function of player i for $i = 1, 2, \dots, N$ of the Nash game is defined by $V^N = (V_i^N : i = 1, 2, \dots, N)$ satisfying the equilibrium condition*

$$V_i^N(y, x^{(N)}) = J_i^N(y, x^{(N)}, \hat{\alpha}_i^{(N)}, \hat{\alpha}_{-i}^{(N)}) \leq J_i^N(y, x^{(N)}, \alpha_i^{(N)},\ \hat{\alpha}_{-i}^{(N)}), \quad \forall \alpha_i^{(N)} \in \mathcal{A}^{(N)}. \tag{1.9}$$

2. *The equilibrium path of the N-player game is the random path $\hat{X}_t^{(N)} = (\hat{X}_{1t}^{(N)}, \dots, \hat{X}_{Nt}^{(N)})$ driven by (1.7) associated to the control $\hat{\alpha}_t^{(N)}$ satisfying the equilibrium condition of (1.9).*

1.2.3 Main result with time-dependent sensitivity $h(y,t)$

We consider the following two functions $F : \mathcal{Y} \times \mathbb{R} \times [0,T] \times \mathcal{P}_2(\mathbb{R}) \mapsto \mathbb{R}, G : \mathcal{Y} \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}) \mapsto \mathbb{R}$ in the cost functional (1.2):

$$F(y, x, t, m) = h(y, t) \int_{\mathbb{R}} (x - z)^2 m(dz), \tag{1.10}$$

and

$$G(y, x, m) = g(y) \int_{\mathbb{R}} (x - z)^2 m(dz), \tag{1.11}$$

for some $h : \mathcal{Y} \times [0,T] \mapsto \mathbb{R}^+, g : \mathcal{Y} \mapsto \mathbb{R}^+$. Note that the cost function F is a function of time variable t which is an extension of the cost function F in [34, Jian, Lai, Song, and Ye]. In this case, the F and G terms in (1.8) of the N-player game can be written by

$$F(Y_t^{(N)}, X_{it}^{(N)}, t, \rho(X_t^{(N)})) = \frac{h(Y_t^{(N)}, t)}{N} \sum_{j=1}^{N} (X_{it}^{(N)} - X_{jt}^{(N)})^2,$$

and

$$G(Y_T^{(N)}, X_{iT}^{(N)}, \rho(X_T^{(N)})) = \frac{g(Y_T^{(N)})}{N} \sum_{j=1}^{N} (X_{iT}^{(N)} - X_{jT}^{(N)})^2,$$

respectively.

Remark 1.2.1. *First, we note that F and G possess the quadratic structures in x. Secondly, the coefficients $h(y,t)$ and $g(y)$ provide the sensitivity to the mean-field effects, which depend on the current CTMC state. For another remark, let us consider the scenario where the number of states is 2 and sensitivities are invariant, for instance*

$$h(0,t) = h(1,t) = h,\ \ g(0) = g(1) = 0.$$

Then the cost function and hence the entire problem is free from the common noise. Interestingly, as shown in the Appendix A.1, there is no global solution for MFGs when $h < 0$, while there is a global solution when $h > 0$.

Moreover, the uniqueness of the MFGs can be achieved under the displacement monotonicity condition. It is easy to check that (1.10)-(1.11) satisfy the displacement monotonicity condition. Note that

$$F_x(y, x, t, m) = 2h(y, t)(x - [m]_1), \quad G_x(y, x, m) = 2g(y)(x - [m]_1),$$

which gives that

$$\begin{aligned}\mathbb{E}\left[\left(F_x(y, X_1, t, m_{X_1}) - F_x(y, X_2, t, m_{X_2})\right)(X_1 - X_2)\right] = \\ 2h(y, t)\left(\mathbb{E}\left[(X_1 - X_2)^2\right] - (\mathbb{E}[X_1] - \mathbb{E}[X_2])^2\right) \geq 0\end{aligned}$$

for all $y \in \mathcal{Y}$ if $h(y, t) > 0$ on $\mathcal{Y}$, where m_{X_1} and m_{X_2} is the law of X_1 and X_2 respectively. Similarly, we can obtain that

$$\mathbb{E}\left[\left(G_x(y, X_1, m_{X_1}) - G_x(y, X_2, m_{X_2})\right)(X_1 - X_2)\right] \geq 0$$

for all $y \in \mathcal{Y}$ if $g(y) > 0$ on $\mathcal{Y}$. Therefore, we require positive values for all sensitivities for simplicity. It is of course an interesting problem to investigate the explosion when some sensitivities are negative.

Wrapping up the above discussions, we impose the following assumptions.

Assumption 1.2.1 $(\tilde{b}_1, \tilde{b}_2, h, g, X_0)$**.**

(A0) $\tilde{b}_1(y, \cdot), \tilde{b}_2(y, \cdot), h(y, \cdot) : [0, T] \mapsto \mathbb{R}$ *are continuous functions for all* $y \in \mathcal{Y}$*.*

(A1) The cost functions are given by (1.10)-(1.11) with $h, g > 0$; The initial X_0 of MFGs satisfies $\mathbb{E}[X_0^2] < \infty$.

(A2) In addition to (A1), the initial $x^{(N)} = (x_1^{(N)}, x_2^{(N)}, \ldots, x_N^{(N)})$ of the N-player game is a vector of i.i.d. random variables in $\Omega^{(N)}$ with the same distribution as the initial $\mathcal{L}(X_0)$ of MFG.

Our objective of this chapter is to understand the Nash equilibrium of MFGs and its connection to the N-player game equilibrium:

(P1) With Assumption (A0),(A1), and (A2), obtain the convergence rate of $(\hat{X}_{1t}^{(N)}, Y^{(N)})$ from the N-player game of Definition 1.2.2 to $(\hat{X}_t, Y)$ from MFGs of Definition 1.2.1 in distribution.

To answer (P1), it is critical to have a solid understanding of the joint distribution $(\hat{X}_t, Y)$ for the underlying MFG, which yields another question:

(P2) With Assumption (A0) and (A1), characterize the MFG equilibrium path $\hat{X}$, as well as associated equilibrium measure $\hat{m}$ along the Definition 1.2.1.

For the main result, we first answer (P2) via the following Riccati system for unknowns $(a_y, b_y, c_y, k_y : y \in \mathcal{Y})$:

$$\begin{cases} a_y'(t) + 2\tilde{b}_{1y}(t)a_y(t) - 2\tilde{b}_{2y}^2(t)a_y^2(t) + \sum_{i=1}^{\kappa} q_{y,i}a_i(t) + h_y(t) = 0, \\ b_y'(t) + \left(2\tilde{b}_{1y}(t) - 4\tilde{b}_{2y}^2(t)a_y(t)\right) b_y(t) + \sum_{i=1}^{\kappa} q_{y,i}b_i(t) + h_y(t) = 0, \\ c_y'(t) + a_y(t) + b_y(t) + \sum_{i=1}^{\kappa} q_{y,i}c_i(t) = 0, \\ k_y'(t) - 2\tilde{b}_{2y}^2(t)a_y^2(t) + 4\tilde{b}_{2y}^2(t)a_y(t)b_y(t) + 2\tilde{b}_{1y}(t)k_y(t) + \sum_{i=1}^{\kappa} q_{y,i}k_i(t) = 0, \\ a_y(T) = b_y(T) = g_y \ , \ c_y(T) = k_y(T) = 0, \end{cases} \tag{1.12}$$

where $h_y(t) = h(y,t)$, $g_y = g(y)$ for $y \in \mathcal{Y}$.

Theorem 1.2.1 (MFG equilibrium)**.** *Under assumptions (A0)-(A1), there exists a unique solution $(a_y, b_y, c_y, k_y : y \in \mathcal{Y})$ to the Riccati system (1.12). With these solutions, the MFG equilibrium path $\hat{X} = \hat{X}[\hat{m}]$ is given by*

$$d\hat{X}_t = \left(\tilde{b}_1(Y_t,t)\hat{X}_t - 2\tilde{b}_2^2(Y_t,t)a_{Y_t}(t)\left(\hat{X}_t - \hat{\mu}_t\right)\right)dt + dW_t, \quad \hat{X}_0 = X_0, \tag{1.13}$$

with equilibrium control

$$\hat{\alpha}_t = -2\tilde{b}_2(Y_t,t)a_{Y_t}(t)\left(\hat{X}_t - \hat{\mu}_t\right), \tag{1.14}$$

where

$$d\hat{\mu}_t = \tilde{b}_1(Y_t,t)\hat{\mu}_t dt, \quad \hat{\mu}_0 = \mathbb{E}[X_0].$$

Moreover, the value function U is

$$U(m_0,y,x) = a_y(0)x^2 - 2a_y(0)x[m_0]_1 + k_y(0)[m_0]_1^2 + b_y(0)[m_0]_2 + c_y(0), \quad y \in \mathcal{Y}.$$

Theorem 1.2.2 (Convergence rate)**.** *Under assumptions (A0)-(A1)-(A2), the joint law $(\hat{X}_{1t}^{(N)}, Y_t^{(N)})$ of the N-player game converges in distribution to that of the MFG equilibrium $(\hat{X}_t, Y_t)$ for any $t \in (0,T]$ at the convergence rate*

$$\mathbb{W}_2\left(\mathcal{L}(\hat{X}_{1t}^{(N)}, Y_t^{(N)}), \mathcal{L}(\hat{X}_t, Y_t)\right) = O\left(N^{-\frac{1}{2}}\right), \ \text{as } N \to \infty.$$

1.2.4 Remarks on the main results

One can interpret the main results in plain words: For N-player game with dynamic (1.7) and cost structure (1.8) for large N, the equilibrium control of the generic player can be effectively approximated by steering itself toward the population center $\hat{\mu}_t$ depending only on the function $\tilde{b}_1(\cdot,\cdot)$ and the entire past of the common noise, whose velocity is dependent on only the function $\tilde{b}_2(\cdot,\cdot)$ and the entire past of the common noise. The effectiveness can be quantified by the convergence rate of $1/2$ for the one-dimensional Mean Field Game under LQ structure and CTMC common noise. A natural question is whether the convergence rate can be generalized to more general settings.

This chapter focuses on the one-dimensional problem to avoid unnecessary symbol complexity. Therefore, the main convergence rate $1/2$ still holds for multidimensional problems using the same coupling procedure. For convenience to check, we summarized the computation involved in multidimensional problems in the Appendix A.5.

The current coupling procedure can also be adapted with suitable modifications to the LQ Mean Field Game problem with Brownian common noise, see [35]. In particular, the reduction of the $O(N^3)$-dimensional ODE can be conducted similarly and the convergence rate is still maintained as $1/2$. However, the dependence of the mean and variance process on the common noise and subsequent calculations are significantly different from the current chapter, see Definition 4 of [35] for more details.

Indeed, choosing the CTMC common noise instead of Brownian motion does not simplify the underlying problem, since it preserves the path-dependence feature of the equilibrium measure. On the contrary, the advantage of CTMC common noise is that the applications aim to model less frequently changing environment settings, such as government policies implemented by multiple different regimes. Due to its realistic applications, stochastic control theory perturbed by CTMC is extensively studied in the context of hybrid control problems, see books [45, 65] and the references therein.

We close this chapter with a remark on the uniqueness. The uniqueness of Mean Field Game can be achieved under Lasry-Lions monotonicity [39] or displacement monotonicity [19] and our setting in Section 1.2.2 satisfies the displacement monotonicity. Thus the convergence of Theorem 1.2.2 implies that the unique equilibrium path of N-player game converges to the unique equilibrium path of the limiting MFG, which is characterized by Theorem 1.2.1.

1.3 Main Result of MFG problem

This section is devoted to the proof of the first main result Theorem 1.2.1 on the MFG solution. First, we outline the scheme based on the Markovian structure of the equilibrium by reformulating the MFG problem in Subsection 1.3.1. Next, we solve the underlying control problem in Subsection 1.3.2 and provide the corresponding Riccati system. Finally, Subsection 1.3.3 proves Theorem 1.2.1 by checking the fixed point condition of MFG problem.

1.3.1 Overview

By Definition 1.2.2, to solve for the equilibrium measure, one shall search the infinite-dimensional space of the random measure flows $m : (0,T] \times \Omega \mapsto \mathcal{P}_2(\mathbb{R})$ until a measure flow satisfies the fixed point condition $m_t = \mathcal{L}(\hat{X}_t|Y), \forall t \in (0,T]$, see Figure 1.1, which requires to check the following infinitely many conditions:

$$[m_t]_k = \mathbb{E}[\hat{X}_t^k|Y], \quad \forall k = 1, 2, \ldots,$$

if they exist.

The first observation is that the cost functions F and G in (1.10)-(1.11) are dependent on the measure m only via the first two moments:

$$F(y,x,t,m) = h(y,t)(x^2 - 2x[m]_1 + [m]_2),$$
$$G(y,x,m) = g(y)(x^2 - 2x[m]_1 + [m]_2).$$

Therefore, the underlying stochastic control problem for MFGs can be entirely determined by the input given by $\mathbb{R}^2$ valued random process $\mu_t = [m_t]_1$ and $\nu_t = [m_t]_2$, which implies that the fixed point condition can be effectively reduced to check two conditions only:

$$\mu_t = \mathbb{E}[\hat{X}_t|Y], \ \nu_t = \mathbb{E}[\hat{X}_t^2|Y].$$

This observation effectively reduces our search from the space of random measure-valued processes $m : (0,T] \times \Omega \mapsto \mathcal{P}_2(\mathbb{R})$ to the space of $\mathbb{R}^2$-valued random processes $(\mu,\nu) : (0,T] \times \Omega \mapsto \mathbb{R}^2$.

Note that, if underlying MFGs have no common noise Y, then (μ,ν) is a deterministic mapping $[0,T] \mapsto \mathbb{R}^2$ and the above observation is enough to reduce the original infinite-dimensional MFGs into a finite-dimensional system. However, the following example shows that this is not the case for MFGs with a common noise and it becomes the main drawback to characterizing MFGs via a finite-dimensional system.

Example 1.3.1. *To illustrate, we consider the following uncontrolled mean field dynamics. Let the mean field term $\mu_t := \mathbb{E}[\hat{X}_t|Y]$, where the underlying dynamic is given by*

$$d\hat{X}_t = -\mu_t Y_t dt + dW_t.$$

- μ_t *is path dependent on Y, i.e.*

$$\mu_t = \mu_0 \exp\Big\{ - \int_0^t Y_s ds \Big\}.$$

  This implies that no finite-dimensional system is possible to characterize the process μ_t, since the $(t, Y) \mapsto \mu_t$ is a function on an infinite dimensional domain.

- μ_t *is Markovian, i.e.*

$$d\mu_t = -Y_t \mu_t dt.$$

  It might be possible to characterize μ_t via a function $(t, Y_t, \mu_t) \mapsto \frac{d\mu_t}{dt}$ on a finite dimensional domain.

To solidify the above idea, we need to postulate the Markovian structure for the first and second moments of the MFG equilibrium. More precisely, our search for the equilibrium will be confined to the space $\mathcal{M}$ of measure flows whose first and second moment exhibits Markovian structure.

Definition 1.3.1 (Confined searching space $\mathcal{M}$). *The space $\mathcal{M}$ is the collection of all $\mathcal{F}_t^Y$-adapted measure flows $m : [0,T] \times \Omega \mapsto \mathcal{P}_2(\mathbb{R})$, whose first moment $[m_t]_1 := \mu_t$ and second moment $[m_t]_2 := \nu_t$ satisfy*

$$\begin{aligned} \mu_t &= \mu_0 + \int_0^t \left(w_0(Y_s, s)\mu_s + w_1(Y_s, s)\right) ds, \\ \nu_t &= \nu_0 + \int_0^t \Big(w_2(Y_s, s)\mu_s + w_3(Y_s, s)\nu_s + w_4(Y_s, s)\mu_s^2 + w_5(Y_s, s)\Big) ds, \end{aligned} \tag{1.15}$$

for all $t \in [0,T]$ and for some smooth deterministic functions $(w_i : i = 0, 1, \ldots, 5)$.

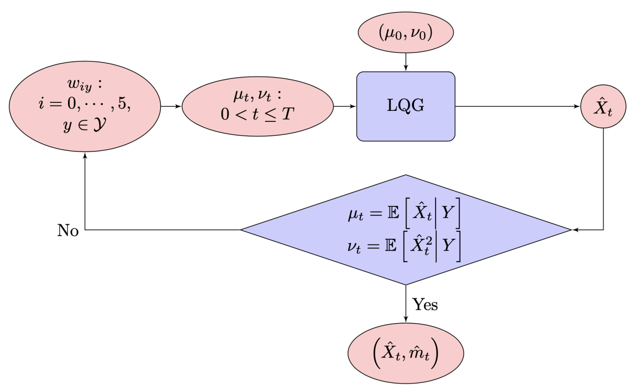


Figure 1.2: Equivalent MFGs diagram: finite dimensional fixed point condition with $\mu_0 = [m_0]_1$ and $\nu_0 = [m_0]_2$.

The flowchart for our equilibrium is depicted in Figure 1.2. Subsection 1.3.2 covers the derivation of the Riccati system for the LQG system with a given population measure flow $m \in \mathcal{M}$, which provides the key building block to MFGs. In Subsection 1.3.3, we check the fixed point condition and provide a finite-dimensional characterization of MFGs, which gives the first main result Theorem 1.2.1.

1.3.2 Generic player's control problem given a population measure

The advantage of the generic player's control problem associated with $m \in \mathcal{M}$ is that its optimal path can be characterized via the following classical stochastic control problem:

- (P3) Given smooth functions $w = (w_i : i = 0, 1, \dots, 5)$, find the optimal value $\bar{V} = \bar{V}[w]$

$$\begin{aligned}\bar{V}(y, x, t, \bar{\mu}, \bar{v}) \quad &= \inf_{\alpha \in \mathcal{A}} \mathbb{E}\left[\int_t^T \left(\frac{1}{2}\alpha_s^2 + \bar{F}(Y_s, X_s, s, \mu_s, \nu_s)\right) ds \right. \\ &\qquad \left. + \bar{G}(Y_T, X_T, \mu_T, \nu_T) \middle| Y_t = y, X_t = x, \mu_t = \bar{\mu}, \nu_t = \bar{\nu}\right]\end{aligned}$$

underlying $\mathbb{R}^4$-valued processes (Y, X, μ, ν) defined through (1.1)-(1.3)-(1.15) with the finite dimensional cost functions: $\bar{F} : \mathbb{R}^2 \times [0, T] \times \mathbb{R}^2 \mapsto \mathbb{R}, \bar{G} : \mathbb{R}^4 \mapsto \mathbb{R}$ given by

$$\bar{F}(y, x, t, \bar{\mu}, \bar{\nu}) = h(y, t)(x^2 - 2x\bar{\mu} + \bar{\nu}),$$

$$\bar{G}(y, x, \bar{\mu}, \bar{\nu}) = g(y)(x^2 - 2x\bar{\mu} + \bar{\nu}),$$

where $\bar{\mu}, \bar{\nu}$ are scalars, while μ, ν are used as processes.

Lemma 1.3.1. *Given $m \in \mathcal{M}$ associated with $w = (w_i : i = 0, 1, \dots, 5)$, the player's value* (1.4) *under assumption (A1) is*

$$U[m_0](y, x) = \bar{V}(y, x, 0, [m_0]_1, [m_0]_2),$$

and the optimal control has a feedback form

$$\hat{\alpha}_t = \bar{\alpha}(Y_t, X_t, t, \mu_t, \nu_t)$$

underlying the processes (Y, X, μ, ν) defined through (1.1)-(1.3)-(1.15)*, whenever there exists a feedback optimal control $\bar{\alpha}$ for the problem (P3).*

Proof. Due to the quadratic cost structure in (1.10)-(1.11), we have enough regularity to all concerned value functions, and the details are omitted. □

Next, we turn to the solution to the control problem (P3).

HJB equation

For the simplicity of notations, for each $i \in \{0, 1, 2, 3, 4, 5\}$ and $y \in \mathcal{Y}$, denote the function $(x, t, \bar{\mu}, \bar{\nu}) \mapsto v(y, x, t, \bar{\mu}, \bar{\nu})$ as v_y, and denote $t \mapsto w_i(y, t)$ as w_{iy}. We apply similar notations for other functions whenever they have a variable $y \in \mathcal{Y}$. Formally, under enough regularity conditions, the value function $\bar{V}$ defined in (P3) is the solution v of the following coupled HJBs

$$\begin{cases}\partial_t v_y + \tilde{b}_{1y} x \partial_x v_y - \dfrac{1}{2}\left(\tilde{b}_{2y}\partial_x v_y\right)^2 + \dfrac{1}{2}\partial_{xx} v_y + \partial_\mu v_y \left(w_{0y}\bar{\mu} + w_{1y}\right) + \\ \qquad \partial_\nu v_y \left(w_{2y}\bar{\mu} + w_{3y}\bar{\nu} + w_{4y}\bar{\mu}^2 + w_{5y}\right) + \displaystyle\sum_{i=1}^{\kappa} q_{y,i} v_i + \bar{F}_y = 0, \\ v_y(x, T, \mu_T, \nu_T) = \bar{G}_y(x, \mu_T, \nu_T), \ \ y \in \mathcal{Y}.\end{cases} \tag{1.16}$$

Furthermore, the optimal control has to admit the feedback form of

$$\hat{\alpha}(t) = -\tilde{b}_2(Y_t, t)\partial_x v(Y_t, \hat{X}_t, t, \mu_t, \nu_t). \tag{1.17}$$

Next, we identify what conditions are needed for equating the control problem (P3) and HJB equation. Denote

$$\mathbb{S} = \left\{ v \in C^\infty : \begin{array}{c} (1+|x|^2)^{-1}(|v| + |\partial_t v|) + \\ (1+|x|)^{-1}(|\partial_x v| + |\partial_\mu v| + |\partial_\nu v|) + |\partial_{xx} v| < K, \\ \forall (y, x, t, \mu, \nu), \text{ for some } K \end{array} \right\}.$$

Lemma 1.3.2 (Verification theorem)**.** *Consider the control problem (P3) with some given smooth functions w. Suppose there exists a solution $v \in \mathbb{S}$ of (1.16). Then, $v_y(x, t, \bar{\mu}, \bar{\nu}) = \bar{V}(y, x, t, \bar{\mu}, \bar{\nu})$ holds, and an optimal control is provided by (1.17).*

Proof. We first prove the verification theorem. Since $v \in \mathbb{S}$, for any admissible $\alpha \in L^4_{\mathbb{F}}$, the process X^α is well defined and one can use Dynkin's formula given by Lemma A.2.1 to write

$$\mathbb{E}\left[v(Y_T, X_T, T, \mu_T, \nu_T)\right] = v(y, x, t, \bar{\mu}, \bar{\nu}) + \mathbb{E}\left[\int_t^T \mathcal{G}^{\alpha(s)} v(Y_s, X_s, s, \mu_s, \nu_s) ds\right],$$

where

$$\mathcal{G}^a f(y, x, s, \bar{\mu}, \bar{\nu}) = \Big(\partial_t + \left(\tilde{b}_{1y} x + \tilde{b}_{2y} a\right)\partial_x + \frac{1}{2}\partial_{xx} + \mathcal{Q} + \left(w_{0y}\bar{\mu} + w_{1y}\right)\partial_{\bar{\mu}} + \\ \left(w_{2y}\bar{\mu} + w_{3y}\bar{\nu} + w_{4y}\bar{\mu}^2 + w_{5y}\right)\partial_{\bar{\nu}}\Big) f(y, x, s, \bar{\mu}, \bar{\nu}).$$

Note that HJB actually implies that

$$\inf_a \left\{\mathcal{G}^a v + \frac{1}{2}a^2\right\} = -\bar{F},$$

which again implies

$$-\mathcal{G}^a v \leq \frac{1}{2}a^2 + \bar{F}.$$

Hence, we obtain that for all $\alpha \in L^4_{\mathbb{F}}$,

$$\begin{aligned} & v(y, x, t, \bar{\mu}, \bar{\nu}) \\ = & \mathbb{E}\left[\int_t^T -\mathcal{G}^{\alpha(s)} v(Y_s, X_s, s, \mu_s, \nu_s) ds\right] + \mathbb{E}\left[v(Y_T, X_T, T, \mu_T, \nu_T)\right] \\ \leq & \mathbb{E}\left[\int_t^T \left(\frac{1}{2}\alpha^2(s) + \bar{F}(Y_s, X_s, s, \mu_s, \nu_s)\right) ds\right] + \mathbb{E}\left[\bar{G}(Y_T, X_T, \mu_T, \nu_T)\right] \\ = & J(y, x, t, \alpha, \bar{\mu}, \bar{\nu}). \end{aligned}$$

In the above, if α is replaced by $\hat{\alpha}$ given by the feedback form (1.17), then since $\partial_x v$ is

Lipschitz continuous in x, there exists corresponding optimal path $\hat{X} \in L^4_{\mathbb{F}}$. Thus, $\hat{\alpha}$ is also in $L^4_{\mathbb{F}}$. One can repeat all above steps by replacing X and α by $\hat{X}$ and $\hat{\alpha}$, and $\leq$ sign by $=$ sign to conclude that v is indeed the optimal value. □

LQG solution

Note that, the costs $\bar{F}$ and $\bar{G}$ of (P3) are quadratic functions in $(x, \bar{\mu}, \bar{\nu})$, while the drift function of the process ν of (1.15) is not linear in $(x, \bar{\mu}, \bar{\nu})$. Therefore, the control problem (P3) does not fall into the standard LQG control framework. Nevertheless, similar to the LQG solution, we guess the value function as a quadratic function in the form of

$$v_y(x, t, \bar{\mu}, \bar{\nu}) = a_y(t)x^2 + d_y(t)x + e_y(t)\bar{\mu} + f_y(t)x\bar{\mu} + k_y(t)\bar{\mu}^2 + b_y(t)\bar{\nu} + c_y(t), \quad y \in \mathcal{Y}. \tag{1.18}$$

With the above setup, for $t \in [0, T]$, the optimal control is

$$\hat{\alpha}_t = -\tilde{b}_2(Y_t, t)\partial_x v(Y_t, \hat{X}_t, t, \mu_t, \nu_t) = -\tilde{b}_2(Y_t, t)\left(2a_{Y_t}(t)\hat{X}_t + d_{Y_t}(t) + f_{Y_t}(t)\mu_t\right), \tag{1.19}$$

and the optimal path $\hat{X}$ is

$$d\hat{X}_t = \left(\tilde{b}_1(Y_t, t)\hat{X}_t - \tilde{b}_2^2(Y_t, t)\left(2a_{Y_t}(t)\hat{X}_t + d_{Y_t}(t) + f_{Y_t}(t)\mu_t\right)\right)dt + dW_t. \tag{1.20}$$

Denote the following ODE systems for $y \in \mathcal{Y}$,

$$\begin{cases} a_y'(t) + 2\tilde{b}_{1y}(t)a_y(t) - 2\tilde{b}_{2y}^2(t)a_y^2(t) + \sum_{i=1}^{\kappa} q_{y,i}a_i(t) + h_y(t) = 0, \\ d_y'(t) + \tilde{b}_{1y}(t)d_y(t) - 2\tilde{b}_{2y}^2(t)a_y(t)d_y(t) + f_y(t)w_{1y}(t) + \sum_{i=1}^{\kappa} q_{y,i}d_i(t) = 0, \\ e_y'(t) - \tilde{b}_{2y}^2(t)d_y(t)f_y(t) + 2k_y(t)w_{1y}(t) + e_y(t)w_{0y}(t) + b_y(t)w_{2y}(t) + \\ \qquad\qquad \sum_{i=1}^{\kappa} q_{y,i}e_i(t) = 0, \\ f_y'(t) + \tilde{b}_{1y}(t)f_y(t) - 2\tilde{b}_{2y}^2(t)a_y(t)f_y(t) + f_y(t)w_{0y}(t) + \sum_{i=1}^{\kappa} q_{y,i}f_i(t) - 2h_y(t) = 0, \\ k_y'(t) - \frac{1}{2}\tilde{b}_{2y}^2(t)f_y^2(t) + 2k_y(t)w_{0y}(t) + b_y(t)w_{4y}(t) + \sum_{i=1}^{\kappa} q_{y,i}k_i(t) = 0, \\ b_y'(t) + b_y(t)w_{3y}(t) + \sum_{i=1}^{\kappa} q_{y,i}b_i(t) + h_y(t) = 0, \\ c_y'(t) + a_y(t) - \frac{1}{2}\tilde{b}_{2y}^2 d_y^2(t) + e_y(t)w_{1y} + b_y(t)w_{5y} + \sum_{i=1}^{\kappa} q_{y,i}c_i(t) = 0, \end{cases} \tag{1.21}$$

with terminal conditions

$$a_y(T) = g_y,\ b_y(T) = g_y,\ c_y(T) = 0,\ d_y(T) = 0,\ e_y(T) = 0,\ f_y(T) = -2g_y,\ k_y(T) = 0. \tag{1.22}$$

Lemma 1.3.3. *Suppose there exists a unique solution $(a_y, b_y, c_y, d_y, e_y, f_y, k_y : y \in \mathcal{Y})$ to the ODE system (1.21)-(1.22) on $[0, T]$. Then the value function of (P3) is*

$$\begin{aligned} &\bar{V}(y, x, t, \bar{\mu}, \bar{\nu}) = v_y(x, t, \bar{\mu}, \bar{\nu}) \\ &= a_y(t)x^2 + d_y(t)x + e_y(t)\bar{\mu} + f_y(t)x\bar{\mu} + k_y(t)\bar{\mu}^2 + b_y(t)\bar{\nu} + c_y(t) \end{aligned} \tag{1.23}$$

for $y \in \mathcal{Y}$ and the optimal control and optimal path are given by (1.19) and (1.20), respectively.

Proof. With the form of value function v_y given in (1.18) and the first and second moment of the conditional population density given in (1.15), we have

$$\begin{aligned} \partial_t v_y &= a'_y(t)x^2 + d'_y(t)x + e'_y(t)\bar{\mu} + f'_y(t)x\bar{\mu} + k'_y(t)\bar{\mu}^2 + b'_y(t)\bar{\nu} + c'_y(t), \\ \partial_x v_y &= 2xa_y(t) + d_y(t) + f_y(t)\bar{\mu}, \\ \partial_{xx} v_y &= 2a_y(t), \\ \partial_{\bar{\mu}} v_y &= e_y(t) + f_y(t)x + 2k_y(t)\bar{\mu}, \\ \partial_{\bar{\nu}} v_y &= b_y(t), \end{aligned}$$

for $y \in \mathcal{Y}$. Plugging them back to the coupled HJBs in (1.16), we get a system of ODEs in (1.21) by equating x, $\bar{\mu}$, $\bar{\nu}$-like terms in each equation.

Therefore, any solution $(a_y, b_y, c_y, d_y, e_y, f_y, k_y : y \in \mathcal{Y})$ of ODE system (1.21) leads to the solution of HJB (1.16) in the form of the quadratic function given by (1.23). Since the $(a_y, b_y, c_y, d_y, e_y, f_y, k_y : y \in \mathcal{Y})$ are differentiable functions on the closed set $[0, T]$, they are also bounded, and the function v meets regularity conditions required by Lemma 1.3.2 to conclude the desired result. □

1.3.3 Proof of Main Theorem 1.2.1

Going back to the ODE system (1.21), there are 7κ equations, while we have total 13κ deterministic functions of $[0, T] \times \mathbb{R}$ to be determined to characterize MFGs. Those are

$$(a_y, b_y, c_y, d_y, e_y, f_y, k_y : y \in \mathcal{Y}) \text{ and } (w_{iy} : i = 0, 1, \dots 5,\ y \in \mathcal{Y}).$$

In the following, we identify the missing 6κ equations by checking the fixed point condition:

$$\mu_s = \mathbb{E}\left[\hat{X}_s \middle| Y\right],\ \nu_s = \mathbb{E}\left[\hat{X}_s^2 \middle| Y\right], \quad \forall s \in [0, T], \tag{1.24}$$

where μ and ν are two auxiliary processes $(\mu,\nu)[w]$ defined in (1.15), see Figure 1.2. This leads to a complete characterization of the equilibrium for the MFG posed by (P2).

Note that based on the dynamic of the optimal $\hat{X}$ defined in (1.20), the fixed point condition (1.24) implies that the first moment $\hat{\mu}_s := \mathbb{E}\left[\hat{X}_s \middle| Y\right]$ and the second moment $\hat{\nu}_s := \mathbb{E}\left[\hat{X}_s^2 \middle| Y\right]$ of the optimal path conditioned on Y satisfy

$$\begin{cases} \hat{\mu}_s = \bar{\mu} + \int_t^s \Big(\Big(\tilde{b}_1(Y_r,r) - \tilde{b}_2^2(Y_r,r)\left(2a_{Y_r}(r) + f_{Y_r}(r)\right)\Big)\hat{\mu}_r - \tilde{b}_2^2(Y_r,r)d_{Y_r}(r)\Big)\,dr, \\ \hat{\nu}_s = \bar{\nu} + \int_t^s \Big(1 + 2\tilde{b}_1(Y_r,r)\hat{\nu}_r - \tilde{b}_2^2(Y_r,r)\left(4a_{Y_r}(r)\hat{\nu}_r + 2d_{Y_r}(r)\hat{\mu}_r + 2f_{Y_r}(r)\hat{\mu}_r^2\right)\Big)\,dr, \end{cases} \tag{1.25}$$

for $s \geq t$. Note that under the optimal control in (1.19), comparing the terms in (1.15) and (1.25), we obtain another 6κ equations:

$$\begin{aligned} &w_{0y} = \tilde{b}_{1y} - 2\tilde{b}_{2y}^2 a_y - \tilde{b}_{2y}^2 f_y,\ w_{1y} = -\tilde{b}_{2y}^2 d_y,\ w_{2y} = -2\tilde{b}_{2y}^2 d_y, \\ &w_{3y} = -4\tilde{b}_{2y}^2 a_y + 2\tilde{b}_{1y},\ w_{4y} = -2\tilde{b}_{2y}^2 f_y,\ w_{5y} = 1, \end{aligned} \tag{1.26}$$

for $y \in \mathcal{Y}$. Using further algebraic structures, one can reduce the ODE system of 13κ equations composed by (1.21) and (1.26) into a system of 4κ equations of the form (1.12) for the MFG characterization in Theorem 1.2.1.

Proof of Theorem 1.2.1. Since a_y $(y \in \mathcal{Y})$ has the same expressions as (1.12), its existence, uniqueness and boundedness are shown in Lemma A.4.3. Given a_y $(y \in \mathcal{Y})$ and smooth bounded w's,

$$(b_y, d_y, e_y, f_y : y \in \mathcal{Y})$$

is a coupled linear system, and their existence, uniqueness and boundedness is shown by Theorem 12.1 in [2]. Similarly, given $(b_y, d_y, f_y : y \in \mathcal{Y})$, $(k_y, c_y : y \in \mathcal{Y})$ is a linear system, and their existence and uniqueness is also guaranteed by Theorem 12.1 in [2].

The ODE system (1.21) can be rewritten by

$$
\begin{cases}
a_y'(t) + 2\tilde{b}_{1y}a_y(t) - 2\tilde{b}_{2y}^2 a_y^2(t) + \sum_{i=1}^{\kappa} q_{y,i}a_i(t) + h_y(t) = 0, \\
d_y'(t) + \tilde{b}_{1y}d_y(t) - 2\tilde{b}_{2y}^2 a_y(t)d_y(t) - \tilde{b}_{2y}^2 f_y(t)d_y(t) + \sum_{i=1}^{\kappa} q_{y,i}d_i(t) = 0, \\
e_y'(t) - \tilde{b}_{2y}^2 d_y(t)f_y(t) - 2\tilde{b}_{2y}^2 k_y(t)d_y(t) + e_y(t)\left(\tilde{b}_{1y} - 2\tilde{b}_{2y}^2 a_y(t) - \tilde{b}_{2y}^2 f_y(t)\right) - 2\tilde{b}_{2y}^2 b_y(t)d_y(t) + \\
\qquad \sum_{i=1}^{\kappa} q_{y,i}e_i(t) = 0, \\
f_y'(t) + \tilde{b}_{1y}f_y(t) - 2\tilde{b}_{2y}^2 a_y(t)f_y(t) + f_y(t)\left(\tilde{b}_{1y} - 2\tilde{b}_{2y}^2 a_y(t) - \tilde{b}_{2y}^2 f_y(t)\right) + \sum_{i=1}^{\kappa} q_{y,i}f_i(t) - 2h_y(t) \\
\qquad = 0, \\
k_y'(t) - \frac{1}{2}\tilde{b}_{2y}^2 f_y^2(t) + 2k_y(t)\left(\tilde{b}_{1y} - 2\tilde{b}_{2y}^2 a_y(t) - \tilde{b}_{2y}^2 f_y(t)\right) - 2\tilde{b}_{2y}^2 b_y(t)f_y(t) + \sum_{i=1}^{\kappa} q_{y,i}k_i(t) = 0, \\
b_y'(t) + b_y(t)\left(-4\tilde{b}_{2y}^2 a_y(t) + 2\tilde{b}_{1y}\right) + \sum_{i=1}^{\kappa} q_{y,i}b_i(t) + h_y(t) = 0, \\
c_y'(t) + a_y(t) - \frac{1}{2}\tilde{b}_{2y}^2 d_y^2(t) - 2\tilde{b}_{2y}^2 d_y(t)e_y(t) + b_y(t) + \sum_{i=1}^{\kappa} q_{y,i}c_i(t) = 0,
\end{cases}
$$

with the terminal conditions

$$a_y(T) = g_y,\ b_y(T) = g_y,\ c_y(T) = 0,\ d_y(T) = 0,\ e_y(T) = 0,\ f_y(T) = -2g_y,\ k_y(T) = 0.$$

Since a_y, b_y $(y \in \mathcal{Y})$ has the same expressions as (1.12), its existence, uniqueness, and boundedness are shown in Lemma A.4.3. Meanwhile, with the given $(a_y, b_y : y \in \mathcal{Y})$, we denote $l_y = 2a_y + f_y$, and then

$$l_y'(t) + 2\tilde{b}_{1y}l_y(t) - \tilde{b}_{2y}^2 l_y^2(t) + \sum_{i=1}^{\kappa} q_{y,i}l_i(t) = 0\ ,\ l_y(T) = 0.$$

By Lemma A.4.1 and Lemma A.4.2 in Appendix, there exists a unique solution for l_y $(y \in \mathcal{Y})$, which is $l_y(t) = 0, y \in \mathcal{Y}$. This gives $f_y(t) = -2a_y(t)$ and $d_y'(t) + \tilde{b}_{1y}d_y(t) + \sum_{i=1}^{\kappa} q_{y,i}d_i(t) = 0$, which implies $d_y(t) = 0, y \in \mathcal{Y}$. Then, the equation for e_y can be simplified as $e_y'(t) + \tilde{b}_{1y}e_y(t) + \sum_{i=1}^{\kappa} q_{y,i}e_i(t) = 0$, which indicates that $e_y(t) = 0, y \in \mathcal{Y}$. For k_y, c_y, with the given of $(a_y, b_y : y \in \mathcal{Y})$, we have

$$k_y'(t) + 2\tilde{b}_{1y}k_y(t) - 2\tilde{b}_{2y}^2 a_y^2(t) + 4\tilde{b}_{2y}^2 a_y(t)b_y(t) + \sum_{i=1}^{\kappa} q_{y,i}k_i(t) = 0\ ,\ k_y(T) = 0,$$

$$c_y'(t) + a_y(t) + b_y(t) + \sum_{i=1}^{\kappa} q_{y,i}c_i(t) = 0\ ,\ c_y(T) = 0.$$

The existence and uniqueness of the solution for k_y, c_y ($y \in \mathcal{Y}$) are yielded by Theorem 12.1 in [2].

Note that in this case, since $2a_y + f_y = 0$ and $d_y = 0$ for $y \in \mathcal{Y}$, from (1.25) we have

$$\hat{\mu}_s = \bar{\mu} + \int_t^s \tilde{b}_1(Y_r, r)\hat{\mu}_r \, dr$$

for all $s \in [t, T]$. Then

$$\hat{\nu}_s = \bar{\nu} + \int_t^s \left(1 + 2\tilde{b}_1(Y_r, r)\hat{\nu}_r - 4\tilde{b}_2^2(Y_r, r)a_{Y_r}(r)\hat{\nu}_r + 4\tilde{b}_2^2(Y_r, r)a_{Y_r}(r)\hat{\mu}_r^2\right) dr.$$

Plugging $d_y = 0$ for $y \in \mathcal{Y}$ back to (1.19), we obtain the optimal control by

$$\hat{\alpha}_s = -2\tilde{b}_2^2(Y_s, s)a_{Y_s}(s)\left(\hat{X}_s - \hat{\mu}_s\right).$$

Since we have $d_y = 0$ for $y \in \mathcal{Y}$, the value function can be simplified from (1.18) to

$$v_y(x, t, \bar{\mu}, \bar{\nu}) = a_y(t)x^2 - 2a_y(t)x\bar{\mu} + k_y(t)\bar{\mu}^2 + b_y(t)\bar{\nu} + c_y(t).$$

By the equivalence Lemma 1.3.1, it yields the value function U of Theorem 1.2.1. Moreover, since $f_y = -2a_y$ and $k_y \neq 0$, the ODE system (1.21) together with (1.26) can be reduced into (1.12). From the Lemma A.4.3, the existence and uniqueness of $(a_y, b_y, c_y, k_y : y \in \mathcal{Y})$ in (1.12) is guaranteed.

□

1.4 Main Result of Convergence of N-player Game to MFGs

In this section, we show the convergence of the N-player game to MFGs. To simplify the presentation, we may omit the superscript (N) for the processes in the probability space $\Omega^{(N)}$, whenever there is no confusion.

First, we solve the N-player game in Subsection 1.4.1, which provides a Riccati system consisting of $O(N^3)$ equations. Subsection 1.4.2 reduces the corresponding Riccati system into an ODE system whose dimension is independent of N. This becomes the key building block of the convergence rate obtained in Subsection 1.4.3. To obtain the convergence rate, Subsection 1.4.3 provides an explicit embedding of some processes in $\Omega^{(N)}$ into the probability space Ω. Note that, $\Omega^{(N)}$ is much richer than Ω since $\Omega^{(N)}$ contains N Brownian motions while Ω has only two Brownian motions. Therefore, careful treatment has to be carried out to some processes of our interest, otherwise, such an embedding is in general implausible.

1.4.1 Characterization of N-player game by Riccati system

The N-player game is indeed an N-coupled stochastic LQG problem by its very own definition, see Subsection 1.2.2. Therefore, the solution can be derived via Riccati system with the existing LQG theory given below: For $i = 1, 2, \dots, N$, $y \in \mathcal{Y}$,

$$\begin{cases} A'_{iy} + 2\tilde{b}_{1y} e_i e_i^\top A_{iy} - 2\tilde{b}_{2y}^2 A_{iy}^\top e_i e_i^\top A_{iy} + \sum_{j \neq i}^{N} \left(2\tilde{b}_{1y} e_j e_j^\top A_{iy} - 4\tilde{b}_{2y}^2 A_{jy}^\top e_j e_j^\top A_{iy} \right) + \\ \qquad \sum_{j=1}^{\kappa} q_{y,j} A_{ij} + \frac{h_y}{N} \sum_{j \neq i}^{N} (e_i - e_j)(e_i - e_j)^\top = 0, \\ B'_{iy} + \sum_{j \neq i}^{N} \left(\tilde{b}_{1y} e_j e_j^\top B_{iy} - 2\tilde{b}_{2y}^2 A_{iy}^\top e_j e_j^\top B_{jy} - 2\tilde{b}_{2y}^2 A_{jy}^\top e_j e_j^\top B_{iy} \right) + \\ \qquad \tilde{b}_{1y} e_i e_i^\top B_{iy} - 2\tilde{b}_{2y}^2 A_{iy}^\top e_i e_i^\top B_{iy} + \sum_{j=1}^{\kappa} q_{y,j} B_{ij} = 0, \\ C'_{iy} - \frac{1}{2}\tilde{b}_{2y}^2 B_{iy}^\top e_i e_i^\top B_{iy} - \sum_{j \neq i}^{N} \tilde{b}_{2y}^2 B_{jy}^\top e_j e_j^\top B_{iy} + \sum_{j=1}^{N} tr(A_{jy}) + \sum_{j=1}^{\kappa} q_{y,j} C_{ij} = 0, \\ A_{iy}(T) = \frac{g_y}{N} \Lambda_i, \ B_{iy}(T) = 0 \cdot \mathbb{1}_N, \ C_{iy}(T) = 0, \end{cases} \tag{1.27}$$

where the solutions consist of $N \times N$ symmetric matrices A_{iy}'s, N-dimensional vectors B_{iy}'s, and $C_{iy} \in \mathbb{R}$. In the above, $\mathbb{1}_N$ is the N-dimensional vector with all entries are 1, Λ_i's are $N \times N$ matrices with diagonal 1 except $(\Lambda_i)_{ii} = N - 1$, $(\Lambda_i)_{ij} = (\Lambda_i)_{ji} = -1$ for any $j \neq i$ and the rest entries as 0, and e_i's are the N-dimensional natural basis.

Lemma 1.4.1. *Suppose $(A_{iy}, B_{iy}, C_{iy} : i = 1, 2, \dots, N, \ y \in \mathcal{Y})$ is the solution of (1.27). Then, the value functions of N-player game defined by (1.9) are*

$$V_i(y, x^{(N)}) = (x^{(N)})^\top A_{iy}(0) x^{(N)} + (x^{(N)})^\top B_{iy}(0) + C_{iy}(0), \quad i = 1, 2, \dots, N.$$

Moreover, the path and the control under the equilibrium are

$$d\hat{X}_{it} = \left(\tilde{b}_1(Y_t, t)\hat{X}_{it} - \tilde{b}_2^2(Y_t, t) \left(2(A_{iY_t})_i^\top \hat{X}_t + (B_{iY_t})_i \right) \right) dt + dW_{it}, \quad i = 1, 2, \dots, N, \tag{1.28}$$

and

$$\hat{\alpha}_{it} = -\tilde{b}_2(Y_t, t) \left(2(A_{iY_t})_i^\top \hat{X}_t + (B_{iY_t})_i \right),$$

where $(A)_i$ denotes the i-th column of matrix A, $(B)_i$ denotes the i-th entry of vector B and $\hat{X}_t = [\hat{X}_{1t}, \hat{X}_{2t}, \dots, \hat{X}_{Nt}]^\top$.

Proof. It is standard that, under enough regularities, the value function of the N-player game $V(y, x^{(N)}) = (V_1, V_2, \dots, V_N)(y, x^{(N)})$ can be lifted to the solution $v_{iy}(x^{(N)}, t)$ of the following system of HJB equations, for $i = 1, 2, \ldots, N$ and $y \in \mathcal{Y}$,

$$\begin{cases} \partial_t v_{iy} + \tilde{b}_{1y} x_i \partial_i v_{iy} - \frac{1}{2}\left(\tilde{b}_{2y}\partial_i v_{iy}\right)^2 + \sum_{j\neq i}^{N}\left(\tilde{b}_{1y}x_j - \tilde{b}_{2y}^2\partial_j v_{jy}\right)\partial_j v_{iy} + \\ \qquad \frac{1}{2}\Delta v_{iy} + \sum_{j=1}^{\kappa} q_{y,j} v_{ij} + \frac{h_y}{N}\sum_{j\neq i}^{N}\left((e_i - e_j)^\top x^{(N)}\right)^2 = 0, \\ v_{iy}(x^{(N)}, T) = \frac{g_y}{N}\sum_{j\neq i}^{N}\left((e_i - e_j)^\top x^{(N)}\right)^2. \end{cases} \tag{1.29}$$

Then, the value functions V of N-player game defined by (1.9) is $V_i(y, x^{(N)}) = v_{iy}(x^{(N)}, 0)$ for all $i = 1, 2, \ldots, N$. Moreover, the path and the control under the equilibrium are

$$d\hat{X}_{it} = \left(\tilde{b}_1(Y_t, t)\hat{X}_{it} - \tilde{b}_2^2(Y_t, t)\partial_i v_{iY_t}(\hat{X}_t, t)\right) dt + dW_{it}, \quad i = 1, 2, \ldots, N,$$

and

$$\hat{\alpha}_{it} = -\tilde{b}_2(Y_t, t)\partial_i v_{iY_t}(\hat{X}_t, t).$$

The proof is the application of Dynkin's formula and the details are omitted here. Due to its LQG structure, the value function leads to a quadratic function of the form

$$v_{iy}(x^{(N)}, t) = (x^{(N)})^\top A_{iy}(t) x^{(N)} + (x^{(N)})^\top B_{iy}(t) + C_{iy}(t).$$

For each $i = 1, 2, \ldots, N$, after plugging V_{iy} into (1.29), and matching the coefficient of variables, we get the desired results. □

1.4.2 Reduced Riccati form for the equilibrium

So far, the N-player game and MFG have been characterized by Lemma 1.4.1 and Theorem 1.2.1, respectively. One of our main objectives is to investigate the convergence of the generic optimal path $\hat{X}_{1t}^{(N)}$ of N-player game generated (1.27)-(1.28) to the optimal path $\hat{X}_t$ of MFG generated by (1.12)-(1.13).

Note that $\hat{X}_t$ relies only on κ functions $(a_y : y \in \mathcal{Y})$ from the simple ODE system (1.12) while $\rho(\hat{X}_t^{(N)})$ depends on $O(N^3)$ functions from $(A_{iy} : i = 1, 2, \ldots, N,\ y \in \mathcal{Y})$ solved from a huge Riccati system (1.27). Therefore, it is almost a hopeless task for a meaningful comparison between these two processes without gaining further insight into the complex structure of the Riccati system (1.27).

To proceed, let us first observe some hidden patterns from a numerical result for the solution of Riccati (1.27). The following matrix shows A_{20} at $t = 1$ for $N = 5$ with the same parameters as in Figure 1.3 and Figure 1.4 in Section 1.5.1:

$$A_{20}(1) = \begin{bmatrix} 0.1319 & -0.1924 & 0.0202 & 0.0202 & 0.0202 \\ -0.1924 & 0.7696 & -0.1924 & -0.1924 & -0.1924 \\ 0.0202 & -0.1924 & 0.1319 & 0.0202 & 0.0202 \\ 0.0202 & -0.1924 & 0.0202 & 0.1319 & 0.0202 \\ 0.0202 & -0.1924 & 0.0202 & 0.0202 & 0.1319 \end{bmatrix}.$$

Interestingly enough, we observe that the entire 25 entries of $A_{20}(1)$ indeed consist of 4 distinct values. Moreover, similar computation with different values of N only yields a larger table depending on N, but always consists of 4 values. Inspired by this accidental discovery from the above numerical example, we may want to believe and prove a pattern of the matrix A_{iy} in the following form:

$$(A_{iy})_{pq} = \begin{cases} a_{1y}(t), & \text{if } p = q = i, \\ a_{2y}(t), & \text{if } p = q \neq i, \\ a_{3y}(t), & \text{if } p \neq q, p = i \text{ or } q = i, \\ a_{4y}(t), & \text{otherwise,} \end{cases} \tag{1.30}$$

for $y \in \mathcal{Y}$. The next result justifies the above pattern: the N^2 entries of the matrix A_{iy} can be embedded to a 2κ-dimensional vector space no matter how big N is.

Lemma 1.4.2. *There exists a unique solution (a^N_{1y}, a^N_{2y}) from the ODE system(1.31)*

$$\begin{cases} a'_{1y}(t) + 2\tilde{b}_{1y}a_{1y}(t) - \dfrac{2(N+1)}{N-1}\tilde{b}^2_{2y}a^2_{1y}(t) + \displaystyle\sum_{j=1}^{\kappa} q_{y,j}a_{1j}(t) + \dfrac{N-1}{N}h_y(t) = 0, \\ a'_{2y}(t) + 2\tilde{b}_{1y}a_{2y}(t) + \dfrac{2}{(N-1)^2}\tilde{b}^2_{2y}a^2_{1y}(t) - \dfrac{4N}{N-1}\tilde{b}^2_{2y}a_{1y}(t)a_{2y}(t) + \displaystyle\sum_{j=1}^{\kappa} q_{y,j}a_{2j}(t) + \dfrac{h_y(t)}{N} = 0, \\ a_{1y}(T) = \dfrac{N-1}{N}g_y, \;\; a_{2y}(T) = \dfrac{g_y}{N}, \end{cases} \tag{1.31}$$

for $y \in \mathcal{Y}$. Moreover, the path and the control of player i under the equilibrium are

$$d\hat{X}^{(N)}_{it} = \left(\tilde{b}_1(Y^{(N)}_t, t)\hat{X}^{(N)}_{it} - 2\tilde{b}^2_2(Y^{(N)}_t, t)a^N_{1Y^{(N)}_t}(t)\left(\hat{X}^{(N)}_{it} - \frac{1}{N-1}\sum_{j\neq i}^{N}\hat{X}^{(N)}_{jt}\right)\right)dt + dW^{(N)}_{it}, \tag{1.32}$$

and

$$\hat{\alpha}^{(N)}_{it} = -2\tilde{b}_2(Y^{(N)}_t, t)a^N_{1Y^{(N)}_t}(t)\left(\hat{X}^{(N)}_{it} - \frac{1}{N-1}\sum_{j\neq i}^{N}\hat{X}^{(N)}_{jt}\right)$$

for $i = 1, 2, \ldots, N$.

Proof. It is obvious to see that in the Riccati system (1.27), $B_{iy} = 0$ for all $i = 1, 2, \ldots, N$

and $y \in \mathcal{Y}$. Note that in this case, for $i = 1, 2, \dots, N$, the optimal control is given by

$$\hat{\alpha}_i^{(N)} = -2\tilde{b}_2(Y_t^{(N)}, t) \sum_{j=1}^{N} (A_{iY_t^{(N)}})_{ij} \hat{X}_{jt}^{(N)} = -2\tilde{b}_2(Y_t^{(N)}, t) \left(A_{iY_t^{(N)}}\right)_i^\top \hat{X}_t^{(N)}.$$

Plugging the pattern (1.30) into the differential equation of A_{iy}, we have

$$a'_{1y} + 2\tilde{b}_{1y}a_{1y} - 2\tilde{b}_{2y}^2 a_{1y}^2 - 4(N-1)\tilde{b}_{2y}^2 a_{3y}^2 + \sum_{j=1}^{\kappa} q_{y,j}a_{1j} + \frac{N-1}{N}h_y = 0,$$

$$a'_{2y} + 2\tilde{b}_{1y}a_{2y} - 2\tilde{b}_{2y}^2 a_{3y}^2 - 4\tilde{b}_{2y}^2 \left(a_{1y}a_{2y} + (N-2)a_{3y}a_{4y}\right) + \sum_{j=1}^{\kappa} q_{y,j}a_{2j} + \frac{h_y}{N} = 0,$$

$$a'_{3y} + 2\tilde{b}_{1y}a_{3y} - 2\tilde{b}_{2y}^2 a_{1y}a_{3y} - 4\tilde{b}_{2y}^2 \left(a_{1y}a_{3y} + (N-2)a_{3y}^2\right) + \sum_{j=1}^{\kappa} q_{y,j}a_{3j} - \frac{h_y}{N} = 0,$$

$$a'_{3y} + 2\tilde{b}_{1y}a_{3y} - 2\tilde{b}_{2y}^2 a_{1y}a_{3y} - 4\tilde{b}_{2y}^2 \left(a_{2y}a_{3y} + (N-2)a_{3y}a_{4y}\right) + \sum_{j=1}^{\kappa} q_{y,j}a_{3j} - \frac{h_y}{N} = 0,$$

$$a'_{4y} + 2\tilde{b}_{1y}a_{4y} - 2\tilde{b}_{2y}^2 a_{3y}^2 - 4\tilde{b}_{2y}^2 \left(a_{2y}a_{3y} + a_{1y}a_{4y} + (N-3)a_{3y}a_{4y}\right) + \sum_{j=1}^{\kappa} q_{y,j}a_{4j} = 0,$$

which gives $a_{1y} + (N-2)a_{3y} = a_{2y} + (N-2)a_{4y}$ since two expressions for a_{3y} should be identical. This implies that $(a_{1y} + (N-2)a_{3y})' = (a_{2y} + (N-2)a_{4y})'$ or

$$\begin{aligned}
&-2\tilde{b}_{1y}a_{1y} + 2\tilde{b}_{2y}^2 a_{1y}^2 + 4(N-1)\tilde{b}_{2y}^2 a_{3y}^2 - \frac{N-1}{N}h_y - \sum_{j=1}^{\kappa} q_{y,j}a_{1j} \\
&+ (N-2)\left(-2\tilde{b}_{1y}a_{3y} + 2\tilde{b}_{2y}^2 a_{1y}a_{3y} + 4\tilde{b}_{2y}^2 \left(a_{2y}a_{3y} + (N-2)a_{3y}a_{4y}\right) - \sum_{j=1}^{\kappa} q_{y,j}a_{3j} + \frac{h_y}{N}\right) \\
= &-2\tilde{b}_{1y}a_{2y} + 2\tilde{b}_{2y}^2 a_{3y}^2 + 4\tilde{b}_{2y}^2 \left(a_{1y}a_{2y} + (N-2)a_{3y}a_{4y}\right) - \sum_{j=1}^{\kappa} q_{y,j}a_{2j} - \frac{h_y}{N} \\
&+ (N-2)\left(-2\tilde{b}_{1y}a_{4y} + 2\tilde{b}_{2y}^2 a_{3y}^2 + 4\tilde{b}_{2y}^2 \left(a_{1y}a_{4y} + a_{2y}a_{3y} + (N-3)a_{3y}a_{4y}\right) - \sum_{j=1}^{\kappa} q_{y,j}a_{4j}\right).
\end{aligned}$$

After combining terms and substituting $a_{2y} + (N-2)a_{4y}$ with $a_{1y} + (N-2)a_{3y}$, we get $a_{1y}^2 + (N-2)a_{1y}a_{3y} - (N-1)a_{3y}^2 = 0$, which yields $a_{3y} = a_{1y}$ or $a_{3y} = -\frac{1}{N-1}a_{1y}$. Note that $a_{3y} \neq a_{1y}$ due to their different differential equations. Hence, we can conclude that

$a_{3y} = -\frac{1}{N-1} a_{1y}$. In conclusion, for $i = 1, 2, \dots, N$, A_{iy} ($y \in \mathcal{Y}$) has the following expressions:

$$(A_{iy})_{pq} = \begin{cases} a_{1y}(t), & \text{if } p = q = i, \\ a_{2y}(t), & \text{if } p = q \neq i, \\ -\frac{1}{N-1} a_{1y}(t), & \text{if } p \neq q, p = i \text{ or } q = i, \\ \frac{1}{(N-1)(N-2)} a_{1y}(t) - \frac{1}{N-2} a_{2y}(t), & \text{otherwise.} \end{cases}$$

The existence and uniqueness of (1.27) is equivalent to the existence and uniqueness of (1.31). For a_{1y}, the existence and uniqueness can be deduced from Lemma A.4.1 and A.4.2. Given a_{1y}'s, a_{2y}'s are linear equations, thus their existence and uniqueness are guaranteed by Theorem 12.1 in [2]. Together with previous discussions, we conclude the results. □

1.4.3 Proof of convergence rate

Based on the current progress, let us reiterate our goal (P1) for convergence. Our objective is the convergence of the joint distribution $\mathcal{L}(\hat{X}_{1t}^{(N)}, Y_t^{(N)})$ of N-player game generated (1.31)-(1.32) in the probability space $\Omega^{(N)}$ to the distribution $\mathcal{L}(\hat{X}_t, Y_t)$ of MFG generated by (1.12)-(1.13) in the probability space Ω. More precisely, we want to find a number $\eta > 0$ satisfying

$$\mathbb{W}_2 \left(\mathcal{L}(\hat{X}_{1t}^{(N)}, Y_t^{(N)}), \mathcal{L}(\hat{X}_t, Y_t) \right) = O\left(N^{-\eta} \right), \tag{1.33}$$

where $\mathbb{W}_2$ is the 2-Wasserstein metric. This procedure is given in the following two steps:

1. We will construct a process Z^N in the probability space Ω, who provides exact copy of the joint distribution in the sense of

   $$\mathcal{L}(\hat{X}_{1t}^{(N)}, Y_t^{(N)}) = \mathcal{L}(Z^N, Y).$$

   Note that, the (1.32) shows that $\hat{X}_{1t}^{(N)}$ correlates to N many Brownian motions $\{W_i^{(N)} : i = 1, 2, \dots, N\}$ from a much richer space $\Omega^{(N)}$ while Ω is a much smaller space having only two Brownian motions W and B. Therefore, such an embedding essentially requires to represent $\hat{X}_{1t}^{(N)}$ by two independent Brownian motions and is in general not possible. However, due to the symmetric structure of MFG (or the nature of the mean-field effect), the embedding is possible and the details are provided in Lemma 1.4.3.

2. By Proposition 1.2.1, we can use distribution copy (Z^N, Y) in Ω to write

   $$\mathbb{W}_2^2 \left(\mathcal{L}(\hat{X}_{1t}^{(N)}, Y_t^{(N)}), \mathcal{L}(\hat{X}_t, Y_t) \right) \leq \mathbb{E}\left[\left| Z_t^N - \hat{X}_t \right|^2 \right]. \tag{1.34}$$

   To obtain the estimate of the above right-hand side, we shall compare the (1.35) of Z^N and (1.13) of $\hat{X}$, and it becomes essential to obtain the convergence rate of the ODE system (1.31) towards the ODE system (1.12). The details are provided in Lemma 1.4.4.

Lemma 1.4.3. *Let $\{X_0^i : i \in \mathbb{N}\}$ be i.i.d. random variables in Ω independent to (W, B, Y) with $X_0^1 = X_0$. Let Z^N be the solution of*

$$Z_t^N = X_0 + \int_0^t \tilde{b}_1(Y_s, s) Z_s^N ds - \int_0^t 2\tilde{b}_2^2(Y_s, s)\hat{a}_{1Y_s}^N(s)\left(Z_s^N - \bar{X}_s^N\right) ds + W_t, \tag{1.35}$$

where

$$d\bar{X}_t^N = \tilde{b}_1(Y_t, t)\bar{X}_t^N dt + \frac{\sqrt{N-1}}{N} dB_t + \frac{1}{N} dW_t, \quad \bar{X}_0^N = \frac{1}{N}\sum_{i=1}^N X_0^i,$$

and

$$\hat{a}_{1y}^N = \frac{N}{N-1} a_{1y}^N,$$

where a_{1y}^N is from the ODE system(1.31). Then, (Z_t^N, Y_t) in $(\Omega, \mathcal{F}_T, \mathbb{P})$ has the same distribution as $(\hat{X}_{1t}^{(N)}, Y_t^{(N)})$ in $(\Omega^{(N)}, \mathcal{F}_T^{(N)}, \mathbb{P}^{(N)})$.

Proof. Continued from the Lemma 1.4.2, player i's path in the N-player game follows

$$\hat{X}_{it}^{(N)} = x_i^{(N)} + \int_0^t \tilde{b}_1(Y_s^{(N)}, s)\hat{X}_{is}^{(N)} ds - \int_0^t 2\tilde{b}_2^2(Y_s^{(N)}, s) a_{1Y_s^{(N)}}^N(s)\left(\hat{X}_{is}^{(N)} - \frac{1}{N-1}\sum_{j\neq i}^N \hat{X}_{js}^{(N)}\right) ds + W_{it}^{(N)}.$$

With the notation

$$\bar{X}_s^{(N)} = \frac{1}{N}\sum_{i=1}^N \hat{X}_{is}^{(N)},$$

one can rewrite the path by

$$\hat{X}_{it}^{(N)} = x_i^{(N)} + \int_0^t \tilde{b}_1(Y_s^{(N)}, s)\hat{X}_{is}^{(N)} ds - \int_0^t 2\tilde{b}_2^2(Y_s^{(N)}, s)\hat{a}_{1Y_s^{(N)}}^N(s)\left(\hat{X}_{is}^{(N)} - \bar{X}_s^{(N)}\right) ds + W_{it}^{(N)}. \tag{1.36}$$

By adding up the above equations (1.36) indexed by $i = 1$ to N, one can have

$$\begin{aligned} \bar{X}_t^{(N)} &= \bar{x}^{(N)} + \int_0^t \tilde{b}_1(Y_s^{(N)}, s)\bar{X}_s^{(N)} ds + \frac{1}{N}\sum_{i=1}^N W_{it}^{(N)} \\ &= \bar{x}^{(N)} + \int_0^t \tilde{b}_1(Y_s^{(N)}, s)\bar{X}_s^{(N)} ds + \frac{\sqrt{N-1}}{N}\left(\sqrt{N-1}\bar{W}_{-it}^{(N)}\right) + \frac{1}{N} W_{it}^{(N)}, \end{aligned} \tag{1.37}$$

where $\bar{W}_{-it}^{(N)} := \frac{1}{N-1}\sum_{j\neq i} W_{jt}^{(N)}$.

Next, we define solution maps of (1.36) and (1.37):

$$\bar{G}_t(x, \phi, W_1, W_2) = \mathcal{E}_t(\phi)\left(x + \int_0^t \mathcal{E}_s(-\phi) d(W_{1s} + W_{2s})\right) \tag{1.38}$$

and

$$G_t(x,\phi_1,\phi_2,\phi_3,W) = x\mathcal{E}_t(\phi_1-\phi_2) + \mathcal{E}_t(\phi_1-\phi_2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2)\left(\phi_2(s)\phi_3(s)ds + dW_s\right), \tag{1.39}$$

where

$$\mathcal{E}_t(\phi) = \exp\left\{\int_0^t \phi_s ds\right\}.$$

Now, we can rewrite $\bar{X}_t^{(N)}$ of (1.37) and $\hat{X}_{1t}^{(N)}$ of (1.36) as

$$\bar{X}_t^{(N)} = \bar{G}_t\left(\frac{1}{N}\sum_{i=1}^N x_i^{(N)}, \tilde{b}_1(Y_\cdot^{(N)},\cdot), \frac{\sqrt{N-1}}{N}\left(\sqrt{N-1}\bar{W}_{-1}^{(N)}\right), \frac{1}{N}W_1^{(N)}\right),$$

and

$$\hat{X}_{1t}^{(N)} = G_t\left(x_1^{(N)}, \tilde{b}_1(Y_\cdot^{(N)},\cdot), 2\tilde{b}_2(Y_\cdot^{(N)},\cdot)\hat{a}_1^N(Y_\cdot^{(N)},\cdot), \bar{X}^{(N)}(\cdot), W_1^{(N)}\right)$$

Meanwhile, $(Z^N, \bar{X}^N)$ of (1.35) can also be written in the form of

$$\bar{X}_t^N = \bar{G}_t\left(\frac{1}{N}\sum_{i=1}^N X_0^i, \tilde{b}_1(Y_\cdot,\cdot), \frac{\sqrt{N-1}}{N}B, \frac{1}{N}W\right),$$

and

$$Z_t^N = G_t\left(X_0, \tilde{b}_1(Y_\cdot,\cdot), 2\tilde{b}_2(Y_\cdot,\cdot)\hat{a}_1^N(Y_\cdot,\cdot), \bar{X}^N(\cdot), W\right) \tag{1.40}$$

Finally, the fact that the distribution of (Z^N, Y) in the space Ω is identical distribution to $(\hat{X}_1^{(N)}, Y^{(N)})$ in $\Omega^{(N)}$ comes from the followings:

- $\tilde{b}_1, \tilde{b}_2, \hat{a}_1^N$ are deterministic functions.
- The random processes $(\sqrt{N-1}\bar{W}_{-1}^{(N)}, W_1^{(N)}, Y^{(N)})$ are independent mutually in $\Omega^{(N)}$, while the random elements (B, W, Y) are also independent triples. Moreover, two random triples have identical joint distributions.
- Initial states are generated from identical joint distributions $\{x_i^{(N)} : i = 1, 2, \ldots, N\}$ and $\{X_0^i : i = 1, 2, \ldots, N\}$.

Therefore, (Z^N, Y) and $(\hat{X}_1^{(N)}, Y^{(N)})$ have the same distributions. This completes the proof. □

In view of (1.34), we shall estimate the second moment $\mathbb{E}\left[\left|Z_t^N - \hat{X}_t\right|^2\right]$. First, we can rewrite $\hat{X}$ of (1.13) using above representations via G_t:

$$\hat{X}_t = G_t\left(X_0, \tilde{b}_1(Y_\cdot,\cdot), 2\tilde{b}_2(Y_\cdot,\cdot)a(Y_\cdot,\cdot), \hat{\mu}(\cdot), W\right),$$

which leads to a better comparison with Z^N in the form of (1.40). To proceed, the following properties of G_t are useful for the estimate of the second moment, whose proof is relegated to the Appendix A.3. Throughout the proof of the next lemma, we will use K in various places as a generic constant which varies from line to line.

Lemma 1.4.4. *The convergence rate under the Wasserstein metric $\mathbb{W}_2(\cdot,\cdot)$ is*

$$\mathbb{W}_2\left(\mathcal{L}(\hat{X}_{1t}^{(N)}, Y_t^{(N)}), \mathcal{L}(\hat{X}_t, Y_t)\right) = O\left(N^{-\frac{1}{2}}\right).$$

Proof. In view of (1.34), we start with

$$\begin{aligned}
&\mathbb{W}_2^2\left(\mathcal{L}(\hat{X}_{1t}^{(N)}, Y_t^{(N)}), \mathcal{L}(\hat{X}_t, Y_t)\right) \le \mathbb{E}\left[\left|Z_t^N - \hat{X}_t\right|^2\right] \\
=&\mathbb{E}\left[\left|G_t\left(X_0, \tilde{b}_1(Y_\cdot,\cdot), 2\tilde{b}_2(Y_\cdot,\cdot)\hat{a}_1^N(Y_\cdot,\cdot), \bar{X}^N(\cdot), W\right) - G_t\left(X_0, \tilde{b}_1(Y_\cdot,\cdot), 2\tilde{b}_2(Y_\cdot,\cdot)a(Y_\cdot,\cdot), \hat{\mu}(\cdot), W\right)\right|^2\right] \\
:=&\mathbb{E}\left[|I_1(t) - I_2(t)|\right].
\end{aligned}$$

Applying the Lipschitz continuity of $(\phi_2, \phi_3) \mapsto G_t(x, \phi_1, \phi_2, \phi_3, W)$ by Appendix A.3 on the conditional expectation $\mathbb{E}\left[|I_1(t) - I_2(t)| \,\middle|\, Y\right]$, we have

$$\begin{aligned}
\mathbb{E}|Z_t^N - \hat{X}_t|^2 &\le K\mathbb{E}\left[\sup_{0\le t\le T}\left(2\tilde{b}_2(Y_t,t)\hat{a}_{1Y_t}^N(t) - 2\tilde{b}_2(Y_t,t)a_{Y_t}(t)\right)^2 + \sup_{0\le t\le T}\left(\bar{X}^N(t) - \hat{\mu}(t)\right)^2\right] \\
&\le K\mathbb{E}\left[\sup_{0\le t\le T}\left|\tilde{b}_2(Y_t,t)\right|^2 \sup_{0\le t\le T}\left|\hat{a}_{1Y_t}^N(t) - a_{Y_t}(t)\right|^2 + \sup_{0\le t\le T}\left|\bar{X}^N(t) - \hat{\mu}(t)\right|^2\right] \\
&\le K\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{a}_{1Y_t}^N(t) - a_{Y_t}(t)\right|^2 + \sup_{0\le t\le T}\left|\bar{X}^N(t) - \hat{\mu}(t)\right|^2\right]
\end{aligned}$$

From the dynamic of $\bar{X}^N$ and $\hat{\mu}$,

$$\begin{cases}
d\left(\bar{X}_t^N - \hat{\mu}_t\right) = \tilde{b}_1(Y_t,t)\left(\bar{X}_t^N - \hat{\mu}_t\right)dt + \dfrac{\sqrt{N-1}}{N}dB_t + \dfrac{1}{N}dW_t, \\
\bar{X}_0^N - \hat{\mu}_0 = \dfrac{1}{N}\displaystyle\sum_{i=1}^N X_0^i - \hat{\mu}_0,
\end{cases}$$

which can be written in terms of $\bar{G}_t$ of (1.38):

$$\bar{X}^N(t) - \hat{\mu}(t) = \bar{G}_t\left(\frac{1}{N}\sum_{i=1}^N X_0^i - \hat{\mu}_0, \tilde{b}_1(Y_\cdot,\cdot), \frac{\sqrt{N-1}}{N}B, \frac{1}{N}W\right).$$

Using the fact of $\left|\tilde{b}_{1y}\right|_\infty < \infty$ and Ito's isometry, this yields the following estimation:

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\bar{X}^N(t)-\hat{\mu}(t)\right|^2\right] \le K\left(\frac{1}{N}+\mathbb{E}\left|\frac{1}{N}\sum_{i=1}^N X_0^i-\hat{\mu}_0\right|^2\right).$$

Note that, by the central limit theorem, we have

$$N\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^N X_0^i-\hat{\mu}_0\right|^2\right] = \mathbb{E}\left[\left|\frac{\sum_{i=1}^N(X_0^i-\hat{\mu}_0)}{\sqrt{N}}\right|^2\right] \to Var(X_0^1)<\infty, \quad N\to\infty,$$

and we conclude that

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\bar{X}^N(t)-\hat{\mu}(t)\right|^2\right] = O(N^{-1}). \tag{1.41}$$

Next, we investigate the boundness of

$$\sup_{0\le t\le T}\left|\hat{a}^N_{1Y_t}(t)-a_{Y_t}(t)\right|^2.$$

From (1.31) and $\hat{a}^N_{1y}=\frac{N}{N-1}a^N_{1y}$, we have

$$\begin{cases} \left(\hat{a}^N_{1y}\right)'+2\tilde{b}_{1y}\hat{a}^N_{1y}-\dfrac{2(N+1)}{N}\tilde{b}^2_{2y}\left(\hat{a}^N_{1y}\right)^2+\displaystyle\sum_{i=1}^{\kappa}q_{y,i}\hat{a}^N_{1i}+h_y(t)=0 \\ \hat{a}^N_{1y}(T)=g_y. \end{cases}$$

Define $u_y=a_y-\hat{a}^N_{1y}$, let $\tau=T-t$ and denote $u_y(\tau):=u_y(T-t)$, we have

$$\begin{cases} u_y'(\tau)=2\tilde{b}_{1y}(\tau)u_y(\tau)-2\tilde{b}^2_{2y}(\tau)\left(a_y(\tau)+\hat{a}^N_{1y}(\tau)\right)u_y(\tau)+\dfrac{2}{N}\tilde{b}^2_{2y}(\tau)\left(\hat{a}^N_{1y}(\tau)\right)^2+\displaystyle\sum_{i=1}^{\kappa}q_{y,i}u_i(\tau) \\ u_y(0)=0, \end{cases} \tag{1.42}$$

which gives that

$$u_y(\tau)=\int_0^\tau\left(2\tilde{b}_{1y}(s)u_y(s)-2\tilde{b}^2_{2y}(s)\left(a_y(s)+\hat{a}^N_{1y}(s)\right)u_y(s)+\frac{2}{N}\tilde{b}^2_{2y}(s)\left(\hat{a}^N_{1y}(s)\right)^2+\sum_{i=1}^{\kappa}q_{y,i}u_i(s)\right)ds.$$

Thus for $\tau\in[0,T]$,

$$\begin{aligned} |u_y(\tau)| \le \int_0^\tau\Bigg(&2\left|\tilde{b}_{1y}\right|_\infty|u_y(s)|+2\left|\tilde{b}_{2y}\right|^2_\infty\left(|a_y|_\infty+\left|\hat{a}^N_{1y}\right|_\infty\right)|u_y(s)| \\ &+\frac{2}{N}\left|\tilde{b}_{2y}\right|^2_\infty\left|\hat{a}^N_{1y}\right|^2_\infty+\sum_{i=1}^{\kappa}|q_{y,i}||u_i(s)|\Bigg)ds. \end{aligned}$$

Let $\left(\left|\tilde{b}_{1y}\right|_\infty, \left|\tilde{b}_{2y}\right|_\infty, |a_y|_\infty, \left|\hat{a}^N_{1y}\right|_\infty, \sup_{i\in\mathcal{Y}}|q_{y,i}|\right) \le K_1$, then

$$|u_y(\tau)| \le \frac{2}{N}K_1^4T + \int_0^\tau \left(\left(2K_1 + 4K_1^3\right)|u_y(s)| + K_1\sum_{i=1}^\kappa |u_i(s)|\right) ds.$$

By adding up the above equation indexed by $y = 1$ to κ, one can have

$$\sum_{y=1}^\kappa |u_y(\tau)| \le \frac{2\kappa K_1^4 T}{N} + \left(2K_1 + 4K_1^3 + \kappa K_1\right)\int_0^\tau \sum_{y=1}^\kappa |u_y(s)|ds.$$

Let $K_2 = 2\kappa K_1^4 T$ and $K_3 = 2K_1 + 4K_1^3 + \kappa K_1$, by the Grönwall's inequality,

$$\sum_{y=1}^\kappa |u_y(\tau)| \le \frac{K_2}{N}e^{K_3\tau} \le \frac{K_2}{N}e^{K_3 T}, \quad \forall \tau \in [0, T],$$

which implies that

$$\sum_{y=1}^\kappa |u_y(\tau)| \le \frac{K}{N}, \quad \forall \tau \in [0, T].$$

Thus, we have

$$\sup_{0\le t\le T}\left|\hat{a}^N_{1Y_t}(t) - a_{Y_t}(t)\right|^2 \le \frac{K}{N^2}, \text{ almsot surely }. \tag{1.43}$$

Therefore, the convergence is obtained from (1.41) and (1.43):

$$\mathbb{W}_2^2\left(\mathcal{L}(Z_t^N), \mathcal{L}(\hat{X}_t)\right) \le K\mathbb{E}\left[\sup_{0\le t\le T}\left|\hat{a}^N_{1Y_t}(t) - a_{Y_t}(t)\right|^2 + \sup_{0\le t\le T}\left|\bar{X}^N(t) - \hat{\mu}(t)\right|^2\right] = O(N^{-1}).$$

□

1.5 Numerical Simulations

In this section, we present numerical illustrations of Mean Field Games (MFGs) and the convergence of the N-player game towards its limiting MFGs. Subsection 1.5.1 depicts simulations involving the Riccati system, value function, and optimal control of a generic player in MFGs. Subsection 1.5.2 exhibits simulations showcasing the convergence of the N-player game towards MFGs across various values of N.

1.5.1 Simulations of a generic player in MFGs

We have derived a 4κ dimensional Riccati ODE system (1.12) to determine the parameter functions

$$(a_y, b_y, c_y, k_y : y \in \mathcal{Y})$$

needed for the characterization of the equilibrium and the value function. Meanwhile, we also show the solvability of the Riccati ODE system in Section 1.3.

As mentioned earlier, different from the MFG characterization with the common noise, the derived Riccati system is essentially finite-dimensional. In this subsection, we present a numerical experiment and show some numerical results for solving the Riccati system to demonstrate its computational advantages.

For the illustration purpose, assume the finite time horizon is given with $T = 5$, the number of states $\kappa = 2$, and that the coefficients of the dynamic equation are listed below:

$$\begin{aligned}
&\mathcal{Y} = \{0, 1\}, \\
&Q = \begin{bmatrix} -0.5 & 0.5 \\ 0.6 & -0.6 \end{bmatrix}, \\
&\tilde{b}_1 = 0, \ \tilde{b}_2 = 1, \\
&h_0 = 2, \ h_1 = 5, \ g_0 = 3, \ g_1 = 1, \\
&\mu_0 = 0, \nu_0 = 2.
\end{aligned}$$

Firstly, using the forward Euler's method with the step size $\delta = 10^{-2}$, we obtain trajectories of $(a_y, b_y, c_y : y \in \mathcal{Y})$, which is the solution of ODE system (1.12) in Figure 1.3. Next, using the trajectories of the parameter functions and Markov chain Y_t, we achieve the simulations for $\hat{\alpha}_t$ and $\hat{X}_t$ in Figure 1.4.

As shown in the Figure 1.4, people tend to centralize since the conditional second moment of the population density ν_t is always decreasing.

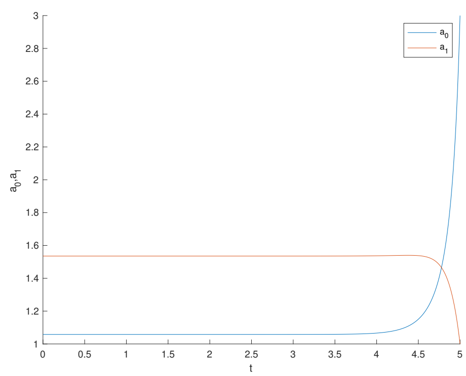
1.5.2 Simulations of convergence of the N-player game to MFGs

In section 1.4, we showed that the generic player's path for the N-player game is convergent to the generic player's path for MFGs. In this subsection, we demonstrate the convergence of the conditional first moment, conditional second moment, and the value functions of the N-player game to the corresponding terms of the generic player in the Mean Field Game setup by using some numerical examples.

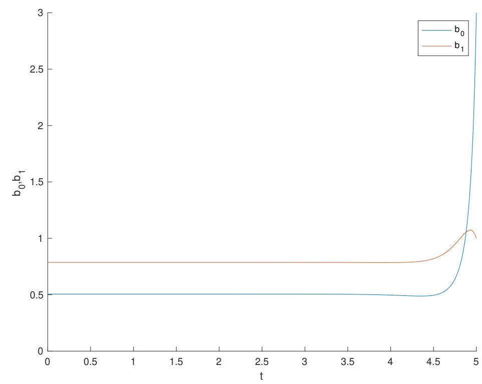
The following figures show the value functions (see Figure 1.7), $\mu^{(N)}$ (see Figure 1.5) and $\nu^{(N)}$ (see Figure 1.6) under $N \in \{10, 20, 50, 100\}$ with the same parameters' settings as in Figure 1.3 and Figure 1.4 in section 1.5.1. We can see the convergence to the solution of the generic player as N gets larger.

1.6 Conclusion and Future Work

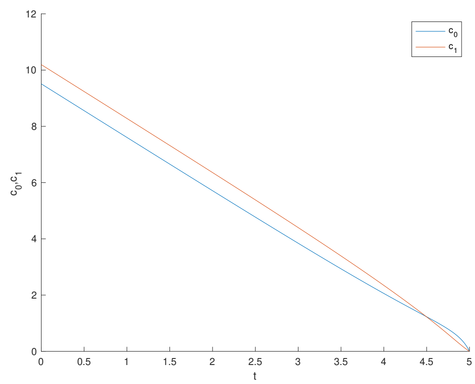
Chapter 1 investigates the convergence rate of the N-player game, governed by a Markov chain common noise, towards its asymptotic MFG under the Linear-Quadratic-Gaussian



(a) $a_y(t)$



(b) $b_y(t)$



(c) $c_y(t)$

Figure 1.3: Simulations for a_y, b_y and c_y, the solution to Riccati system (1.12).

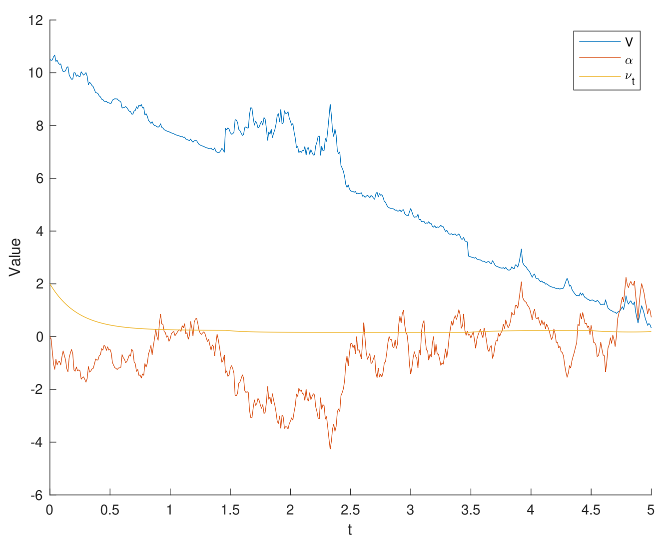


Figure 1.4: Simulations for value function V, optimal control α, and conditional second moment ν

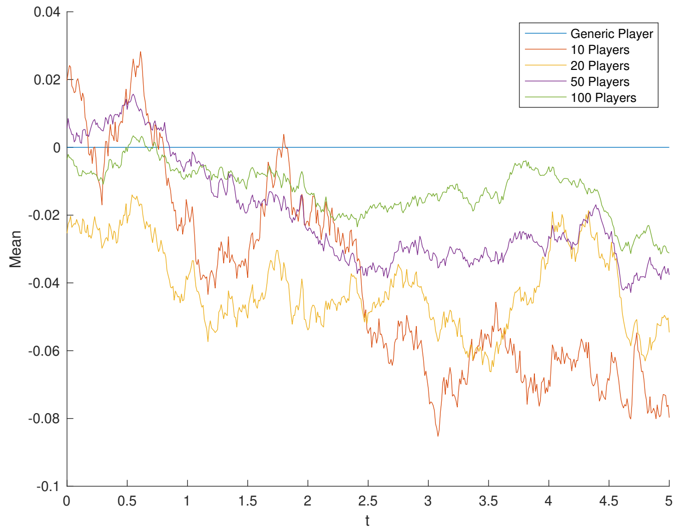


Figure 1.5: μ_t: conditional mean of the population density

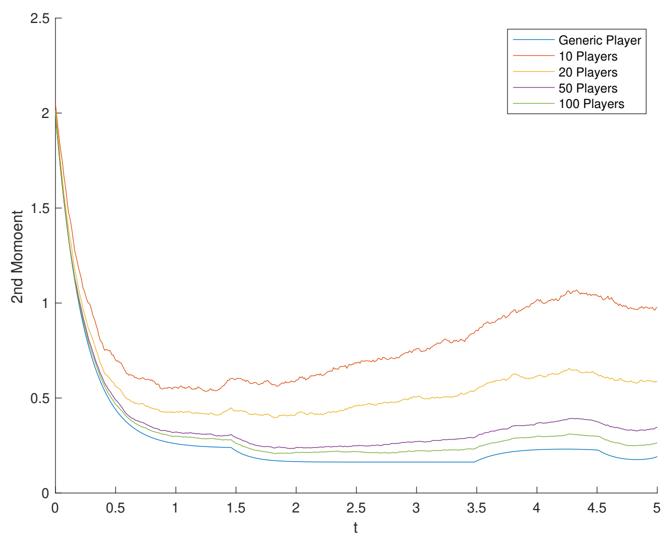


Figure 1.6: ν_t: conditional 2nd moment of the population density

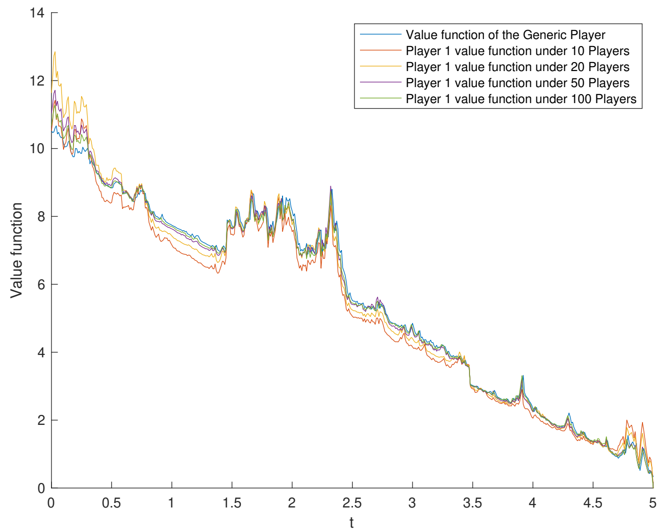


Figure 1.7: Simulation of player 1's optimal value function V.

structure. To achieve this, firstly, we introduce a Markovian structure using two auxiliary processes for the first and second moments of the MFG equilibrium and employ the fixed point condition in MFG. By doing so, we characterize the equilibrium measure in MFG with a finite-dimensional Riccati system of ODEs. Consequently, we obtain the equilibrium path, equilibrium control, and the value function in MFG. Subsequently, we address the N-player game under the LQG structure, and we characterize its equilibrium path, equilibrium control, and the value function through a Riccati system of ODEs with a dimension of $O(N^3)$. Leveraging the N-invariant algebraic structure of this system of ODEs, we establish a dimension reduction result, facilitating a comparison between the equilibrium path $\hat{X}_1^{(N)}$ in the N-player game and the equilibrium path $\hat{X}$ in the MFG.

To demonstrate the convergence between the two equilibrium paths, we embed $\hat{X}_1^{(N)}$ from $\Omega^{(N)}$ to Ω using a distribution copy $Z^N \in \Omega$, leading to the achievement of the convergence result and the computation of the convergence rate. Lastly, some numerical simulations are presented to demonstrate the convergence result.

1.6.1 Future work

Future explorations can encompass broader considerations within the Mean Field Game (MFG) framework. Specifically, extensions could involve incorporating additional complexities, such as time delays and Poisson jumps, into the MFG model. Furthermore, beyond the confines of the Linear Quadratic Gaussian (LQG) structure, investigations into the convergence behavior of MFG with common noise under more generalized structural frameworks are warranted.

Moreover, our study in this project imposes a constraint mandating positive sensitivities in the cost functional. Our findings reveal a noteworthy observation: the absence of a global solution for MFG instances wherein the coefficient of the cost functional assumes a negative value, whereas a global solution is attainable under positive coefficients.

2

Robust Relative Performance for Portfolio Optimization

We conduct an analysis of robust portfolio management problems under competition and relative performance criteria, focusing on portfolio managers, herein referred to as agents, with Constant Relative Risk Aversion (CRRA) utilities trading in log-Gaussian markets within a common investment horizon $[0, T]$. We develop explicit constant robust strategies tailored for finite populations of competitive agents. Additionally, we delve into the concept of information value, which entails contrasting portfolio outcomes of terminal wealth achieved under robust strategies in situations of incomplete information with those attained under Nash equilibrium strategies in situations of complete information.

2.1 Introduction and Literature Review

In incomplete information situations, the robust relative performance problem involves finding the optimal robust strategy $\hat{\pi}_1$ for agent 1 to maximize the robust utility:

$$\sup_{\pi_1 \in \mathcal{A}} \inf_{\pi \in \mathcal{A}^n} J(\pi_1, \pi)$$

over one closed convex admissible strategy set $\mathcal{A}$. In this chapter, we contemplate J as a particular instance of power utility of the Constant Relative Risk Aversion (CRRA) type.

In contrast to scenarios of complete information, where investors possess full knowledge regarding each other's actions, this model operates under the assumption of limited information access. The goal is to hedge against the worst-case scenario by evaluating the adverse outcomes resulting from other investors' actions for every feasible strategy. The optimal robust strategy maximizes utility across all worst-case scenarios.

The concept of the value of information (VOI) plays a crucial role in decision-making under uncertainty, providing a quantitative framework for assessing the utility of acquiring additional information. In various fields such as economics, finance, engineering, and healthcare, decision-makers often face situations where the acquisition of information entails costs but can lead to better-informed decisions and improved outcomes. The value of information captures the incremental benefit gained from reducing uncertainty and making more informed choices.

Robust portfolio optimization is a critical aspect of financial decision-making, particularly in the realm of investment strategies where uncertainty and risk are inherent. Relative performance criteria offer a nuanced perspective on portfolio optimization, focusing not only on absolute returns but also on the performance relative to a benchmark or peer group. In this context, robust relative performance portfolio optimization seeks to develop strategies that are resilient to market fluctuations and uncertainties, ensuring consistent performance regardless of prevailing conditions.

The literature on robust relative performance portfolio management has witnessed significant growth in recent years, driven by the need for adaptive investment strategies in volatile financial markets. Researchers have explored various aspects of this domain, ranging from theoretical frameworks to practical implementations. One prominent line involves the consideration of different utility functions to model investor preferences. Studies by Markowitz (1952) [46] and Sharpe (1964) [57] laid the foundation for modern portfolio theory, which emphasizes the trade-off between risk and return. Building upon this framework, researchers have extended the analysis to incorporate relative performance metrics, as highlighted by the works of Roll (1978) [55] and Grinold and Kahn (1999) [12].

Furthermore, the emergence of behavioral finance has provided valuable insights into investor behavior and decision-making processes. Prospect theory, proposed by Kahneman and Tversky (1979) [36], challenges the traditional assumptions of rationality in financial decision-making and underscores the importance of framing effects and loss aversion. Integrating behavioral insights into robust portfolio management strategies has been a subject of interest for researchers aiming to capture the nuances of investor behavior and sentiment.

In addition to theoretical developments, empirical studies have contributed to the advancement of robust relative performance portfolio management. Empirical research by Daniel et al. (1997) [13] and Fama and French (1993) [16] has provided empirical evidence on the efficacy of various factors in explaining portfolio returns, paving the way for factor-based investing strategies. Moreover, the advent of machine learning and artificial intelligence techniques has enabled researchers to develop sophisticated models for portfolio optimization and risk management, as demonstrated in studies by Gu et al. (2020) [23] and Liang et al. (2024) [11].

The paper of Huang et al. (2010) [26] considers the relative robust conditional value-at-risk portfolio selection problem where the underlying probability distribution of portfolio return is only known to belong to a certain set. They construct a robust portfolio with multiple experts (priors) by solving a sequence of linear programs or a second-order cone program. Georgantas (2021) [21] provides comprehensive empirical assessments of the performance of different types of robust optimization (RO) models based on popular risk measures, using data from the US market during the period 2005–2020. For the optimal portfolio and robust ranking issue, Nguyen and Lo (2012) [53] found a weight vector that maximizes some generic objective function for the worst realization of the ranking.

To this end, the literature on robust relative performance portfolio management underscores the importance of adaptive and resilient investment strategies in navigating dynamic financial markets. By integrating insights from various disciplines, including finance, economics, and behavioral science, researchers aim to develop robust frameworks that can withstand market uncertainties and deliver consistent performance over time.

The literature on the value of information spans multiple disciplines and has been a subject of interest for researchers for decades. Early works by Blackwell (1953) [5] and Lindley (1956) [41] laid the theoretical groundwork for the Bayesian approach to decision theory, which emphasizes the role of prior beliefs and posterior probabilities in quantifying the value of information. These seminal contributions provided foundational insights into the optimal allocation of resources for information acquisition and decision-making under uncertainty.

Subsequent research efforts have extended the value of information framework to various applied domains, including finance, healthcare, environmental science, and engineering. In finance, studies by Hirshleifer and Riley (1992) [25] and Grossman and Stiglitz (1980) [22] have explored the role of information in asset pricing and market efficiency, highlighting the impact of asymmetric information on market dynamics and investor behavior.

In addition to theoretical developments, empirical studies have provided insights into the practical applications of the value of information framework. Research by Yokota and Thompson (2004) [66] has examined decision-making processes in environmental health risk management, demonstrating how the value of information analysis can inform strategic risk management efforts.

Overall, the literature on the value of information underscores its significance as a decision-making tool in various domains. By quantifying the benefits of reducing uncertainty and making more informed choices, the value of information framework offers valuable insights for policymakers, managers, and other decision-makers facing complex and uncertain situations.

2.1.1 Outline

The rest of this chapter is outlined as follows: Section 2.2 presents a precise formulation of the problem. Section 2.3 is devoted to the derivation of our main result: the robust strategy of the relative performance problem. In Section 2.4, we first show the Nash equilibrium strategy under the complete information setting and investigate the value of information. Section 2.5 demonstrates the robust strategy by some numerical examples. The conclusion and some potential future works are summarized in Section 2.6.

2.2 Problem Setup

In this section, we collect the notations and give the research problem.

The finite-population case consists of $(n+1)$ agents trading in a common risk-free asset S_0 and an individual stock asset. For convenience, we may assume that the risk-free rate $r=0$. The common invest horizon is the interval $[0,T]$. Let $(\Omega, \mathcal{F}_T, \mathbb{P}, \mathbb{F}=\{\mathcal{F}_t : 0 \le t \le T\})$ be a filtered probability space satisfying usual conditions, on which $B_t, W_t^1, ..., W_t^{n+1}$ are independent one-dimensional standard Brownian motions.

The price of each stock asset S_i is modeled as a log-Gaussian process driven by two independent Brownian motions. The first Brownian motion B_t is the same for all prices, representing a common market noise, while the second W_t^i is idiosyncratic, specific to each individual stock. Precisely, the i-th agent specializes in stock i, $i=1,...,(n+1)$, whose price S_i is given by

$$\frac{dS_i}{S_i} = \mu_i dt + \nu_i dW_t^i + \sigma_i dB_t.$$

We refer to B_t as the common noise and W_t^i as an idiosyncratic noise.

The investment strategy $\pi_i(t)$ is taken to be the fraction of wealth that agent i invests in the stock S_i at time t, $i=1,...,(n+1)$. The discounted wealth of agent i is given by

$$dX_i = \pi_i X_i(\mu_i dt + \nu_i dW_t^i + \sigma_i dB_t) \tag{2.1}$$

with initial value $X_i^0 = x_i^0$. The class of admissible strategies is the set $\mathcal{A}$ of self-financing $\mathbb{F}$-progressively measurable processes $\pi_i(t)$ satisfying $\mathbb{E}\int_0^T |\pi_i(t)|^2 dt < \infty$.

We aim to analyze the robust optimization for X_1, the controlled wealth process of agent one. The same robust procedure can be executed similarly by any other agent.

Let $\pi = (\pi_2, ...\pi_{n+1})^\top, \mu = (\mu_2, ..., \mu_{n+1})^\top, \nu = (\nu_2, ..., \nu_{n+1})^\top, \sigma = (\sigma_2, ..., \sigma_{n+1})^\top$ be n-dimensional vectors and $M = \text{diag}(\nu_2^2, ..., \nu_{n+1}^2)$, $N = \text{diag}(\sigma_2^2, ..., \sigma_{n+1}^2)$, $\Sigma = (\sigma_i \sigma_j)_{i,j \in \{2,...,n+1\}}$ be $n \times n$-dimensional matrices.

Now we consider the relative performance to the geometric mean of the wealth of all agents except for agent i, $\overline{X}_{-i} = (\prod_{k=1, k\neq i}^{(n+1)} X_k)^{\frac{1}{n}}$, with initial value $\overline{X}_{-i}^0 = y_0$. Let $Z_t = \dfrac{X_1}{\overline{X}_{-1}^\theta}$, whose initial value is z_0, the wealth ratio of the agent one over the geometric mean of the wealth of all other agents with a constant power parameter $\theta \ge 0$. Here, θ serves as a sensitivity parameter, with $\theta > 1$ indicating competition among agents, $0 < \theta < 1$ indicating collaboration, and $\theta = 0$ denoting a scenario of non-relative performance.

In this chapter, we consider the utility a function $U : \mathbb{R}_+ \to \mathbb{R}$ of CRRA utility. Thus the performance functional J is the particular case of power utility of CRRA type,

$$J(\pi_1, \pi) = \mathbb{E}\left[U\left(\frac{X_1(T)}{\overline{X}_{-1}^\theta(T)}\right)\right] = \mathbb{E}[U(Z_T)] := \begin{cases} p^{-1}\mathbb{E}\left[Z_T^p\right], & p \neq 0 \\ \mathbb{E}\left[\log Z_T\right], & p = 0 \end{cases} \tag{2.2}$$

where $p < 1$ denotes a constant parameter, and it is recognized that $1 - p$ represents the relative risk aversion parameter.

Since $\overline{X}_{-1} = (\prod_{i=2}^{(n+1)} X_i)^{\frac{1}{n}}$, by It$\hat{o}$' formula we have in vector form

$$\frac{d\overline{X}_{-1}}{\overline{X}_{-1}} = (\frac{1}{n}\mu^\top\pi + \frac{1-n}{2n^2}\pi^\top M\pi - \frac{1}{2n}\pi^\top N\pi + \frac{1}{2n^2}\pi^\top\Sigma\pi)dt + \frac{1}{n}\sum_{i=2}^{(n+1)}\nu_i\pi_i dW^i + \frac{1}{n}\sigma^\top\pi dB. \quad (2.3)$$

Next we consider $Z_t = \frac{X_1}{\overline{X}_{-1}^\theta}$ with $\theta \geq 0$, applying It$\hat{o}$' formula to function $f(x,y) = \frac{x}{y^\theta}$, therefore

$$\begin{aligned}\frac{dZ_t}{Z_t} &= [\mu_1\pi_1 - \frac{\theta}{n}\mu^\top\pi - \frac{(1-n)\theta}{2n^2}\pi^\top M\pi + \frac{\theta}{2n}\pi^\top N\pi - \frac{\theta}{2n^2}\pi^\top\Sigma\pi - \frac{\theta}{n}\sigma_1\pi_1\sigma^\top\pi \\ &+ \frac{(\theta+1)\theta}{2n^2}\pi^\top(M+\Sigma)\pi]dt + \nu_1\pi_1 dW_t^1 + \frac{\theta}{n}\sum_{i=2}^{(n+1)}\nu_i\pi_i dW^i + (\sigma_1\pi_1 + \frac{\theta}{n}\sigma^\top\pi)dB.\end{aligned} \quad (2.4)$$

Research Problem: Find $\hat{\pi}_1$ such that the robust utility maximization

$$\sup_{\pi_1}\inf_{\pi} J(\pi_1, \pi) \quad (2.5)$$

is obtained over the admissible set $\mathcal{A}$.

2.3 Main Results of Robust Relative Performance Problem with Incomplete Information

In this section, we find a unique constant robust strategy in Theorem 2.3.1, which we subsequently specialize to the single stock case in Corrolary 2.3.1. Before we state the main results, we denote A:= $\frac{(p\theta+n)}{n}M + N + \frac{p\theta}{n}\Sigma$ for convenience, and examine various cases for A as discussed in Lemma 2.3.1 and 2.3.2.

Lemma 2.3.1. *If A is positive-definite, and if ψ satisfies ordinary differential equation*

$$\psi^{'}(t) + p\lambda_t(\pi_1)\psi(t) = 0, \qquad \psi(T) = 1 \quad (2.6)$$

where

$$\lambda_t(\pi_1) = \frac{\theta}{2n}\inf_{\pi}[\pi^\top A\pi - 2(\mu^\top + p\sigma_1\pi_1\sigma^\top)\pi] + \mu_1\pi_1 + (p-1)\frac{\nu_1^2+\sigma_1^2}{2}\pi_1^2,$$

then we obtain a solution $\psi(t) = \mathbb{E}[\exp(\int_t^T p\lambda_s(\pi_1)ds)|\mathcal{F}_t]$ to the differential equation (2.6).

Proof of Lemma 2.3.1. It is easy to check that the terminal condition is satisfied, $\psi(T) =$

$\mathbb{E}[\exp(\int_T^T p\lambda_s(\pi_1)ds)|\mathcal{F}_t] = 1$. Fixed $\{\omega\} \in \mathcal{F}_t$, the differential equation yields

$$[\log \psi^\omega(t)]' = \frac{\psi'(t)}{\psi(t)} = -p\lambda(\pi_1^\omega).$$

Hence integrating from t to T gives

$$-\log \psi^\omega(t) = \int_t^T d(\log \psi^\omega(s)) = -p\int_t^T \lambda(\pi_1^\omega(s))ds.$$

Therefore $\psi^\omega(t) = \exp(p\int_t^T \lambda(\pi_1^\omega(s))ds)$ for $\{\omega\} \in \mathcal{F}_t$ and thus we solve equation (2.6) with solution

$$\psi(t) = \mathbb{E}[\exp(p\int_t^T \lambda(\pi_1(s,\omega))ds)|\mathcal{F}_t].$$

□

Lemma 2.3.2. *If A is not a positive-definite matrix, then the map $\pi \mapsto \pi^\top A\pi$, where $\pi \in \mathcal{A}^n$, attains its infimum of $-\infty$.*

Proof of Lemma 2.3.2. Assume that the diagonal decomposition of A is $P^{-1}AP = B = diag(\lambda_2, ..., \lambda_{n+1})$ where λ_i are the eigenvalue of A, and P is a unitary matrix with each column being the unite eigenvector. If A is not a positive-definite matrix, at least one eigenvalue of A, say λ_2 is negative; then in the map $\pi \mapsto \pi^\top A\pi = \pi^\top PBP^{-1}\pi$ where $\pi \in \mathcal{A}^n$, there is one map $f_2 : \pi_2 \mapsto \lambda_2\pi_2^2$ where $\pi_2 \in \mathcal{A}$ is strictly concave on the closed convex set $\mathcal{A}$ since $\lambda_2 < 0$, therefore the minimum image value f_2 is $-\infty$. It follows that in the whole map $\pi \mapsto \pi^\top A\pi = \pi^\top PBP^{-1}\pi$, the inf value among all the map values is $-\infty$. □

We interpret Lemma 2.3.2 in plain words, intuitively it means that at least one agent can take a strategy that could ruin X_1.

2.3.1 Main theorem and remarks

Theorem 2.3.1 (Main result)**.** *Assume that for all $i = 1, ..., (n+1)$, we have $x_0^i > 0, p < 1, \theta \geq 0, \mu_i > 0, \sigma_i \geq 0, \nu_i \geq 0$ and $\sigma_i + \nu_i > 0$. Define the constants*

$$\Phi_n := \frac{1}{n}\sigma^\top A^{-1}\mu$$

and

$$\Psi_n := \frac{1}{n}\sigma^\top A^{-1}\sigma.$$

If A is a positive-definite matrix, there exists a unique constant robust strategy, given by the solution to the robust problem with

$$\hat{\pi}_1 = \frac{\mu_1 - \theta p\sigma_1\Phi_n}{(1-p)(\nu_1^2 + \sigma_1^2) + \theta p^2\sigma_1^2\Psi_n}. \tag{2.7}$$

If A is not a positive-definite matrix, the robust utility problem is 0 if $0 < p < 1$ or $-\infty$ if $p < 0$ for whatever π_1 is.

Corollary 2.3.1 (Single stock)**.** *Assume that for all $i = 1, ..., (n+1)$ we have $\mu_i = \mu > 0, \sigma_i = \sigma > 0$, and $\nu_i = 0$. Then $A = N + \frac{p\theta}{n}\Sigma$ with eigenvalues $\{\sigma^2, ..., \sigma^2, (1+p\theta)\sigma^2\}$ is positive-definite if $p \in (-\frac{1}{\theta}, 1)$. Moreover, the robust strategy is given by*

$$\hat{\pi}_1 = \frac{\mu - \theta p \sigma \Phi_n}{(1-p)\sigma^2 + \theta p^2 \sigma^2 \Psi_n}.$$

Remark 2.3.1. *If $\theta = 0$, the robust strategy is reduced to the Merton case, $\hat{\pi}_1 = \frac{\mu_1}{(1-p)(\nu_1^2 + \sigma_1^2)}$. If $\theta \neq 0$, there exists relative performance between X_1 and other agents.*

Remark 2.3.2 (No common noise)**.** *If the case $\sigma_i = 0$ for all $i = 1, ...(n+1)$, the matrix $A = \frac{(p\theta + n)}{n} M$, hence A is positive-definite if p is $(-\frac{n}{\theta}, 1)$. Moreover, the robust strategy is reduced to the Merton case, $\hat{\pi}_1 = \frac{\mu_1}{(1-p)(\nu_1^2 + \sigma_1^2)}$.*

2.3.2 Proof of the Main Theorem 2.3.1

Proof of Theorem 2.3.1. Firstly, we fix π_1 in the inf problem (2.5). Now the aim is to find $\hat{\pi}^{\pi_1}$ such that the value function $v(t, x, y)$ satisfies $v^{\pi_1}(0, x_1^0, y_0) = \inf_\pi J(\pi_1, \pi)$, and $v(t, x, y)$ satisfies the HJB equation

$$\begin{aligned} v_t + \min_\pi [\frac{1}{2}(\frac{(\theta+1)\theta}{2n^2}\pi^\top (M + \Sigma)\pi) y^2 v_{yy} + \frac{\theta}{n}\sigma_1 \pi_1 \sigma^\top \pi x y v_{xy} \\ +(\frac{1}{n}\mu^\top \pi + \frac{1-n}{2n^2}\pi^\top M \pi - \frac{1}{2n}\pi^\top N \pi + \frac{1}{2n^2}\pi^\top \Sigma \pi) y v_y] \\ +\frac{1}{2}(\sigma_1^2 + \nu_1^2)\pi_1^2 x^2 v_{xx} + \mu_1 \pi_1 x v_x = 0, \end{aligned} \tag{2.8}$$

for $(t, x, y) \in [0, T] \times \mathbb{R}_+ \times \mathbb{R}_+$, with terminal condition

$$v(T, x, y) = U(\frac{x}{y}) = p^{-1}(\frac{x}{y})^p.$$

We claim here that the homogeneity of v holds, i.e., $v(t, \frac{x}{k}, \frac{y}{k}) = v(t, x, y)$ for nonzero coefficient k, hence we can simplify the above HJB equation (2.8) by letting $z = \frac{x}{y}$. Additionally, we have the relation $v(t, x, y) = v(t, \frac{x}{y}, \frac{y}{y}) = v(t, \frac{x}{y}, 1) = v(t, z)$.

From now on, we simplify and proceed based on the fact of homogeneity, turning the aim to find $\hat{\pi}^{\pi_1}$ such that the value function $v(t, z)$ satisfies $v^{\pi_1}(0, z_0) := \inf_\pi J(\pi_1, \pi)$.

Define an operator $\mathcal{L}^\pi v = b(z,\pi)v_z + \frac{1}{2}\sigma^2(z,\pi)v_{zz}$ where $b(z,\pi) = z(\mu_1\pi_1 - \frac{\theta}{n}\mu^\top\pi - \frac{(1-n)\theta}{2n^2}\pi^\top M\pi + \frac{\theta}{2n}\pi^\top N\pi - \frac{\theta}{2n^2}\pi^\top\Sigma\pi - \frac{\theta}{n}\sigma_1\pi_1\sigma^\top\pi + \frac{(\theta+1)\theta}{2n^2}\pi^\top(M+\Sigma)\pi)$ and $\sigma^2(z,\pi) = z^2(\nu_1^2\pi_1^2 + \frac{\theta^2}{n^2}\pi^\top M\pi + \sigma_1^2\pi_1^2 + \frac{\theta^2}{n^2}\pi^\top\Sigma\pi - \frac{2\theta}{n}\sigma_1\pi_1\sigma^\top\pi)$. Therefore from (2.8), $v(t,z)$ satisfies the HJB equation (2.9)

$$\begin{aligned} 0 &= v_t + \inf_\pi[\mathcal{L}^\pi v] \\ &= v_t + \inf_\pi[z(\mu_1\pi_1 - \frac{\theta}{n}\mu^\top\pi - \frac{(1-n)\theta}{2n^2}\pi^\top M\pi + \frac{\theta}{2n}\pi^\top N\pi - \frac{\theta}{2n^2}\pi^\top\Sigma\pi - \frac{\theta}{n}\sigma_1\pi_1\sigma^\top\pi \\ &+ \frac{(\theta+1)\theta}{2n^2}\pi^\top(M+\Sigma)\pi)v_z + \frac{1}{2}z^2(\nu_1^2\pi_1^2 + \frac{\theta^2}{n^2}\pi^\top M\pi + \sigma_1^2\pi_1^2 + \frac{\theta^2}{n^2}\pi^\top\Sigma\pi - \frac{2\theta}{n}\sigma_1\pi_1\sigma^\top\pi)v_{zz}] \end{aligned} \tag{2.9}$$

for $(t,z)\in[0,T]\times\mathbb{R}_+$, with terminal condition

$$v(T,z) = U(z) = p^{-1}z^p. \tag{2.10}$$

We can find explicitly a smooth solution to (2.9)-(2.10). In the following, we prove the Theorem 2.6 in two cases: $p\neq 0$ and $p=0$.

Case I: If $p\neq 0$, we consider a candidate solution

$$w(t,z) = \psi(t)U(z) = \psi(t)p^{-1}z^p \tag{2.11}$$

for some positive smooth function ψ. Hence for $w(t,z)$ we derive

$$w_t(t,z) = \psi^{'}(t)p^{-1}z^p, \quad w_z(t,z) = \psi(t)z^{p-1}, \quad w_{zz}(t,z) = (p-1)\psi(t)z^{p-2}. \tag{2.12}$$

Substituting (2.12) into (2.9)-(2.10) and reordering the terms related to π, we get

$$z^p(\frac{\psi^{'}(t)}{p} + \psi(t)(\inf_\pi[\frac{\theta}{n}\pi^\top A\pi - \frac{\theta}{n}(\mu^\top + p\sigma_1\pi_1\sigma^\top)\pi] + p\mu_1\pi_1 + (p-1)p\frac{\nu_1^2+\sigma_1^2}{2}\pi_1^2) = 0. \tag{2.13}$$

If A is positive-definite, we derive that ψ should satisfy differential equation

$$\psi^{'}(t) + p\lambda_t(\pi_1)\psi(t) = 0, \qquad \psi(T) = 1, \tag{2.14}$$

where

$$\lambda_t(\pi_1) = \frac{\theta}{2n}\inf_\pi[\pi^\top A\pi - 2(\mu^\top + p\sigma_1\pi_1\sigma^\top)\pi] + \mu_1\pi_1 + (p-1)\frac{\nu_1^2+\sigma_1^2}{2}\pi_1^2. \tag{2.15}$$

From Lemma 2.3.1, $\psi(t) = \mathbb{E}[\exp(\int_t^T p\lambda_s(\pi_1)ds)|\mathcal{F}_t]$ is a solution to equation (2.14) and

$$w(t,z) = \psi(t)p^{-1}z^p = p^{-1}z^p\mathbb{E}\exp(\int_t^T p\lambda_s(\pi_1)ds), (t,z) \in [0,T] \times \mathbb{R}_+. \tag{2.16}$$

Hence from (2.16), $w(t,z)$ is strictly increasing and concave in z and is a smooth solution to (2.9)-(2.10). Furthermore, if the matrix A is positive-definite, the function $\pi \mapsto \pi^\top A\pi$ is strictly convex on the closed convex set $\mathcal{A}^n$ and thus attains its minimum at some $\hat{\pi}$. By construction, $\hat{\pi}$ attains the infimum of the operator $\inf_\pi[\mathcal{L}^\pi v]$.

To solve $\hat{\pi}^{\pi_1}$, from the equation (2.15), the min value could be achieved at the point where its first-order derivative is zero. Then we get the minimizer

$$\hat{\pi}^{\pi_1} = A^{-1}[\mu + p\sigma_1\pi_1\sigma]. \tag{2.17}$$

Hence by substituting back into (2.15) we solve

$$\lambda_t(\pi_1) = -\frac{\theta}{2n}[\mu^\top + p\sigma_1\pi_1\sigma^\top]A^{-1}[\mu + p\sigma_1\pi_1\sigma] + \mu_1\pi_1 + (p-1)\frac{\nu_1^2+\sigma_1^2}{2}\pi_1^2. \tag{2.18}$$

In the sup problem (2.5), given $\hat{\pi}^{\pi_1}$ from (2.17) in the inf problem, the aim is to find $\hat{\pi}_1$ such that

$$\sup_{\pi_1} v(0,z_0) = \sup_{\pi_1} p^{-1}{z_0}^p\mathbb{E}\exp(\int_0^T p\lambda_s(\pi_1)ds) \tag{2.19}$$

where $\lambda(\pi_1)$ is in equation (2.18). Since the max value is achieved at the point where its first-order derivative is zero, we get the solution that π_1 satisfies

$$\hat{\pi}_1 = \frac{\mu_1 - \theta p\sigma_1\Phi_n}{(1-p)(\nu_1^2+\sigma_1^2) + \theta p^2\sigma_1^2\Psi_n}. \tag{2.20}$$

It follows that the robust strategy $\hat{\pi}_1$ is a unique constant strategy, hence the value function

$$J(\hat{\pi}_1,\hat{\pi}) = \sup_{\pi_1} v(0,z_0) = p^{-1}{z_0}^p\exp(p\lambda(\hat{\pi}_1)T). \tag{2.21}$$

For this functional value (2.21), if we take its inverse of the utility function, we derive the process

$$Z_T = z_0\exp(\lambda(\hat{\pi}_1)T). \tag{2.22}$$

We refer $\lambda(\hat{\pi}_1)$ as the robust growth rate.

If A is not a positive-definite matrix, from Lemma 2.3.2, the inf value in (2.15) is $-\infty$, and it leads to the following:

$$\inf_\pi J(\pi_1,\pi) = \begin{cases} 0, & 0<p<1, \\ -\infty, & p<0. \end{cases}$$

Therefore if A is not a positive-definite matrix, for whatever π_1 is, the robust utility is

$$\sup_{\pi_1}\inf_{\pi} J(\pi_1,\pi) = \begin{cases} 0, & 0<p<1, \\ -\infty, & p<0. \end{cases} \tag{2.23}$$

In plain words, if A is not a positive-definite matrix, there is no robust strategy that could hedge the worst case.

Case II: If $p=0$, the matrix $A=M+N$ is always positive-definite, and a candidate solution is in the form

$$w(t,z)=\psi(t)U(z)=\psi(t)+\log z \tag{2.24}$$

for some positive function ψ. We derive ψ should satisfy the ordinary differential equation

$$\psi^{'}(t)=-\lambda, \qquad \psi(T)=1 \tag{2.25}$$

where

$$\lambda(\pi_1)=\frac{\theta}{2n}\inf_{\pi}[\pi^\top A\pi-2\mu^\top\pi]+\mu_1\pi_1-\frac{\nu_1^2+\sigma_1^2}{2}\pi_1^2. \tag{2.26}$$

We then obtain a solution $\psi(t)=\mathbb{E}[\int_t^T\lambda(\pi_1(s))ds|\mathcal{F}_t]$ and

$$w(t,z)=\log z+\mathbb{E}[\int_t^T\lambda(\pi_1(s))ds|\mathcal{F}_t],(t,z)\in[0,T]\times\mathbb{R}_+. \tag{2.27}$$

Hence, $w(t,z)$ is strictly increasing and concave in z and is a smooth solution to (2.9)-(2.10). Furthermore, since A is positive-definite, the function $\pi\mapsto\pi^\top A\pi-2\mu^\top\pi$ is strictly convex on the closed convex set $\mathcal{A}^n$ and thus attains its minimum at some $\hat{\pi}$. By construction, $\hat{\pi}$ attains the infimum of $\inf_\pi[\mathcal{L}^\pi v]$.

Solving $\hat{\pi}^{\pi_1}$, from the equation (2.26), the min value is achieved at the point where its first-order derivative is zero. Then we derive

$$\hat{\pi}^{\pi_1}=A^{-1}\mu. \tag{2.28}$$

Hence by substituting into equation (2.26) we have

$$\lambda(\pi_1)=-\frac{\theta}{2n}\mu^\top A^{-1}\mu+\mu_1\pi_1-\frac{\nu_1^2+\sigma_1^2}{2}\pi_1^2. \tag{2.29}$$

In the second layer, given $\hat{\pi}^{\pi_1}$, the aim is to find $\hat{\pi}_1$ such that

$$\sup_{\pi_1} v(0,z_0)=\sup_{\pi_1}(\log z_0+\mathbb{E}[\int_0^T\lambda(\pi_1(s))ds|\mathcal{F}_0]). \tag{2.30}$$

Hence we search the critical point of $\lambda(\pi_1)$ by taking first-order derivative and second-order derivative:

$$\lambda^{'}(\pi_1)=\mu_1-(\sigma_1^2+\nu_1^2)\pi_1, \tag{2.31}$$

$$\lambda^{''}(\pi_1) = -(\sigma_1^2 + \nu_1^2) < 0. \tag{2.32}$$

It is easy to verify that we indeed have $\sup_{\pi_1} \lambda(\pi_1)$. Solving $\hat{\pi}_1$, we derive

$$\hat{\pi}_1 = \frac{\mu_1}{\sigma_1^2 + \nu_1^2}. \tag{2.33}$$

Finally, combining the two cases, we proved the Theorem 2.3.1. □

2.3.3 Verification for the case $p \neq 0$

For the **Case I**, $p \neq 0$, to verify that $\hat{\pi}$ indeed achieves the minimum, pick arbitrary control $\pi \in \mathcal{A}^n$. Firstly apply the Itô' formula to $w(u, z) = w(u, Z_u)$ between $t \in [0, T)$ and T,

$$w(T, Z_T^{t,x}) = w(t,x) + \int_t^T w_t(u, Z_u^{t,x})du + \int_t^T w_z(u, Z_u^{t,x})dZ_u + \frac{1}{2}\int_t^T w_{zz}(u, Z_u^{t,x}) < dZ_u, dZ_u > .$$

Substituting equations (2.4) inside, we have

$$\begin{aligned}
w(T, Z_T^{t,x}) &= w(t,x) + \int_t^T w_t(u, Z_u^{t,x})du + \int_t^T w_z(u, Z_u^{t,x})dZ_u + \frac{1}{2}\int_t^T w_{zz}(u, Z_u^{t,x}) < dZ_u, dZ_u > \\
&= w(t,x) + \int_t^T w_t(u, Z_u^{t,x})du + \int_t^T w_z(u, Z_u^{t,x})Z_u([\mu_1\pi_1 - \frac{\theta}{n}\mu^\top\pi - \frac{(1-n)\theta}{2n^2}\pi^\top M\pi \\
&+ \frac{\theta}{2n}\pi^\top N\pi - \frac{\theta}{2n^2}\pi^\top\Sigma\pi - \frac{\theta}{n}\sigma_1\pi_1\sigma^\top\pi + \frac{(\theta+1)\theta}{2n^2}\pi^\top(M+\Sigma)\pi]du + \nu_1\pi_1 dW_u^1 \\
&+ \frac{\theta}{n}\sum_{i=2}^{(n+1)}\nu_i\pi_i dW^i + (\sigma_1\pi_1 + \frac{\theta}{n}\sigma^\top\pi)dB) + \frac{1}{2}\int_t^T w_{zz}(u, Z_u^{t,x})Z_u^2(\nu_1^2\pi_1^2 du \\
&+ \frac{\theta^2}{n^2}\pi^\top M\pi du + (\sigma_1^2\pi_1^2 + \frac{\theta^2}{n^2}\pi^\top\Sigma\pi + \frac{2\theta}{n}\sigma_1\pi_1\sigma^\top\pi))du.
\end{aligned}$$

Since $\mathbb{E}\int_0^T |\pi_i(t)|^2 dt < \infty$ for $i = 1, ...(n+1)$, we know that $\mathbb{E}\int_0^T |\nu_1\pi_1 w_z(u, Z_u^{t,x})Z_u|^2 du < \infty, \mathbb{E}\int_0^T |\frac{\theta}{n}\sum_{i=2}^{(n+1)}\nu_i\pi_i w_z(u, Z_u^{t,x})Z_u|^2 du < \infty$ and $\mathbb{E}\int_0^T |(\sigma_1\pi_1 + \frac{\theta}{n}\sigma^\top\pi)w_z(u, Z_u^{t,x})Z_u|^2 du < \infty$. So take expectations on both sides in the above equality

$$\begin{aligned}
\mathbb{E}w(T, Z_T^{t,x}) = w(t,x) + \mathbb{E}\int_t^T &[w_t + z w_z(\mu_1\pi_1 - \frac{\theta}{n}\mu^\top\pi - \frac{(1-n)\theta}{2n^2}\pi^\top M\pi + \frac{\theta}{2n}\pi^\top N\pi \\
&- \frac{\theta}{2n^2}\pi^\top\Sigma\pi - \frac{\theta}{n}\sigma_1\pi_1\sigma^\top\pi + \frac{(\theta+1)\theta}{2n^2}\pi^\top(M+\Sigma)\pi) + \frac{1}{2}z^2 w_{zz}(\nu_1^2\pi_1^2 + \frac{\theta^2}{n^2}\pi^\top M\pi \\
&+ \sigma_1^2\pi_1^2 + \frac{\theta^2}{n^2}\pi^\top\Sigma\pi - \frac{2\theta}{n}\sigma_1\pi_1\sigma^\top\pi)]du.
\end{aligned}$$

From equation (2.9), the integrand is greater than or equal to 0 and equals 0 only when the control is the minimum control $\hat{\pi}$ and denotes the corresponding process $\hat{Z}$. Therefore we derive

$$\mathbb{E}w(T,\hat{Z}_T^{t,x}) \leq \mathbb{E}w(T,Z_T^{t,x}), \qquad \forall \pi \in \mathcal{A}^n.$$

Hence from the equation (2.10), we know for all $t \in [0,T)$,

$$\mathbb{E}w(T,\hat{Z}_T^{t,x}) = \mathbb{E}U(\hat{Z}_T^{t,x}), \mathbb{E}w(T,Z_T^{t,x}) = \mathbb{E}U(Z_T^{t,x}).$$

So if t equals 0, we derive

$$\mathbb{E}U(\hat{Z}_T) \leq \mathbb{E}U(Z_T),$$

which shows that $J(\pi_1,\hat{\pi}) \leq J(\pi_1,\pi), \forall \pi \in \mathcal{A}^n$. Thus it has been verified that $\hat{\pi}$ is indeed the optimal control and $v = w$.

2.4 Value of Information

In this section, we first derive the unique constant Nash equilibrium strategy π_i^* in Subsection 2.4.1. After that, in Subsection 2.4.2 we conduct analysis on the value of the information.

2.4.1 Analysis of Nash equilibrium strategy π_i^* with complete information

We derive the equilibrium of the model using the method in the paper of Lacker D. and Zariphopoulou T. (2019) [38].

For the notations, let $\pi_{-i} = (\pi_1,...,\pi_{i-1},\pi_{i+1},...\pi_{n+1})^\top, \mu_{-i} = (\mu_1,...,\mu_{i-1},\mu_{i+1},...,\mu_{n+1})^\top$, $\nu_{-i} = (\nu_1,...,\nu_{i-1},\nu_{i+1},...,\nu_{n+1})^\top, \sigma_{-i} = (\sigma_1,...,\sigma_{i-1},\sigma_{i+1},...,\sigma_{n+1})^\top$ be n-dimensional vectors and $M_{-i} = \mathrm{diag}(\nu_1^2,...,\nu_{i-1}^2,\nu_{i+1}^2,...,\nu_{n+1}^2)$, $N_{-i} = \mathrm{diag}(\sigma_1^2,...,\sigma_{i-1}^2,\sigma_{i+1}^2,...,\sigma_{n+1}^2)$, $\Sigma_{-i} = (\sigma_i\sigma_j)_{i,j\in\{1,...,i-1,i+1,...,n+1\}}$ be $n \times n$-dimensional matrix.

Theorem 2.4.1 (Nash equilibrium strategy)**.** *Assume that for all $i = 1,...,(n+1)$ we have $x_0^i > 0, p < 1, \theta > 0, \mu_i > 0, \sigma_i \geq 0, \nu_i \geq 0$ and $\sigma_i + \nu_i \geq 0$. Define the constants*

$$\Phi_{n+1} := \frac{1}{n+1}\sum_{k=1}^{n+1}\frac{\sigma_k\mu_k}{\nu_k^2(1-p) - \sigma_k^2(1-p(1+\dfrac{\theta}{n}))}$$

and

$$\Psi_{n+1} := \frac{p\theta}{n}\sum_{k=1}^{n+1}\frac{\sigma_k^2}{\nu_k^2(1-p) - \sigma_k^2(1-p(1+\dfrac{\theta}{n}))}.$$

There exists a unique constant equilibrium, given by

$$\pi_i^* = \frac{\mu_i - (1+\frac{1}{n})p\theta\sigma_i\frac{\Phi_{n+1}}{1+\Psi_{n+1}}}{\nu_i^2(1-p) - \sigma_i^2(1-p(1+\frac{\theta}{n}))}. \tag{2.34}$$

Moreover, we have the identity

$$\frac{1}{n+1}\sum_{k=1}^{n+1}\sigma_k\pi_k^* = \frac{\Phi_{n+1}}{1+\Psi_{n+1}}.$$

Proof of Theorem 2.4.1. Fix an agent i and constant strategy $\alpha_k \in \mathbb{R}$ for $k \neq i$. Since $\overline{X}_{-i} = (\prod_{i=1,k\neq i}^{(n+1)} X_k)^{\frac{1}{n}}$, then by Itô formula we have

$$\frac{d\overline{X}_{-i}}{\overline{X}_{-i}} = (\frac{1}{n}\mu_{-i}^\top\pi + \frac{1-n}{2n^2}\pi^\top M_{-i}\pi - \frac{1}{2n}\pi^\top N_{-i}\pi + \frac{1}{2n^2}\pi^\top\Sigma_{-i}\pi)dt + \frac{1}{n}\sum_{k=1,k\neq i}^{(n+1)}\nu_k\pi_k dW^k + \frac{1}{n}\sigma_{-i}^\top\pi dB. \tag{2.35}$$

The utility is a function $U:\mathbb{R}_+\to\mathbb{R}$ and the performance functional J is the particular case of power utility of CRRA type,

$$\sup_{\pi_i} J(\pi_i,\pi_{-i}) = \sup_{\pi_i}\mathbb{E}\left[U\left(\frac{X_i(T)}{\overline{X}_{-i}^\theta(T)}\right)\right] = \begin{cases} p^{-1}\sup_{\pi_i}\mathbb{E}\left[\left(\frac{X_i(T)}{\overline{X}_{-i}^\theta(T)}\right)^p\right], & p\neq 0 \\ \sup_{\pi_i}\mathbb{E}\left[\log\left(\frac{X_i(T)}{\overline{X}_{-i}^\theta(T)}\right)\right], & p=0. \end{cases} \tag{2.36}$$

Then we obtain that the value (2.36) is equal to $\hat{v}(X_0^i, \overline{X}_0^{-i}, 0)$, where $\hat{v}(x,y,t)$ solves the HJB equation

$$\begin{aligned} 0 &= \hat{v}_t + \sup_{\pi_i}[\mathcal{L}^\pi\hat{v}] \\ &= \hat{v}_t + \sup_{\pi_i}[\frac{\nu_i^2+\sigma_i^2}{2}\pi_i^2x^2\hat{v}_{xx} + \pi_i\mu_i x\hat{v}_x + \frac{1}{n}\pi_i\sigma_i\sigma_{-i}^\top\pi_{-i}xy\hat{v}_{xy} \\ &+ \frac{1}{2n^2}(\pi_{-i}^\top N_{-i}\pi_{-i} + \pi_{-i}^\top\Sigma_{-i}\pi_{-i})y^2\hat{v}_{yy} \\ &+ (\frac{1}{n}\mu_{-i}^\top\pi_{-i} + \frac{1-n}{2n^2}\pi_{-i}^\top M_{-i}\pi_{-i} - \frac{1}{2n}\pi_{-i}^\top N_{-i}\pi_{-i} + \frac{1}{2n^2}\pi_{-i}^\top\Sigma_{-i}\pi_{-i})y\hat{v}_y], \end{aligned} \tag{2.37}$$

for $(x,y,t)\in\mathbb{R}_+^2\times[0,T]$, with terminal condition when $p\neq 0$

$$\hat{v}(x,y,T) = U(\frac{x}{y^\theta}) = p^{-1}(\frac{x}{y^\theta})^p. \tag{2.38}$$

We can find explicitly a smooth solution to (2.37)-(2.38). We are looking for a candidate solution in the form

$$\hat{w}(x,y,t) = \psi(t)U(\frac{x}{y}) = \psi(t)p^{-1}(\frac{x}{y^\theta})^p \tag{2.39}$$

for some positive function ψ. Hence we derive

$$\begin{aligned}0 = (\frac{x}{y^\theta})^p(\frac{\psi^{'}(t)}{p} + \psi(t)(\sup_{\pi_i}[(p-1)\frac{\nu_i^2+\sigma_i^2}{2}\pi_i^2 + (\mu_i - \frac{p\theta}{n}\sigma_{-i}^\top\pi_{-i}\sigma_i)\pi_i \\ + \frac{\theta(1+p\theta)}{2n^2}(\pi_{-i}^\top N_{-i}\pi_{-i} + \pi_{-i}^\top\Sigma_{-i}\pi_{-i}) \\ - \theta(\frac{1}{n}\mu_{-i}^\top\pi_{-i} + \frac{1-n}{2n^2}\pi_{-i}^\top M_{-i}\pi_{-i} - \frac{1}{2n}\pi_{-i}^\top N_{-i}\pi_{-i} + \frac{1}{2n^2}\pi_{-i}^\top\Sigma_{-i}\pi_{-i})])).\end{aligned} \tag{2.40}$$

And we derive that ψ should satisfy the differential equation

$$\psi^{'}(t) + p\rho\psi(t) = 0, \qquad \psi(T) = 1 \tag{2.41}$$

where

$$\begin{aligned}\rho = \sup_{\pi_i}[(p-1)\frac{\nu_i^2+\sigma_i^2}{2}\pi_i^2 + (\mu_i - \frac{p\theta}{n}\sigma_{-i}^\top\pi_{-i}\sigma_i)\pi_i + \frac{\theta(1+p\theta)}{2n^2}(\pi_{-i}^\top N_{-i}\pi_{-i} + \pi_{-i}^\top\Sigma_{-i}\pi_{-i}) \\ - \theta(\frac{1}{n}\mu_{-i}^\top\pi_{-i} + \frac{1-n}{2n^2}\pi_{-i}^\top M_{-i}\pi_{-i} - \frac{1}{2n}\pi_{-i}^\top N_{-i}\pi_{-i} + \frac{1}{2n^2}\pi_{-i}^\top\Sigma_{-i}\pi_{-i})].\end{aligned} \tag{2.42}$$

We then obtain $\psi(t) = \mathbb{E}\exp(\int_t^T p\rho_s ds)$ and

$$\hat{w}(x,y,t) = p^{-1}(\frac{x}{y^\theta})^p\mathbb{E}\exp(\int_t^T p\rho_s ds), (x,y,t) \in \mathbb{R}_+^2 \times [0,T]. \tag{2.43}$$

Hence, $\hat{w}$ is strictly increasing and concave in $\frac{x}{y}$, and is a smooth solution to (2.37)-(2.38). Furthermore, the function $\pi_i \mapsto (p-1)\frac{\nu_i^2+\sigma_i^2}{2}\pi_i^2 + (\mu_i - \frac{p\theta}{n}\sigma_{-i}^\top\pi_{-i}\sigma_i)\pi_i$ is strictly concave on the closed convex set $\mathcal{A}$, and thus attains its maximum at some π_i^*. By construction, π_i^* attains the supremum of $\sup_{\pi_i}[\mathcal{L}^\pi\hat{v}]$.

Solving π_i^*, from the equation (2.42), the max value could be achieved at the point where its first-order derivative is zero. Then we get the maximizer

$$\pi_i^* = \frac{\mu_i - \frac{p\theta}{n}\sigma_i\sigma_{-i}^\top\pi_{-i}}{(\nu_i^2+\sigma_i^2)(1-p)}. \tag{2.44}$$

We want to get rid of the term $\sigma_{-i}^\top\pi_{-i}$. For $(\alpha_1, ..., \alpha_{n+1})$ to be a constant equilibrium, we must have $\pi_i^* = \alpha_i$, for each $i = 1, ...(n+1)$. Using (2.44) and abbreviating

$$\overline{\sigma\alpha} := \frac{1}{n+1}\sum_{k=1}^{n+1}\sigma_k\alpha_k = \frac{1}{n+1}\sigma_i\alpha_i + \frac{1}{n+1}\sigma_{-i}^\top\alpha_{-i}, \tag{2.45}$$

we deduce that we must have

$$\alpha_i = \frac{\mu_i - (1+\frac{1}{n})p\theta\sigma_i\overline{\sigma\alpha}}{\nu_i^2(1-p) - \sigma_i^2(1-p(1+\frac{\theta}{n}))}. \tag{2.46}$$

Multiplying both sides by σ_i and averaging over $i = 1, ..., (n+1)$ give

$$\overline{\sigma\alpha} = \Phi_{n+1} - \Psi_{n+1}\overline{\sigma\alpha}, \tag{2.47}$$

where Φ_{n+1}, Ψ_{n+1} are as in (2.4.1) and (2.4.1). We then deduce from (2.46) that equilibrium strategy $\alpha_i = \pi_i^*$ is given by (2.34). □

As a byproduct, we obtain the value function

$$v(t,x,y) = p^{-1}(\frac{x}{y^\theta})^p\mathbb{E}\exp(\int_t^T p\rho_s ds), (t,x,y) \in [0,T]\times\mathbb{R}_+^2, \tag{2.48}$$

where

$$\begin{aligned}\rho = \sup_{\pi_i}[(p-1)\frac{\nu_i^2+\sigma_i^2}{2}\pi_i^2 + (\mu_i - \frac{p\theta}{n}\sigma_{-i}^\top\pi_{-i}\sigma_i)\pi_i + \frac{\theta(1+p\theta)}{2n^2}(\pi_{-i}^\top N_{-i}\pi_{-i} + \pi_{-i}^\top\Sigma_{-i}\pi_{-i})\\ - \theta(\frac{1}{n}\mu_{-i}^\top\pi_{-i} + \frac{1-n}{2n^2}\pi_{-i}^\top M_{-i}\pi_{-i} - \frac{1}{2n}\pi_{-i}^\top N_{-i}\pi_{-i} + \frac{1}{2n^2}\pi_{-i}^\top\Sigma_{-i}\pi_{-i})].\end{aligned} \tag{2.49}$$

It follows that the Nash strategy π_1^* is constant, hence the value function is

$$J(\pi_1^*, \pi_{-1}) = \sup_{\pi_1} v(0,x_0,y_0) = p^{-1}(\frac{x_0}{y_0})^p\exp(p\rho(\pi_1^*)T) = p^{-1}{z_0}^p\exp(p\rho(\pi_1^*)T). \tag{2.50}$$

For this functional value, if we take its inverse of the utility function, we derive that

$$\tilde{Z}_T = z_0\exp(\rho(\pi_1^*)T)). \tag{2.51}$$

We refer $\rho(\pi_1^*)$ as the Nash equilibrium growth rate.

2.4.2 Value of the Information

We denote the ratio between the wealth under the equilibrium in complete information and the wealth under the robust strategy in incomplete information as the value of the information of knowledge on the behavior of other agents in the market. Thus the value of the information is captured by

$$\text{Value of Infor} = \frac{\tilde{Z}_T}{Z_T} = \exp\{[\rho(\pi_1^*) - \lambda(\hat{\pi}_1)]T\}. \tag{2.52}$$

Equivalently, the difference between the certainty equivalent rates $\rho(\pi_1^*)$ and $\lambda(\hat{\pi}_1)$ is

$$\rho(\pi_1^*) - \lambda(\hat{\pi}_1) = \frac{\log \tilde{Z}_T - \log Z_T}{T}. \tag{2.53}$$

2.5 Numerical Simulations

In this section, we study the robust model of the single stock case. We first list the input constants here: the yearly yield $\mu_i = \mu = 16\%$, yearly volatility $\sigma_i = \sigma = 20\%, \nu_i = 0$, the investment horizon is half a year $T = 0.5$, and the initial relative wealth $z_0 = 1$. The three parameters are p, θ, n with constrains $p < 1, \theta > 0, n \in \mathbb{N}^+$. Then from Remark 2.3.1 we know that $A = N + \frac{p\theta}{n}\Sigma$ has eigenvalues $\{0.447, \cdots, 0.447, (1+p\theta)0.447\}$. So one more constraint is $1 + p\theta > 0$ to let A be positive definite.

- plot robust strategy $\hat{\pi}_1$, robust growth rate λ and terminal wealth Z_T given $n = 60, \theta = 1.2 > 1$ with the range for p is $(-0.83333, 1)$. We see all three functions are decreasing as p increases, see Figure 2.1.

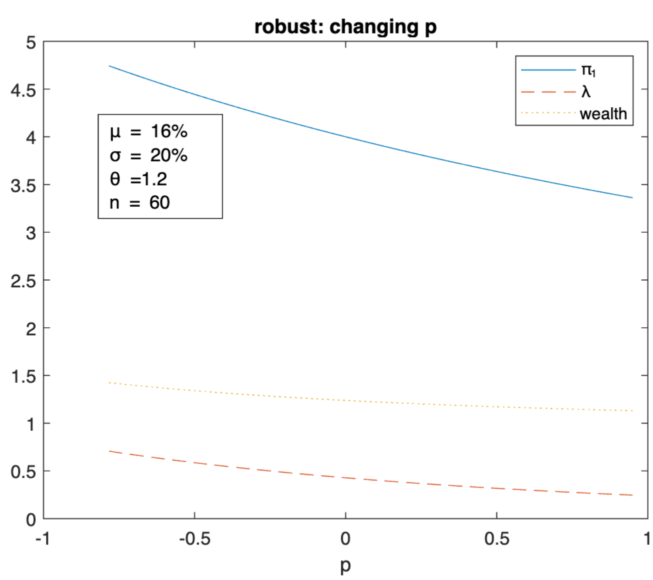


Figure 2.1: Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, \theta = 1.2$ given $\nu_i = 0$

- plot robust strategy $\hat{\pi}_1$, robust growth rate λ and terminal wealth Z_T given $n = 60, \theta = 0.7 < 1$ with the range for p is $(-1.42857, 1)$. We see all three functions are increasing as p increases, see Figure 2.2.

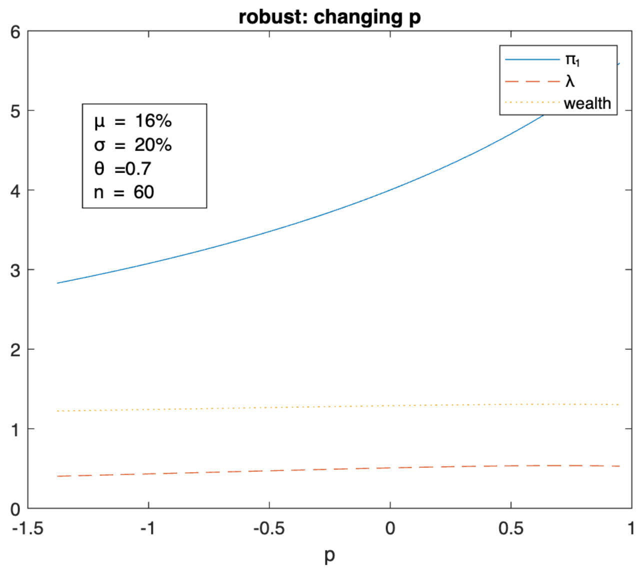


Figure 2.2: Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, \theta = 0.7$ given $\nu_i = 0$

- plot robust strategy $\hat{\pi}_1$, robust growth rate λ and terminal wealth Z_T given $n = 60, p = 0.5$ positive with the range for θ is $(0, \infty)$. We see all three functions are decreasing as p increases, see Figure 2.3.

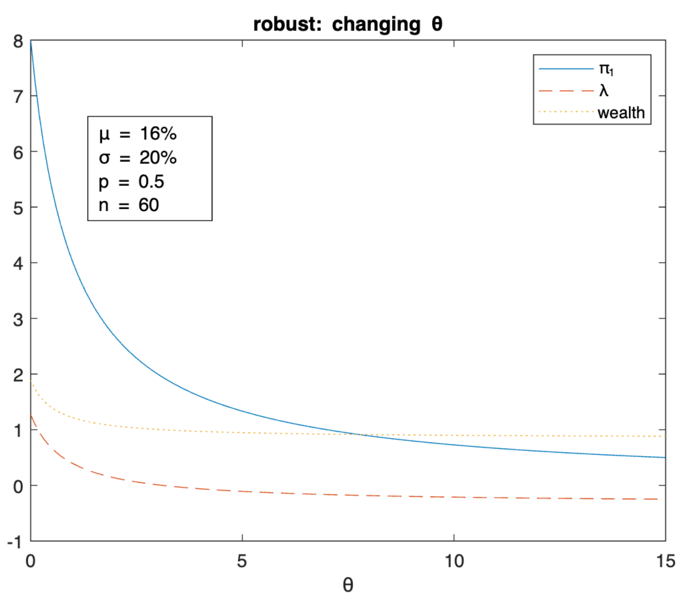


Figure 2.3: Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, p = 0.5$ given $\nu_i = 0$

- plot robust strategy $\hat{\pi}_1$, robust growth rate λ and terminal wealth Z_T given $n = 60, p = -0.5$ negative with the range for θ is $(0, 2)$. We see all three functions are increasing as p increases, see Figure 2.4.

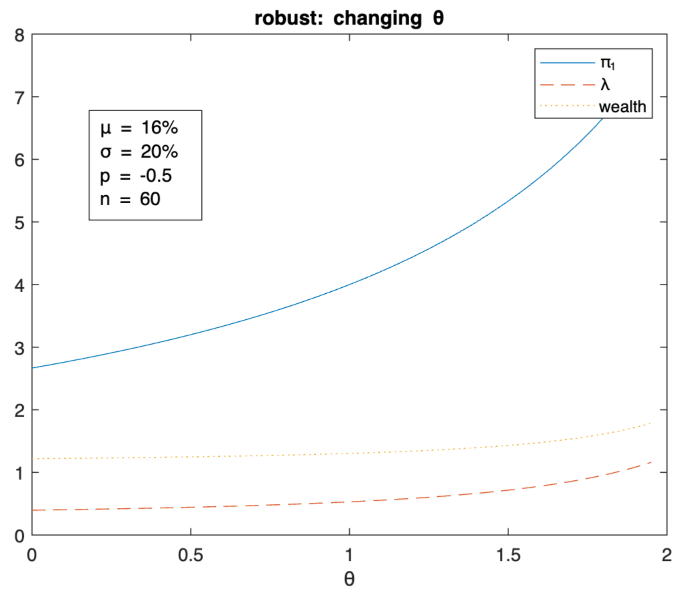


Figure 2.4: Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $n = 60, p = -0.5$ given $\nu_i = 0$

- plot robust strategy $\hat{\pi}_1$, robust growth rate λ and terminal wealth Z_T given $p = 0.5, \theta = 0.7$. We see all three functions are constant lines since all the Greek constants Φ_n and Ψ_n are the same value for whatever n. The robust strategy $\hat{\pi}_1 = 4.7059$, robust growth rate $\lambda = 0.53367$ and terminal wealth $Z_T = 1.3058$, see Figure 2.5.

Now we assume that $\mu_i = 0.16, \sigma_i = 0.2, \nu_i = 0.2$, and plot 3-dimensional figures for the robust growth rate λ and the Nash equilibrium growth rate ρ with fixed p, note that the xy-plane are variables n and θ.

2.6 Conclusion and Future Work

Firstly we can extend the current framework to incorporate dynamic competition dynamics among investors. Investigate how strategic interactions evolve over time and how this affects the robustness of portfolio management strategies. Next, we may investigate dynamic risk management strategies within the context of robust portfolio management. We suggest to explore how adaptive risk allocation and hedging techniques can enhance portfolio resilience and mitigate downside risks in competitive environments. Lastly, we may validate the robust portfolio management framework using real-world financial data. Conduct empirical studies to assess the performance of robust strategies under different market conditions and validate their effectiveness in practice.

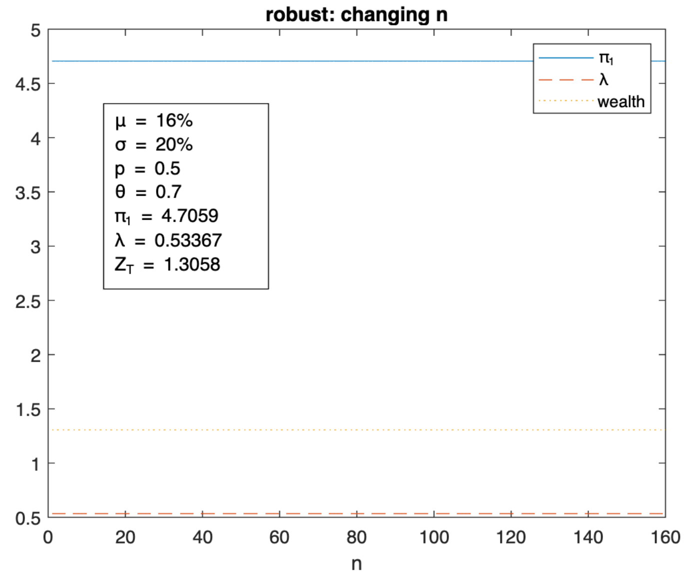


Figure 2.5: Simulations of $\hat{\pi}_1, \lambda, Z_T$ when $\theta = 0.7, p = 0.5$ given $\nu_i = 0$

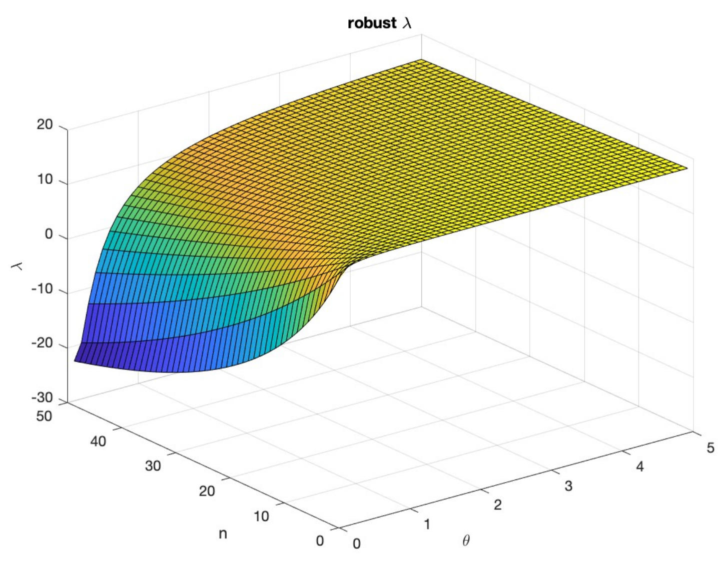


Figure 2.6: Simulations of robust growth rate λ of θ and n, fix p given $\nu_i = 0.2$

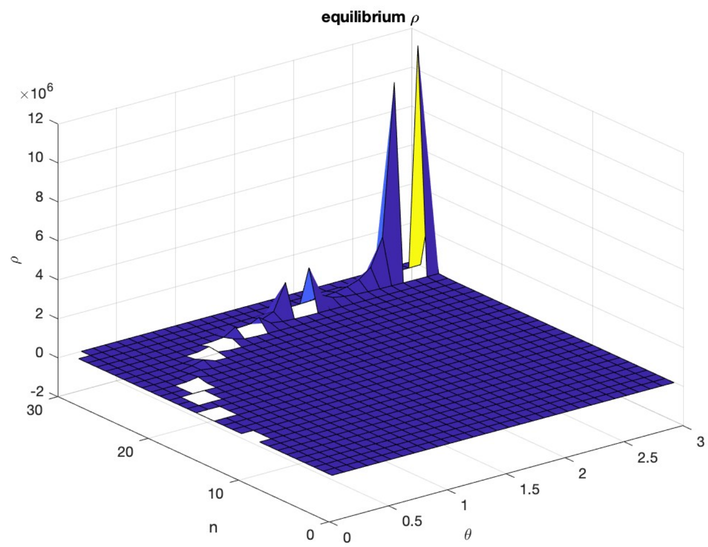


Figure 2.7: Simulations of Nash equilibrium growth rate ρ of variant θ and n, fix p given $\nu_i = 0.2$,

3

Random Matrix Perturbation Bounds for Low-Rank Approximation

In this chapter, we apply the Dyson Bessel process and the theory of stochastic differential equations to find a tighter perturbation bound for a rectangular matrix when the perturbation is a Gaussian noise, under the Frobenius-norm distance. The primary method employed is the Dyson-Bessel process, enabling us to monitor the evolution of the right singular vectors of the perturbed matrix $\hat{A}$ as it progresses through time. Two main applications for the derived perturbation bound are considered. One application for this perturbation bound is the subspace recovery problem and another application is the rank-k matrix approximation problem.

3.1 Introduction

Random matrix perturbation bounds represent a fundamental concept in the field of matrix analysis, with broad applications in various domains, including statistics [67], randomized numerical linear algebra (see [15, 24, 48, 58, 61]), machine learning, and data privatization [44]. These bounds quantify the effect of perturbations of the eigenvectors and singular vectors of matrices, providing valuable insights into the stability and robustness of numerical algorithms and statistical estimators. In recent years, there has been growing interest in developing tighter perturbation bounds for random matrix models. This chapter provides an introduction to random matrix perturbation bounds, explores their theoretical foundations, and reviews the existing literature on their applications in diverse fields.

Let $A \in \mathbb{R}^{m\times d}$ be a rectangular matrix, and denote by $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$ its singular values. For instance, A may be a data matrix where each row represents a datapoint, and each column represents a feature variable. In numerous practical scenarios, the task arises of identifying a rank-k matrix closely resembling A. This situation is evident in various contexts, such as the rank-k matrix approximation problem, aimed at determining a rank-k matrix that minimizes a specified distance from A. Additionally, it manifests in the subspace recovery problem, wherein the objective is to calculate a rank-k projection matrix $V_k V_k^\top$, where V_k denotes the $d \times k$ matrix comprising the top-k singular vectors of A.

In the context of matrix analysis, given an initial matrix A, the matrix A may be perturbed by adding another matrix E with the same dimensions as A. This operation yields a perturbed matrix denoted as $\hat{A}$, defined as the sum of matrices A and E,

$$\hat{A} = A + E. \tag{3.1}$$

One natural question is how to obtain tight bounds for low-rank approximations to A? Perturbation bounds are important for its broad applications of low-rank approximation in statistics [67], for instance, Principle Component Analysis (PCA) [7, 37, 40, 71], randomized numerical linear algebra [15, 24, 48, 58, 61], and many more fields. In the setting when the matrix E may be any deterministic matrix, tight perturbation bounds on the eigenvalues and eigenvectors have been obtain in multiple previous works [14, 63, 64]. Weyl's perturbation theorem gives a deterministic perturbation bound on the singular values $\sigma(A), \sigma(\hat{A})$ [64]. The Davis-Kahan-Wedin sine-Theta theorem [14, 63] gives a perturbation bound on the matrices $V_k(A), \hat{V}_k(\hat{A})$ whose columns are the top k singular vectors under deterministic perturbation. Specifically, the Davis-Kahan-Wedin sine-Theta theorem gives the bound

$$\|V_k V_k^\top - \hat{V}_k \hat{V}_k^\top\|_2 \leq \frac{\|E\|_2}{\sigma_k - \sigma_{k+1}}$$

where $V_k, \hat{V}_k$ are the matrices with columns of the top k singular vectors respectively, σ_k, σ_{k+1} are singular values of matrix A. The bound given above is tight for the worst case when the perturbation E is deterministic.

However, oftentimes the observed data matrices are corrupted by random noise. For many applications, the noise is white noise (i.e., Gaussian) $E = G$ where G is a matrix whose entries are independent and identically distributed Gaussian random variables [54, 56],

$$\hat{A} = A + G. \tag{3.2}$$

However, the classical perturbation bound given by Davis-Kahan-Wedin [14, 63] is not tight with respect to random perturbations. This raises the question of how to obtain a tight perturbation bound for random perturbations which are Gaussian distributed. Our main result, Theorem (3.2.1), gives perturbation bounds for low-rank approximations which are tight with respect to Gaussian random matrix perturbations. Specifically, let $B(t) \in \mathbb{R}^{m \times d}$ be a matrix-valued Brownian motion, where each entry undergoes an independent standard Brownian motion. To prove our bounds, we analyze the incorporation of Gaussian noise into a matrix structure as a matrix-valued stochastic process– the Dyson Bessel process:

$$\Phi(t) = A + \sqrt{t}G = A + B(t). \tag{3.3}$$

This approach offers several advantages. Firstly, this approach allows us to bypass higher-order terms that arise in Taylor expansion of deterministic perturbations, as these terms vanish in the stochastic derivative when the perturbation is a Brownian motion, due to the independence of random noise additions at each time step of the Brownian motion. Secondly, unlike in deterministic perturbations where the singular vector perturbations may add up in a worse-case manner, the distribution of Gaussian noise has rotational symmetry, potentially leading to cancellations when the perturbations are added at each time step of the Brownian motion.

Towards this end, we represent the perturbed matrix as a matrix-valued Dyson Bessel process $\Phi(t)$, with the initial condition $\Phi(0) = A$ and terminal condition $\Phi(T) = \hat{A}$. This perspective allows us to apply stochastic analysis tools to investigate the singular value processes $\sigma_i(t)$ and the right singular vector processes $v_i(t)$ of the Dyson Bessel process $\Phi(t)$.

3.1.1 Previous works

The literature on random matrix perturbation bounds spans several decades and encompasses a wide range of theoretical and applied research. Early works by Davis and Kahan (1970) [14] and Wedin (1972) [63] laid the groundwork for perturbation theory in numerical linear algebra, providing rigorous bounds on the sensitivity of eigenvalues and singular values to matrix perturbations. These seminal results formed the basis for subsequent developments in the field, including extensions to more general perturbation models (for example [58]) and applications in statistical estimation and machine learning.

In recent years, there has been a surge of interest in developing perturbation bounds for random matrix models, driven by the increasing prevalence of stochastic processes in modern data analysis and scientific computing. Studies by Higham (2002) [24] and Martinsson et al. (2019) [47] have investigated the effects of random perturbations, such as Gaussian noise, on the spectral properties of matrices, highlighting the importance of robust numerical algorithms in the presence of uncertainty.

Moreover, researchers have explored the applications of random matrix perturbation bounds in various domains, including subspace recovery, low-rank approximation, and spectral clustering. For example, Zou et al. (2006) [71] demonstrated the error of perturbation bounds in sparse principal component analysis, while Mahoney (2011) [43] utilized random matrix theory to develop efficient algorithms for large-scale numerical linear algebra problems.

3.1.2 Outline

The rest of this chapter is outlined as follows: Section 3.2 presents a precise formulation of the problem, main results, and two applications to matrix perturbation theory. Section 3.3 is devoted to the preliminaries that are needed for the proof of the main theorem. In Section 3.4, we first show the outline of the proof and then present in detail every step to achieve the main theorem and the proofs of two applications. Section 3.5 investigates the tightness of our bounds with numerical examples. The conclusion and some potential future works are summarized in Section 3.6. Appendix A.6 gives additional comparisons to a previous bound.

3.2 Problem Formulation

Suppose we are given a set of values $\gamma_1 \geq \cdots \geq \gamma_k \geq 0$ for some $k \in [d]$. Let $\Gamma := \operatorname{diag}\{\gamma_1, \cdots, \gamma_d\}$, where $\gamma_i := 0$ for $i > k$. Given a rectangular matrix $A \in \mathbb{R}^{m \times d}$ with singular values $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$ and corresponding right-singular vectors $V = [v_1, \cdots, v_d] \in \mathbb{R}^{d \times d}$, we consider the perturbed matrix $\hat{A} = A + \sqrt{T}G$ for some $T > 0$, where G is a matrix with independent and identically distributed (i.i.d.) standard normal entries. For the perturbed matrix $\hat{A}$, let its right-singular vectors be denoted as $\hat{V} = [\hat{v}_1, \cdots, \hat{v}_d] \in \mathbb{R}^{d \times d}$.

We consider the problem of obtaining a bound on the Frobenius norm perturbation error $\mathbb{E}\left[\|\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top\|_F^2\right]$,

$$\mathbb{E}\left[\|\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top\|_F^2\right] \leq O(m, d, k, T, \gamma_i, \sigma_i)? \tag{3.4}$$

Throughout this chapter, in our notation, denoted by O, we denote a multiple of C, where C denotes a positive constant. For simplicity, we refrain from explicitly stating the constant C.

3.2.1 Notations

Before we present our results, we give the notation used throughout this chapter. Without loss of generality, we may assume $m \geq d$.

- $(\gamma_1, \cdots, \gamma_d)$, a set of specified values
- $(\sigma_1, \cdots, \sigma_d)$, singular values of matrix A
- $(\sigma_1(t), \cdots, \sigma_d(t))$, singular value processes of $\Phi(t) = A + \sqrt{t}G$
- $(\hat{\sigma}_1, \cdots, \hat{\sigma}_d)$, singular values of perturbed matrix $\hat{A} = A + \sqrt{T}G = \Phi(T)$
- $\Delta_{ij}(t) = \sigma_i(t) - \sigma_j(t), i < j$, singular value gaps of $\Phi(t) = A + \sqrt{t}G$
- $(v_1, \cdots, v_d)$, right singular vectors of matrix A

- $(v_1(t), \cdots, v_d(t))$, right singular vector processes of $\Phi(t) = A + \sqrt{t}G$
- $(\hat{v}_1, \cdots, \hat{v}_d)$, right singular vectors of perturbed matrix $\hat{A} = A + \sqrt{T}G = \Phi(T)$

3.2.2 Main results of matrix perturbation bounds

The main result (Theorem 3.2.1) gives a new upper bound under the Frobenius-norm error of the Gaussian perturbation for the rectangular random matrix approximation problem.

Before we state the main result, we present the assumption on the gaps in the top $k+1$ singular values $\sigma_1 \geq ... \geq \sigma_{k+1}$ of the matrix A, required by our results. As illustrated in Figure 3.1, this assumption is satisfied on many real-world datasets.

Assumption 3.2.1 ($(A, k, T, \sigma, \gamma)$ Singular value gaps)**.** *The gaps in the top $k+1$ singular values $\sigma_1 \geq ... \geq \sigma_d$ of the matrix $A \in \mathbb{R}^{m\times d}$ satisfy $\sigma_i - \sigma_{i+1} \geq 8T\sqrt{d}\log(\frac{1}{\delta})$ for every $i \in [k]$, where $\delta := \frac{1}{8d\gamma_1^2} \times \frac{\gamma_1^2 - \gamma_d^2}{(\sigma_1 - \sigma_d)^2}$. Further, we assume $\sigma_k - \sigma_j \geq 4T\sqrt{d}$ for any $j > k$.*

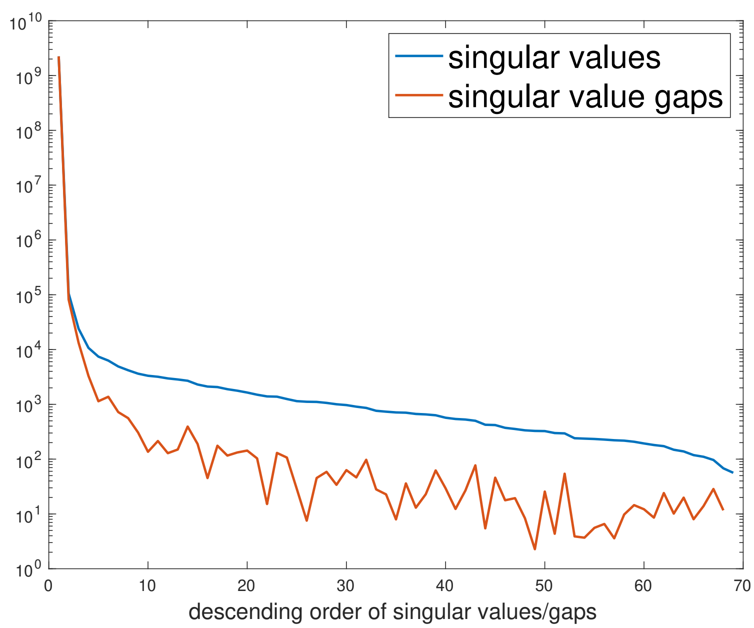


Figure 3.1: US census 1990 dataset (data source see [49]): the singular values decay *exponentially* fast. The horizontal axis shows the descending order of singular values, and the vertical axis shows a log plot of corresponding singular values

Let $A = U\Sigma V^\top$ be a singular value decomposition of A, and let $\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^\top$ be a singular value decomposition of $\hat{A}$. Note that the eigenvalue decompositions of $A^\top A = V\Sigma^\top\Sigma V^\top$ and of $\hat{A}^\top\hat{A} = \hat{V}\hat{\Sigma}^\top\hat{\Sigma}\hat{V}^\top$ involve only right singular vector matrices V, $\hat{V}$.

We now state our main result.

Theorem 3.2.1 (Main result)**.** *Let $T > 0$. Given a rectangular matrix $A \in \mathbb{R}^{m\times d}$ with singular values $\sigma_1 \geq ... \geq \sigma_d \geq 0$ and corresponding orthonormal right-singular vectors $v_1, ...v_d$ (denote $V = [v_1, ...v_d] \in \mathbb{R}^{d\times d}$). Let G be a matrix with i.i.d. $N(0, 1)$ entries, and consider the perturbed matrix $\hat{A} := A + \sqrt{T}G \in \mathbb{R}^{m\times d}$.*

Define $\hat{\sigma}_1 \geq ... \geq \hat{\sigma}_d \geq 0$ to be the singular values of $\hat{A}$ with corresponding orthonormal right-singular vectors $\hat{v}_1, ...\hat{v}_d$ (denote $\hat{V} = [\hat{v}_1, ...\hat{v}_d]$).

Let $\gamma_1 \geq ... \geq \gamma_d \geq 0$ and $k \in [d]$ be any numbers such that $\gamma_i = 0$ for $i > k$, and define $\Gamma :=$ diag $(\gamma_1, ..., \gamma_d)$. Then if A satisfies Assumption 3.2.1 for $(A, k, T, \sigma, \gamma)$, we have

$$\mathbb{E}\left[\|\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top\|_F^2\right] \leq O\left(\sum_{i=1}^{k}\sum_{j=i+1}^{d}\frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\right)T. \tag{3.5}$$

3.2.3 Two applications to matrix theory

Application to subspace recovery.

For the subspace recovery problem, we plug in $\gamma_i = 1$ for all $i \leq k$, and $\gamma_i = 0$ for all $i > k$ to the Theorem 3.2.1. The perturbed matrix indeed gives a projection matrix, and we obtain error bounds for the subspace recovery problem.

Corollary 3.2.1 (Subspace recovery)**.** *Let $T > 0$. Given a rectangular matrix $A \in \mathbb{R}^{m\times d}$ with singular values $\sigma_1 \geq ... \geq \sigma_d \geq 0$ and right-singular vectors $v_1, ...v_d$ (denoted $V \in \mathbb{R}^{d\times d}$). Let G be a matrix with i.i.d. $N(0, 1)$ entries, and consider the perturbed matrix $\hat{A} = A + \sqrt{T}G$.*

For any $k \in [d]$, define the $d \times k$ matrices $V_k = [v_1, ...v_k]$ and $\hat{V}_k = [\hat{v}_1, ...\hat{v}_k]$ where $\hat{v}_1, ...\hat{v}_d$ are corresponding orthonormal right-singular vectors of $\hat{A}$. Then if A satisfies Assumption 3.2.1 for $(A, k, T, \sigma, \gamma)$ where $\gamma = (1, \cdots, 1, 0, \cdots, 0)$ is the vector with the first k entries equal to 1, we have

$$\mathbb{E}\left[\|\hat{V}_k\hat{V}_k^\top - V_kV_k^\top\|_F\right] \leq O(\frac{\sqrt{kd}}{(\sigma_k-\sigma_{k+1})})\sqrt{T}. \tag{3.6}$$

Moreover, if we further have that $\sigma_i - \sigma_{i+1} \geq \Omega(\sigma_k - \sigma_{k+1})$ for all $i \leq k$, then

$$\mathbb{E}\left[\|\hat{V}_k\hat{V}_k^\top - V_kV_k^\top\|_F\right] \leq O(\frac{\sqrt{d}}{(\sigma_k-\sigma_{k+1})})\sqrt{T}. \tag{3.7}$$

Theore 3.2.1 improves by $\sqrt{k}\frac{\sqrt{m}}{\sqrt{d}}$ (in expectation) on the bound implied by the Davis-Kahan-Wedin sine theorem [14, 63] which, in the Gaussian case, is

$$||\hat{V}_k\hat{V}_k^\top - V_kV_k^\top||_F \leq O(\frac{\sqrt{km}}{\sigma_k-\sigma_{k+1}}),$$

with high probability.

Further when $\frac{\sqrt{r}}{\sqrt{d}} \geq (\frac{\sigma_{k+1}}{\sigma_k} - \frac{k-1}{k})$, i.e. $(\frac{\sigma_{k+1}}{\sigma_k} - \frac{k-1}{k})^2 d \leq r \leq d$, or when the matrix

A is full rank, i.e. $r = d$, the bound in Corollary 3.2.1 improves by a factor of at least $\sqrt{k}$ (in expectation) on the bound of [54] (their Theorem 18), which says that for any $t > 0$, rank $r > 0$, with high probability

$$\|\hat{V}_k\hat{V}_k^\top - V_kV_k^\top\|_F \leq 4\sqrt{2}k\left(\frac{t\sqrt{r}}{\sigma_k - \sigma_{k+1}} + \frac{4m}{\sigma_k(\sigma_k - \sigma_{k+1})} + \frac{2\sqrt{m}}{\sigma_k}\right). \tag{3.8}$$

Moreover, in contrast to their bound (3.8) which grows linearly with m, our bound only grows linearly with $\sqrt{d}$, which may be much smaller than m since in many applications $d << m$. Furthermore, when $\frac{\sqrt{r}}{\sqrt{d}} \geq \frac{\sigma_{k+1}}{\sigma_k}$, or when the matrix A is full rank, our Corollary 3.2.1 improves by a factor of k on the bound shown in Theorem 18 of [54], and also improves the dependence on m to a dependence on d.

Application to rank-k matrix approximation.

For the rank-k matrix approximation problem, we plug in $\gamma_i = \sigma_i$ for all $i \leq k$, and $\gamma_i = \sigma_i$ for all $i > k$. The perturbed matrix in Theorem 3.2.1 outputs rank-k matrix approximation.

Corollary 3.2.2 (Rank-k matrix approximation)**.** *Let $T > 0$. Given a rectangular matrix $A \in \mathbb{R}^{m\times d}$ with singular values $\sigma_1 \geq ... \geq \sigma_d \geq 0$ and with right-singular vectors $v_1, ...v_d$, where we define $V := [v_1, ...v_d] \in \mathbb{R}^{d\times d}$. Let G be a matrix with i.i.d. $N(0, 1)$ entries, and consider the perturbed matrix that outputs $\hat{A} = A + \sqrt{T}G$.*

For any $k \in [d]$, define $\Sigma_k :=$ diag $(\sigma_1, ..., \sigma_k, 0, ...0)$. Define $\hat{\sigma}_1 \geq ... \geq \hat{\sigma}_d \geq 0$ to be the singular values of $\hat{A}$ with corresponding orthonormal right-singular vectors $\hat{v}_1, ...\hat{v}_d$, where we define $\hat{V} := [\hat{v}_1, ...\hat{v}_d]$, and define $\hat{\Sigma}_k :=$ diag $(\hat{\sigma}_1, ..., \hat{\sigma}_k, 0, ...0)$. Then if A satisfies Assumption 3.2.1 for $(A, k, T, \sigma, \gamma)$ for $\gamma = (\sigma_1, \cdots, \sigma_k, 0, \cdots, 0)$, we have

$$\mathbb{E}\left[||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - V\Sigma_k^\top\Sigma_kV^\top||_F^2\right] \leq O\left(d\|\Sigma_k\|_F^2 + k\sum_{j=k+1}^{d}\left(\sigma_k\frac{\sigma_k}{\sigma_k - \sigma_j}\right)^2\right)T. \tag{3.9}$$

In particular, Corrollary 3.2.2 implies that

$$\sqrt{\mathbb{E}\left[||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - V\Sigma_k^\top\Sigma_kV^\top||_F^2\right]} \leq O\left(\sqrt{k}\sqrt{d}\left(\sigma_1 + \sigma_k\frac{\sigma_k}{\sigma_k - \sigma_{k+1}}\right)\right)T.$$

Corollary 3.2.2 improves by a factor of k on the bound $||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - V\Sigma_k^\top\Sigma_kV^\top||_F \leq O(\sqrt{k}\sqrt{m}(k + \frac{\sigma_k}{\sigma_k - \sigma_{k+1}}))$ implied by the Davis-Kahan-Wedin sine theorem [14, 63] (see Appendix A.6 for details). Further, when the rank r of the matrix A satisfies $r \geq (\frac{\sigma_{k+1}}{\sigma_k})^2 d$ or when the matrix A is full rank, applying the perturbation bound of Theorem 18 of [54] gives with high probability,

$$||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - V\Sigma_k^\top\Sigma_kV^\top||_F \leq 4k\sqrt{m}(k + \frac{\sigma_k}{\sigma_k - \sigma_{k+1}}), \tag{3.10}$$

(see Appendix A.6 for details). Hence, Corollary 3.2.2 improves by a factor of $k^{1.5}$ on the bound implied by Theorem 18 [54].

Remark 3.2.1 (Tightness for full-rank special case)**.** *In the special case where $k = d$, we have $\|(A+\sqrt{T}G)^\top(A+\sqrt{T}G) - A^\top A\|_F = \|\sqrt{T}A^\top G + \sqrt{T}G^\top A + TG^\top G\|_F = \Theta(||A^\top G||_F\sqrt{T}) = \Theta(||\Sigma_d||_F\sqrt{d}\sqrt{T})$ with high probability. Thus, Corollary 3.2.2 is tight for this special case.*

The last equality above holds w.h.p. because $||A^\top G||_F^2 = \mathrm{tr}(G^\top AA^\top G) = \mathrm{tr}(G^\top \Sigma_d \Sigma_d^\top G) = \mathrm{tr}(\Sigma_d\Sigma_d^\top GG^\top) = \|\Sigma_d\|_F^2 d$ w.h.p., where we may assume without loss of generality that A is a diagonal matrix because the distribution of G is invariant w.r.t. multiplication by orthogonal matrices.

3.3 Preliminaries

In this section, we present the preliminary materials as needed in the main theorems and main proofs. We discuss the previously mentioned Dyson Bessel process in Section 3.3.1. Next, we present the diffusions of right-singular vectors in Section 3.3.2. In Section 3.3.3, we show deterministic perturbation bounds derived by Weyl, Davis-Kahan, and Wedin.

3.3.1 Dyson Bessel process

We express the perturbed matrix as a matrix-valued Dyson Bessel process $\Phi(t)$ valued at a certain time T with the initial condition $\Phi(0) = A$. Through this perspective, we leverage stochastic analysis tools to the singular value processes $\sigma_i(t)$ and the right singular vector processes $v_i(t)$ of the Dyson Bessel process $\Phi(t)$.

At every time $t > 0$ the singular values $\sigma_1(t), \dots, \sigma_d(t)$ of $\Phi(t)$ are distinct with probability 1 and equation (3.3) induces a stochastic process on the singular values $\sigma_i(t)$ and singular vectors $v_i(t)$. The process can be expressed via the following diffusion equations. The dynamics of the singular values $\sigma_i(t)$ of the Dyson-Bessel process are given by (see Theorem 1 in [6]),

$$\mathrm{d}\sigma_i(t) = \mathrm{d}\beta_{ii}(t) + \left(\frac{1}{2\sigma_i(t)}\sum_{j\neq i}\frac{(\sigma_i(t))^2 + (\sigma_j(t))^2}{(\sigma_i(t))^2 - (\sigma_j(t))^2} + \frac{m-1}{2\sigma_i(t)}\right)\mathrm{d}t, 1 \le i \le d, \tag{3.11}$$

where $\beta_{ii}, 1 \le i \le d$ is a family of independent one-dimensional Brownian motions.

3.3.2 Right singular vector SDE

The dynamics of right singular vectors $v_i(t)$ of the Dyson-Bessel process are given by (see Theorem 2 in [6]),

$$\begin{aligned} \mathrm{d}v_i(t) &= \sum_{j\neq i} v_j(t)\sqrt{\frac{(\sigma_j(t))^2+(\sigma_i(t))^2}{((\sigma_j(t))^2-(\sigma_i(t))^2)^2}}\mathrm{d}\beta_{ji}(t) - \frac{1}{2}v_i(t)\sum_{j\neq i}\frac{(\sigma_j(t))^2+(\sigma_i(t))^2}{((\sigma_j(t))^2-(\sigma_i(t))^2)^2}\mathrm{d}t \\ &= \sum_{j\neq i} v_j(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t) - \frac{1}{2}v_i(t)\sum_{j\neq i}c_{ij}^2(t)\mathrm{d}t, \end{aligned} \tag{3.12}$$

where $\beta_{ij}, 1 \le i < j \le d$ is a family of independent one-dimensional Brownian motions, and β_{ij} is skew-symmetry, i.e. $\beta_{ij}(t) = -\beta_{ji}(t)$. For convenience, in above equation (3.12), we denote $c_{ij}(t) = \sqrt{\frac{(\sigma_j(t))^2+(\sigma_i(t))^2}{((\sigma_j(t))^2-(\sigma_i(t))^2)^2}} = c_{ji}(t)$, i.e. $c_{ij}(t)$ is symmetric.

3.3.3 Previous perturbation bounds

In this section, we present previous perturbation bounds for comparison. In his work in 1912, Weyl [64] gives a deterministic perturbation bound for singular values $\sigma(A), \sigma(\hat{A})$ as follows.

Lemma 3.3.1 (Weyl's [64] deterministic bound)**.** *If E is a deterministic matrix,*

$$\max_{1\le i\le d} |\sigma_i(A) - \sigma_i(\hat{A})| \le \|E\|_2.$$

In the following, the Davis-Kahan-Wedin sine theorem addresses the perturbation bound on the singular vectors $v(A), v(\hat{A})$ under deterministic perturbation.

Lemma 3.3.2 (Davis-Kahan [14], Wedin [63] sine theorem)**.** *If E is a deterministic matrix,*

$$\sin\angle(v_i(A), v_i(\hat{A})) \le \frac{\|E\|_2}{\sigma_i(A) - \sigma_{i+1}(\hat{A})}.$$

When the perturbation matrix E is random, it is more natural to deal with $\sigma_i(A) - \sigma_{i+1}(A)$, the singular value gap of A instead of $\sigma_i(A) - \sigma_{i+1}(\hat{A})$, Lemma 3.3.2 implies the following bound.

Lemma 3.3.3 (Modified sine theorem)**.** *For a deterministic or random matrix E,*

$$\sin\angle(v_i(A), v_i(\hat{A})) \le \frac{\|E\|_2}{\sigma_i(A) - \sigma_{i+1}(A)}.$$

In the modified version, the upper bound involves only the singular values of matrix A.

3.4 Proofs of Main Results

We present an overview of the proof of Theorem 3.2.1 along with the main technical lemmas used to prove this result. Section 3.4.1 outlines the different steps in the proof. In Steps 1 and 2 we express the perturbed matrix as a matrix-values diffusion used in the proof. Steps 3, 4, and 5 present the main technical lemmas, and we complete the proof in Step 6. In Section 3.4.2, we show the full proof.

3.4.1 Outline of proof of Main Theorem 3.2.1

In this subsection, we give the proof outline of the main result Theorem 3.2.1.

1. **Step 1: Expressing the perturbed matrix as a matrix-valued Dyson Bessel process.** In order to establish the error bound, we adopt a strategy of representing the perturbed matrix as a matrix-valued Dyson Bessel process. This involves defining $\Phi(t) := A + \sqrt{t}G = A + B(t), \forall t \geq 0$, where the initial input matrix is denoted as $A = \Phi(0)$. Upon running this process for a duration of time T, the resulting output matrix is designated as $\hat{A} = \Phi(T)$, with G representing a matrix comprising independent and identically distributed standard normal entries.

2. **Step 2: Given $\Gamma =$ diag $(\gamma_1, ..., \gamma_d)$, expressing "$\Gamma^2$-*projected*" perturbed matrix as a matrix diffusion $\Psi(t)$.** Note that $A = U\Sigma V^\top$ and $\hat{A} = \hat{U}\hat{\Sigma}\hat{V}^\top$ are the singular value decompositions of A and $\hat{A}$, Our goal is to bound $\|\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top\|_F$, where $A^\top A = V\Sigma^\top\Sigma V^\top$ and $\hat{A}^\top\hat{A} = \hat{V}\hat{\Sigma}^\top\hat{\Sigma}\hat{V}^\top$ are eigenvalue decompositions of $A^\top A$ and $\hat{A}^\top\hat{A}$. To bound it, we first define a stochastic process $\Psi(t)$ such that $\Psi(0) = V\Gamma^\top\Gamma V^\top$ and $\Psi(T) = \hat{V}\Gamma^\top\Gamma\hat{V}^\top$. Then we bound the Frobenius-norm distance $\|\Psi(T) - \Psi(0)\|_F$ by integrating the stochastic derivative of $\Psi(t)$ over the time period $[0, T]$.

   At every time t, let $\Phi(t) = V(t)\Sigma(t)V(t)^\top$ be a singular value decomposition of the rectangular matrix $\Phi(t)$, where $\Sigma(t)$ is a diagonal matrix with entries $\sigma_1(t) \geq \cdots \geq \sigma_d(t)$ on the diagonal that are the singular values of $\Phi(t)$, and $V(t) = [v_1(t), \cdots, v_d(t)]$ is a $d \times d$ orthogonal matrix whose columns $v_1(t), \cdots, v_d(t)$ are an orthogonal basis of the right singular vectors of $\Phi(t)$. At every time, denote $\Psi(t)$ to be the symmetric matrix with given eigen values $\Gamma(t)^\top\Gamma(t)$ and eigenvectors given by the columns of $V(t)$: $\Psi(t) := V(t)\Gamma(t)^\top\Gamma(t)V(t)^\top, \forall t \in [0, T]$.

3. **Step 3: Computing the stochastical derivative $\mathrm{d}\Psi(t) = \sum_{i=1}^d \gamma_i^2 \mathrm{d}(v_i(t)v_i^\top(t))$.** To bound the expected squared Frobenius distance $\mathbb{E}\left[||\Psi(T) - \Psi(0)||_F^2\right]$, we first compute the stochastic derivative $\mathrm{d}\Psi(t)$ of the matrix diffusion $\Psi(t)$ (Lemma 3.4.2),

$$\mathrm{d}\Psi(t) = \sum_{i=1}^d \sum_{j\neq i} (\gamma_i^2 - \gamma_j^2)\left[\frac{c_{ij}(t)}{2}\mathrm{d}\beta_{ji}(t)(v_i(t)v_j^\top(t) + v_j(t)v_i^\top(t)) - c_{ij}^2(t)\mathrm{d}t(v_i(t)v_i^\top(t))\right].$$

where $c_{ij}(t)$ are

4. **Step 4: Bounding the singular value gaps.** The above equation for the derivative $\mathrm{d}\Psi(t)$ includes $c_{ij}(t)$, terms with magnitude proportional to the inverse of the singular value gaps $\sigma_i(t) - \sigma_j(t)$ for each $i, j \in [d]$. In order to bound these terms, we showed that the gaps $\Delta_{ij}(t)$ in the top $k+1$ singular values for any $i < j$, for time $t \in [0, T]$ satisfy (Lemma 3.4.3),
$$\Delta_{ij}(t) \geq \frac{1}{2}(\sigma_i - \sigma_j),$$
provided that the initial gaps are sufficiently large enough, see Assumption 3.2.1.

5. **Step 5: Integrating the stochastic derivative of $\mathrm{d}\Psi(t)$ over the time interval $[0, T]$.** Next we express the expected squared Frobenius distance $\mathbb{E}\left[||\Psi(T) - \Psi(0)||_F^2\right]$ as an integral $\mathbb{E}\left[||\Psi(T) - \Psi(0)||_F^2\right] = \mathbb{E}\left[|| \int_0^T \mathrm{d}\Psi(t)\|_F^2\right]$. Then we apply Itô's Lemma to derive an upper bound for this integral. In addition, the upper bound we derive (Lemma 3.4.5) is
$$\mathbb{E}\left[\|\Psi(T) - \Psi(0)\|_F^2\right] \leq 32 \int_0^T \mathbb{E}\left[\sum_{i=1}^d \sum_{j \neq i} \frac{(\gamma_i^2 - \gamma_j^2)^2}{(\sigma_i - \sigma_j)^2}\right] \mathrm{d}t + 16T \int_0^T \mathbb{E}\left[\sum_{i=1}^d \left(\sum_{j \neq i} \frac{(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2}\right)^2\right] \mathrm{d}t. \quad (3.13)$$

6. **Step 6: Complete the proof.** Plugging the singular value gap bound $\Delta_{ij}(t) \geq \frac{1}{2}(\sigma_i - \sigma_j)$ into the above expression (3.13), and noting that the second term on the right-hand side of (3.13)is at least as small as the first term, we obtain the bound in Theorem 3.2.1.

3.4.2 Proof of Main Theorem 3.2.1

Step 3: Computing the stochastical derivative $\mathrm{d}\Psi(t)$

$\Psi(t)$ is itself a matrix-valued diffusion. We first decomposite the matrix $\Psi(t)$ as a sum of its right singular vectors: $\Psi(t) = \sum_{i=1}^d \gamma_i^2 (v_i(t) v_i^\top(t))$. Thus we have
$$\mathrm{d}\Psi(t) = \sum_{i=1}^d \gamma_i^2 \mathrm{d}(v_i(t) v_i^\top(t)) \quad (3.14)$$

We begin by computing the stochastic derivative $\mathrm{d}v_i(t)v_i^\top(t)$ for each $i \in [d]$, by applying the formula in (3.12), together with Ito's Lemma.

The following lemma gives the dynamic of the right singular vectors $v_i(t)v_i(t)^\top$, note that $\lambda_i(t)$ are the eigenvalues of noisy covariance matrix $M(t) = \hat{A}^\top \hat{A}$, and $\beta_{ij}, 1 \leq i < j \leq d$ is a family of independent Brownian motion where β_{ij} is skew-symmetry, i.e. $\beta_{ij}(t) = -\beta_{ji}(t)$. For convenience, we denote $c_{ij}(t) = \sqrt{\frac{\lambda_j(t)+\lambda_i(t)}{(\lambda_j(t)-\lambda_i(t))^2}} = c_{ji}(t)$, i.e. symmetric.

The dynamic of $v_i(t)v_i(t)^\top$ is used to derive the dynamic of $\Psi(t)$, the perturbed matrix projected on the plane whose elements have the given singular values γ_i, see definition below.

Lemma 3.4.1 (Stochastic derivative of $v_i(t)v_i(t)^\top$). *For all $t \in [0, T]$,*

$$\mathrm{d}\left(v_i(t)v_i^\top(t)\right) = \sum_{j\neq i} v_j(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t) - \frac{1}{2}v_i(t)\sum_{j\neq i} c_{ij}^2(t)\mathrm{d}t.$$

Proof. The dynamic of right singular vectors [6] are the following:

$$\begin{aligned}
\mathrm{d}v_i(t) &= \sum_{j\neq i} v_j(t)\sqrt{\frac{\lambda_j(t)+\lambda_i(t)}{(\lambda_j(t)-\lambda_i(t))^2}}\mathrm{d}\beta_{ji}(t) - \frac{1}{2}v_i(t)\sum_{j\neq i}\frac{\lambda_j(t)+\lambda_i(t)}{(\lambda_j(t)-\lambda_i(t))^2}\mathrm{d}t \\
&= \sum_{j\neq i} v_j(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t) - \frac{1}{2}v_i(t)\sum_{j\neq i} c_{ij}^2(t)\mathrm{d}t.
\end{aligned}$$

Thus we have

$$\begin{aligned}
&\mathrm{d}\left(v_i(t)v_i^\top(t)\right) = \left(v_i(t) + \mathrm{d}v_i(t)\right)\left(v_i(t) + \mathrm{d}v_i(t)\right)^\top - v_i(t)v_i^\top(t) \\
&= \left(v_i + \sum_{j\neq i} v_j c_{ij}\mathrm{d}\beta_{ji}(t) - \frac{1}{2}v_i\sum_{j\neq i} c_{ij}^2\mathrm{d}t\right)\left(v_i^\top + \sum_{j\neq i} v_j^\top c_{ij}\mathrm{d}\beta_{ji}(t) - \frac{1}{2}v_i^\top\sum_{j\neq i} c_{ij}^2\mathrm{d}t\right) - v_i v_i^\top \\
&= v_i(t)\left(\sum_{j\neq i} v_j^\top(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right) - \frac{1}{2}v_i(t)v_i^\top(t)\sum_{j\neq i} c_{ij}^2(t)\mathrm{d}t + \left(\sum_{j\neq i} v_j(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right)v_i^\top(t) \\
&+ \left(\sum_{j\neq i} v_j(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right)\left(\sum_{j\neq i} v_j^\top(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right) - \frac{1}{2}v_i(t)v_i^\top(t)\sum_{j\neq i} c_{ij}^2(t)\mathrm{d}t + o(\mathrm{d}t) \\
&= v_i(t)\left(\sum_{j\neq i} v_j^\top(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right) + \left(\sum_{j\neq i} v_j(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right)v_i^\top(t) - v_i(t)v_i^\top(t)\sum_{j\neq i} c_{ij}^2(t)\mathrm{d}t \\
&+ \sum_{k\neq i}\sum_{j\neq i} v_k(t)v_j^\top(t)c_{ik}(t)c_{ij}(t)\mathrm{d}\beta_{ki}(t)\mathrm{d}\beta_{ji}(t) \\
&= v_i(t)\left(\sum_{j\neq i} v_j^\top(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right) + \left(\sum_{j\neq i} v_j(t)c_{ij}(t)\mathrm{d}\beta_{ji}(t)\right)v_i^\top(t) - v_i(t)v_i^\top(t)\sum_{j\neq i} c_{ij}^2(t)\mathrm{d}t \\
&+ \sum_{k\neq i}\sum_{j\neq i} v_k(t)v_j^\top(t)c_{ik}(t)c_{ij}(t)\delta_{kj}\delta_{ii}\mathrm{d}t \\
&= \sum_{j\neq i} c_{ij}(t)\mathrm{d}\beta_{ji}(t)(v_i(t)v_j^\top(t) + v_j(t)v_i^\top(t)) - \sum_{j\neq i} c_{ij}^2(t)\mathrm{d}t(v_i(t)v_i^\top(t) - v_j(t)v_j^\top(t)).
\end{aligned}$$

□

For a specific set of singular values γ_i, denote

$$\Psi(t) = \sum_{i=1}^{d} \gamma_i^2 (v_i(t) v_i^\top(t)) \tag{3.15}$$

and the dynamic of $\Psi(t)$ is the following. The dynamic of $\Psi(t)$ plays a crucial role in the next since its integration from $[0, T]$ gives the desired upper bound.

Lemma 3.4.2 (Stochastic derivative of $\Psi(t)$)**.** *For all $t \in [0, T]$, we have that*

$$\mathrm{d}\Psi(t) = \sum_{i=1}^{d} \sum_{j \neq i} (\gamma_i^2 - \gamma_j^2) \left[\frac{c_{ij}(t)}{2} \mathrm{d}\beta_{ji}(t)(v_i(t)v_j^\top(t) + v_j(t)v_i^\top(t)) - c_{ij}^2(t)\mathrm{d}t(v_i(t)v_i^\top(t)) \right].$$

Proof.

$$\begin{aligned}
\mathrm{d}\Psi(t) &= \sum_{i=1}^{d} \gamma_i^2 \mathrm{d}(v_i(t)v_i^\top(t)) \\
&= \sum_{i=1}^{d} \gamma_i^2 \left(\sum_{j \neq i} c_{ij}(t)\mathrm{d}\beta_{ji}(t)(v_i(t)v_j^\top(t) + v_j(t)v_i^\top(t)) - \sum_{j \neq i} c_{ij}^2(t)\mathrm{d}t(v_i(t)v_i^\top(t) - v_j(t)v_j^\top(t)) \right) \\
&= \sum_{i=1}^{d} \sum_{j \neq i} \gamma_i^2 c_{ij}(t)\mathrm{d}\beta_{ji}(t)(v_j(t)v_i^\top(t) + v_i(t)v_j^\top(t)) - \sum_{i=1}^{d} \sum_{j \neq i} \gamma_i^2 c_{ij}^2(t)\mathrm{d}t(v_i(t)v_i^\top(t) - v_j(t)v_j^\top(t)) \\
&= \frac{1}{2} \sum_{i=1}^{d} \sum_{j \neq i} (\gamma_i^2 - \gamma_j^2) c_{ij}(t)\mathrm{d}\beta_{ji}(t)(v_i v_j^\top + v_j v_i^\top) - \frac{1}{2} \sum_{i=1}^{d} \sum_{j \neq i} (\gamma_i^2 - \gamma_j^2) c_{ij}^2(t)\mathrm{d}t(v_i v_i^\top - v_j v_j^\top) \\
&= \frac{1}{2} \sum_{i=1}^{d} \sum_{j \neq i} (\gamma_i^2 - \gamma_j^2) c_{ij}(t)\mathrm{d}\beta_{ji}(t)(v_i v_j^\top + v_j v_i^\top) - \sum_{i=1}^{d} \sum_{j \neq i} (\gamma_i^2 - \gamma_j^2) c_{ij}^2(t)\mathrm{d}t v_i v_i^\top,
\end{aligned}$$

note that the last two equations are because of these observations:

$$\begin{aligned}
c_{ij}(t)\mathrm{d}\beta_{ij}(t)(v_j v_i^\top + v_i v_j^\top) &= -c_{ij}(t)\mathrm{d}\beta_{ji}(t)(v_j v_i^\top + v_i v_j^\top) \\
c_{ij}^2(t)\mathrm{d}t(v_i v_i^\top - v_j v_j^\top) &= -c_{ij}^2(t)\mathrm{d}t(v_j v_j^\top - v_i v_i^\top) \\
(\gamma_i^2 - \gamma_j^2)c_{ij}^2(t)\mathrm{d}t(v_i v_i^\top - v_j v_j^\top) &= (\gamma_j^2 - \gamma_i^2)c_{ij}^2(t)\mathrm{d}t(v_j v_j^\top - v_i v_i^\top).
\end{aligned}$$

□

Step 4: Bounding the singular value gaps

From the above equation, we need bound estimations for the singular gaps $\Delta_{ij}(t) = \sigma_i(t) - \sigma_j(t)$ and coefficients $c_{ij}(t)$ to analyze the bound when we integrate. Next, we show the uniform boundness of singular value gaps $\Delta_{ij}(t) = \sigma_i(t) - \sigma_j(t)$ of noisy matrix $\hat{A} = A + \sqrt{t}G$ by half of the initial singular value gaps.

Lemma 3.4.3 (Bound of singular gaps $\Delta_{ij}(t)$:)**.** *Suppose that Assumption 3.2.1 for* $(A, k, T, \sigma, \gamma)$ *is satisfied. Then for all* $t \in [0, T]$, *with probability* $1-\delta$ *where* $\delta := \frac{1}{8d\gamma_1^2} \times \frac{\gamma_1^2 - \gamma_d^2}{(\sigma_1 - \sigma_d)^2}$, *we have* $|\Delta_{ij}(t)| \geq \frac{1}{2}(\sigma_i - \sigma_j)$ *for any* $i < j$.

Proof. With probability $1-\delta$, by Theorem 4.4.5 of [62], we have $\|G\|_2 = 2\sqrt{\max\{m, d\}}\log(\frac{1}{\delta}) = 2\sqrt{m}\log(\frac{1}{\delta})$. We know $|\sigma_i(t) - \sigma_i| \leq \sigma_i + ||G||_2 = \sigma_i + 2\sqrt{m}\log(\frac{1}{\delta})$ for any i, therefore, we bound $|\Delta_{ij}(t)| = |\sigma_i(t) - \sigma_j(t)| \geq \sigma_i - \sigma_j - 4\sqrt{m}\log(\frac{1}{\delta}) \geq \frac{1}{2}(\sigma_i - \sigma_j)$ for any $i < j$ and any $t \in [0, T]$. □

The following proposition shows that the symmetric coefficients $c_{ij}(t)$ are bounded by the reciprocal of the initial singular value gaps.

Lemma 3.4.4 (Bound of coefficients $c_{ij}(t)$)**.** *Suppose that Assumption 3.2.1 for* $(A, k, T, \sigma, \gamma)$ *is satisfied. Then for all* $t \in [0, T]$, *with probability* $1-\delta$ *where* $\delta := \frac{1}{8d\gamma_1^2} \times \frac{\gamma_1^2 - \gamma_d^2}{(\sigma_1 - \sigma_d)^2}$, *we have*

$$c_{ij}(t) \leq \frac{4}{\sigma_i - \sigma_j}, \quad \textit{for any } i < j.$$

Proof. From the above bounds of coefficients $c_{ij}(t)$, we have with probability $1-\delta$

$$\begin{aligned} c_{ij}(t) &= \frac{\sqrt{\sigma_j^2(t) + \sigma_i^2(t)}}{|\sigma_j^2(t) - \sigma_i^2(t)|} \\ &\leq 2\frac{\sigma_j(t) + \sigma_i(t)}{|\sigma_j(t) - \sigma_i(t)|(\sigma_i(t) + \sigma_i(t))} \\ &= \frac{2}{|\sigma_j(t) - \sigma_i(t)|} = \frac{2}{|\Delta_{ij}(t)|} \leq \frac{4}{\sigma_i - \sigma_j}, \quad \text{for any } i < j. \end{aligned}$$

□

Step 5: Integrating the stochastic differential equation

Combining the two bounds propositions above, next, we turn to bound the error of the project matrix process $\Psi(t)$. The following lemma shows that the Frobenius-norm distance between $\Psi(T)$ and $\Psi(0)$ can be bounded by two integrals.

Lemma 3.4.5 (Bound the Frobenius error as an integral of $\Psi(t)$)**.**

$$\begin{aligned} &\mathbb{E}\left[\|\Psi(T) - \Psi(0)\|_F^2\right] \\ &\leq 64\int_0^T \mathbb{E}\left[\sum_{i=1}^d \sum_{j\neq i} \frac{(\gamma_i^2 - \gamma_j^2)^2}{(\sigma_i - \sigma_j)^2}\right] \mathrm{d}t + 32T\int_0^T \mathbb{E}\left[\sum_{i=1}^d \left(\sum_{j\neq i} \frac{(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2}\right)^2\right] \mathrm{d}t. \end{aligned} \tag{3.16}$$

Proof. Let E be the event that $|\Delta_{ij}(t)| \geq \frac{1}{2}(\sigma_i - \sigma_j)$ for any $i < j$. By Lemma 3.4.3, we have $\mathbb{P}(E) \geq 1 - \delta$.

We truncate the following error into two integrals. Whenever the event E holds, we have,

$$
\begin{aligned}
&||\Psi(T) - \Psi(0)||_F^2 = ||\int_o^T \mathrm{d}\Psi(t)||_F^2 \\
&= ||\frac{1}{2}\int_0^T \sum_{i=1}^d \sum_{j\neq i} |\gamma_i^2 - \gamma_j^2||c_{ij}(t)|\mathrm{d}\beta_{ji}(t)(v_i v_j^\top + v_j v_i^\top) - \int_0^T \sum_{i=1}^d \sum_{j\neq i} (\gamma_i^2 - \gamma_j^2)c_{ij}^2(t)\mathrm{d}t v_i v_i^\top ||_F^2 \\
&\leq \frac{1}{2}||\int_0^T \sum_{i=1}^d \sum_{j\neq i} \frac{4|\gamma_i^2 - \gamma_j^2|\mathrm{d}\beta_{ji}(t)}{\sigma_i - \sigma_j}(v_i v_j^\top + v_j v_i^\top)||_F^2 + ||\int_0^T \sum_{i=1}^d \sum_{j\neq i} \frac{16(\gamma_i^2 - \gamma_j^2)\mathrm{d}t}{(\sigma_i - \sigma_j)^2} v_i v_i^\top ||_F^2 \\
&:= \frac{1}{2} I_1 + I_2. \qquad (3.17)
\end{aligned}
$$

For the first integral I_1, define $X(t) = \int_0^T \sum_{i=1}^d \sum_{j\neq i} |\gamma_i^2 - \gamma_j^2||c_{ij}(t)|\mathrm{d}\beta_{ji}(t)(v_i v_j^\top + v_j v_i^\top)$, we know that

$$
\mathrm{d}X(t) = \sum_{i=1}^d \sum_{j\neq i} |\gamma_i^2 - \gamma_j^2||c_{ij}(t)|\mathrm{d}\beta_{ji}(t)(v_i v_j^\top + v_j v_i^\top) := \sum_{i=1}^d \sum_{j\neq i} R_{ji}(t)\mathrm{d}\beta_{ji}(t)
$$

where $R_{ji}(t) = |\gamma_i^2 - \gamma_j^2||c_{ij}(t)|(v_i v_j^\top + v_j v_i^\top)$, so its $[l, r]$ component is

$$
\mathrm{d}X(t)[l, r] = \sum_{i=1}^d \sum_{j\neq i} R_{ji}(t)[l, r]\mathrm{d}\beta_{ji}(t).
$$

Thanks to the function $f(X) := ||X||_F^2 := \sum_{l=1}^d \sum_{r=1}^d X^2[l, r]$ and Ito's Lemma, we have

$$
\begin{aligned}
\mathrm{d}f(X) &= \sum_{l=1}^d \sum_{r=1}^d 2X(t)[l, r]\mathrm{d}X(t)[l, r] + \frac{1}{2}\sum_{l=1}^d \sum_{r=1}^d 2 < \mathrm{d}X(t)[l, r], \mathrm{d}X(t)[l, r] > \\
&= \sum_{l=1}^d \sum_{r=1}^d 2X(t)[l, r] \sum_{i=1}^d \sum_{j\neq i} R_{ji}(t)[l, r]\mathrm{d}\beta_{ji}(t) + \sum_{l=1}^d \sum_{r=1}^d \sum_{i=1}^d \sum_{j\neq i} R_{ji}^2(t)[l, r]\mathrm{d}t.
\end{aligned}
$$

And then

$$\begin{aligned}
\mathbb{E}(I_1 \times 1_E) &= \mathbb{E}\left[(f(X(T)) - f(X(0)) \times 1_E\right] = 0 + \mathbb{E}[\int_0^T \sum_{l=1}^d \sum_{r=1}^d \sum_{i=1}^d \sum_{j\neq i} R_{ji}^2(t)[l,r]\mathrm{d}t \times 1_E] \\
&= \mathbb{E}[\int_0^T \sum_{l=1}^d \sum_{r=1}^d \sum_{i=1}^d \sum_{j\neq i} \left(|\gamma_i^2 - \gamma_j^2||c_{ij}(t)|(v_i v_j^\top + v_j v_i^\top)[l,r]\right)^2 \mathrm{d}t \times 1_E] \\
&= \mathbb{E}[\int_0^T \sum_{i=1}^d \sum_{j\neq i} \sum_{l=1}^d \sum_{r=1}^d \left(|\gamma_i^2 - \gamma_j^2||c_{ij}(t)|(v_i v_j^\top + v_j v_i^\top)[l,r]\right)^2 \mathrm{d}t \times 1_E] \\
&= \mathbb{E}[\int_0^T \sum_{i=1}^d \sum_{j\neq i} |||\gamma_i^2 - \gamma_j^2||c_{ij}(t)|(v_i v_j^\top + v_j v_i^\top))||_F^2 \mathrm{d}t \times 1_E] \\
&= 2\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j\neq i} (\gamma_i^2 - \gamma_j^2)^2 c_{ij}^2(t)||(v_i v_j^\top + v_j v_i^\top))||_F^2 \mathrm{d}t \times 1_E] \\
&\leq 2\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j\neq i} \frac{16(\gamma_i^2 - \gamma_j^2)^2}{(\sigma_i - \sigma_j)^2} 2\mathrm{d}t] = 64\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j\neq i} \frac{(\gamma_i^2 - \gamma_j^2)^2}{(\sigma_i - \sigma_j)^2}\mathrm{d}t], \qquad (3.18)
\end{aligned}$$

where the last inequality holds by Inequality (3.17).

For the second integral I_2, we have

$$\begin{aligned}
I_2 &= ||\int_0^T \sum_{i=1}^d \sum_{j\neq i} (\gamma_i^2 - \gamma_j^2) \frac{16\mathrm{d}t}{(\sigma_i - \sigma_j)^2} v_i v_i^\top ||_F^2 \\
&= ||\int_0^T \sum_{i=1}^d \sum_{j\neq i} \frac{16(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2} v_i v_i^\top \times 1\mathrm{d}t||_F^2 \\
&\leq \int_0^T ||\sum_{i=1}^d \sum_{j\neq i} \frac{16(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2} v_i v_i^\top ||_F^2 \mathrm{d}t \times \int_0^T 1^2 \mathrm{d}t \\
&= T\int_0^T ||\sum_{i=1}^d \sum_{j\neq i} \frac{16(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2} v_i v_i^\top ||_F^2 \mathrm{d}t \\
&= T\int_0^T \sum_{i=1}^d || \left(\sum_{j\neq i} \frac{16(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2} \right)^2 v_i v_i^\top ||_F^2 \mathrm{d}t \\
&= T\int_0^T \sum_{i=1}^d \left(\sum_{j\neq i} \frac{16(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2} \right)^2 ||v_i v_i^\top||_F^2 \mathrm{d}t \\
&= 16T\int_0^T \sum_{i=1}^d \left(\sum_{j\neq i} \frac{(\gamma_i^2 - \gamma_j^2)}{(\sigma_i - \sigma_j)^2} \right)^2 \mathrm{d}t. \qquad (3.19)
\end{aligned}$$

We can express $\mathbb{E}[||\Psi(T)-\Psi(0)||_F^2]$ as the following sum,

$$\mathbb{E}[||\Psi(T)-\Psi(0)||_F^2] = \mathbb{E}[||\Psi(T)-\Psi(0)||_F^2 \times 1_E] + \mathbb{E}[||\Psi(T)-\Psi(0)||_F^2 \times 1_{E^c}] \tag{3.20}$$

Combining (3.18) and (3.19), it follows that

$$\begin{aligned}
\mathbb{E}[||\Psi(T)-\Psi(0)||_F^2 \times 1_E] &\leq \mathbb{E}[||\Psi(T)-\Psi(0)||_F^2] \\
&\leq \mathbb{E}[\frac{1}{2} I_1 \times 1_E + I_2 \times 1_E] \\
&= \frac{1}{2}\mathbb{E}[I_1 \times 1_E] + \mathbb{E}[I_2 \times 1_E] \\
&\leq 32\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j\neq i} \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\mathrm{d}t] + 16T\int_0^T \mathbb{E}[\sum_{i=1}^d \left(\sum_{j\neq i}\frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2]\mathrm{d}t.
\end{aligned} \tag{3.21}$$

Moreover, we have

$$\begin{aligned}
\mathbb{E}[||\Psi(T)-\Psi(0)||_F^2 \times 1_{E^c}] &\leq \mathbb{P}(E^c) \\
&\leq \mathbb{E}[4||\Psi(T)\|_F^2 + 4\|\Psi(0)||_F^2 \times 1_{E^c}] \\
&\leq 8d\gamma_1^2\mathbb{P}(E^c) \\
&\leq 8d\gamma_1^2 \times \delta \\
&\leq \frac{\gamma_1^2-\gamma_d^2}{(\sigma_1-\sigma_d)^2} \\
&\leq 32\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j\neq i} \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\mathrm{d}t] + 16T\int_0^T \mathbb{E}[\sum_{i=1}^d \left(\sum_{j\neq i}\frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2]\mathrm{d}t,
\end{aligned} \tag{3.22}$$

where the fifth inequality holds since $\delta \leq \frac{1}{8d\gamma_1^2} \times \frac{\gamma_1^2-\gamma_d^2}{(\sigma_1-\sigma_d)^2}$.

Therefore, plugging (3.21) and (3.22) into (3.20), we have

$$\mathbb{E}[||\Psi(T)-\Psi(0)||_F^2] \leq 64\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j\neq i} \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\mathrm{d}t] + 32T\int_0^T \mathbb{E}[\sum_{i=1}^d \left(\sum_{j\neq i}\frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2]\mathrm{d}t.$$

□

Step 6: Completing the proof of Main Theorem

The main result (Theorem 3.2.1) gives a new upper bound under the Frobenius-norm error of the perturbed matrix for the rectangular random matrix approximation problem.

We now complete the proof of the main result.

Proof of Theorem 3.2.1. To obtain a bound for the error $\mathbb{E}\left[||\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top||_F^2\right]$, we need to set a specific T such that $\Psi(T) = \hat{V}\Gamma^\top\Gamma\hat{V}^\top$ and $\Psi(0) = V\Gamma^\top\Gamma V^\top$.

From (3.16), we have

$$
\begin{aligned}
&\mathbb{E}\left[||\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top||_F^2\right] = \mathbb{E}\left[||\Psi(T) - \Psi(0)||_F^2\right] \le \mathbb{E}\left[||Z(T) - Z(0)||_F^2\right] \\
&\le 32\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j\neq i} \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\mathrm{d}t] + 16T\int_0^T \mathbb{E}[\sum_{i=1}^d \left(\sum_{j\neq i} \frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2]\mathrm{d}t \\
&\le 64\int_0^T \mathbb{E}[\sum_{i=1}^d \sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\mathrm{d}t] + 32T\int_0^T \mathbb{E}[\sum_{i=1}^d \left(\sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2]\mathrm{d}t \\
&\le 64\int_0^T \mathbb{E}[\sum_{i=1}^k \sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\mathrm{d}t] + 32T\int_0^T \mathbb{E}[\sum_{i=1}^k \left(\sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2]\mathrm{d}t \\
&\le 64T\sum_{i=1}^k \sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2} + 32T^2\sum_{i=1}^k \left(\sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2 \\
&\le O\left(\sum_{i=1}^k \sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2} + T\sum_{i=1}^k \left(\sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2\right)T. \qquad (3.23)
\end{aligned}
$$

By the Cauchy-Schwarz inequality, we have that

$$
\begin{aligned}
\left(\sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)}{(\sigma_i-\sigma_j)^2}\right)^2 &= \left(\sum_{j=i+1}^d \frac{1}{|\sigma_i-\sigma_j|} \times \frac{|\gamma_i^2-\gamma_j^2|}{|\sigma_i-\sigma_j|}\right)^2 \\
&\le \left(\sum_{j=i+1}^d \frac{1}{(\sigma_i-\sigma_j)^2}\right) \times \left(\sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\right) \\
&\le \left(\sum_{j=i+1}^d \frac{1}{(\sqrt{d})^2}\right) \times \left(\sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\right) \\
&\le \sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}. \qquad (3.24)
\end{aligned}
$$

Finally, we have

$$
\mathbb{E}\left[||\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top||_F^2\right] \le O\left(\sum_{i=1}^k \sum_{j=i+1}^d \frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\right)T.
$$

□

3.4.3 Proof of subspace recovery bound

For the subspace recovery problem, if we plug in $\gamma_i = 1$ for all $i \leq k$, and $\gamma_i = 0$ for all $i > k$ to the Theorem 3.2.1, the mechanism outputs a projection matrix.

Proof of Corollary 3.2.1. To prove Corollary 3.2.1, we plug in $\gamma_1 = \cdots = \gamma_k = 1$ and $\gamma_{k+1} = \cdots = \gamma_d = 0$ to Theorem 3.2.1. There are two cases.

In the first case, A satisfies Assumption 3.2.1, plug in $\gamma_1 = \cdots = \gamma_k = 1$ and $\gamma_{k+1} = \cdots = \gamma_d = 0$ to Theorem 3.2.1 and we get

$$\begin{aligned}
\mathbb{E}\left[||\hat{V}_k\hat{V}_k^\top - V_kV_k^\top||_F^2\right] &= \mathbb{E}\left[||\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top||_F^2\right] \\
&\leq O\left(\sum_{i=1}^{k}\sum_{j=i+1}^{d}\frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\right)T \\
&= O\left(\sum_{i=1}^{k}\sum_{j=k+1}^{d}\frac{1}{(\sigma_i-\sigma_j)^2}\right)T \\
&\leq O\left(\sum_{i=1}^{k}\sum_{j=k+1}^{d}\frac{1}{(\sigma_k-\sigma_{k+1})^2}\right)T \\
&\leq O(\frac{kd}{(\sigma_k-\sigma_{k+1})^2})T \qquad (3.25)
\end{aligned}$$

where the first inequality holds by Theorem 3.2.1 and the second equality holds in that $\gamma_1 = \cdots = \gamma_k = 1$ and $\gamma_{k+1} = \cdots = \gamma_d = 0$.

Thanks to Jensen's Inequality, we have that

$$\mathbb{E}\left[||\hat{V}_k\hat{V}_k^\top - V_kV_k^\top||_F\right] \leq \sqrt{\mathbb{E}\left[||\hat{V}_k\hat{V}_k^\top - V_kV_k^\top||_F^2\right]} \leq O(\frac{\sqrt{kd}}{(\sigma_k-\sigma_{k+1})})\sqrt{T}.$$

In the second case, A satisfies Assumption 3.2.1 and $\sigma_i - \sigma_{i+1} \geq \Omega(\sigma_k - \sigma_{k+1})$ for all

$i \le k$, then we have

$$\begin{aligned}
\mathbb{E}\left[||\hat{V}_k\hat{V}_k^\top - V_kV_k^\top||_F^2\right] &= \mathbb{E}\left[||\hat{V}\Gamma^\top\Gamma\hat{V}^\top - V\Gamma^\top\Gamma V^\top||_F^2\right] \\
&\le O\left(\sum_{i=1}^{k}\sum_{j=i+1}^{d}\frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\right)T \\
&= O\left(\sum_{i=1}^{k}\sum_{j=k+1}^{d}\frac{1}{(\sigma_i-\sigma_j)^2}\right)T \\
&\le O\left(\sum_{i=1}^{k}\sum_{j=k+1}^{d}\frac{1}{(i-k-1)^2(\sigma_k-\sigma_{k+1})^2}\right)T \\
&\le O\left(\sum_{i=1}^{k}\frac{d}{(i-k-1)^2(\sigma_k-\sigma_{k+1})^2}\right)T \\
&\le O\left(\frac{d}{(\sigma_k-\sigma_{k+1})^2}\sum_{i=1}^{k}\frac{1}{i^2}\right)T \\
&\le O\left(\frac{d}{(\sigma_k-\sigma_{k+1})^2}\right)T
\end{aligned} \tag{3.26}$$

where the first inequality holds by Theorem 3.2.1 and the second equality holds in that $\gamma_1 = \cdots = \gamma_k = 1$ and $\gamma_{k+1} = \cdots = \gamma_d = 0$, the second inequality holds since $\sigma_i - \sigma_{i+1} \ge \Omega(\sigma_k - \sigma_{k+1})$ for all $i \le k$, and the last inequality holds in that $\sum_{i=1}^{k}\frac{1}{i^2} \le \sum_{i=1}^{\infty}\frac{1}{i^2} = O(1)$.

Thanks to Jensen's Inequality, we have that

$$\mathbb{E}\left[||\hat{V}_k\hat{V}_k^\top - V_kV_k^\top||_F\right] \le \sqrt{\mathbb{E}\left[||\hat{V}_k\hat{V}_k^\top - V_kV_k^\top||_F^2\right]} \le O(\frac{\sqrt{d}}{(\sigma_k-\sigma_{k+1})})\sqrt{T}.$$

□

3.4.4 Proof of rank-k matrix approximation bound

For the rank-k matrix approximation problem, we plug in $\gamma_i = \sigma_i$ for all $i \le k$, and $\gamma_i = \sigma_i$ for all $i > k$, the mechanism in Theorem 3.2.1 outputs rank-k matrix approximation.

Proof of Corollary 3.2.2. We first bound the quantity $\mathbb{E}\left[||\hat{V}\Sigma_k^\top\Sigma_k\hat{V}^\top - V\Sigma_k^\top\Sigma_k V^\top||_F\right]$.

Set $\gamma_i = \sigma_i$ for $i \le k$ and $\gamma_i = 0$ for $i > k$. Then by Theorem 3.2.1 we have

$$\begin{aligned}
\mathbb{E}\left[||\hat{V}\Sigma_k^\top\Sigma_k\hat{V}^\top - V\Sigma_k^\top\Sigma_k V^\top||_F^2\right] &\le O\left(\sum_{i=1}^{k}\sum_{j=i+1}^{d}\frac{(\gamma_i^2-\gamma_j^2)^2}{(\sigma_i-\sigma_j)^2}\right)T \\
&= O\left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{(\sigma_i^2-\sigma_j^2)^2}{(\sigma_i-\sigma_j)^2} + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\left(\frac{\sigma_i^2-0^2}{\sigma_i-\sigma_j}\right)^2\right)T \\
&= O\left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(\sigma_i+\sigma_j)^2 + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\left(\frac{\sigma_i^2-\sigma_k^2}{\sigma_i-\sigma_j}+\frac{\sigma_k^2}{\sigma_i-\sigma_j}\right)^2\right)T \\
&\le O\left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(\sigma_i+\sigma_j)^2 + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\left(\sigma_i+\frac{\sigma_k^2}{\sigma_i-\sigma_j}\right)^2\right)T \\
&= O\left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(\sigma_i+\sigma_j)^2 + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\left(\sigma_i+\sigma_k\frac{\sigma_k}{\sigma_i-\sigma_j}\right)^2\right)T \\
&\le O\left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(\sigma_i+\sigma_j)^2 + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\left(\sigma_i+\sigma_k\frac{\sigma_k}{\sigma_k-\sigma_j}\right)^2\right)T \\
&\le O\left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}(\sigma_i+\sigma_j)^2 + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\sigma_i^2 + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\left(\sigma_k\frac{\sigma_k}{\sigma_k-\sigma_j}\right)^2\right)T \\
&\le O\left(d\|\Sigma_k\|_F^2 + \sum_{i=1}^{k}\sum_{j=k+1}^{d}\left(\sigma_k\frac{\sigma_k}{\sigma_k-\sigma_j}\right)^2\right)T \\
&\le O\left(d\|\Sigma_k\|_F^2 + k(d-k)\left(\sigma_k\frac{\sigma_k}{\sigma_k-\sigma_{k+1}}\right)^2\right)T. \qquad (3.27)
\end{aligned}$$

We next bound the quantity $\mathbb{E}\left[||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - \hat{V}\Sigma_k^\top\Sigma_k\hat{V}^\top||_F\right]$.

In the case E_1 where $||G|| > \sqrt{\max(m,d)}\log(1/\delta)$, with probability δ, since $||\Sigma_k||_F \le \sqrt{k}\sigma_1$ and $||\hat{\Sigma}_k||_F < \sqrt{k}\sigma_1(t)$, we use the "worst-case/deterministic" bound $||\Sigma_k^\top\Sigma_k - \hat{\Sigma}_k^\top\hat{\Sigma}_k||_F < ||\Sigma_k^\top\Sigma_k||_F + ||\hat{\Sigma}_k^\top\hat{\Sigma}_k||_F < k\sigma_1 + k\sigma_1^2(t) < 4k\sigma_1^2$ and hence $\mathbb{E}[||\Sigma_k^\top\Sigma_k - \hat{\Sigma}_k^\top\hat{\Sigma}_k||_F * 1_{E_1}] < 2\sqrt{k}\sigma_1 * P(E_1) < 4k\sigma_1^2 * \delta$.

Choose $\delta < \frac{1}{k\sigma_1^2}$, then

$$\mathbb{E}[||\Sigma_k^\top\Sigma_k - \hat{\Sigma}_k^\top\hat{\Sigma}_k||_F * 1_{E_1}] < 4$$

In the other Case E_2 where $||G|| < \sqrt{\max(m,d)}\log(1/\delta)$, with probability $(1-\delta)$, $\mathbb{E}[||\Sigma_k^\top\Sigma_k - \hat{\Sigma}_k^\top\hat{\Sigma}_k||_F * 1_{E_2}] < \mathbb{E}[||(\Sigma_k - \hat{\Sigma}_k)(\Sigma_k + \hat{\Sigma}_k)||_F * 1_{E_2}] < \mathbb{E}[\sqrt{T}||G_k|| * (||\Sigma_k||_F + ||\hat{\Sigma}_k||_F) * 1_{E_2}] < \mathbb{E}[2\sqrt{kT}\sigma_1||G_k|| * 1_{E_2}] < 2\sqrt{kd}\sigma_1\log(1/\delta)\sqrt{T}$.

Finally, put the two cases together:

$$\begin{aligned}\mathbb{E}\left[||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - \hat{V}\Sigma_k^\top\Sigma_k\hat{V}^\top||_F\right] &= \mathbb{E}[||\Sigma_k - \hat{\Sigma}_k||_F] \\ &= \mathbb{E}[||\Sigma_k - \hat{\Sigma}_k||_F * 1_{E_1}] + \mathbb{E}[||\Sigma_k - \hat{\Sigma}_k||_F * 1_{E_2}] \\ &< 4 + 2\sqrt{kd}\sigma_1 \log(1/\delta)\sqrt{T} \\ &= O(\sqrt{kd}\sigma_1 \log(1/\delta))\sqrt{T}. \end{aligned} \tag{3.28}$$

Combined (3.27) and (3.28), we have

$$\begin{aligned}&\mathbb{E}\left[||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - V\Sigma_k^\top\Sigma_k V^\top||_F\right] \\ &\leq \mathbb{E}\left[||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - \hat{V}\Sigma_k^\top\Sigma_k\hat{V}^\top||_F\right] + \mathbb{E}\left[||\hat{V}\Sigma_k^\top\Sigma_k\hat{V}^\top - V\Sigma_k^\top\Sigma_k V^\top||_F\right] \\ &\leq O\left(\sqrt{d}\|\Sigma_k\|_F + \sqrt{k(d-k)}\left(\sigma_k\frac{\sigma_k}{\sigma_k - \sigma_{k+1}}\right)\right)\sqrt{T} + O(\sqrt{kd}\sigma_1\log(1/\delta))\sqrt{T} \\ &\leq O\left(\sqrt{d}\|\Sigma_k\|_F + \sqrt{k(d-k)}\left(\sigma_k\frac{\sigma_k}{\sigma_k - \sigma_{k+1}}\right)\right)\sqrt{T}. \end{aligned} \tag{3.29}$$

□

3.5 Numerical Simulations

In this section, we present numerical simulations that illustrate the theoretical results in Theorem 3.2.1, and investigate the extent to which the bounds in Theorem 3.2.1 are tight. Through numerical simulations, we show that the squared Frobenius norm error for the rank-k approximation problem is linear in the column dimension d but otherwise independent of row dimension m, confirming that the dependence on d and m in Corollary 3.2.2 is tight. On the other hand, for the subspace recovery problem, we find that the squared Frobenius norm error does not always grow with the column dimension d. This suggests that it may be possible to obtain tighter bounds for the subspace recovery problem for some classes of input matrices.

3.5.1 Simulations of rank-k matrix approximation

In this section, we present the simulations of the rank-k matrix approximation problem. Our simulations suggest that the dependence on d of our bound in Corollary 3.2.2 is tight up to a constant factor.

Specifically, in the simulation of Figure 3.2, the parameters are the following: $d = 15, k = 5$ where the x-axis is the number of dimension m. When we simulate, m starts from 20 with step size 500. For every value of m, the number of simulations is 1000.

From Figure 3.2, we learn that as m increases, the ratio of the left-hand side and right-hand side in Corollary 3.2.2 inequality (3.9) is nearly constant,

$$\frac{||\hat{V}\hat{\Sigma}_k^\top\hat{\Sigma}_k\hat{V}^\top - V\Sigma_k^\top\Sigma_k V^\top||_F^2}{d\|\Sigma_k\|_F^2 + k\sum_{j=k+1}^d(\sigma_k\dfrac{\sigma_k}{\sigma_k-\sigma_j})^2} \leq O(1). \tag{3.30}$$

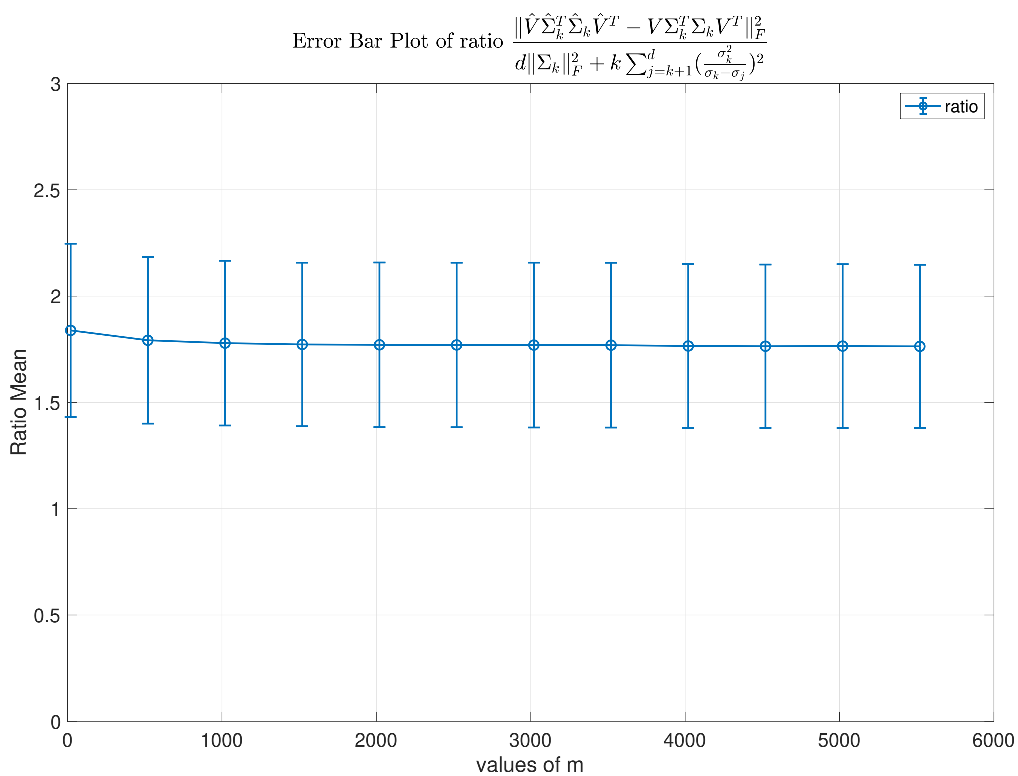


Figure 3.2: Simulation of the ratio of l.h.s. and r.h.s. of the bound in Corollary 3.2.2, when $k = 15, d = 15$.

In the simulation of Figure 3.3, the parameters are the following: $m = 2150, k = 5$ where the x-axis is the number of dimension d. When we simulate, d starts from 20 with step size 300. For every value of d, the number of simulations is 1000.

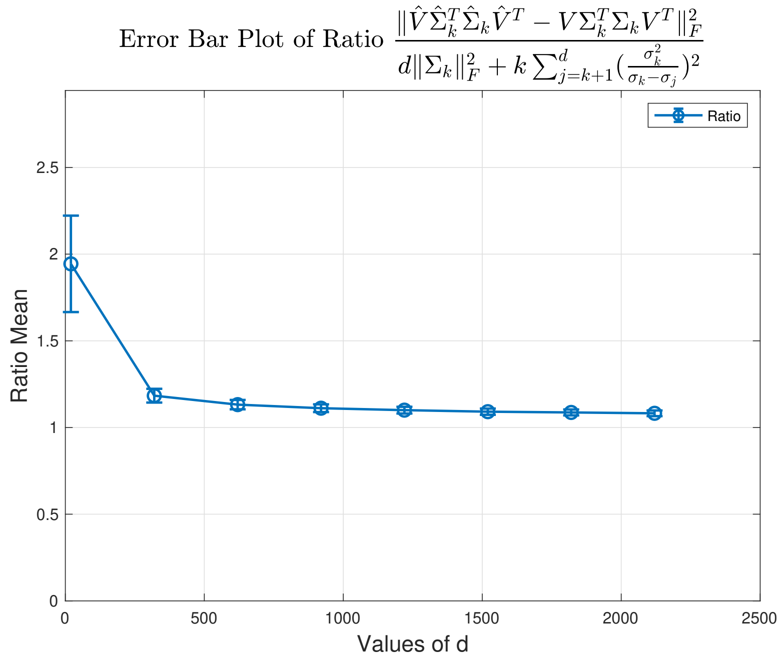


Figure 3.3: Simulation of the ratio of l.h.s. and r.h.s. of the bound in Corollary 3.2.2, when $k = 10, m = 2150$.

From Figure 3.3, we learn that as d increases, the ratio of the left-hand side and right-hand side in Corollary 3.2.2 inequality (3.9) is nearly constant.

In the simulation of Figure 3.4, the parameters are the following: $m = 850, d = 800$. When we simulate, k starts from 1 with step size 50, and the number of simulations is 1000.

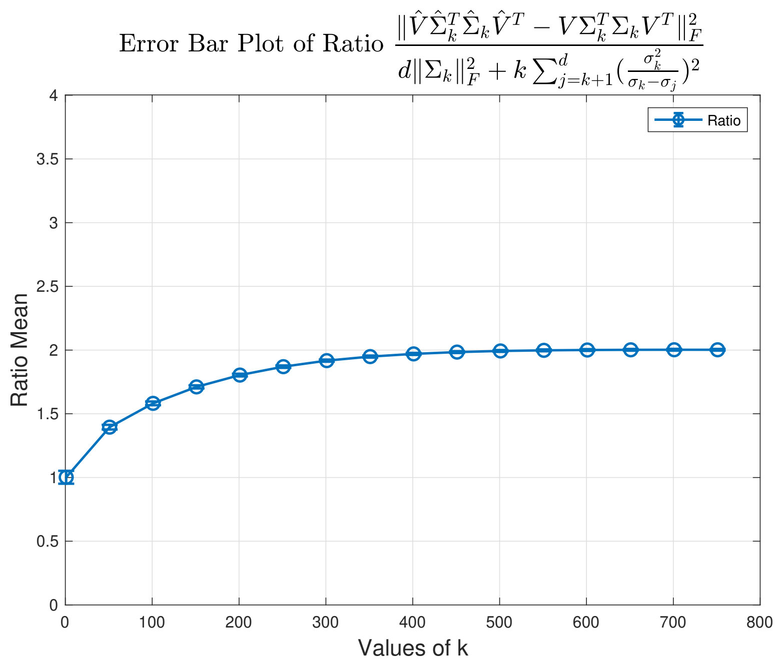


Figure 3.4: Simulation of the ratio of l.h.s. and r.h.s. of the bound in Corollary 3.2.2, when $m = 850, d = 800$.

The results suggest that the dependence of our upper bound for the rank-k matrix approximation problem in Corollary 3.2.2 is tight in m, d, and k.

3.5.2 Simulations of subspace recovery

In this section, we present the simulations of the subspace recovery problem. In the simulation of Figure 3.5, the parameters are the following: $d = 15, k = 9, m \geq 20$ where the x-axis is the number of dimension m. When we simulate, m starts from 20 with step size 500. For every value of m, the number of simulations is 1000. From Figure 3.5, we learn that the quantity

$$(\sigma_k - \sigma_{k+1})\left(\|\hat{V}_k\hat{V}_k^\top - V_kV_k^\top\|_F\right) \leq O(\sqrt{d}) \tag{3.31}$$

does not grow with the row dimension m, which is consistent with the bound in Corollary 3.2.1.

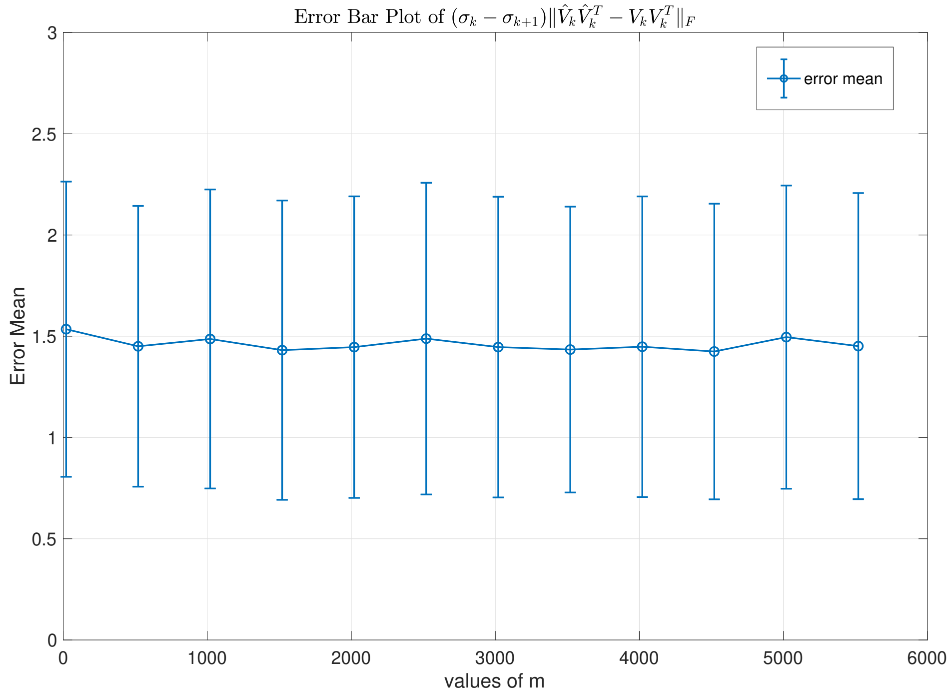


Figure 3.5: Simulation of the error of variable m when $k = 9, d = 15$.

In the simulation of Figure 3.6, the parameters are the following: $m = 850, d = 800$ where the x-axis is the number of dimension k. When we simulate, k starts from 1 with step size 50. For every value of k, the number of simulations is 1000.

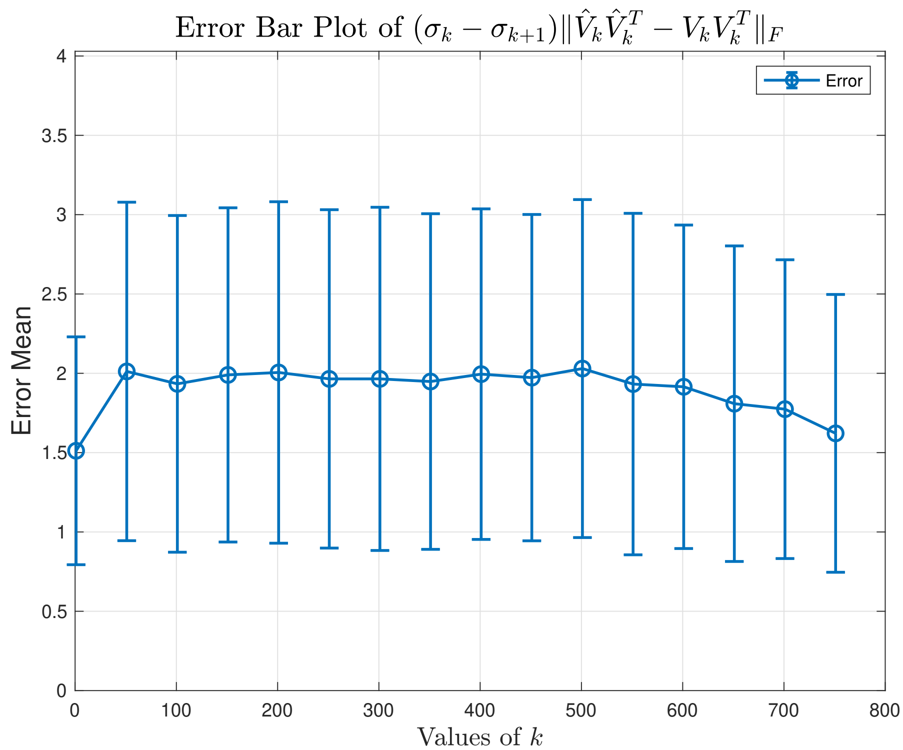


Figure 3.6: Simulation of the error of variable k when $m = 850, d = 800$.

The results suggest that the dependence of our upper bound for the subspace recovery problem in Corollary 3.2.1 is tight in both m and k.

In the simulation of Figure 3.7, the parameters are the following: $m = 2350, k = 10$ where the x-axis is the number of dimension d. When we simulate, d starts from 20 with step size 400. For every value of d, the number of simulations is 1000.

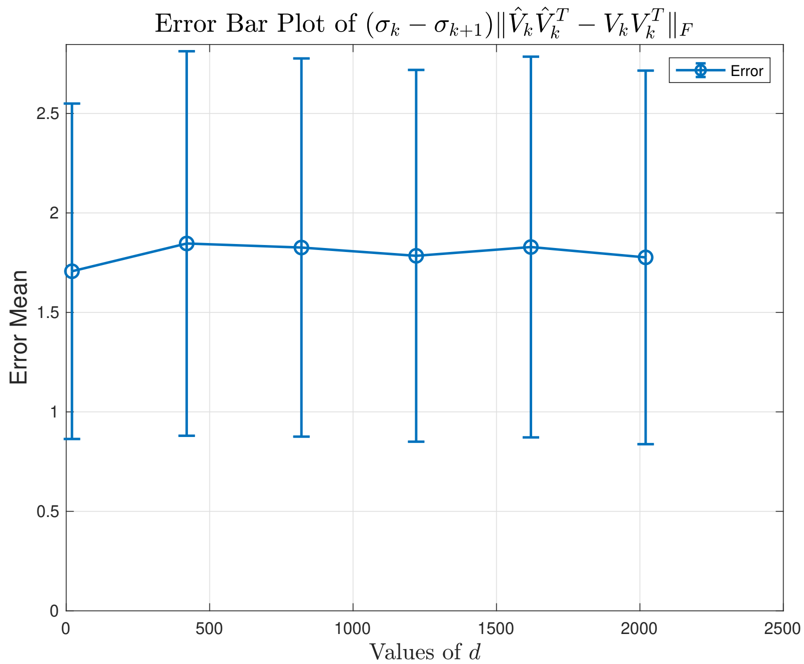


Figure 3.7: Simulation of the error of variable d when $k = 10, m = 2350$.

However, this experiment of Figure 3.7 shows that the estimated result may be improved because our bound suggests the error mean should go as nearly $\sqrt{d}$ as d increases, but from Figure 3.7, the error mean is nearly constant.

Therefore, one may consider one open question, for what kind of input matrix A will the perturbation bound be independent of dimension d?

3.6 Conclusion and Future Work

Firstly, an expansion of the analysis is warranted to encompass scenarios involving $A + GC$, where the Gaussian noises exhibit non-independence. This question is not a trivial extension of considering the perturbed matrix $AC^{-1} + \sqrt{t}G$ since the covariance matrix C will re-weight the top k singular vectors, of no use to capture the top k singular vectors of $A + GC$. In this case, we need to investigate the dynamics of singular vectors of $A + GC$. Secondly, there is scope for extending the investigation to include *rectangular* matrix perturbation bounds, thereby broadening the scope of applicability. Thirdly, efforts can be directed towards identifying illustrative examples that showcase diverse patterns of singular value decay, thus enriching the understanding of the phenomenon. Additionally, there is potential for relaxing the assumption concerning singular value gaps, thereby accommodating a wider range of practical scenarios. Lastly, exploring the implications of incorporating more general distributions for the noise, beyond the Gaussian distribution, into the analysis, presents an intriguing avenue.

This page is intentionally left blank.

A

Appendix

A.1 Some explicit solutions on LQG-MFGs

In this part, we provide explicit solutions to some LQG-MFGs without the common noise.

Suppose the position of a generic player X_t follows

$$dX_t = \alpha_t dt + \sigma dW_t,\ X_0 \sim \mathcal{N}(0,1).$$

The goal of the generic player is to minimize the running cost

$$\inf_{\alpha\in\mathcal{A}} \mathbb{E}\left[\int_0^T \left(\frac{1}{2}\alpha_t^2 + h\int_{\mathbb{R}} (X_t - y)^2 m(t,dy)\right) dt\right],$$

subject to

$$m_t = \mathcal{L}aw(X_t), \quad \forall t \in [0,T],$$

where $h \in \mathbb{R}$ is a constant.

Denote

$$V(x,t) = \inf_{\alpha} \mathbb{E}\left[\int_t^T \left(\frac{1}{2}\alpha_s^2 + h\int_{\mathbb{R}} (X_s - y)^2 m(s,dy)\right) ds \middle| X_t = x\right].$$

Note that the model can be characterized by the Hamilton-Jacobian-Bellman equation coupled by Fokker-Planck-Kolmogorov equation:

$$\begin{cases} \partial_t V + \frac{1}{2}\sigma^2 \partial_{xx} V - \frac{1}{2}(\partial_x V)^2 + F(x,m) = 0, & (t,x) \in [0,T]\times\mathbb{R}, \\ \partial_t m - \frac{1}{2}\sigma^2\partial_{xx} m - \partial_x(m\partial_x V) = 0, & (t,x)\in[0,T]\times\mathbb{R}, \\ m_0 \sim \mathcal{N}(0,1), V(x,T) = 0, & x\in\mathbb{R}, \end{cases}$$

where $F(x,m) = h\int_{\mathbb{R}}(x-y)^2 m(dy)$.

The monotonicity condition on the source term F in the variable m plays a crucial role in the uniqueness of the MFG system. A monotone function $f : \mathbb{R} \mapsto \mathbb{R}$ is said to be increasing if it satisfies $(f(x_1) - f(x_2))(x_1 - x_2) \geq 0$, and decreasing if $-f$ is increasing. This definition can be generalized to an infinite-dimensional function $F(x,m)$.

Definition A.1.1. *The real function F on $\mathbb{R} \times \mathcal{P}_2(\mathbb{R})$ is said to be monotone, if, for all $m \in \mathcal{P}_2(\mathbb{R})$, the mapping $\mathbb{R} \ni x \mapsto F(x,m)$ is at most of quadratic growth, and for all m_1, m_2 it satisfies*

$$\int_{\mathbb{R}} (F(x,m_1) - F(x,m_2))\, d(m_1 - m_2)(x) \geq 0.$$

F is said to be anti-monotone, if $(-F)$ is monotone.

According to [8], if F is monotone, then MFGs have at most one solution. Interestingly, the monotonicity of F is dependent on the sign of h.

Lemma A.1.1. *$F(x,m) = h\int_{\mathbb{R}}(x-y)^2 m(dy)$ is monotone if $h < 0$, and anti-monotone if $h > 0$.*

A natural question is how the MFG system behaves differently to the monotonicity of F?

Case I: $h > 0$

Lemma A.1.2. *For $h > 0$, there exists a solution (may not be unique) to the MFG system in the form of $V(x,t) = f_1(t)x^2 + f_3(t)$ and $m(t) \sim \mathcal{N}(0, \gamma(t))$, where*

$$f_1(t) = \sqrt{\frac{h}{2}}\frac{1 - e^{-2\sqrt{2h}(T-t)}}{1 + e^{-2\sqrt{2h}(T-t)}} \ , \ \gamma(t) = e^{-\int_0^t 4f_1(s)ds}\left(1 + \int_0^t \sigma^2 e^{\int_0^s 4f_1(u)du} ds\right),$$
$$f_3(t) = \int_t^T (\sigma^2 f_1(s) + h\gamma(s))ds.$$

Case II: $h < 0$

Lemma A.1.3. *For $h < 0$, there exists a unique solution in $(t_0, T]$ to the MFG system in the form of $V(x,t) = g_1(t)x^2 + g_3(t)$ and $m(t) \sim \mathcal{N}(0, \lambda(t))$, where*

$$g_1(t) = -\sqrt{-\frac{h}{2}}\tan\left(\sqrt{-2h}(T-t)\right) \ , \ \lambda(t) = e^{-\int_0^t 4g_1(s)ds}\left(1 + \int_0^t \sigma^2 e^{\int_0^s 4g_1(u)du} ds\right),$$
$$g_3(t) = \int_t^T (\sigma^2 g_1(s) + h\lambda(s))ds \ , \ t_0 = max\left(0, T - \frac{1}{\sqrt{-2h}}\frac{\pi}{2}\right).$$

Remark

When $h > 0$, the cost is anti-monotone, and there exists at least one global solution. When $h < 0$, the cost is monotone, and there exists at most one solution. Unfortunately, this solution lives in a short period. Lemma A.1.3 coincides with the notes in Section 3.8 of [10] saying that due to the opposite time evolution of the system of HJB-FPK, the existence of the solution may exist for only a short period.

A.2 Dynkin's formula for a regime-switching diffusion with a quadratic function

Since the running cost (1.10) has a quadratic growth in the state variable, the value function $V[\hat{m}](y,x,t)$ is expected to possess similar growth. Next, we present a version of Dynkin's formula for the functions of quadratic growth, which is sufficient for our purpose. Throughout this subsection, we will use K in various places as a generic constant that varies from line to line. The notions of this subsection are independent of other parts of the paper.

Lemma A.2.1. *Let X be the $\mathbb{R}^d$-valued process satisfying*

$$X_t = X_0 + \int_0^t \left(\tilde{b}_1(Y_s,s)X_s + \tilde{b}_2(Y_s,s)\alpha_s\right) ds + \int_0^t \sigma(s) dW_s,$$

where Y is CTMC with a generator

$$Y \sim Q = (q_{ij})_{i,j=1,2,\ldots,\kappa},$$

Suppose $\sigma(\cdot)$, $\tilde{b}_1(y,\cdot)$ and $\tilde{b}_2(y,\cdot)$ are continuous functions on $[0,T]$ for every $y \in \mathcal{Y} := \{1,2,\ldots,\kappa\}$. If $X_0 \in L^4$, $\alpha \in L^4_{\mathbb{F}}$ and $f : \mathcal{Y} \times R^d \times \mathbb{R} \mapsto \mathbb{R}$ satisfies, for some large K

$$\sup_{y\in\mathcal{Y}, t\in[0,T]} \{|f(y,x,t)|+(1+|x|)|\nabla f(y,x,t)|+(1+|x|)^2|\Delta f(y,x,t)|+|\partial_t f(y,x,t)|\} \leq K(|x|^2+1),$$

then the following identity holds for all $t \in [0,T]$:

$$\mathbb{E}\left[f(Y_t,X_t,t)\right] = \mathbb{E}\left[f(Y_0,X_0,0)\right] + \mathbb{E}\left[\int_0^t (\partial_t + \mathcal{L}^{\alpha_s} + \mathcal{Q}) f(Y_s,X_s,s) ds\right],$$

where

$$\mathcal{L}^a f(y,x,s) = \left(\frac{1}{2} Tr\left(\sigma_s \sigma_s^\top \Delta\right) + \left(\tilde{b}_{1y} x + \tilde{b}_{2y} a\right) \cdot \nabla_x\right) f(y,x,s)$$

and

$$\mathcal{Q} f(y,x,s) = \sum_{i=1}^n q_{y,i} f(i,x,s).$$

Proof. It's enough to show that the local martingale defined by Itô's formula

$$M_t^f = f(Y_t,X_t,t) - f(Y_0,X_0,0) - \int_0^t (\partial_t + \mathcal{L}^{\alpha_s} + \mathcal{Q}) f(Y_s,X_s,s) ds \qquad \text{(A.1)}$$

is uniformly integrable, hence is a true martingale.

First, note that from the assumptions on X_0 and α, we have

$$\begin{aligned}\mathbb{E}\left[\|X_t\|^4\right] &\le K\mathbb{E}\left[\|X_0\|^4 + \int_0^t \|\tilde{b}_1(Y_s,s)X_s + \tilde{b}_2(Y_s,s)\alpha_s\|^4 ds + \int_0^t \|\sigma_s W_s\|^4 ds\right]\\ &\le K\mathbb{E}\left[\|X_0\|^4 + \int_0^t \|X_s\|^4 ds + \int_0^t \|\alpha_s\|^4 ds + \int_0^t \|\sigma_s W_s\|^4 ds\right]\\ &\le K + K\int_0^t \mathbb{E}\left[\|X_s\|^4\right] ds,\end{aligned}$$

where K is a generic constant that varies from line to line. Then, by the Grönwall's inequality,

$$\mathbb{E}\left[\|X_t\|^4\right] \le Ke^{Kt} \le K,$$

which implies that $\{X_t : 0 \le t \le T\}$ is L^4 bounded uniformly in t.

On the other hand, since $x \mapsto f(y,x,t)$ is at most quadratic growth uniformly in (y,t), we conclude that $f(Y_t, X_t, t)$ is uniformly L^2 bounded from the fact

$$\sup_{t\in[0,T]} \mathbb{E}\left[f^2(Y_t,X_t,t)\right] \le K \sup_{t\in[0,T]} \mathbb{E}\left[\|X_t\|^4\right] + K \le K.$$

The uniform L^2-boundedness of $\int_0^t \partial_t f(Y_s,X_s,s)ds$ follows from our assumption on $\partial_t f$. Similarly, since $\mathcal{Q}f$ has a quadratic growth uniformly in y and t, and

$$\left\{\int_0^t \mathcal{Q}f(Y_s,X_s,s)ds : 0 \le t \le T\right\}$$

is L^2 bounded. At last, we have

$$\begin{aligned}&\mathbb{E}\left[\left(\int_0^t \mathcal{L}^{\alpha_s} f(Y_s,X_s,s)ds\right)^2\right]\\ &\le K\mathbb{E}\left[\int_0^t \left(\left(\tilde{b}_1(Y_s,s)X_s + \tilde{b}_2(Y_s,s)\alpha_s\right)\cdot\nabla f + \frac{1}{2}\mathrm{Tr}\left(\sigma_s\sigma_s^\top \Delta f\right)\right)^2 (Y_s,X_s,s)ds\right]\\ &\le K\mathbb{E}\left[\int_0^t \|\tilde{b}_1(Y_s,s)X_s + \tilde{b}_2(Y_s,s)\alpha_s\|^2 \|\nabla f\|^2 (Y_s,X_s,s)ds\right]\\ &\qquad + K\mathbb{E}\left[\int_0^t \frac{1}{4}\|\mathrm{Tr}\left(\sigma_s\sigma_s^\top\Delta f\right)\|^2 (Y_s,X_s,s)ds\right]\\ &\le K\mathbb{E}\left[\int_0^t \|\alpha_s\|^4 ds\right] + K\mathbb{E}\left[\int_0^t \|X_s\|^4 ds\right] + K\mathbb{E}\left[\int_0^t |\nabla f|^4 (Y_s,X_s,s)ds\right]\\ &\qquad + K\mathbb{E}\left[\int_0^t \frac{1}{4}\|\mathrm{Tr}\Delta f\|^2 (Y_s,X_s,s)\, ds\right].\end{aligned}$$

Since ∇f is linear growth in x, the second term $\sup_{t\in[0,T]} \mathbb{E}\left[\int_0^t \|\nabla f\|^4(Y_s,X_s,s)ds\right]$ is finite. Together with assumptions on Δf and α, we have uniform L^2-boundedness of $\int_0^t \mathcal{L}^{\alpha_s} f(Y_s,X_s,s)ds$.

As a result, each term of the right-hand side of (A.1) is uniform L^2-bounded in t, and thus M_t^f belongs to $L^2_{\mathbb{F}}$ and this implies the uniform integrability. $\square$

A.3 Proof of the property of G

Lemma A.3.1. *Define*

$$\mathcal{E}_t(\phi) = \exp\left\{\int_0^t \phi_s ds\right\},$$

and

$$G_t(x, \phi_1, \phi_2, \phi_3, W) = x\mathcal{E}_t(\phi_1 - \phi_2) + \mathcal{E}_t(\phi_1 - \phi_2)\int_0^t \mathcal{E}_s(-\phi_1 + \phi_2)\left(\phi_2(s)\phi_3(s)ds + dW_s\right),$$

where x is a given constant, ϕ_1, ϕ_2, ϕ_3 are RCLL functions on $[0, T]$. Then

$$\begin{aligned}&\mathbb{E}\left[\left|G_t(x^1, \phi_1, \phi_2^1, \phi_3^1, W) - G_t(x^2, \phi_1, \phi_2^2, \phi_3^2, W)\right|^2\right]\\ \leq& K\left(|x^1 - x^2|^2 + \sup_{0\leq t\leq T}|\phi_2^1(t) - \phi_2^2(t)|^2 + \sup_{0\leq t\leq T}|\phi_3^1(t) - \phi_3^2(t)|^2\right).\end{aligned}$$

Proof. Firstly, it can be shown that $G(\cdot, \phi_1, \phi_2, \phi_3, W)$ is Lipschitz continuous with respect to x

$$\begin{aligned}\mathbb{E}\left[\left|G_t(x^1, \phi_1, \phi_2, \phi_3, W) - G(x^2, \phi_1, \phi_2, \phi_3, W)\right|\right] &\leq \left|x^1\mathcal{E}_t(\phi_1 - \phi_2) - x^2\mathcal{E}_t(\phi_1 - \phi_2)\right|\\ &\leq \mathcal{E}_t(\phi_1 - \phi_2)|x^1 - x^2|\\ &\leq K(|\phi_1|_\infty, |\phi_2|_\infty, T)|x^1 - x^2|.\end{aligned}$$

Next, we have

$$\begin{aligned}&\mathbb{E}\left[\left|G_t(x, \phi_1, \phi_2, \phi_3^1, W) - G(x, \phi_1, \phi_2, \phi_3^2, W)\right|^2\right]\\ =&\left|\mathcal{E}_t(\phi_1 - \phi_2)\int_0^t \mathcal{E}_s(\phi_1 - \phi_2)\phi_2(s)(\phi_3^1(s) - \phi_3^2(s))ds\right|^2\\ \leq&\mathcal{E}_t(2\phi_1 - 2\phi_2)\left(\int_0^t \mathcal{E}_s(\phi_1 - \phi_2)|\phi_2(s)||(\phi_3^1(s) - \phi_3^2(s))|ds\right)^2\\ \leq&K(|\phi_1|_\infty, |\phi_2|_\infty, T)\left(\int_0^T |\phi_3^1(s) - \phi_3^2(s)|ds\right)^2\\ \leq&K(|\phi_1|_\infty, |\phi_2|_\infty, T)\sup_{0\leq t\leq T}\left|\phi_3^1(t) - \phi_3^2(t)\right|^2.\end{aligned}$$

Similarly, for $\phi_2^1(\cdot), \phi_2^2(\cdot) \in C([0,T])$,

$$
\begin{aligned}
&\mathbb{E}\left[\left|G_t(x,\phi_1,\phi_2^1,\phi_3,W) - G(x,\phi_1,\phi_2^2,\phi_3,W)\right|^2\right] \\
\leq& K\left|x\mathcal{E}_t(\phi_1-\phi_2^1) - x\mathcal{E}_t(\phi_1-\phi_2^2)\right|^2 \\
&+ K\left|\mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds - \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)\phi_2^2(s)\phi_3(s)ds\right|^2 \\
&+ K\mathbb{E}\left[\left|\mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)dW_s - \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)dW_s\right|^2\right] \\
:=& K(J_1+J_2+J_3).
\end{aligned}
$$

Note that by the mean-value theorem and the continuity of ϕ_1, ϕ_2^1 and ϕ_2^2 on $[0,T]$, we can get

$$
\begin{aligned}
J_1 &= \left|x\mathcal{E}_t(\phi_1-\phi_2^1) - x\mathcal{E}_t(\phi_1-\phi_2^2)\right|^2 \\
&= x^2\left(e^{\int_0^t(\phi_1(s)-\phi_2^1(s))ds} - e^{\int_0^t(\phi_1(s)-\phi_2^2(s))ds}\right)^2 \\
&\leq K\left(x, \left|\phi_2^1\right|_\infty, \left|\phi_2^2\right|_\infty, T\right) e^{\int_0^t 2\phi_1(s)ds}\left|\phi_2^1-\phi_2^2\right|_\infty^2 \\
&\leq K\left(x, |\phi_1|_\infty, \left|\phi_2^1\right|_\infty, \left|\phi_2^2\right|_\infty, T\right)\left|\phi_2^1-\phi_2^2\right|_\infty^2,
\end{aligned}
$$

and

$$
\begin{aligned}
J_3 &= \mathbb{E}\left[\left|\mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)dW_s - \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)dW_s\right|^2\right] \\
&= \mathbb{E}\Bigg[\Bigg|\mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)dW_s - \mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)dW_s \\
&\qquad + \mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)dW_s - \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)dW_s\Bigg|^2\Bigg] \\
&\leq 2\mathcal{E}_t(2\phi_1-2\phi_2^1)\int_0^t\left(\mathcal{E}_s(-\phi_1+\phi_2^1) - \mathcal{E}_s(-\phi_1+\phi_2^2)\right)^2 ds \\
&\qquad + 2\left(\mathcal{E}_t(\phi_1-\phi_2^1) - \mathcal{E}_t(\phi_1-\phi_2^2)\right)^2\int_0^t \mathcal{E}_s(-2\phi_1+2\phi_2^2)ds \\
&\leq K\left(|\phi_1|_\infty, \left|\phi_2^1\right|_\infty, \left|\phi_2^2\right|_\infty, T\right)\left|\phi_2^1-\phi_2^2\right|_\infty^2.
\end{aligned}
$$

Lastly, using a similar argument, we have

$$
\begin{aligned}
J_2 &= \left|\mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds - \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)\phi_2^2(s)\phi_3(s)ds\right|^2 \\
&= \Bigg|\mathcal{E}_t(\phi_1-\phi_2^1)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds - \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds \\
&\qquad + \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds - \mathcal{E}_t(\phi_1-\phi_2^2)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)\phi_2^2(s)\phi_3(s)ds\Bigg|^2 \\
&\leq 2\left|\left(\mathcal{E}_t(\phi_1-\phi_2^1)-\mathcal{E}_t(\phi_1-\phi_2^2)\right)\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds\right|^2 \\
&\qquad + 2\left|\mathcal{E}_t(\phi_1-\phi_2^2)\left(\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds - \int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)\phi_2^2(s)\phi_3(s)ds\right)\right|^2 \\
&\leq K\left(|\phi_1|_\infty, \left|\phi_2^1\right|_\infty, \left|\phi_2^2\right|_\infty, |\phi_3|_\infty, T\right)\left|\phi_2^1-\phi_2^2\right|_\infty^2 \\
&\qquad + 2\Bigg|\mathcal{E}_t(\phi_1-\phi_2^2)\left(\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^1)\phi_2^1(s)\phi_3(s)ds - \int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)\phi_2^1(s)\phi_3(s)ds\right) \\
&\qquad + \mathcal{E}_t(\phi_1-\phi_2^2)\left(\int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)\phi_2^1(s)\phi_3(s)ds - \int_0^t \mathcal{E}_s(-\phi_1+\phi_2^2)\phi_2^2(s)\phi_3(s)ds\right)\Bigg|^2 \\
&\leq K\left(|\phi_1|_\infty, \left|\phi_2^1\right|_\infty, \left|\phi_2^2\right|_\infty, |\phi_3|_\infty, T\right)\left|\phi_2^1-\phi_2^2\right|_\infty^2 .
\end{aligned}
$$

Sum up the above inequalities for J_1, J_2 and J_3, then

$$
\mathbb{E}\left[\left|G_t(x,\phi_1,\phi_2^1,\phi_3,W)-G(x,\phi_1,\phi_2^2,\phi_3,W)\right|^2\right] \leq K\left(x, |\phi_1|_\infty, \left|\phi_2^1\right|_\infty, \left|\phi_2^2\right|_\infty, |\phi_3|_\infty, T\right)\left|\phi_2^1-\phi_2^2\right|_\infty^2 .
$$

Thus, we can obtain the desired result.

□

A.4 Proof of the existence and uniqueness of the ODE system

Consider the following ODE system

$$
\begin{cases}
a_y' + C_1\tilde{b}_{1y}a_y - C_2\tilde{b}_{2y}^2 a_y^2 + \sum_{i=1}^{\kappa} q_{y,i}a_i + h_y(t) = 0, \\
a_y(T) = g_y,
\end{cases}
\tag{A.2}
$$

for $y \in \mathcal{Y} = \{1, 2, \ldots, \kappa\}$, where C_1, C_2, h_y, g_y are in $\mathbb{R}^+$. We need to show the existence and uniqueness of the solution to (A.2). Define $T_y^{(N)}$ as

$$T_y^{(N)}[a](t) = \left[\left(g_y + \int_t^T \left(h_y(s) + C_1\tilde{b}_{1y}(s)a_y(s) - C_2\tilde{b}_{2y}^2(s)a_y^2(s) + \sum_{i=1}^{\kappa} q_{y,i}a_i(s)\right) ds\right) \wedge N\right] \vee 0,$$

where $a = [a_1, a_2, \ldots, a_\kappa]^\top$. Let $D = \{f \in C([0,T]) : 0 \le \sup_{t\in[0,T]} f(t) \le N\}$. Note that $T_y^{(N)}(y \in \mathcal{Y})$ maps D^κ to D^κ.

Lemma A.4.1. *For fixed N, there exists a unique solution in $C([0,T])$ to*

$$a = T_y^{(N)}[a]. \tag{A.3}$$

Proof. Denote the norm $\|f\|_k = \left\|e^{kt}\max_{y\in\mathcal{Y}}|f_y|\right\|_\infty$, where k needs to be determined later and f is a κ dimensional vector with entry of $f_y, y \in \mathcal{Y}$, which is equivalent to the infinite norm. Define the iteration rule $a_y^{(n+1)} = T_y^{(N)}\left[a_y^{(n)}\right]$ for $y \in \mathcal{Y}$. Note that

$$\begin{aligned}
&\left\|e^{kt}\left(a_y^{(n+1)}(t) - a_y^{(n)}(t)\right)\right\|_\infty \\
\le& \sup_{t\in[0,T]} e^{kt}\int_t^T \left(C_1\left|\tilde{b}_{1y}\right|_\infty \left|a_y^{(n)}(s) - a_y^{(n-1)}(s)\right| + C_2\left|\tilde{b}_{2y}\right|_\infty^2 \left|\left(a_y^{(n)}(s)\right)^2 - \left(a_y^{(n-1)}(s)\right)^2\right|\right. \\
&\qquad\qquad \left. + \sum_{i=1}^{\kappa} q_{y,i}\left|a_i^{(n)}(s) - a_i^{(n-1)}(s)\right|\right) ds \\
\le& \sup_{t\in[0,T]} e^{kt}\int_t^T \left(C_1\left|\tilde{b}_{1y}\right|_\infty \left|a_y^{(n)}(s) - a_y^{(n-1)}(s)\right| + 2NC_2\left|\tilde{b}_{2y}\right|_\infty^2 \left|a_y^{(n)}(s) - a_y^{(n-1)}(s)\right|\right. \\
&\qquad\qquad \left. + \sum_{i=1}^{\kappa} q_{y,i}\left|a_i^{(n)}(s) - a_i^{(n-1)}(s)\right|\right) ds \\
\le& \sup_{t\in[0,T]} e^{kt}\int_t^T e^{-ks}\left(C_1\left|\tilde{b}_{1y}\right|_\infty + 2NC_2\left|\tilde{b}_{2y}\right|_\infty^2 + \kappa\max_{i\in\mathcal{Y}}|q_{y,i}|\right)\left\|a^{(n)} - a^{(n-1)}\right\|_k ds \\
\le& \frac{C_1\left|\tilde{b}_{1y}\right|_\infty + 2NC_2\left|\tilde{b}_{2y}\right|_\infty^2 + \kappa\max_{i\in\mathcal{Y}}|q_{y,i}|}{k}\left\|a^{(n)} - a^{(n-1)}\right\|_k.
\end{aligned}$$

Choose $k > C_1\left|\tilde{b}_{1y}\right|_\infty + 2NC_2\left|\tilde{b}_{2y}\right|_\infty^2 + \kappa\max_{i\in\mathcal{Y}}|q_{y,i}|$, then

$$\left\|a^{(n+1)} - a^{(n)}\right\|_k \le \frac{C_1\left|\tilde{b}_{1y}\right|_\infty + 2NC_2\left|\tilde{b}_{2y}\right|_\infty^2 + \kappa\max_{i\in\mathcal{Y}}|q_{y,i}|}{k}\left\|a^{(n)} - a^{(n-1)}\right\|_k,$$

which gives us a contraction mapping from D^κ to D^κ. Hence, by the Banach fixed point theorem, there exists a unique solution to (A.3). □

Next, we want to show that for large enough N, the solution to (A.3) is also the solution to (A.2).

Lemma A.4.2. *For*

$$N \geq e^{KT} \left(\sum_{y=1}^{\kappa} g_y + \sum_{y=1}^{\kappa} \int_0^T h_y(s) ds \right),$$

where $K := C_1 \max_{y \in \mathcal{Y}} \left| \tilde{b}_{1y} \right|_\infty + \max_{i \in \mathcal{Y}} \sum_{y=1}^{\kappa} |q_{y,i}|$*, the solution* $a^{(N)}$ *to* (A.3) *satisfies the inequalities*

$$0 \leq g_y + \int_t^T \left(h_y(s) + C_1 \tilde{b}_{1y}(s) a_y^{(N)}(s) - C_2 \tilde{b}_{2y}^2(s) \left(a_y^{(N)}(s) \right)^2 + \sum_{i=1}^{\kappa} q_{y,i} a_i^{(N)}(s) \right) ds \leq N \tag{A.4}$$

for all $t \in [0, T]$*, where* $y \in \mathcal{Y}$*.*

Proof. For simplicity of notations, a_y is used instead of $a_y^{(N)}$ for $y \in \mathcal{Y}$ if there is no confusion.

First, for $y \in \mathcal{Y}$, we prove the positiveness of a_y by contradiction. Suppose a_y ($y \in \mathcal{Y}$) are not positive functions on $[0, T]$. Since a_1 is continuous and $a_1(T) = g_1 > 0$, there exists some $\tau_1 \in [0, T]$ as the closest time to T such that $a_1(\tau_1) = 0$. Note that finding such a τ_1 is possible. Let $t_n \in [0, T]$ be a non-decreasing sequence such that $a_1(t_n) = 0$, there exists some τ_1 such that $t_n \to \tau_1 < T$ as $n \to \infty$ since a_1 is continuous and $a_1(T) = g_1 > 0$. By the continuity of a_1, we have $a_1(\tau_1) = 0$, which gives the desirable point τ_1. Then for all $t \in (\tau_1, T]$, $a_1(t) > 0$ and it implies that $a_1'(\tau_1) > 0$. In this case, plugging $t = \tau_1$ to (A.2), we have

$$a_1'(\tau_1) = -h_1(\tau_1) - \sum_{i \neq 1}^{\kappa} q_{1,i} a_i(\tau_1) > 0,$$

which implies there is some $y \in \mathcal{Y}$ and $y \neq 1$ such that $a_y(\tau_1) < 0$. Without loss of generality, we let $a_2(\tau_1) < 0$. Since a_2 is continuous on $[0, T]$ and $a_2(T) = g_2 > 0$, from the intermediate value theorem, there exists some $\tau_2 \in (\tau_1, T)$ such that $a_2(\tau_2) = 0$ and $a_2'(\tau_2) > 0$. This indicates that $a_2'(\tau_2) = -h_2(\tau_2) - \sum_{i \neq 2}^{\kappa} q_{2,i} a_i(\tau_2) > 0$ by plugging $t = \tau_2$ back to (A.2), and it implies that there is some $y \in \mathcal{Y}$ and $y \neq 1, 2$ such that $a_y(\tau_2) < 0$ since we already know $a_1(\tau_2) > 0$. Without loss of generality, we can let $a_3(\tau_2) < 0$. By induction with the same argument, there is a $\tau_\kappa \in (\tau_{\kappa-1}, T)$ such that $a_\kappa(\tau_\kappa) = 0$ and $a_\kappa'(\tau_\kappa) > 0$, which gives

$$a_\kappa'(\tau_\kappa) + h_\kappa(\tau_k) + \sum_{i \neq \kappa}^{\kappa} q_{\kappa,i} a_i(\tau_\kappa) = 0.$$

But it contradicts the fact that

$$a_\kappa'(\tau_\kappa) > 0, \;\; h_\kappa(\tau_k) > 0, \;\; q_{\kappa,i} > 0, \;\; a_i(\tau_\kappa) > 0$$

for $i \in \{1, 2, \ldots, \kappa - 1\}$. Thus the positiveness of a_y on $[0, T]$ for all $y \in \mathcal{Y}$ is obtained.

Next, we prove the upper boundness for the integral in (A.4). Note that for all $t \in [0, T]$ and $y \in \mathcal{Y}$, let $\tau = T - t$, we have

$$a'_y(\tau) = h_y(\tau) + C_1 \tilde{b}_{1y}(\tau) a_y(\tau) - C_2 \tilde{b}_{2y}^2(\tau) a_y^2(\tau) + \sum_{i=1}^{\kappa} q_{y,i} a_i(\tau),$$

and thus

$$\begin{aligned}
\sum_{y=1}^{\kappa} a'_y(\tau) &= \sum_{y=1}^{\kappa} h_y(\tau) + C_1 \sum_{y=1}^{\kappa} \tilde{b}_{1y}(\tau) a_y(\tau) - C_2 \sum_{y=1}^{\kappa} \tilde{b}_{2y}^2(\tau) a_y^2(\tau) + \sum_{y=1}^{\kappa} \sum_{i=1}^{\kappa} q_{y,i} a_i(\tau) \\
&\le \sum_{y=1}^{\kappa} h_y(\tau) + C_1 \max_{y \in \mathcal{Y}} \left| \tilde{b}_{1y} \right|_{\infty} \sum_{y=1}^{\kappa} a_y(\tau) + \sum_{y=1}^{\kappa} \sum_{i=1}^{\kappa} |q_{y,i}| a_i(\tau) \\
&\le \sum_{y=1}^{\kappa} h_y(\tau) + \sum_{i=1}^{\kappa} \left(C_1 \max_{y \in \mathcal{Y}} \left| \tilde{b}_{1y} \right|_{\infty} + \sum_{y=1}^{\kappa} |q_{y,i}| \right) a_i(\tau) \\
&\le \sum_{y=1}^{\kappa} h_y(\tau) + K \sum_{i=1}^{\kappa} a_i(\tau),
\end{aligned}$$

where

$$K := C_1 \max_{y \in \mathcal{Y}} \left| \tilde{b}_{1y} \right|_{\infty} + \max_{i \in \mathcal{Y}} \sum_{y=1}^{\kappa} |q_{y,i}|$$

with $\sum_{y=1}^{\kappa} a_y(T) = \sum_{y=1}^{\kappa} g_y$. By Grönwall's inequality, for all $\tau \in [0, T]$,

$$\sum_{y=1}^{\kappa} a_y(\tau) \le e^{KT} \left(\sum_{y=1}^{\kappa} g_y + \sum_{y=1}^{\kappa} \int_0^T h_y(s) ds \right).$$

Hence $a_y(t) \le e^{KT} \left(\sum_{y=1}^{\kappa} g_y + \sum_{y=1}^{\kappa} \int_0^T h_y(s) ds \right)$ for all $t \in [0, T]$ and $y \in \mathcal{Y}$. Hence, when

$$e^{KT} \left(\sum_{y=1}^{\kappa} g_y + \sum_{y=1}^{\kappa} \int_0^T h_y(s) ds \right) \le N,$$

(A.4) holds. □

Lemma A.4.3. *With the given of $h_y, g_y \in \mathbb{R}^+$, $y \in \mathcal{Y}$, there exists a unique solution to the Riccati system* (1.12)*.*

Proof. The existence, uniqueness, and boundedness of the solution to a_y $(y \in \mathcal{Y})$ are shown in Lemma A.4.1 and Lemma A.4.2. Given $(a_y : y \in \mathcal{Y})$, the coefficient functions b_y $(y \in \mathcal{Y})$ form a linear ordinary differential equation system. Their existence and uniqueness are guaranteed by Theorem 12.1 in [2]. Similarly, with the given of $(a_y, b_y : y \in \mathcal{Y})$, the coefficient functions c_y, k_y $(y \in \mathcal{Y})$ also form a linear ordinary differential equation system. Applying the Theorem 12.1 in [2], we can obtain the existence and uniqueness of c_y, k_y $(y \in \mathcal{Y})$. □

A.5 Multidimensional Problem on LQG-MFGs

In this subsection, we consider the multidimensional problem, which is a straightforward extension of the previous one-dimensional setup. The same type of Ricatti system to characterize the equilibrium and the value function is obtained, and we have a similar result as the Theorem 1.2.1.

Suppose that X_t, W_t and α_t take values in $\mathbb{R}^d$, and all components of W_t are independent. Suppose that the dynamic of the generic player is given by

$$X_t = X_0 + \int_0^t \left(\tilde{b}_1(Y_s,s)X_s + \tilde{b}_2(Y_s,s)\alpha_s\right) ds + W_t.$$

Consider the cost function

$$\begin{aligned}
&J[m](y,x,t,\bar{\mu},\bar{\nu})\\
=&\mathbb{E}\left[\int_t^T \left(\frac{1}{2}\|\alpha_s\|_2^2 + h(Y_s,s)\int_{\mathbb{R}^d}\|X_s - z\|_2^2 m(dz)\right) ds + \right.\\
&\left. g(Y_T)\int_{\mathbb{R}^d}\|X_T - z\|_2^2 m(dz)\right| X_t = x, Y_t = y, \mu_t = \bar{\mu}, \nu_t = \bar{\nu}\Bigg]\\
=&\mathbb{E}\left[\int_t^T \left(\frac{1}{2}\alpha_s^\top\alpha_s + h(Y_s,s)\left(X_s^\top X_s - 2\mu_s^\top X_s + \nu_s\cdot \mathbb{1}_d\right)\right) ds + \right.\\
&\left. g(Y_T)\left(X_T^\top X_T - 2\mu_T^\top X_T + \nu_T\cdot\mathbb{1}_d\right)\right| X_t = x, Y_t = y, \mu_t = \bar{\mu}, \nu_t = \bar{\nu}\Big],
\end{aligned}$$

where m is the joint density function in $\mathbb{R}^d$, and μ, ν take value in $\mathbb{R}^d$. For $y \in \mathcal{Y}$, define the Riccati system

$$\begin{cases}
a_y' + 2\tilde{b}_{1y}a_y - 2\tilde{b}_{2y}^2 a_y^2 + \sum_{i=1}^{\kappa} q_{y,i}a_i + h_y(t) = 0,\\
b_y' + \left(2\tilde{b}_{1y} - 4\tilde{b}_{2y}^2 a_y\right) b_y + \sum_{i=1}^{\kappa} q_{y,i}b_i + h_y(t) = 0,\\
c_y' + da_y + db_y + \sum_{i=1}^{\kappa} q_{y,i}c_i = 0,\\
k_y' - 2\tilde{b}_{2y}^2 a_y^2 + 4\tilde{b}_{2y}^2 a_y b_y + 2\tilde{b}_{1y}k_y + \sum_{i=1}^{\kappa} q_{y,i}k_i = 0,\\
a_y(T) = b_y(T) = g_y\ ,\ c_y(T) = k_y(T) = 0.
\end{cases} \tag{A.5}$$

Theorem A.5.1 (Verification theorem for MFGs)**.** *There exists a unique solution $(a_y, b_y, c_y, k_y : y \in \mathcal{Y})$ for the Riccati system (A.5). With these solutions, for $t \in [0,T]$, the MFG equilibrium path follows $\hat{X} = \hat{X}[\hat{m}]$ is given by*

$$d\hat{X}_t = \left(\tilde{b}_1(Y_t,t)\hat{X}_t - 2\tilde{b}_2^2(Y_t,t)a_{Y_t}(t)\left(\hat{X}_t - \hat{\mu}_t\right)\right) dt + dW_t, \quad \hat{X}_0 = X_0,$$

with equilibrium control $\hat{\alpha}_t = -2\tilde{b}_2(Y_t,t)a_{Y_t}(t)\left(\hat{X}_t - \hat{\mu}_t\right)$, *where*

$$d\hat{\mu}_t = \tilde{b}_1(Y_t,t)\hat{\mu}_t dt, \quad \hat{\mu}_0 = \mathbb{E}[X_0].$$

Moreover, the value function U *is*

$$U(m_0,y,x) = a_y(0)x^\top x - 2a_y(0)x^\top[m_0]_1 + k_y(0)[m_0]_1^\top[m_0]_1 + b_y(0)[m_0]_2^\top \mathbb{1}_d + c_y(0)$$

for $y \in \mathcal{Y}$.

The proof is similar to the one-dimensional problem, and we don't show the details here.

A.6 Comparison to bound (3.10) in Theorem 18 of S. O'Rourke et al.[35]

When $\frac{\sqrt{r}}{\sqrt{d}} \geq (\frac{\sigma_{k+1}}{\sigma_k} - \frac{k-1}{k})$, we have $(\frac{\sigma_{k+1}}{\sigma_k} - \frac{k-1}{k})\sqrt{d} \leq \sqrt{r}$ and

$$\begin{aligned}
\frac{\sqrt{d}}{k} &\leq \sqrt{r} + \sqrt{d} - \frac{\sigma_{k+1}}{\sigma_k}\sqrt{d} \\
&= \sqrt{r} + \sqrt{d}(\frac{\sigma_k - \sigma_{k+1}}{\sigma_k}) \\
&\leq \sqrt{r} + \sqrt{m}(\frac{\sigma_k - \sigma_{k+1}}{\sigma_k}) + \frac{m}{\sigma_k}
\end{aligned}$$

rearrange terms leading to

$$\frac{\sqrt{d}}{\sigma_k - \sigma_{k+1}} \leq k(\frac{\sqrt{r}}{\sigma_k - \sigma_{k+1}} + \frac{\sqrt{m}}{\sigma_k} + \frac{m}{\sigma_k(\sigma_k - \sigma_{k+1})}).$$

Therefore we proved that when $\frac{\sqrt{r}}{\sqrt{d}} \geq (\frac{\sigma_{k+1}}{\sigma_k} - \frac{k-1}{k})$, i.e. $(\frac{\sigma_{k+1}}{\sigma_k} - \frac{k-1}{k})^2 d \leq r \leq d$, or when the matrix A is full rank, i.e. $r = d$, the bound obtained by Corollary 3.2.1 inequality (3.7) is at least as good as bound shown in Theorem 18 of S. O'Rourke et al. [54].

Bibliography

[1] Saran Ahuja. *Mean field games with common noise.* Stanford University, 2015.

[2] Panos J Antsaklis and Anthony N Michel. *Linear systems.* Springer Science & Business Media, 2006.

[3] Martino Bardi. **Explicit solutions of some linear-quadratic mean field games**. *Networks & Heterogeneous Media*, **7**(2):243, 2012.

[4] Martino Bardi and Fabio S Priuli. **LQG mean-field games with ergodic cost**. In *52nd IEEE Conference on Decision and Control*, pages 2493–2498. IEEE, 2013.

[5] David Blackwell. **Equivalent comparisons of experiments**. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

[6] Marie-France Bru. **Diffusions of perturbed principal component analysis**. *Journal of Multivariate Analysis*, **29**(1):127–136, 1989.

[7] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. **Robust principal component analysis?** *Journal of the ACM (JACM)*, **58**(3):1–37, 2011.

[8] Pierre Cardaliaguet. **Notes on mean field games**. Technical report, 2010.

[9] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. *The master equation and the convergence problem in mean field games:(AMS-201)*, **201**. Princeton University Press, 2019.

[10] René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications I-II.* Springer, 2018.

[11] Luyang Chen, Markus Pelger, and Jason Zhu. **Deep learning in asset pricing**. *Management Science*, **70**(2):714–750, 2024.

[12] Gregory Connor. *Active portfolio management: a quantitative approach to providing superior returns and controlling risk*, **13**. Oxford University Press, 2000.

[13] Kent Daniel, David Hirshleifer, and Avanidhar Subrahmanyam. **Investor psychology and security market under-and overreactions**. *The Journal of Finance*, **53**(6):1839–1885, 1998.

[14] Chandler Davis and William Morton Kahan. **The rotation of eigenvectors by a perturbation. III**. *SIAM Journal on Numerical Analysis*, **7**(1):1–46, 1970.

[15] Petros Drineas and Michael W Mahoney. **RandNLA: randomized numerical linear algebra**. *Communications of the ACM*, **59**(6):80–90, 2016.

[16] Eugene F Fama and Kenneth R French. **Common risk factors in the returns on stocks and bonds**. *Journal of Financial Economics*, **33**(1):3–56, 1993.

[17] Xinwei Feng, Jianhui Huang, and Zhenghong Qiu. **Mixed social optima and Nash equilibrium in linear-quadratic-gaussian mean-field system**. *IEEE Transactions on Automatic Control*, **67**(12):6858–6865, 2021.

[18] Dena Firoozi, Sebastian Jaimungal, and Peter E Caines. **Convex analysis for LQG systems with applications to major-minor LQG mean-field game systems**. *Systems & Control Letters*, **142**:104734, 2020.

[19] Wilfrid Gangbo, Alpár R Mészáros, Chenchen Mou, and Jianfeng Zhang. **Mean field games master equations with nonseparable Hamiltonians and displacement monotonicity**. *The Annals of Probability*, **50**(6):2178–2217, 2022.

[20] Shuang Gao, Peter E Caines, and Minyi Huang. **LQG graphon mean field games: Graphon invariant subspaces**. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 5253–5260. IEEE, 2021.

[21] Antonios Georgantas, Michalis Doumpos, and Constantin Zopounidis. **Robust optimization approaches for portfolio selection: a comparative analysis**. *Annals of Operations Research*, pages 1–17, 2021.

[22] Sanford J Grossman and Joseph E Stiglitz. **On the impossibility of informationally efficient markets**. *The American Economic Review*, **70**(3):393–408, 1980.

[23] Shihao Gu, Bryan Kelly, and Dacheng Xiu. **Empirical asset pricing via machine learning**. *The Review of Financial Studies*, **33**(5):2223–2273, 2020.

[24] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2002.

[25] Jack Hirshleifer and John G Riley. *The analytics of uncertainty and information*. Cambridge University Press, 1992.

[26] Dashan Huang, Shushang Zhu, Frank J Fabozzi, and Masao Fukushima. **Portfolio selection under distributional uncertainty: A relative robust CVaR approach**. *European Journal of Operational Research*, **203**(1):185–194, 2010.

[27] Jianhui Huang and Minyi Huang. **Mean field LQG games with model uncertainty**. In *52nd IEEE Conference on Decision and Control*, pages 3103–3108. IEEE, 2013.

[28] Jianhui Huang, Xun Li, and Tianxiao Wang. **Mean-field linear-quadratic-Gaussian (LQG) games for stochastic integral systems**. *IEEE Transactions on Automatic Control*, **61**(9):2670–2675, 2015.

[29] Minyi Huang. **Large-population LQG games involving a major player: the Nash certainty equivalence principle**. *SIAM Journal on Control and Optimization*, **48**(5):3318–3353, 2010.

[30] Minyi Huang, Peter E Caines, and Roland P Malhamé. **Social optima in mean field LQG control: centralized and decentralized strategies**. *IEEE Transactions on Automatic Control*, **57**(7):1736–1751, 2012.

[31] Minyi Huang, Roland P Malhamé, Peter E Caines, et al. **Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle**. *Communications in Information & Systems*, **6**(3):221–252, 2006.

[32] Minyi Huang and Xuwei Yang. **Linear quadratic mean field games: Decentralized O(1/N)-Nash equilibria**. *Journal of Systems Science and Complexity*, **34**(5):2003–2035, 2021.

[33] Joe Jackson and Ludovic Tangpi. **Quantitative convergence for displacement monotone mean field games with controlled volatility**. *Mathematics of Operations Research*, 2023.

[34] Jiamin Jian, Peiyao Lai, Qingshuo Song, and Jiaxuan Ye. **The convergence rate of the equilibrium measure for the hybrid LQG Mean Field Game**. *Nonlinear Analysis: Hybrid Systems*, **52**:101454, 2024.

[35] Jiamin Jian, Qingshuo Song, and Jiaxuan Ye. **Convergence Rate of LQG Mean Field Games with Common Noise**. *arXiv preprint arXiv:2307.00695*, 2023.

[36] DANIEL Kai-Ineman and Amos Tversky. **Prospect theory: An analysis of decision under risk**. *Econometrica*, **47**(2):363–391, 1979.

[37] Ravindran Kannan, Santosh Vempala, et al. **Spectral algorithms**. *Foundations and Trends® in Theoretical Computer Science*, **4**(3–4):157–288, 2009.

[38] Daniel Lacker and Thaleia Zariphopoulou. **Mean field and n-agent games for optimal investment under relative performance criteria**. *Mathematical Finance*, **29**(4):1003–1038, 2019.

[39] Jean-Michel Lasry and Pierre-Louis Lions. **Mean field games**. *Japanese Journal of Mathematics*, **2**(1):229–260, 2007.

[40] Brigitte Le Roux and Henry Rouanet. *Geometric data analysis: from correspondence analysis to structured data analysis*. Springer Science & Business Media, 2004.

[41] Dennis V Lindley. **On a measure of the information provided by an experiment**. *The Annals of Mathematical Statistics*, **27**(4):986–1005, 1956.

[42] Siyu Lv, Jie Xiong, and Xin Zhang. **Linear quadratic leader–follower stochastic differential games for mean-field switching diffusions**. *Automatica*, **154**:111072, 2023.

[43] Michael W Mahoney et al. **Randomized algorithms for matrices and data**. *Foundations and Trends® in Machine Learning*, **3**(2):123–224, 2011.

[44] Oren Mangoubi and Nisheeth Vishnoi. **Re-analyze Gauss: Bounds for private matrix approximation via Dyson Brownian motion**. *Advances in Neural Information Processing Systems*, **35**:38585–38599, 2022.

[45] Xuerong Mao and Chenggui Yuan. *Stochastic differential equations with Markovian switching*. Imperial College Press, 2006.

[46] Harry M Markowits. **Portfolio selection**. *Journal of Finance*, **7**(1):71–91, 1952.

[47] Per-Gunnar Martinsson. **Randomized methods for matrix computations**. *The Mathematics of Data*, **25**(4):187–231, 2019.

[48] Per-Gunnar Martinsson and Joel A Tropp. **Randomized numerical linear algebra: Foundations and algorithms**. *Acta Numerica*, **29**:403–572, 2020.

[49] Chris Meek, Bo Thiesson, and David Heckerman. **US Census Data (1990)**. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5VP42.

[50] Son L. Nguyen and Minyi Huang. **Mean field LQG games with mass behavior responsive to a major player**. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 5792–5797. IEEE, 2012.

[51] Son L. Nguyen, Dung T Nguyen, and George Yin. **A stochastic maximum principle for switching diffusions using conditional mean-fields with applications to control problems**. *ESAIM: Control, Optimisation and Calculus of Variations*, **26**:69, 2020.

[52] Son L Nguyen, George Yin, and Tuan A Hoang. **On laws of large numbers for systems with mean-field interactions and Markovian switching**. *Stochastic Processes and their Applications*, **130**(1):262–296, 2020.

[53] Tri-Dung Nguyen and Andrew W Lo. **Robust ranking and portfolio optimization**. *European Journal of Operational Research*, **221**(2):407–416, 2012.

[54] Sean O'Rourke, Van Vu, and Ke Wang. **Random perturbation of low rank matrices: Improving classical bounds**. *Linear Algebra and its Applications*, **540**:26–59, 2018.

[55] Richard Roll. **Ambiguity when performance is measured by the securities market line**. *The Journal of Finance*, **33**(4):1051–1069, 1978.

[56] Andrey A Shabalin and Andrew B Nobel. **Reconstruction of a low-rank matrix in the presence of Gaussian noise**. *Journal of Multivariate Analysis*, **118**:67–76, 2013.

[57] William F Sharpe. **Capital asset prices: A theory of market equilibrium under conditions of risk**. *The Journal of Finance*, **19**(3):425–442, 1964.

[58] Gilbert W Stewart and Ji-guang Sun. *Matrix perturbation theory.* Boston: Academic Press, 1990.

[59] Rinel Foguen Tchuendom. **Uniqueness for linear-quadratic mean field games with common noise**. *Dynamic Games and Applications*, **8**(1):199–210, 2018.

[60] Ky Tran, George Yin, and Le Yi Wang. **A generalized Goodwin business cycle model in random environment**. *Journal of Mathematical Analysis and Applications*, **438**(1):311–327, 2016.

[61] Lloyd N Trefethen and David Bau. *Numerical linear algebra*, **181**. SIAM, 2022.

[62] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, **47**. Cambridge University Press, 2018.

[63] Per-Åke Wedin. **Perturbation bounds in connection with singular value decomposition**. *BIT Numerical Mathematics*, **12**:99–111, 1972.

[64] Hermann Weyl. **Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)**. *Mathematische Annalen*, **71**(4):441–479, 1912.

[65] George Yin and Chao Zhu. *Hybrid switching diffusions: properties and applications*, **63**. Springer Science & Business Media, 2009.

[66] Fumie Yokota and Kimberly M Thompson. **Value of information analysis in environmental health risk management decisions: past, present, and future**. *Risk Analysis: An International Journal*, **24**(3):635–650, 2004.

[67] Yi Yu, Tengyao Wang, and Richard J Samworth. **A useful variant of the Davis–Kahan theorem for statisticians**. *Biometrika*, **102**(2):315–323, 2015.

[68] Qing Zhang, George Yin, and El-Kébir Boukas. **Controlled Markov chains with weak and strong interactions: asymptotic optimality and applications to manufacturing**. *Journal of Optimization Theory and Applications*, **94**:169–194, 1997.

[69] Xun Yu Zhou and George Yin. **Markowitz's mean-variance portfolio selection with regime switching: A continuous-time model**. *SIAM Journal on Control and Optimization*, **42**(4):1466–1482, 2003.

[70] Chao Zhu and George Yin. **On competitive Lotka–Volterra model in random environments**. *Journal of Mathematical Analysis and Applications*, **357**(1):154–170, 2009.

[71] Hui Zou, Trevor Hastie, and Robert Tibshirani. **Sparse principal component analysis**. *Journal of Computational and Graphical Statistics*, **15**(2):265–286, 2006.