University of Tennessee, Knoxville
Trace: Tennessee Research and Creative Exchange

Doctoral Dissertations Graduate School

8-2016

Stochastic and Optimal Distributed Control for Energy Optimization and Spatially Invariant Systems

Jin Dong
University of Tennessee, Knoxville, jdong@vols.utk.edu

Recommended Citation
Dong, Jin, "Stochastic and Optimal Distributed Control for Energy Optimization and Spatially Invariant Systems. " PhD diss., University of Tennessee, 2016.
https://trace.tennessee.edu/utk_graddiss/3908

This Dissertation is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Jin Dong entitled "Stochastic and Optimal Distributed Control for Energy Optimization and Spatially Invariant Systems." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Seddik M. Djouadi, Major Professor

We have read this dissertation and recommend its acceptance:

Kevin Tomsovic, Jan Rosinski, Donatello Materassi

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Stochastic and Optimal Distributed Control for Energy Optimization and Spatially Invariant Systems

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Jin Dong

August 2016

© by Jin Dong, 2016
All Rights Reserved.

To my parents, Aihong Xiao and Xianglong Dong

whose love, support, and encouragement
over the years have to a large extent
made the writing of this dissertation possible.

Acknowledgements

It has been five years full of discovery, excitement, and achievement, all of which I appreciate with a grateful heart. My debt of gratitude must go to my advisor, Dr. Seddik M. Djouadi. His professionalism and expertise taught me to survive. Working with him has opened a whole new world to me. He introduced me to this amazing research area, enlightened me through the tough moments, encouraged me to think freely, and helped me build a path to the future. As a researcher, he is the role model. As a human being, his talent and personality inspires me in many ways that will influence me for the rest of my life. It is a great honor to be his student and work with him. He has taught me how to be a scientist, but more importantly he has showed me all of the greatest attributes in a mentor and friend. I also want to thank him for his humor and kindness which warmed my heart throughout the difficulties that I faced during this pursuit.

My acknowledgments also go to my committee, Dr. Kevin Tomsovic, Dr. Jan Rosinski and Dr. Donatello Materassi for their support, guidance and helpful comments, criticism and suggestions. Their help with subject data, research ideas, and manuscript reviews provided all the foundation for my study. Without their energy and time I never would have been able to put together the whole material I needed for this dissertation. I also need to thank Dr. Husheng Li, Dr. Yilu Liu, Dr. Douglas Birdwell, Dr. Vasileios Maroulas, Dr. Andreas Malikopoulos, Dr. James Nutaro, Dr. Teja Kuruganti and Dr. Mohammed Olama. They provided me many ideas and new thoughts to accomplish the dissertation. They were there as I needed

them every step of the way and taught me much about research, collaborations, and mentoring.

I am thankful to all my colleagues and friends at UTK who made my academic life exciting. In particular, I would like to thank my friends Xiao Ma, Samir Sahyoun, Ouassim Bara, Anis Drira, Yichen Zhang, Xiaoyang Pan, Mohammadehsan Raoufat, Chris Winstead, Dianne Bull, Liguo Wang, Wenqiang Feng, Yukun Li and Xiaoyu Shi for the camaraderie and the good time that we had. I enjoyed many conversations and discussions that have had a tremendous impact on my research and myself as a person. I would like to deeply appreciate Oak Ridge National Laboratory (ORNL) for their support of my work and for sharing their experiences with me. This collaboration inspired me to develop interesting real-world problems and to propagate the collaborative spirit.

I must express my appreciation to my family, far away in China, for their enormous support and completely selfless love. It would not have been possible to complete this research or write this thesis without the unfailing support of my parents and parents-in-law. They gave me the confidence to come here to pursue the Ph.D degree. I owe them everything and wish I could show them just how much I love and appreciate them.

Last but definitely not least, I wish to thank my wonderful wife, Chen Wen, who accompanied with me for the past five years of my pursuit. There are no words to adequately describe the depth of my appreciation for the sacrifices she has made while I pursued my PhD. I love you!

Thank you all!

It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again; the never satisfied man is so strange if he has completed a structure, then it is not in order to dwell in it peacefully, but in order to begin another. I imagine the world conqueror must feel thus, who, after one kingdom is scarcely conquered, stretches out his arms again for others.
- by Carl Friedrich Gauss.

Now what is science?... It is before all a classification, a manner of bringing together facts which appearances separate, though they were bound together by some natural and hidden kinship. Science, in other words, is a system of relations.
- by Henri Poincaré.

A stochastic process is any process running along in time and controlled by probability laws...more precisely any family of random variables where a random variable...is simply a measurable function... - by Joseph Doob.

Abstract

Improving energy efficiency and grid responsiveness of buildings requires sensing, computing and communication to enable stochastic decision-making and distributed operations. Optimal control synthesis plays a significant role in dealing with the complexity and uncertainty associated with the energy systems.

The dissertation studies general area of complex networked systems that consist of interconnected components and usually operate in uncertain environments. Specifically, the contents of this dissertation include tools using stochastic and optimal distributed control to overcome these challenges and improve the sustainability of electric energy systems.

The first tool is developed as a unifying stochastic control approach for improving energy efficiency while meeting probabilistic constraints. This algorithm is applied to demonstrate energy efficiency improvement in buildings and improving operational efficiency of virtualized web servers, respectively. Although all the optimization in this technique is in the form of convex optimization, it heavily relies on semidefinite programming (SP). A generic SP solver can handle only up to hundreds of variables. This being said, for a large scale system, the existing off-the-shelf algorithms may not be an appropriate tool for optimal control. Therefore, in the sequel I will exploit optimization in a distributed way.

The second tool is itself a concrete study which is optimal distributed control for spatially invariant systems. Spatially invariance means the dynamics of the system do not vary as we translate along some spatial axis. The optimal H^2 [H-2] decentralized control problem is solved by computing an orthogonal projection on a class of Youla parameters with a decentralized structure. Optimal H^∞ [H-infinity] performance is posed as a distance minimization in a general L^∞ [L-infinity] space from a vector function to a subspace with a mixed L^∞ and H^∞ space structure. In this framework, the dual and pre-dual formulations lead to finite dimensional convex optimizations which approximate the optimal solution within desired accuracy. Furthermore, a mixed L^2 [L-2] $/H^\infty$ synthesis problem for spatially invariant systems as trade-offs between transient performance and robustness. Finally, we pursue to deal with a more general networked system, i.e. the Non-Markovian decentralized stochastic control problem, using stochastic maximum principle via Malliavin Calculus.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The dissertation studies the general area of complex networked systems that consist of interconnected components and usually operate in uncertain environments and with incomplete information.

The work in this dissertation is inspired by two facts in today's power system. Firstly, significant potential for energy savings exist by optimally controlling building systems to reduce consumption while maintaining comfort constraints. Secondly, problems associated with complex networked systems are typically large-scale and computationally intractable.

The goal of this dissertation is to develop foundational theories and tools to exploit those structures that can lead to computationally-efficient and distributed solutions, and apply them to improve systems operations and energy efficiency. We present two tools for stochastic and distributed control to overcome these challenges and improve the sustainability of electric power systems.

Specifically, this dissertation focuses on two concrete areas. The first one is to design stochastic control algorithms to manage energy efficient buildings and virtualized web servers; the second one is to design distributed optimization rules for large-scale interconnected engineering systems.

1.1 Problem Descriptions

Different components of the system are linked together with communication paths and sensor nodes to provide interoperability between them, e.g., distribution, transmission and other substations, such as residential, commercial, and industrial sites.

Due to increasing price of fossil fuels, environmental impact and energy security concerns coupled with improvements in wind turbines and solar panels, there has been a great increase in the penetration level of renewable energy around the world. However, the deployment of intermittent renewable sources will inevitably lead to frequent imbalance between supply and demand, as exemplified by the difficulties in maintaining system balance due to renewable energy variability. Advanced control algorithm is certain to play a significant role in dealing with the complexity and uncertainty associated with energy efficient buildings, with stochastic (distributed) control being a natural choice.

1.1.1 Preliminaries of building climate control

Buildings consume up to 40% of the energy produced in the US [3]. Advanced sensors and controls have the potential to reduce the energy consumption of buildings by 20-40% [4, 5]. Heating, Ventilation and Air Conditioning (HVAC) systems play a fundamental role in maintaining a comfortable temperature environment in buildings and account for 50% of building energy consumption [3]. Significant potential for energy savings exist by optimally controlling HVAC systems to reduce consumption while maintaining comfort constraints.

1.1.2 Preliminaries of virtual machines

Recently, energy-efficient computing is becoming a hot topic in the design and operation of enterprise servers and modern data centers. It is well known that the Internet services (e.g., online banking and online business, etc.), data storage and telecommunications industries are crucial parts of the American information economy, the repaid growth of these industries have led to a huge increase in electricity use [1]. As reported by the U.S. Department of Energy, the energy consumption for a data center can be 100 times higher than that of a typical commercial building. Furthermore, it is estimated that energy consumption will contribute more than 50 % of the IT budget in the next few years. Virtual machine (VM) or virtualized web servers (VWS) technology, as an important enabler of cloud computing [3], has been widely employed in modern data centers. Thus, the ultimate goal of this work is to improve those sectors energy efficiency, which will provide significant energy and cost saving, as well as reduce the carbon pollution.

1.1.3 Preliminaries of Distributed control

There has been resonant interest in analysis and synthesis of distributed coordination and control algorithms for spatially interconnected systems. For recent work on this class and some of the background for the present work, for example, we refer the reader to [6, 7, 8, 9, 10], and the references therein. We consider distributed parameter systems where the underlying dynamics are spatially invariant, and where the controls and measurements are spatially distributed. These systems arise in many applications such as the control of vehicular platoons, flow control, micro-electromechanical systems (MEMS), smart structures, and systems described by partial differential equations with constant coefficients and distributed controls and measurements [6].

1.1.4 Applications and Challenges

- **Control (Energy Saving)**: Heating, Ventilation and Air Conditioning (HVAC) systems play a fundamental role in maintaining a comfortable temperature environment in buildings and account for 50% of building energy consumption [3]. Significant potential for energy savings exist by optimally controlling HVAC systems to reduce consumption while maintaining comfort constraints. The virtualization technology can support environment isolation, fault isolation for multi VMs. These VMs can run different operating systems and applications as if they are running on the physical separate machines, but the contention of shared resource (i.e. processor, last level cache and bandwidth) among multiple customer applications leads to a performance interference effect [11]. Moreover, the workloads of web applications show highly volatile variation and unexpected burst in terms of request arrival rate [12]. It is infeasible to decide the resource allocation solution (e.g. how many resources should be allocated and how to assign them) in a fixed way. In such environments, the control approach needs to be robust and adaptive to dynamic variations in workload.

- **Optimization (Distributed Control)**: A networked system is a collection of dynamic units that interact over an information exchange network. Such systems are ubiquitous in diverse areas of science and engineering [13]. There are many important problems that have been cast in the form of a large-scale finite-dimensional or an infinite-dimensional constraint optimization problem [14]. Such problems can range from physical, biological to mechanical and social systems [15, 16, 17, 18, 19]. Distributed control has become a successful strategy to handle such design issues as coordinated control, formation control and synchronization of multi-agent systems [20, 21, 22, 23, 24, 25].

While the technique is promising, there are a lot of challenging issues for these applications.

Typical building controls are set-point based, where zone-level temperature measurement is used for taking control action to keep the zone temperature in a comfortable range. To make buildings more energy and cost efficient, intelligent predictive automation can be used instead of conventional automation. For instance, the predictive automation controllers can operate the buildings passive thermal storage, based on predicted future disturbances (e.g. weather forecast), by making use of low cost energy sources [26]. The goal is to design an optimal controller that can realize the temperature requirement and minimize energy consumptions.

One of the most productive recent development is that of the theory of optimal control of systems with distributed parameters. This class of systems is much broader than simple systems with only lumped parameters. Actually, many of the real problems in control and design in airframe, shipbuilding, electronics, nucleonics, and other engineering fields are, in essence, problems of control of systems with distributed parameters. Spatial invariance is a strong property of a given system, which means that the dynamics of the system do not vary as we translate along some spatial axis [27].

As mentioned in [28], the main difference between distributed parameter systems and lumped parameter systems is that the former are characterized not by a finite set of quantities, coordinates of the object, that vary only in time, but generally by a set of functions that show the dependence of the parameters on the time and spatial variables or any combination of them. In the majority of cases, ideal control designed for lumped systems is not realizable because of the presence of additional constraints imposed on state functions and controlling actions. In particular, these constraints are related with spatial variables on which the state and control functions depend. The impossibility of realizing a perfect control process posed the problem of deriving an optimal process according to a definite, preassigned criterion. Hence, an optimal control problem of distributed parameter systems obviously leads to the powerful apparatus of functional analysis [28].

Even if the subsystems interact locally, the optimal controller will need global information to produce the feedback signal. Standard control design techniques are inadequate since most optimal control techniques cannot handle systems of very high dimensions and with a large number of inputs and outputs. A preferred alternative is to have control signals computed using only local communication among neighboring subsystems as motivated in [6] and [7]. Using an approach based on spatial Fourier transforms and operator theory, Bamieh *et al.* discussed the optimal control of linear spatially invariant systems with standard linear quadratic (LQ) criterion in [6].

What's more, synchronization of multiple heterogeneous linear systems has been studied in [8, 10], where the interconnection topology is represented by an arbitrary graph, using a conservative analysis - LMI. A similar problem is investigated under both fixed and switching communication topologies heterogeneous spatially distributed systems [29, 30]. Using an approach based on spatial Fourier transforms and operator theory, Bamieh *et al.* [6] discussed the optimal control of linear spatially invariant systems with standard linear quadratic(LQ) criterion. Although rigorous, as pointed out in [31, 32], this approach is valid when system operator could generate a semigroup on $L_2(\mathbb{R}^n)$ for a given state-space, which is a non-trivial task.

1.2 Previous Works

1.2.1 Building climate control

Building climate control leads naturally to probabilistic constraints as current standards explicitly state, zone temperatures should be kept within a comfort range with a predefined probability [33, 34]. In order to address this issue and explicitly account for system uncertainties, some efforts have been made for studying a stochastic version of MPC (SMPC) including probabilistic constraints. [35] employed stochastic MPC technique to compute the control strategy for a cost function which was linear in the control variable for the thermal dynamics in a linear state-space model, which described thermal energy and temperatures. [36] proposed a tractable approximation method for the problem. Both schemes in [35] and [36] considered chance constraints and solved them by using affine disturbance feedback.

1.2.2 Virtualized Web Servers

Feedback control approaches have been widely used to guarantee the performance and power request in virtualized environment [37, 38, 39, 40]. For example, Wang et al. introduce two-layer feedback control architecture to achieve the goal of response time optimization and power saving by controlling CPU resource allocation and CPU frequency in separate loops [37]. However, the models of their systems are designed in a way that heavily relies on off-line system identification for specific workloads. Although these approaches can theoretically guarantee the system stability and control accuracy within a range, they cannot adapt to a time varying workload.

1.2.3 Distributed control

After the recent advances in communication technologies, the design of distributed controllers for physically interconnected systems has become an attractive and fruitful research direction [9, 41]. A body of literature has been worked out for the spatially distributed systems, where all signals are functions of both spatial and temporal variables. The linear matrix inequality (LMI) conditions for spatially interconnected systems consisting of homogeneous units are introduced in [9, 42]. Control synthesis results have employed consensus-based observer to guarantee leaderless synchronization of multiple identical linear dynamic systems under switching communication topologies[43]; neighbor-based observer to solve the synchronization problem for general linear time-invariant systems [44]; and individual-based observer with low-gain technique to synchronize a group of linear systems [45]. Synchronization of multiple heterogeneous linear systems has been studied in [8, 10], where the interconnection topology is represented by an arbitrary graph, using a conservative analysis-LMI. A similar problem is investigated under both fixed and switching communication topologies heterogeneous spatially distributed systems [29, 30].

Recently, another theoretical mechanism for decentralized stochastic control was proposed in [46]. A general model of decentralized stochastic control called partial history sharing information structure is presented. In this model, at each step the controller of each individual agent shares part of their observation and control history with each neighboring agent. The optimal control problem at the coordinator level is assumed to follow a *partially observable Markov decision process* (POMDP), i.e. a "memoryless" process.

1.3 Motivations and Contributions

Energy optimization and control design is a critical issue in electrical power systems for security and energy efficient applications, which cannot be guaranteed using traditional deterministic algorithms. Although lots of work have been done on protection and control for smart grid, however, most of them mainly focus on trivial model based or involving impractical constraints and relaxations, little effort has been made to explore how to efficiently use various estimation and control techniques to guarantee reliability and better performance in smart grid with complicated models. Specifically, finding the optimal minimum for the distributed control problem is very difficult, since it is a non-convex optimization problem under infinite dimension. Researchers will therefore have to resort to suboptimal methods, which yield solutions satisfying the constraints while trying to minimize the norm of the model error. However, none of these suboptimal solutions could guarantee how far are the suboptimal solutions away from the true optimal solutions.

In this dissertation, we study a quadratic cost function in terms of temperature errors and control inputs, which is subject to several constraints on the room temperature and control input. In particular, we only consider the case where we assume that the disturbance is *Gaussian* and the problem is formulated to minimize the expected cost subject to a linear constraint on control input and a *probabilistic constraint* on the state. We also propose an efficient algorithm to reduce the probabilistic constraint to a hard constraint on the control input exactly [47]. The problems are formulated into semidefinite optimization problems which may be solved through semidefinite programming (SDP) for the optimal solutions efficiently.

Despite the importance of stochastic control for single building systems or Virtualized web servers, little is known about doing distributed control for power gird or energy-efficient buildings. We consider the problem of optimal distributed control of spatially invariant systems. We have been investigating optimal performance of this problem by utilization of norm based criteria. We develop an input-output framework for problems of this class. Spatially invariant systems are viewed as multiplication operators from a particular Hilbert function space into itself in the Fourier domain. Optimal distributed performance is then posed as a distance minimization in a general $\mathcal{H}^2, \mathcal{L}^\infty$ and $\mathcal{L}^2/\mathcal{H}^\infty$ spaces, respectively. In this framework, a generalized version of the Youla parametrization plays a central role. Our approach is purely input-output and does not use any state space realization.

We start by the optimal H^2 **control** problem for spatially invariant systems. In particular, the H^2 optimal control problem is solved via the computation of an orthogonal projection of a tensor Hilbert space onto a particular subspace. The optimal H^2 decentralized control problem is solved by computing an orthogonal projection on a class of Youla parameters with a decentralized structure. Furthermore, we use Riesz projections after invoking a particular L^2-basis.

After that, we form a centralized distance minimization in **a mixed $\mathcal{L}^\infty$ and $\mathcal{H}^\infty$ space** structure. The duality structure of the problem is characterized by computing various dual and pre-dual spaces. The annihilator and pre-annihilator subspaces are also calculated for the dual and pre-dual problems. We show that these spaces together with the pre-annihilator and annihilator subspaces can be realized explicitly as specific tensor spaces and subspaces, respectively. The tensor space formulation leads to a solution in terms of an operator given by a tensor product. Specifically, the optimal distributed control performance for spatially invariant systems is equal to the operator induced norm of this operator. Meanwhile, we show that these spaces together with the pre-annihilator and annihilator subspaces can be realized explicitly as specific tensor spaces and subspaces, respectively. The tensor space formulation leads to a solution in terms of an operator given by a tensor product. The results bridge the gap between control theory and the metric theory of tensor product spaces. Henceforth the dual and pre-dual formulations lead to finite dimensional convex optimizations which approximate the optimal solution within desired accuracy.

Then we investigate minimizing the **mixed $\mathcal{L}^2/\mathcal{H}^\infty$ norm** of the spatial, temporal closed loop systems, respectively. Such a mixed norm is induced by the aforementioned disturbances, which allows for more flexible and accurate specification of the desirable closed-loop behavior We solve this problem by utilizing the orthogonal projection techniques. Provided θ fixed, the counterpart involving $\mathcal{H}^\infty$-norm could be achieved by following standard techniques in solving *model matching problem*.

Finally, we turn our attention to a stochastic version of the distributed control problem. Instead of using the solution from Markov Decision Process, we consider a controlled Itô-Lévy process where the **'Markovian' property doesn't hold any more**. It should be noted that this is a more complicated situation than the case where standard stochastic maximum principle would fail. Therefore, we need to apply a Malliavin calculus approach to derive a maximum principle, where the adjoint processes are explicitly expressed by the parameters and the states of the system.

1.4 Dissertation Outline

The dissertation is organized as follows: In Chapter 2, we analyze a unifying stochastic control approach for achieving joint performance and power control of Energy Efficient Buildings and Virtualized Web Servers. Starting from Chapter 3, we start our second topic - distributed control for spatially-invariant systems. We first introduce some of the notations that will be used throughout the following chapters. It is followed by an optimal distributed H^2 control problem for spatially-invariant systems in Chapter 4. In Chapter 5 we analyze the open problem in Optimal Distributed Control by using a novel Duality and Operator Theoretic Approach. Distributed Mixed $\mathcal{L}^2/\mathcal{H}^\infty$ control problem Synthesis for Spatially Invariant Systems in Chapter 6. Furthermore, an initial attempt of using Malliavin calculus based stochastic maximum principle to solve the decentralized stochastic control problem has been explained in Chapter 7. Finally, we conclude this dissertation with a summary of work and directions for future work in Chapter 8.

Part I

Stochastic Control for Energy Optimization

Chapter 2

Stochastic Control for Energy Optimization in Energy Efficient Buildings and Virtualized Web Servers

Building climate control and data centers consume most of the power in smart grids. In this chapter, we propose a unifying stochastic control approach for achieving joint performance and power control of Energy Efficient Buildings and Virtualized Web Servers. We employ a constrained Stochastic Linear Quadratic Control (cSLQC) by minimizing a quadratic cost function with a disturbance assumed to be Gaussian. The problem is formulated to minimize the expected cost subject to a linear constraint and a probabilistic constraint. By using cSLQC, the problem is reduced to a semidefinite optimization problem, where the optimal control can be computed efficiently by Semidefinite programming (SDP). Simulation results are provided to demonstrate the effectiveness and power efficiency by utilizing the proposed control approach.

Parts of the results in this chapter appeared in [48, 49, 50].

2.1 Introduction

The key goal in energy efficient buildings is to reduce energy consumption of Heating, Ventilation, and Air-Conditioning (HVAC) systems while maintaining a comfortable temperature and humidity in the building. Virtualization machine (VM) technology, as an important enabler of cloud computing [51], has been widely used in the modern data centers. Both HVAC units and data centers are energy consuming monsters in modern smart grids. Buildings alone consume up to 40% of the energy produced in the US [3]. Advanced sensors and controls have the potential to reduce the energy consumption of buildings by 20-40% [4, 5]. Heating, Ventilation and Air Conditioning (HVAC) systems play a fundamental role in maintaining a comfortable temperature environment in buildings and account for 50% of building energy consumption [3]. Furthermore, the data processing, data storage and telecommunications industries are crucial parts of the American information economy, the rapid growth of these industries have led to an increase in electricity use. Therefore, significant potential for energy savings exist by optimally controlling HVAC systems data centers to reduce consumption, as well as reducing the carbon pollution.

2.1.1 Building Climate Control

Typical building controls are set-point based, where zone-level temperature measurement is used for taking control action to keep the zone temperature in a comfortable range. To make buildings more energy and cost efficient, intelligent predictive automation can be used instead of conventional automation. For instance, the predictive automation controllers can operate the buildings passive thermal storage, based on predicted future disturbances (e.g. weather forecast), by making use of low cost energy sources [26]. The goal is to design an optimal controller that can realize the temperature requirement and minimize energy consumptions.

The accuracy of the controller heavily depends on the assumption that the sensor always provides exact temperature measurement. However, this assumption is not always valid due to the measurement error or real-world environment noise. Consequently, an effective controller for HVAC systems should incorporate time-dependent energy costs, bounds on the control actions, noise from the sensors, as well as account for system uncertainties, i.e., weather conditions and occupancy. Compared with the deterministic control approaches, a key advantage of stochastic control approaches is that a noise term is considered in the model, which represents the unknown and uncertain elements in the system.

Building climate control leads naturally to probabilistic constraints as current standards explicitly state, zone temperatures should be kept within a comfort range with a predefined probability [33, 34]. In order to address this issue and explicitly account for system uncertainties, some efforts have been made for studying a stochastic version of MPC (SMPC) including probabilistic constraints. [35] employed stochastic MPC technique to compute the control strategy for a cost function which was linear in the control variable for the thermal dynamics in a linear state-space model, which described thermal energy and temperatures. [36] proposed a tractable approximation method for the problem. Both schemes in [35] and [36] considered chance constraints and solved them by using affine disturbance feedback.

2.1.2 Virtualized Web Servers

Similaryly, an energy-efficient virtualized server should guarantee the desired application performance by dynamically adjusting the power states of the processor. The virtualization technology can support environment isolation, fault isolation for multi virtual machines (VMs), these VMs can run different operating systems and applications as if they are running on the physical separate machines, but the contention of shared resource (i.e. processor, last level cache and bandwidth) among multiple customer applications leads to a performance interference effect

[11]. Moreover, the workloads of web applications show highly volatile variation and unexpected burst in terms of request arrival rate [12]. It is infeasible to decide the resource allocation solution (e.g. how many resources should be allocated and how to assign them) in a fixed way. In such environments, the control approach needs to be robust and adaptive to dynamic variations in workload.

It aims at addressing the instability and inefficiency issues due to dynamic Web workloads. It features a coordinated control architecture that optimizes the resource allocation and minimizes the overall power consumption while guaranteeing the server level agreements (SLAs). Due to the interference effect among the co-located virtualized web servers and time-varying workloads, the relationship between the hardware resource assignment to different virtual servers and the web applications' performance is considered as a coupled Multi-Input-Multi-Output (MIMO) system and formulated as a robust optimization problem.

The goal of proposed control system is to dynamically select the resource allocations and scale the CPU frequency to reduce power consumption without performance degradation. The processer is our control objective because it is the main contributor to the total power consumption of a server.

2.1.3 Unifying Approach

Now it is clear that both the two services share similar performance requirements and challenges from the control point of view. In this chapter, we consider a unifying stochastic control approach for achieving joint performance and power control of Energy Efficient Buildings and Virtualized Web Servers. We study a quadratic cost function in terms of temperature errors and control inputs, which is subject to several constraints on the room temperature and control input. In particular, we only consider the case where we assume that the disturbance is **Gaussian** and the problem is formulated to minimize the expected cost subject to a linear constraint on control input and a **probabilistic constraint** on the state. The latter constraint can be reduced to a hard constraint on control input exactly [47]. It should be remarked that the power of this proposed control technique could be extended to a more general **norm-bounded** case with distribution unknown and the problem is formulated as a min-max problem. By using the cSLQC approach proposed in [47, 52], the optimal solutions of problems in both cases may be solved via semidefinite programming exactly. Such control technique has already been applied in doubly-fed induction generator (DFIG) to help make full use of wind energy as well as producing and absorbing reactive power [53, 54, 55, 56, 57].

The rest of the chapter is organized as follows: In section 2.2, a general Constrained Stochastic Linear-quadratic control (cSLQC) problem is formulated. In section 2.3, we introduce the control techniques used in this work and solve the problems. Section 2.4 presents the simulation results to show the performance of the methods in controlling the building climate. In Section 2.5, numerical examples are provided to verify the effectiveness of the proposed methods for virtualized servers. Section 2.6 concludes this chapter.

2.2 Problem Formulation

We apply the cSLQC theory [47] to design the controller. cSLQC is a tractable control technique that can deal with stochastic discrete-time linear systems in the presence of control and state constraints. This characteristic makes the cSLQC well suited for building climate control.

2.2.1 Cost Function

We consider the problem where the temperature t_1 is required to remain within certain bounds of a constant in the presence of the disturbance vector d. Moreover, we can assign setpoints for t_1, t_2 and t_3, but without any other constraints on t_2 and t_3. Thus, we can regulate the output error $e_k := x_k - x_r$ at time k, where x_r is the setpoint vector of x. We hope to minimize the error e to keep the temperature t_1 close to the desired value. Meanwhile, we also hope to use as less power as we can to save energy. Thus, our objective is to find for the system (2.14) discretized, the M-control sequence $\{u_0, \cdots, u_{M-1}\}$, where $u_i := u(t_i)$, $i = 0, \cdots, M$; M is an integer large enough, $t_i = i\Delta T$, where $\triangle T$ is the sampling period; and corresponding state sequence $\{x_0, \cdots, x_{M-1}\}$ and error sequence $\{e_0, \cdots, e_{M-1}\}$, that minimize the finite horizon objective function:

$$V_N(e_0, \mathbf{u}, \omega) := \frac{1}{2}[(x_N - x_r)^T P(x_N - x_r) + \sum_{k=1}^{N-1} e_k^T Q e_k + \sum_{k=0}^{N-1} {u_k}^T R u_k] \tag{2.1}$$

where $P \geq 0$, $Q \geq 0$ (i.e., semi-definite positive matrices), $R > 0$ (i.e., positive definite matrix), N is the prediction horizon, and

$$\begin{aligned} \mathbf{x} &:= [x_0^T, ..., x_N^T]^T \\ \mathbf{u} &:= [u_0^T, ..., u_{N-1}^T]^T \\ \omega &:= [\omega_0^T, ..., \omega_{N-1}^T]^T \end{aligned} \tag{2.2}$$

The differences between the cost function above and those considered in [35] and [36] are that [35] used a linear cost function in the control input and [36] assumed the disturbance was 0 in the cost function which simplified the problem.

2.2.2 Constraints

Due to the unknown disturbances d_k, the state x_k is not exactly known. It is more reasonable to utilize the soft constraints on the state, i.e. we do not require constraints on the response time to be satisfied at all time, but only with a predefined probability. For example, constraint (2.3) requires that the condition $\mathbf{G}_i\mathbf{x} > \mathbf{g}_i$ is fulfilled with probability smaller or equal than α_i. Hence, instead of using hard constraints on the state or no constraint, we use the uncertain linear constraints in a probabilistic sense [47]. Thus, similar as [35] and [36], the constraint on x_k can be described by the so-called chance constraint as follows:

$$P[\mathbf{G}_i\mathbf{x} > \mathbf{g}_i] \;\; \leq \;\; \alpha_i \tag{2.3}$$

The above constraint is non-convex and hard to resolve directly. In the first case when the disturbance is Gaussian, as shown in [35] and [36], the authors took u_k as affine disturbance feedback to approximate and simplify this constraint. However, if we do not assume any form of the control input, we can still simplify the chance constraint to a hard constraint exactly as already shown in [47].

Assume the ω are independent and normally distributed, i.e., $\omega \sim \mathcal{N}(\mu, \Sigma)$, where $\Sigma > 0$. Then, we have the following theorem from [47].

Theorem 2.1. *[47] Consider a linear system with the state written as*

$$\mathbf{x} = \tilde{\mathbf{A}}x_0 + \tilde{\mathbf{B}}\mathbf{u} + \tilde{\mathbf{C}}\omega \tag{2.4}$$

Then, the constraint

$$\mathbf{p}^T\mathbf{u} \leq q \tag{2.5}$$

where $\mathbf{p} = \tilde{\mathbf{B}}^T\mathbf{G}_i$, $q = \mathbf{g}_i - \mathbf{G}_i\tilde{\mathbf{A}}x_0 - \mathbf{G}_i\tilde{\mathbf{C}}\mu - \parallel \Sigma^{\frac{1}{2}}\tilde{\mathbf{C}}^T\mathbf{G}_i \parallel_2 \Phi^{-1}(1-\alpha_i)$ *implies the chance constraint (2.3).*

Meanwhile, the control efforts are required to remain in a certain interval:

$$u_{min} \leq u_k \leq u_{max}. \tag{2.6}$$

Then, the problem can be formulated as follows:

Problem 2.2. *Find*

$$\mathbf{u}(x_0) := arg\ min_{\mathbf{u}}\ \mathbb{E}_\omega V_N \tag{2.7}$$

subject to (2.5), (2.6), and discretized version of the state-space model.

In the previous cost function, $\mathbb{E}_\omega(\bullet)$ denotes the expectation operator with respect to the Gaussian disturbance ω.

In the next section, we employ the technique developed in [47] to transform the problem to a semidefinite optimization problems, where they can be solved efficiently.

2.3 Control Strategies

If there is no constraint, the optimization problem under Gaussian disturbance can be solved by linear quadratic regulator (LQR) through Bellman's recursion. However, with constraints, this approach involves a huge amount of computation to find the optimal solution. To find the optimal values for each problem, we employ the SLQC to find the solution by formulating the problems as semidefinite optimization problems.

2.3.1 SDP Approach for Problem 2.2

In this section, we applied the technique in [47] to formulate Problem 2.2 as a semidefinite optimization problem. Unlike the MPC method, which is quadratic programming, the problem will be converted to an SDP optimization problem. An obvious result about the cost function is given in the following proposition.

Proposition 2.3. *The cost function (2.1) can be written as:*

$$\begin{aligned} V_N\left(e_0, \mathbf{u}, \omega\right) &= 2\mathbf{a}^T e_0 + e_0^T \mathbf{A} e_0 + 2\mathbf{b}^T \mathbf{u} + \mathbf{u}^T \mathbf{B} \mathbf{u} + 2\mathbf{c}^T \omega \\ &+ \omega^T \mathbf{C} \omega + 2\mathbf{u}^T \mathbf{D} \omega + \hat{l} \end{aligned} \tag{2.8}$$

for vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ *and matrices* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ *with appropriate dimensions, and where* $\mathbf{B} > 0, \mathbf{C} \geq 0$.

Proof: The proof can be found in the Appendix A.1. ■

Similarly as [47], let $\mathbf{h} = \mathbf{c} - \mathbf{D}^T \mathbf{B}^{-1} \mathbf{b}$ and $\mathbf{F} = \mathbf{B}^{-1/2} \mathbf{D}$, then by eliminating the constant terms and take $\mathbf{u} = \mathbf{B}^{-1/2} \mathbf{y} - \mathbf{B}^{-1} \mathbf{b}$, the cost function above can be further reduced to be:

$$\tilde{V}_N\left(e_0, \mathbf{y}, \mathbf{w}\right) = \mathbf{y}^T\mathbf{y} + 2\mathbf{h}^T\mathbf{w} + 2\mathbf{y}^T\mathbf{F}\omega + \omega^T\mathbf{C}\omega \tag{2.9}$$

Taking the expectation of the above cost, we have

$$\hat{V}_N\left(e_0, \mathbf{y}, \omega\right) = \mathbf{y}^T\mathbf{y} + 2\mathbf{h}^T\mu + 2\mathbf{y}^T\mathbf{F}\mu + trace(\mathbf{C}\Sigma) \tag{2.10}$$

Again, taking away constant terms, the cost to be minimized is $\hat{V}_N\left(e_0, \mathbf{y}, \omega\right) = \mathbf{y}^T\mathbf{y} + 2\mathbf{y}^T\mathbf{F}\mu$. Then, the problem 2.2 is equivalent to find $\mathbf{u}(x_0) := arg\ min_{\mathbf{u}}\ \hat{V}_N$. This problem can be solved through SDP to obtain the optimal solution, as shown in the next theorem.

Theorem 2.4. *Problem 2.2 may be solved by the following semidefinite optimization problem:*

$$\begin{aligned} &minimize\ z \\ &subject\ to\ (2.18),(2.5) \\ &\begin{bmatrix} \mathbf{I}_N & \mathbf{y} + \mathbf{F}\mu \\ \mathbf{y}^T + \mu^T\mathbf{F}^T & z + (\mathbf{F}\mu)^T\mathbf{F}\mu \end{bmatrix} \geq 0 \end{aligned} \tag{2.11}$$

in decision variables $\mathbf{y}$ *and* z*.*

Proof: The proof is given below by following the technique in *Theorem 3* in [47].

First, note the minimization of $\hat{V}_N\left(e_0, \mathbf{y}, \mathbf{w}\right)$ can be rewritten as

$$\begin{aligned} &minimize\ z \\ &subject\ to\ z - \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{F}\mu \geq 0 \end{aligned} \tag{2.12}$$

The constraint (2.12) can be further written as

$$
\begin{aligned}
& z - \mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{F}\mu - (\mathbf{F}\mu)^T\mathbf{F}\mu + (\mathbf{F}\mu)^T\mathbf{F}\mu \geq 0 \\
\Updownarrow & \\
& z + (\mathbf{F}\mu)^T\mathbf{F}\mu - (\mathbf{y} + \mathbf{F}\mu)^T (\mathbf{y} + \mathbf{F}\mu) \geq 0
\end{aligned} \tag{2.13}
$$

Then, by Schur complement lemma, (2.13) can be formulated as (2.11). Moreover, note that (2.6) and (2.5) are linear constraints on the control input, which can be added without increasing the complexity type. Thus, we obtain the statement. ■

2.3.2 Chance Constraints on the Performance

Another interesting requirement is the performance guarantee. The work in [47] has demonstrated that the probability

$$\mathbf{P}(V_N(e_0, \mathbf{u}, \omega) > v) \leq \epsilon$$

may be implied by a convex quadratic constraint, which can be added to either problem without raising the complexity type.

2.4 Numerical Example I - HVAC Systems

2.4.1 Building Climate Plant

It is well known that the HVAC control can be approached using Model Predictive Control (MPC) strategy. In this section, we describe the model used in this work and formulate the problem. The system model was proposed in [26] and employed in [35]. We briefly describe the model in this section.

Building Model

Consider the following continuous-time Linear Time Invariant (LTI) system based on the dynamics of the room temperature, interior-wall surface temperature, and exterior-wall core temperature:

$$
\begin{aligned}
\dot{t}_1 &= \tfrac{1}{C_1}\left[(K_1 + K_2)(t_2 - t_1) + K_5(t_3 - t_1) + K_3(\delta_1 - t_1) + u_h\right. \\
&\left.+u_c + \delta_2 + \delta_3\right] \\
\dot{t}_2 &= \tfrac{1}{C_2}[(K_1 + K_2)(t_1 - t_2) + \delta_2] \\
\dot{t}_3 &= \tfrac{1}{C_3}[K_5(t_1 - t_3) + K_4(\delta_1 - t_3)]
\end{aligned}
$$

where the parameters used in the above model are defined as:

t_1 : room air temperature $[^\circ F]$

t_2 : interior-wall surface temperature $[^\circ F]$

t_3 : exterior-wall core temperature $[^\circ F\ ]$

u_h : heating power (≥ 0) $[kW]$

u_c : cooling power (≤ 0) $[kW]$

δ_1 : outside air temperature $[^\circ F]$

δ_2 : solar radiation $[kW]$

δ_3 : internal heat sources $[kW]$

with constants chosen as:

$C_1 = 9.356 \cdot 10^5$	$kJ/^\circ F$	$C_2 = 2.970 \cdot 10^6$	$kJ/^\circ$F
$C_w = 6.695 \cdot 10^5$	$kJ/^\circ$F	$K_1 = 16.48$	$kW/^\circ$ F
$K_2 = 108.5$	$kW/^\circ$ F	$K_3 = 5$	$kW/^\circ$F
$K_4 = 30.5$	$kW/^\circ$ F	$K_5 = 23.04$	$kW/^\circ$F

The system states are the room air temperature t_1, interior wall surface temperature t_2, and exterior wall core temperature t_3. The control signals u_h and u_c represent heating and cooling power, and they can be combined as one variable

$u = u_h + u_c$ because heating and cooling are not simultaneous. For more details about this model, please refer to [26, 35].

Define the state vector x, the control signal vector u, and the environment stochastic disturbance vector ω as:

$$x := \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}, \quad u := \begin{bmatrix} u_h \\ u_c \end{bmatrix}, \quad \omega := \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$$

The continuous-time state-space model can then be described compactly as:

$$\dot{x} = A_c x + B_c u + C_c \omega \tag{2.14}$$

where

$$A_c := \begin{bmatrix} -\frac{1}{C_1}(K_1 + K_2 + K_3 + K_5) & \frac{1}{C_1}(K_1 + K_2) & \frac{K_5}{C_1} \\ \frac{K_1+K_2}{C_2} & -\frac{(K_1+K_2)}{C_2} & 0 \\ \frac{K_1}{C_3} & 0 & -\frac{(K_5+K_4)}{C_3} \end{bmatrix}$$

$$B_c := \begin{bmatrix} \frac{1}{C_1} + \frac{1}{C_2} \\ 0 \\ 0 \end{bmatrix}, \qquad C_c := \begin{bmatrix} \frac{K_3}{C_1} & \frac{1}{C_1} & \frac{1}{C_1} \\ 0 & \frac{1}{C_2} & 0 \\ \frac{K_4}{C_3} & 0 & 0 \end{bmatrix}. \tag{2.15}$$

Discretizing system (2.14) with period h_k and applying a zero-order-hold, one obtains:

$$x_{k+1} = A_d x_k + B_d u_k + C_d \omega_k \tag{2.16}$$

where the parameters can be computed from the continuous-time model, and $x_k = [t_{1,k},\, t_{2,k},\, t_{3,k}]^T$.

We assume the following constraints are imposed on the temperatures during a day to satisfy the requirement:

$$68°\text{F} \leq t_{1,k} \leq 80.6°\text{F} \tag{2.17}$$

Additionally, we also assume that the control input is the critical actuator yielding its own working properties and conditions. It is meaningful to set a reasonable bound for the u_k. Otherwise, it would cost a lot to build and drive the actuator.
Therefore the control constraint is assumed to be written in terms of u_k as:

$$-50 \leq u_k \leq 200 \tag{2.18}$$

where $u_k > 0$ means heating and the opposite means cooling.

From above constraints, we can observe that both the room air temperature and control signal are constrained. In the next section, the control problem is formulated.

2.4.2 Description of the experimental setup

In this section, we present simulation results which demonstrate validity of the SLQC method in the above problems. The system model is described in Section **??**. As mentioned before, this model was proposed in [26] and employed in [35]. The desired temperature or reference temperature of the room is set as $22°C = 71.6°F$. The temperature is sampled every 10 minutes, and we plot t_1 and the control input for each method during a period of 10 days in the sequel. The disturbances corresponding to different states are shown in Fig. 2.1.

2.4.3 Hysteresis band

It should be noted that the AC continues to cool the building for a few minutes even after it is turned off because of the dynamics of the heat pump. Specifically, it takes a while for the evaporator that cools the air to warm up, and so it keeps cooling

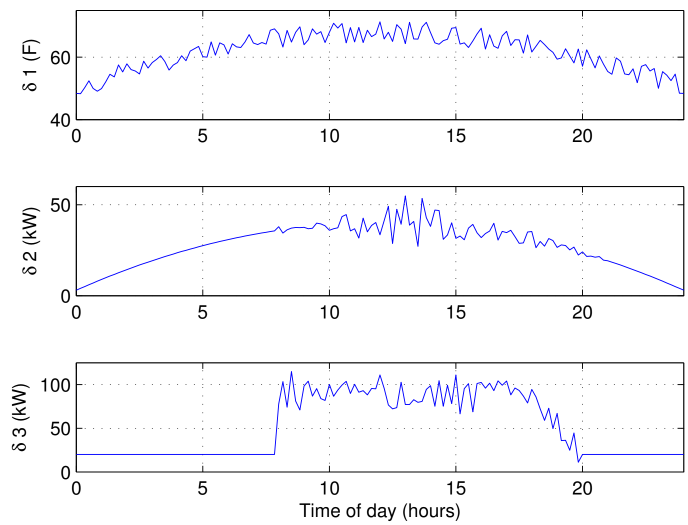


Figure 2.1: Disturbance to the building climate system.

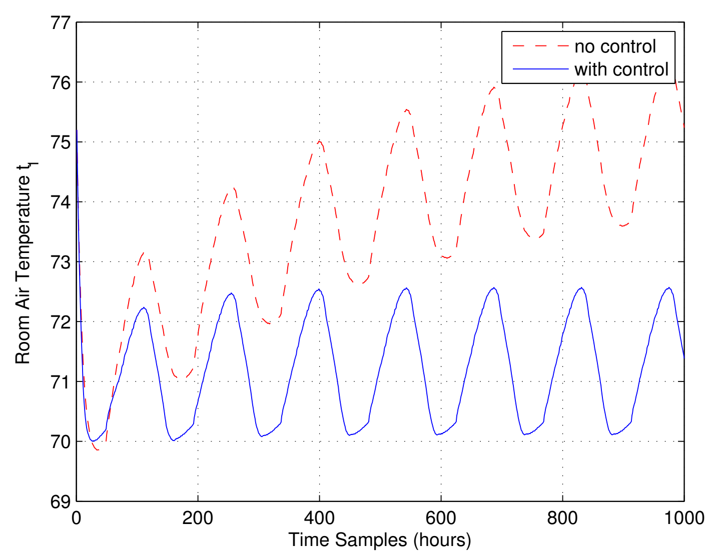


Figure 2.2: Room temperature t_1 comparison with/without control using LQR.

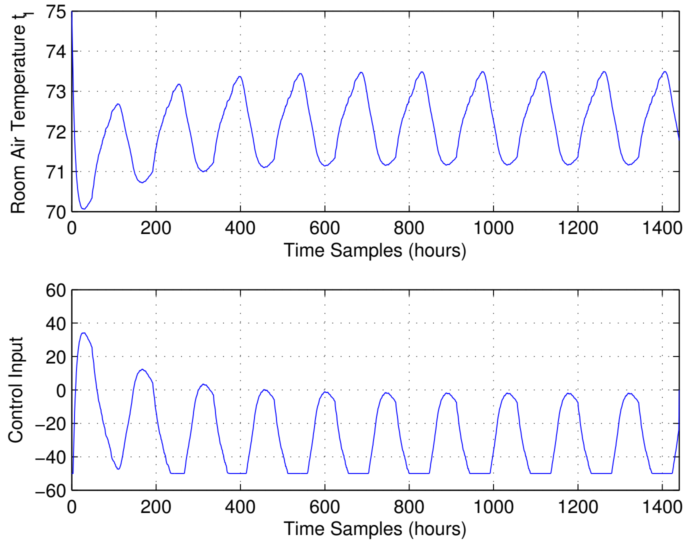


Figure 2.3: Room temperature t_1 and control power in 10 days using cSLQC in Theorem 2.4.

the air for some time after the heat pump is turned off [58]. Therefore, there exists a "delay" in the system model. For the purpose of approximating the real energy consumption of HVAC, we have to consider a 0.5°F hysteresis band ($k = 0.5$°F) in the control scheme.

Practically, we use this hysteresis band to represent the delay of the cooling system for example. In order to cool the room to reach the setting temperature, i.e. x_r, we need to turn off the cooling unit k°F before reaching x_r. This also helps save energy consumption as shown in Sec. 2.4.4.

2.4.4 Summary of the results

The Disturbance Distribution is Gaussian

First, we plot the trend of t_1 using LQR control in Fig. 2.2, and using SDP through (2.11) in Fig. 2.3. It is obvious that both LQR and cSLQC techniques can keep room temperature t_1 in the desired range and close to the reference temperature. Notice that the control input is also bounded below by -50, which is as desired.

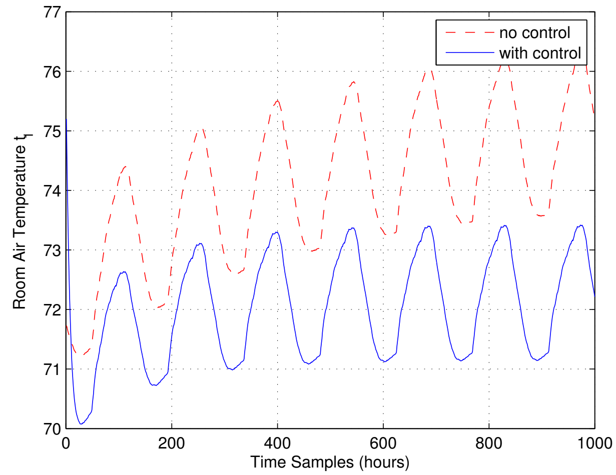


Figure 2.4: Room temperature t_1 comparison with/without control using cSLQC in Theorem 2.4.

To evaluate the control performance for both of the controllers, we compare the room temperature t_1 with and without enabling the controller as shown in Fig. 2.2 and Fig. 2.4.

Simulation results show that without the controller, the temperature will depart far away from the reference temperature.

To further illustrate the difference in energy saving between the two control principles, we compute the total input energy, i.e. the energy cost in all the 10 days denoted by $\| \mathbf{u} \|_2$ (2-norm) for both methods.
For LQR controller, $\| \mathbf{u} \|_2 = 1665.3$; while for cSLQC, $\| \mathbf{u} \|_2 = 1268.3$, which show that the proposed cSLQC technique achieves a more efficient control policy which contributes to reducing energy consumption.

Control with hysteresis band

As discussed in Sec.2.4.3, here we incorporate the proposed cSLQC with a 0.5°F hysteresis band to approximate the real working condition of HVAC as well as pursuing more energy efficiency. In the stage of the experiments we set the control signal to be zero as long as the temperature t_1 reaches the hysteresis band. Basically,

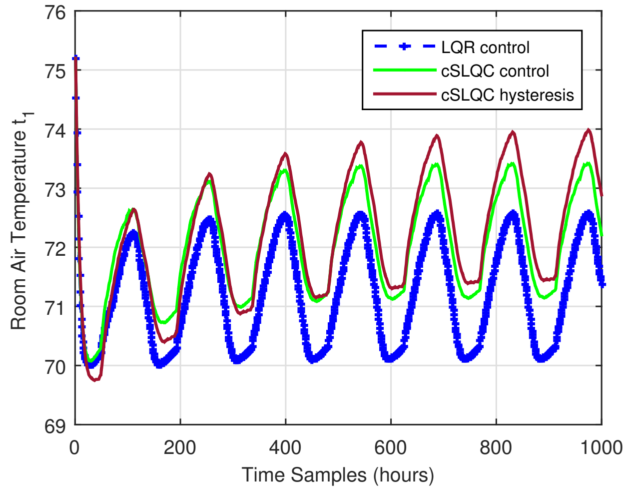


Figure 2.5: Room temperature t_1 comparison.

we compare the energy consumption, thereby the energy cost, of the three control strategies running for 10 days and 30 days separately. It should be remarked that the disturbance is not exactly the same value for each day. In stead, the disturbance follows the same distribution, while the values may not be equal for different days.

The results shown in Fig. 2.5 and Fig. 2.6 clearly indicate that both our controllers outperform the current LQR controller in terms of both energy use and violations of the thermal comfort range. It can be seen from Fig. 2.5 that the cSLQC controller never breaks the desired comfort band constraints, while the LQR controller tends to have violations of the lower bound on the temperature.

Moreover, the temperature variations are smaller with cSLQC, which is a more favorable behavior in terms of comfort. The improvements in energy saving for both short-term and long-term can be explained by Fig. 2.6, where the hysteresis effect is also considered.

2.5 Numerical Example II - Virtualized Web Servers

As green computing is becoming a popular computing paradigm, the issue of energy-efficient data center with performance assurance becomes increasingly important. In

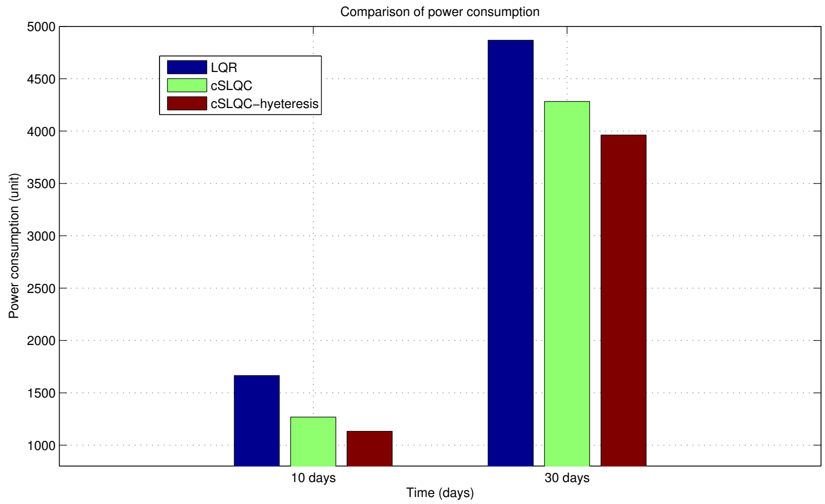


Figure 2.6: Comparison of power consumption for control inputs in 10 days / 30 days.

this example, we apply the same control approach for joint performance and power control of co-located multiple virtualized web servers.

2.5.1 System Modeling

For a virtualized server that hosts n virtual machines, T_1 is the control period. As the control variable, $rt_i(k)$ is the average response time of VM_i in the time interval $[(k-1)T_1, kT_1]$. $r_i(k)$ is the relative response time of VM_i, namely, $r_i(k) = \frac{rt_i(k)}{d_i}$, where d_i is the maximum allowed response time of VM_i. The reference $\hat{r}(k)$ denotes the average relative response time of all VMs, namely, $\hat{r}(k) = \sum_{i=1}^{n} r_i(k)/n$. As the manipulated variable u, we use *weight* to assign the CPU resource to virtual machine. Specifically, the amount of CPU resource allocation to each VM is proportional to the weight value.

Specifically, let $\mathbf{u}(k) = [u_1(t), u_2(t), \ldots, u_n(t)]^T$ be the input vector and $\mathbf{r}(k) = [r_1(t), r_2(t), \ldots, r_n(t)]^T$ be the output vertor. Instead of directly using $\mathbf{u}(k)$ and $\mathbf{r}(k)$, we use their differences $\Delta\mathbf{r}(k) = \mathbf{r}(k) - \hat{\mathbf{r}}(k)$ and $\Delta\mathbf{u}(k) = \mathbf{u}(k) - \hat{\mathbf{u}}$, where $\hat{u}$ is a typical value in Xen. The system model can be described as:

$$\Delta\mathbf{r}(k+1) = \mathbf{A}(k)\Delta\mathbf{r}(k) + \mathbf{B}(k)\Delta\mathbf{u}(k) + \mathbf{C}(k)\omega(k) \tag{2.19}$$

where $\omega(k) \in \mathbb{R}^n$ is the disturbance vector (an unknown quantity), the matrices $\mathbf{A}(k) \in \mathbb{R}^{n\times n}$, $\mathbf{B}(k) \in \mathbb{R}^{n\times n}$, and $\mathbf{C}(k) \in \mathbb{R}^{n\times n}$ respectively. The typical assumption is that the disturbances ω are independent and normally distributed (i.i.d), i.e., $\omega \sim \mathcal{N}(0, I)$.

Hence the finite horizon cost function based on the current state $\Delta\mathbf{r}(0)$ can be described as follow:

$$J_1(\Delta\mathbf{r}(0), \Delta\mathbf{u}, \omega) = \sum_{k=1}^{N} \Delta\mathbf{r}(k)^T Q \Delta\mathbf{r}(k) + \sum_{k=0}^{N-1} \Delta\mathbf{u}(k)^T R \Delta\mathbf{u}(k) \tag{2.20}$$

where Q and R are positive-semidefinite weighting matrices that establish a trade-off between control error and control cost. we choose the resource allocation solution that performs best for the most pessimistic disturbance ω within a reasonable uncertainty set Θ. Specifically, we search for an optimal control $\Delta\mathbf{u}^*$ such that

$$\min_{\Delta\mathbf{u}\in\mathbb{R}^{N\times n}} \{\max_{\omega\in\Theta} J_1(\Delta r(0), \Delta u, \omega)\}. \tag{2.21}$$

2.5.2 Load Balancing

To evaluate the cSLQC controller performance in the load balancing layer, the controller in the second layer temporarily disabled. Fig. 2.7 shows an example of time varying incoming workloads to each of the three virtual machines. The strength of concurrency level exhibits time-of-day variations typical of many enterprise workloads and the concurrency level changes significantly within a vary short time period. The synthetic web workloads used in our experiment, in part, referring to the log files from the Soccer World Cup 1998 Web site [59]. To quantify and compare the performance of cSLQC with a benchmark solution PARTIC [37].

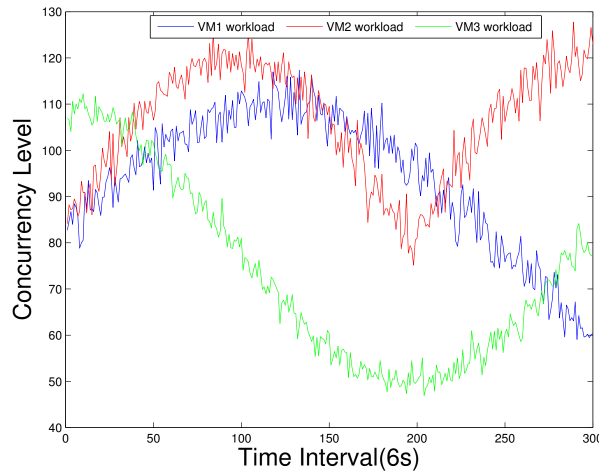
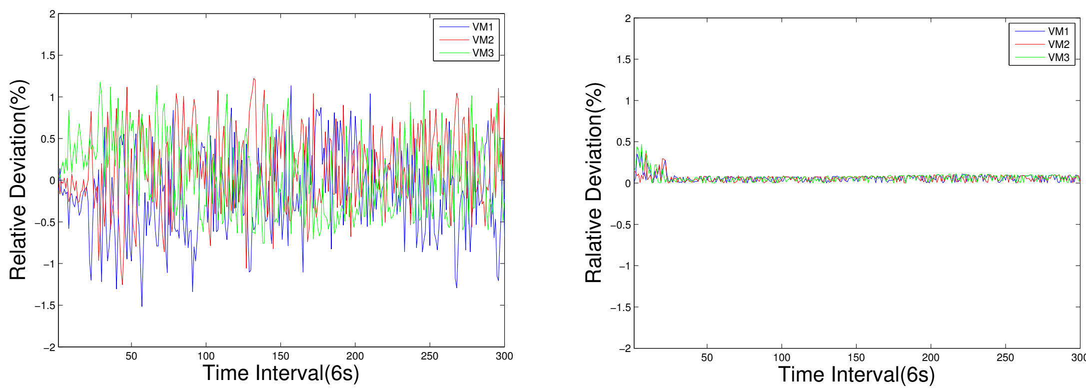


Figure 2.7: High dynamical workloads

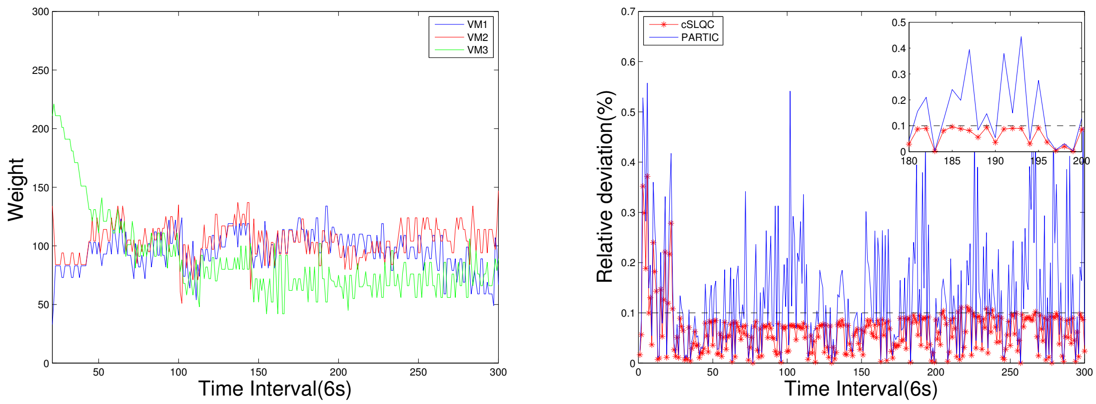


(a) PARTIC method.

(b) cSLQC method.

Figure 2.8: Comparison with different method in term of performance.

Fig. 2.8a is the performance of PARTIC. It shows that the response time of each VM is unstable and exhibiting large oscillation seriously during the whole process, since it is an offline method and the control accuracy is limited to a certain range. Obviously, PARTIC cannot adapt to the high dynamical workload case. cSLQC achieves the performance target by dynamically adjusting the CPU resource allocation among the three virtual machines. From the Fig. 2.8b and 2.9a, we see that the controller gives different portions of CPU resource to the three virtual machines such



(a) CPU allocation.

(b) Performance deviation on VM1.

Figure 2.9: Performance analysis of cSLQC controller.

that each VM converge to the average response time of all VMs. Hence the target is met.

Fig. 2.9b compares the performance assurance capability of cSLQC controller on $VM1$ with that of PARTIC in the face of high dynamical workload. We can see that cSLQC could achieve small relative response time after the first several step. It implies that it can provide performance guarantee in terms of response time when the workload varies significantly. cSLQC controller significantly outperformed the PARTIC in terms of control accuracy and stability, since most of the control error is below 0.1 (i.e., defined limit level).

Fig. 2.10 illustrates the improvement in performance efficiency by cSLQC controller for various constraint violations, compared with PARTIC. It demonstrated that the performance of cSLQC controller degrades when allowing more constraint violations.

2.5.3 System Robustness

In this experiment, we evaluate the performance of stochastic control under a bursty workload. As a case study, we apply a bursty workload to VM2 as shown in Fig. 2.11. The workload of VM1 and VM3 is running at the normal level 60 during the

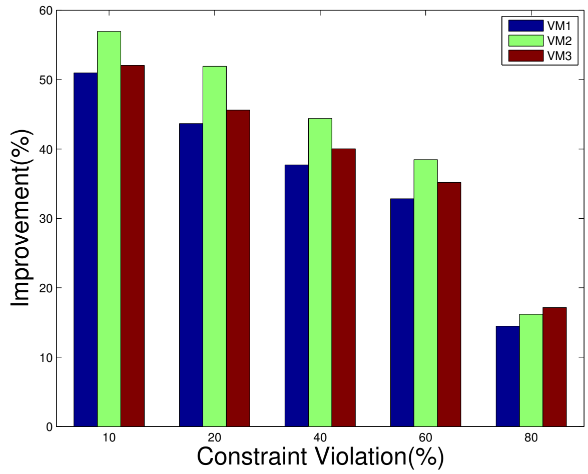


Figure 2.10: Overall control performance.

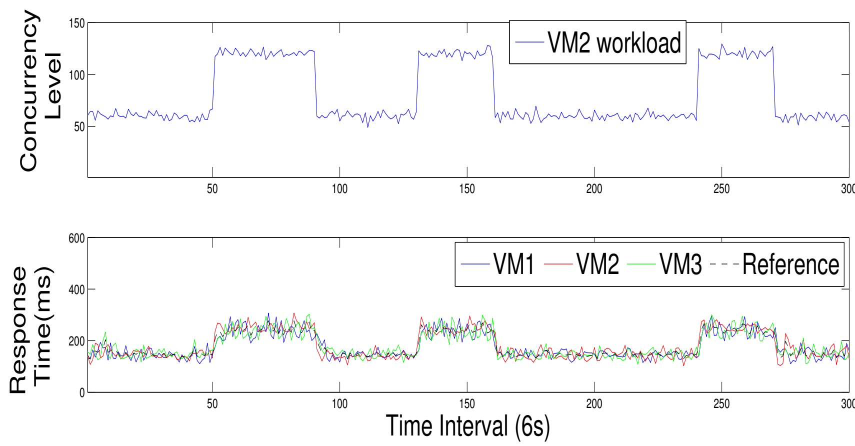


Figure 2.11: Performance of cSLQC under the bursty workload

whole process. Furthermore, we add white Gaussian noise to the concurrency level to simulate the unknown and unpredictable requests. Fig. 2.11 shows that the stochastic control can provide the same response time for all three VMs in the face of the bursty workload case. Clearly, VM1 and VM3 experience a longer response time during the burstiness periods, since cSLQC controller dynamically assigns less resource to VM1 and VM3, and correspondingly assigns more resource to VM2 to cope with the abrupt changes in the workload. As a result, the response time of all three VMs converge to the average response time, which means the desired response time has been guaranteed.

2.6 Summary

In this chapter, we proposed a unifying cSLQC controller for energy efficient buildings and virtualized web servers, aiming at reducing the energy consumption in smart grids. The constrained SLQC approach is employed to solve stochastic optimization problems with chance constraints by SDP. We first present an investigation of constrained quadratic control of room temperature on a dynamic building climate model. Moreover, the proposed cSLQC controller has also been applied onto Virtualized Web Servers.

The mechanism to account for the probabilistic nature of the disturbances affecting the comfort indicators is simplified to a hard constraint exactly without using affine disturbance feedback. Moreover, we consider a stochastic quadratic cost function, which is taken expectation with respect to Gaussian disturbances.

Part II

Optimal Distributed Control for Spatially Invariant Systems

Chapter 3

Spatially Invariant Systems

This chapter introduces the notation that will be used throughout the following chapters. This includes both the functional spaces that will be considered as well as our mathematical performance specification for control systems. One of the main themes in modern control theory is the utilization of norm based criteria to measure the optimal performance of a given control system [60]. We shall first introduce some normed spaces and basic notions of linear operator theory, in particular, the Hardy spaces $\mathcal{H}_2$ and $\mathcal{H}_\infty$. Then we discussed the a special class of large-scale systems, i.e. discrete Spatio-Temporal invariant systems. The original problem will be transformed into a standard feedback configuration which lies the foundation for the following chapters.

3.1 Distributed Parameter Systems

One of the most productive recent development is that of the theory of optimal control of systems with distributed parameters. This class of systems is much broader than simple systems with only lumped parameters. Actually, many of the real problems in control and design in airframe, shipbuilding, electronics, nucleonics, and other engineering fields are, in essence, problems of control of systems with distributed parameters. Spatial invariance is a strong property of a given system, which means that the dynamics of the system do not vary as we translate along some spatial axis [27].

As mentioned in [28], the main difference between distributed parameter systems and lumped parameter systems is that the former are characterized not by a finite set of quantities, coordinates of the object, that vary only in time, but generally by a set of functions that show the dependence of the parameters on the time and spatially variables or any combination of them. In the majority of cases, ideal control designed for lumped systems is not realizable because of the presence of additional constraints imposed on state functions and controlling actions. In particular, these constraints are related with spatial variables on which the state and control functions depend. The impossibility of realizing a perfect control process posed the problem of deriving an optimal process according to a definite, preassigned criterion.

Hence, an optimal control problem of distributed parameter systems obviously leads to the powerful apparatus of functional analysis [28]. In this case, the optimal control processes can be visually and geometrically interpreted in the functional phase space of the system. Then a variational of the state of the controlled system, if it occurs in time, is characterized by a definite point in the functional space of the system, and the transition of the system from one state to another, i.e., evolution in time, is characterized by a trajectory in the functional space. Therefore, in place of the usual finite-dimensional phase space of the system we must use the infinite dimensional functional space. The application of functional analysis methods permits us to generalized the important duality theory for distance minimization problem to systems with parameters distributed in space.

3.2 Hardy Spaces

We use the following notation in this work. The real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. The complex open unit disc is $\mathcal{D}$; its boundary, the unit circle, is $\mathbf{T}$; and the closed unit disc is $\bar{\mathcal{D}}$.

In particular, we denote as following:

$$\begin{aligned}
\mathcal{D} &= \{z;\, |z| < 1\} \\
\mathbf{T} &= \{z;\, |z| = 1\} \\
\bar{\mathcal{D}} &= \{z;\, |z| \leq 1\}
\end{aligned}$$

As mentioned in [61], it is a fact that if $f(x)$ is analytic at z_0 and also at each point in some neighborhood of z_0. Hence a function analytic at z_0 has a power series representation at z_0. Recall that the converse is also true such that a complex-valued function f is analytic in $\mathcal{D}$ provided that it is the sum of a convergent power series [62]

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{3.1}$$

Let V be a vector space over $\mathbb{C}$ (or $\mathbb{R}$) and let $\|\cdot\|$ be a norm defined on V. Then V is a *normed* space. A normed space is said to be *complete* if every Cauchy sequence in V converges in V. A complete normed space is called a *Banach space*. A *Hilbert space* is a complete inner product space with the norm induced by the inner product.

Definition 3.1. *[63] A function $g(z), z \in \mathcal{C}$, is in $\mathcal{H}^p(\mathcal{D})$, $1 \leq p \leq \infty$, if*

1. *: g is analytic in $\mathcal{D}$,*
2. *: it is defined almost everywhere on $\mathbf{T}$, and*
3. *: its $p-$norm defined by*

$$\begin{aligned}
\|g\|_p &= \sup_{r<1} \left(\frac{1}{2\pi} \int_0^{2\pi} |g(re^{j\theta})|^p d\theta \right)^{1/p}, \ (1 \leq p < \infty) \\
&= \left(ess \sup_{r<1,\, \theta \in [0,2\pi]} |g(re^{j\theta})| \right) (p = \infty)
\end{aligned}$$

is finite.

If only the first condition is not satisfied with $r = 1$, then it is in $\mathcal{L}_p(\mathbf{T})$.

Two of the most important norms employed in modern control analysis and design are the $\mathcal{H}_2$-norm and $\mathcal{H}_\infty$-norm.

The set $\mathcal{L}_2(\mathbf{T})$ is the Hilbert space of Lebesgue measurable functions on $\mathbf{T}$, which are square integrable, with inner product

$$\langle F, G \rangle = \frac{1}{2\pi} \int_0^{2\pi} \operatorname{trace}(F^*(e^{j\theta})G(e^{j\theta}))d\theta \tag{3.2}$$

Definition 3.2. *The Hardy space $\mathcal{H}_2$ consists of all analytic functions having power series representations with square-summable coefficients. That is,*

$$\mathcal{H}_2 = \left\{ f \ : \ f(z) = \sum_{n=0}^{\infty} a_n z^n \text{ and } \sum_{n=0}^{\infty} |a_n|^2 < \infty \right\}. \tag{3.3}$$

Theorem 3.3. *Every function in $\mathcal{H}_2$ is analytic on the open unit disk.*

It should be noticed that converse of the above theorem is not true since we could find such a counter-example as $f(z) = \frac{1}{1-z}$. It's obvious that $\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n$, while the coefficients of f are not square-summable.

Theorem 3.4. *[64] For every $z_0 \in \mathcal{D}$, the mapping $f \mapsto f(z_0)$ is a bounded linear functional on $\mathcal{H}_2$.*

As is standard, $\mathcal{H}_2$ denotes the Hardy space of functions analytic outside the closed unit disc, and at infinity, with square-summable power series.

$$\mathcal{H}_2 = \left\{ f : \{\infty\} \cup \mathbb{C} \setminus \bar{\mathcal{D}} \longrightarrow \mathbb{C} \,|\, \exists x \in \mathcal{L}_2(\mathbb{Z}_+) \text{ s.t. } f(z) = \sum_{k=0}^{\infty} x_k z^{-k} \right\} \tag{3.4}$$

The set $\mathcal{H}_2^\perp$ is the orthogonal complement of $\mathcal{H}_2$ in $\mathcal{L}_2$. The prefix $\mathcal{R}$ indicates the subsets of *proper real rational* functions. That is, $\mathcal{RL}_2$ is the set of transfer functions with no poles on $\mathbf{T}$, and $\mathcal{RH}_2$ is the set of transfer functions with no poles on or outside $\mathbf{T}$.

Theorem 3.5. *[64] Let f be analytic on $\mathcal{D}$. Then $f \in \mathcal{H}_2$ if and only if*

$$\sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta < \infty \tag{3.5}$$

Moreover, for $f \in \mathcal{H}_2$,

$$\|f\|^2 = \sup_{0<r<1} \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta \tag{3.6}$$

Definition 3.6. *[64] The space $\mathcal{H}_\infty$ consists of all the functions that are analytic and bounded on the open unit disk. The vector operations are the usual point-wise addition of functions and multiplication by complex scalars. the norm of a function f in $\mathcal{H}_\infty$ is defined by $\|f\|_\infty = \sup\{|f(z)| \,:\, z \in \mathcal{D}\}$.*

Since convergence in the norm on $\mathcal{H}_\infty$ implies uniform convergence on the disk, it is easily seen that $\mathcal{H}_\infty$ is a Banach space. Furthermore, every function in $\mathcal{H}_\infty$ is also in $\mathcal{H}_2$.

Definition 3.7. *A function $m \in \mathcal{H}_\infty(\mathcal{D})$ is called inner if $|m(z)| \leq 1$ for all $z \in \mathcal{D}$ and $|m(e^{j\theta})| = 1$ a.e. $\theta \in [0, 2\pi]$.*

This can be treated as all pass transfer functions since the inner functions always carry constant magnitude a.e. on $\mathbf{T}$.

Definition 3.8. *A function $g \in \mathcal{H}_\infty(\mathcal{D})$ is called outer if the closure of $g\mathcal{L}_+$ in $\mathcal{H}_2(\mathcal{D})$ is the whole space $\mathcal{H}_2(\mathcal{D})$, where $\mathcal{L}_+ = \left\{\sum_{k=0}^n a_k z^k,\ a_k \in \mathbb{C}, n \geq 0\right\}$.*

Note that the outer functions don't have a zero in $\mathcal{D}$, however may have zeros on $\mathbf{T}$, this would guarantee g to be invertible in $\mathcal{H}_\infty(\mathcal{D})$.

We denote $\mathbb{R}^{m\times n}$ as the set of $m \times n$ matrices in $\mathbb{R}$. This notation will also be used to denote $m \times n$ block matrices, where the dimensions of the blocks are implied by the context.

Also, we denote the subspace $\mathcal{L}_\infty(\mathbf{T})$ as the set of Lebesgue measurable functions which are bounded on $\mathbf{T}$. Similarly, $\mathcal{H}_\infty$ is the subspace of $\mathcal{L}_\infty$ with functions analytic outside of $\mathbf{T}$, and $\mathcal{H}_\infty^-$ is the subspace of $\mathcal{L}_\infty$ with functions analytic inside $\mathbf{T}$. Consequently, $\mathcal{RH}_\infty$ is the set of transfer functions with no poles outside of $\mathbf{T}$. Note that, in this case, $\mathcal{RH}_2 = \mathcal{RH}_\infty$; we will use these spaces interchangeably.

For any Hilbert spaces S, T and bounded operator $G: S \longrightarrow T$, we let $G^* : T \longrightarrow S$ denote its *adjoint operator*. The special case is when G is a real matrix; in which case, G^T denotes its transpose. Also, the following notation denotes the image of G.

$$GS = \{G(F) \in T \,|\, F \in S\} \tag{3.7}$$

Some useful facts about Hardy spaces which we will make use of in this work are :

- if $G \in \mathcal{L}_\infty$, then G$\mathcal{L}_2 \subset \mathcal{L}_2$
- if $G \in \mathcal{H}_\infty$, then G$\mathcal{H}_2 \subset \mathcal{H}_2$
- if $G \in \mathcal{H}_\infty^-$, then G$\mathcal{H}_2^\perp \subset \mathcal{H}_2^\perp$

For transfer functions $F \in \mathcal{RL}_2$, we use the notation

$$F(z) = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = C(zI - A)^{-1}B + D \tag{3.8}$$

Lastly, we define $P_{\mathcal{H}_2} : \mathcal{L}_{\in} \to \mathcal{H}_2$ as the orthogonal projection onto $\mathcal{H}_2$.

Definition 3.9. *[61] Two normed spaces V_1 and V_2 are said to be linearly isometric, denoted by $V_1 \simeq V_2$, if there exists a linear operator T of V_1 onto V_2 such that*

$$\|Tx\| = \|x\|$$

for all x in V_1. In this case, the mapping T is said to be an isometric isomorphism.

Let $\mathcal{H}$ be a Hilbert space and $M \subset \mathcal{H}$ a subset. Then the orthogonal complement of M, denoted by $M^\perp$ or $\mathcal{H} \ominus M$, is defined as

$$M^\perp = \{x : \langle x, y\rangle = 0, \forall y \in M, x \in \mathcal{H}\}. \tag{3.9}$$

Definition 3.10. *[61] Let M and N be subspaces of a vector space V. V is said to be the direct sum of M and N, written $V = M \oplus N$, if $M \cap N = \{0\}$, and every element $v \in V$ can be expressed as $v = x + y$ with $x \in M$ and $y \in M$.*

Theorem 3.11. *[61] Let $\mathcal{H}$ be a Hilbert space, and let M be a closed subspace of $\mathcal{H}$. Then for each vector $v \in \mathcal{H}$, there exist unique vectors $x \in M$ and $y \in M^{\perp}$ such that $v = x + y$, i.e., $\mathcal{H} = M \oplus M^{\perp}$.*

3.3 Discrete Spatio-Temporal Invariant Systems

The spatially invariant systems are composed of *identical* subsystems connected to their nearest neighbors. The primary motivation for this kind of system is to explore control design and analysis for interconnected systems. These systems are comprised of many similar units that interact directly with their nearest neighbors, and that have sensing and actuating capabilities at every unit. The resulting interconnected systems often display rich and complex behavior, even when the units have tractable models and interact with their neighbors in a simple and predictable manner.

In addition to formation flight problem [65, 66], there are many examples of such engineered systems, including automated highway systems [67] and trajectory optimization for formation control of vehicles [68, 69]; Cross-Directional control in the paper processing applications and chemical process industry [70, 71]; decentralized control of power systems [72, 73, 74] and micro-cantilever array control for massively parallel data storage [75]. One can also consider lumped approximations of partial differential equations (PDEs) examples include the deflection of beams, plates, and membranes, and the temperature distribution of thermally conductive materials [76, 77]. A general structure of spatially invariant systems is captured in Figure 3.1.

Spatial invariance has been recognized as a powerful tool for simplifying the design of a controller for large-scale systems. For example, the early works treated spatially invariant systems as systems over modules [78, 79]. Later, the papers [80, 81, 82, 76] have used Fourier techniques or algebraic transformations to derive implementable

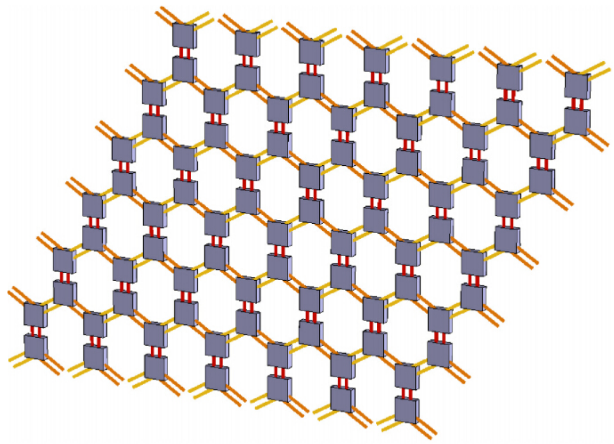


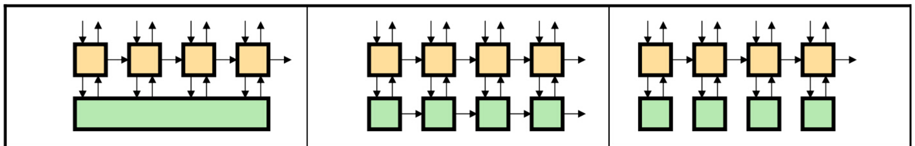
Figure 3.1: Infinite interconnection of identical subsystems in a hexagonal lattice.

and scalable optimal control algorithms, which could even handle interconnected systems consisting of infinite number of subsystems. This is feasible because spatial-invariance yields a tractable model for analysis and obtaining analytical performance bounds as it allows ys to apply multi-dimensional Discrete Fourier Transform to generate algebraic relationships between the system's performance and the network and controller parameters [83]. More recently, spatially invariant systems have been widely studied for their role in modelling, analysis and control of large-scale systems [81, 84, 85, 25]. It also seems to be appropriate for power grid, in particular, spatially-invariant DC power grid model is discussed in [83].

In general, the characteristic and advantage of doing distributed control are summarized in Figure 3.2.

3.3.1 Preliminaries

Following [84, 81], we consider signals that are both functions of discrete time t and discrete space i, denoted $u(t, i)$. Spatially invariant spatio-temporal systems act on signals by convolution. If $y(t, i)$ denotes the output of a spatially invariant systems



Centralized control	Distributed control	Decentralized control
Best performance		Worst performance
Depends on formation	Independent of formation	Independent of formation
Heavy communication	Light communication	No communication
Demanding calculation	Light calculation	Light calculation
Difficult to synthesize	Easy to synthesize	Easy to synthesize

Figure 3.2: Comparisons of Centralized, Distributed, and Decentralized Control Design and Implementation [1].

G then $y = Gu$, where

$$y(t,i) = \sum_{j=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} \hat{g}(t-\tau, i-j)u(\tau, j), \tag{3.10}$$

where $\hat{g}(t,i)$ denotes the spatio-temporal impulse response of G.

We assume temporal causality [84, 81], that is,

$$\hat{g}(t,i) = 0, \quad \text{for} \quad t < 0. \tag{3.11}$$

The λ-transform of $\hat{g}(t,i)$ is defined as

$$g(\lambda, i) = \sum_{t=0}^{\infty} \hat{g}(t,i)\lambda^t. \tag{3.12}$$

The spatial-temporal transfer functions is given by:

$$G_{22}(z,\lambda) := \sum_{i=-\infty}^{\infty} g(\lambda, i)z^i. \tag{3.13}$$

The input-output relationship is given by the expression:

$$Y(z, \lambda) = G(z, \lambda)U(z, \lambda), \tag{3.14}$$

where $U(z, \lambda)$ is the transform of $u(t, i)$, and $Y(z, \lambda)$ is the transform of $y(t, i)$.

The system $G(z, \lambda)$ can be viewed as a multiplication operator on $\mathcal{L}_2(\mathbf{T}, \bar{\mathcal{D}})$ where $\mathbf{T}$ is the unit circle and $\bar{\mathcal{D}}(\mathcal{D})$ is the closed (open) unit disc of the complex domain $\mathbb{C}$. If we assume that $G(z, \lambda)$ is stable, then [84, 81]:

$$\begin{aligned} G(z, \lambda): \ \mathcal{L}_2(\mathbf{T}, \bar{\mathcal{D}}) \quad &\longrightarrow \quad \mathcal{L}_2(\mathbf{T}, \bar{\mathcal{D}}), \\ u \quad &\longrightarrow \quad Gu = G(e^{i\theta}, \lambda)u(e^{i\theta}, \lambda), \end{aligned}$$

where $\theta \in [0, 2\pi), \ |\lambda| \leq 1$.

Then we define the operator induced norm as [84, 81]:

$$\begin{aligned} \|G\| &= \sup_{\|u\|_2 \leq 1} \|Gu\|_2 = \|G\|_\infty \\ &= ess\sup_{\substack{0 \leq \theta < 2\pi \\ \|\lambda\| \leq 1}} |G(e^{i\theta}, \lambda)| < \infty, \end{aligned}$$

where

$$\|u\|_2^2 = \int_{\bar{\mathcal{D}}} \int_{[0, 2\pi)} |G(e^{i\theta}, \lambda)u(e^{i\theta}, \lambda)|^2 d\theta d\lambda.$$

3.4 Summary

This is a transient chapter which focuses on the preliminary knowledge about functional spaces and the mathematical performance specification for control systems. A class of Spatio-Temporal invariant systems is introduced at the end of this chapter. Computing finite horizon centralized controllers for this particular class of systems is equivalent to solve a finite-dimensional convex problem, so there exist many methods for finding an optimal lower triangular matrix Q. However, in the sequel, we turn to infinite-dimensional problems, in which we consider transfer functions, greater care is required to find optimal solutions.

Chapter 4

Optimal Distributed $\mathcal{H}^2$ Control of Spatially Invariant Systems

In this chapter, the optimal $\mathcal{H}^2$ control problem for spatially invariant systems is considered. The framework adopted is borrowed from the work pioneered by Voulgaris et al. In particular, the $\mathcal{H}^2$ optimal control problem is solved via the computation of an orthogonal projection of a tensor Hilbert space onto a particular subspace. Next, the optimal $\mathcal{H}^2$ decentralized control problem is solved by computing an orthogonal projection on a class of Youla parameters with a decentralized structure. The latter uses Riesz projections after invoking a particular $\mathcal{L}^2$-basis. Moreover, a numerical example is included to explain the performance of the developed controller.
Parts of the results in this chapter appeared in [86].

4.1 Introduction

This chapter is concerned with designing distributed coordination and control algorithms for spatially interconnected systems. The goal is to perform distributed computations over a communication network to implicitly solve a global optimization problem. The theory of such systems has been worked out in some detail. We consider only spatially invariant distributed systems. Early work for this class of systems has been done in the control systems community, we refer the reader to [6, 7, 8, 9, 10] and references therein.

A networked system is a collection of dynamic units that interact over an information exchange network. Even if the subsystems interact locally, the optimal controller will need global information to produce the feedback signal. It's challenging to inherit the standard control design techniques for systems of very high dimensions and with a large number of inputs and outputs. A preferred alternative is to have control signals computed using only local communication among neighboring subsystems as motivated in [6] and [7].

After the recent advances in communication technologies, the design of distributed controllers for physically interconnected systems has become an attractive and fruitful research direction [9, 41]. In the literature, much attention has been paid to consensus of first order or second order integrator dynamics [87]. However, for a large scale networked systems, it is more interesting to study synchronization of general linear systems [88, 44, 45, 89], which includes the integrator dynamics (of any order) as a special case. For general linear time-invariant plants with either an H_2 objective, there is no known method of computing the optimal distributed controller.

The origin of the synthesis problem of distributed control lie in the *team decision* problems, which were introduced by Radner [90]. For systems with feedback, the well-known example of Witsenhausen showed that the optimal H_2 controller is in general nonlinear. Moreover, many of such problems are shown to be computational intractable in [91]. The optimal distributed controllers of spatially invariant systems were stated by Bamieh *et al.* [6], who claimed that the dependence of a controller on information coming from other parts of the system decays exponentially as one moves away from that controller at least for state operators which generate semigroups in $\mathcal{L}^2$ [32]. An alternative algorithm for the problem is to reformulate the optimal control problems using a model-matching framework [92]. The approach was based on Youla parameterization and worked under the assumption that a closed-loop transfer function is affine in the Youla parameter [93]. However, as claimed in [85], in the distributed setup, the nonlinearity of the mapping from the controller to the Youla parameter removes the convexity of the constraint set. In other words, a convex constraint set for controller is not always mapped to a convex constraint set for the parameter.

Due to the non-convexity of the design problem with respect to some state-space design parameters, only suboptimal controllers could be achieved after employing certain relaxations and numerical optimization algorithms [85]. To ensure the closed-loop maps are convex, a broader class of problems, called *quadratically-invariant*, was developed in [94]. In [94], the result was derived under the fact that the H_2 norm does not change when a matrix transfer function is vectorized. However, it may only apply to systems with very few states due to the numerical issues discussed in [95].

Certain subspaces of localized systems which remain invariant under this nonlinear mapping have been characterized. The pioneering work by Bamieh and Voulgaris [96] and Voulgaris et al. [84] introduce the important ideas such as subspaces of cone causal and funnel causal systems, respectively. A similar but more general characterization, termed quadratic invariance, was introduced in [97]. It is important to note that constructs such as cone and funnel causality lead to optimal control problems that are convex in the Markov (i.e., impulse response) parameters of the Youla variable, but not in the state-space parameters.

Most of the early work on the optimal distributed controller design problem dealt with the non-convex problem using a variety of relaxations and numerical optimization algorithms to the original problem, thereby rendering suboptimal controllers. Using an approach based on spatial Fourier transforms and operator theory, Bamieh *et al.* discussed the optimal control of linear spatially invariant systems with standard linear quadratic (LQ) criterion in [6]. Although rigorous, as pointed out in [31, 32], this approach is valid when system operator could generate a semigroup on $L_2(\mathbb{R}^n)$ for a given state-space, which is a non-trivial task. Therefore, one is still faced with solving a realization problem for a distributed system.

In the previous chapter, we used the fact that spatially invariant systems can be viewed as multiplication operators from a particular Hilbert function space into itself in the Fourier domain. We have successfully posed the optimal distributed performance as a distance minimization in a general $\mathcal{L}^\infty$ space, from a vector function to a subspace with a mixed $\mathcal{L}^\infty$ and $\mathcal{H}^\infty$ space structure via a spatial-temporal Youla parametrization [84]. In [98], we explained that the dual and pre-dual formulations lead to finite dimensional convex optimizations which approximate the optimal solution *within desired tolerance*. In [99], the duality structure of the problem is characterized in terms of tensor product spaces. We show that these spaces together with the pre-annihilator and annihilator subspaces can be realized explicitly as specific tensor spaces and subspaces, respectively. Specifically, the optimal distributed control performance for spatially invariant systems is equal to the operator induced norm of an operator given by a tensor product.

In this chapter, the $\mathcal{H}^2$ control problem for spatially invariant systems in the framework proposed by Voulgaris et al [84, 81] is considered. First, the optimal $\mathcal{H}^2$ control problem is solved by an orthogonal projection from a tensor Hilbert space of $\mathcal{L}^2$ and $\mathcal{H}^2$ onto a particular subspace. Next, the optimal $\mathcal{H}^2$ decentralized control problem is solved by constructing an orthogonal projection onto the Youla parameters with a particular decentralized structure. This is achieved by using $\{z^n\}_{n=-\infty}^{\infty}$ as a basis for $\mathcal{L}^2$ and invoking Riesz projections.

The rest of the chapter is organized as follows. In Section 4.2, we introduce mathematical preliminaries for discrete spatio-temporal invariant systems. We define the $\mathcal{H}^2$-distributed control problem and shed light on how to solve this problem in Section 4.3. In Section 4.4 the optimal decentralized $\mathcal{H}^2$ control problem is solved. In Section 4.5, we present a numerical experiment through which we show that the proposed approach achieves a better performance. Finally, some concluding remarks are drawn in Section 4.6.

4.2 Discrete Spatio-Temporal Invariant Systems

Following [84, 81], we consider signals that are both functions of discrete time t and discrete space i, denoted $u(t,i)$. Spatially invariant spatio-temporal systems act on signals by convolution. If $y(t,i)$ denotes the output of a spatially invariant systems G then $y = Gu$, where

$$y(t,i) = \sum_{j=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} \hat{g}(t-\tau, i-j)u(\tau, j) \tag{4.1}$$

where $\hat{g}(t,i)$ denotes the spatio-temporal impulse response of G.

We assume temporal causality [84, 81], that is,

$$\hat{g}(t,i) = 0, \quad \text{for} \quad t < 0 \tag{4.2}$$

The λ-transform of $\hat{g}(t,i)$ is defined as

$$g(\lambda, i) = \sum_{t=0}^{\infty} \hat{g}(t,i)\lambda^t \tag{4.3}$$

The spatial-temporal transfer functions is given by:

$$G_{22}(z,\lambda) := \sum_{i=-\infty}^{\infty} g(\lambda, i)z^i \tag{4.4}$$

The input-output relationship is given by the expression:

$$Y(z,\lambda) = G(z,\lambda)U(z,\lambda) \tag{4.5}$$

where $U(z,\lambda)$ is the transform of $u(t,i)$, and $Y(z,\lambda)$ is the transform of $y(t,i)$.

The system $G(z,\lambda)$ can be viewed as a multiplication operator on $\mathcal{L}_2(\mathbf{T}, \bar{\mathcal{D}})$ where $\mathbf{T}$ is the unit circle and $\bar{\mathcal{D}}(\mathcal{D})$ is the closed (open) unit disc of the complex domain

$\mathbb{C}$. If we assume that $G(z, \lambda)$ is stable, then [84, 81]:

$$\begin{aligned} G(z,\lambda):\ \mathcal{L}_2(\mathbf{T},\bar{\mathcal{D}}) &\longrightarrow \quad \mathcal{L}_2(\mathbf{T},\bar{\mathcal{D}}), \\ u &\longrightarrow \quad Gu = G(e^{i\theta},\lambda)u(e^{i\theta},\lambda), \end{aligned}$$

where $\theta \in [0, 2\pi),\ |\lambda| \leq 1$.

Following [81], the ℓ_2-norm of G can be defined as

$$\|G\|_2 = \left(\sum_{i=-\infty}^{\infty}\sum_{t=0}^{\infty}|\hat{g}(t,i)|^2\right)^{\frac{1}{2}}, \tag{4.6}$$

and the $\mathcal{H}^2$-norm of its transform $G(z, \lambda)$ is given by

$$\|G\|_{\mathcal{H}^2} = \left[\left(\frac{1}{2\pi}\right)^2\int_{\theta\in[0,2\pi)}\int_{w\in[0,2\pi)}\left|G(e^{i\theta},e^{iw})\right|^2 dw d\theta\right]^{\frac{1}{2}} \tag{4.7}$$

And

$$\|G\|_2 = \|G\|_{\mathcal{H}^2} \tag{4.8}$$

The system G is said to be stable if its uniform norm

$$\|G\|_\infty = ess\sup_{\substack{0\leq\theta<2\pi \\ \|\lambda\|\leq 1}} |G(e^{i\theta},\lambda)| < \infty \tag{4.9}$$

In the next section, the $\mathcal{H}^2$ optimal control problem for spatially invariant systems is formulated and solved. Note that the optimal solution does not necessarily result in distributed or decentralized controllers.

4.3 $\mathcal{H}^2$ Optimal Control Problem Formulation

Consider the standard feedback configuration of Figure 4.1, where w is the external disturbance, z is the controlled output, y is the measurement signal, and u is the control for all spatio-temporal sequences. The plant G and controller K are spatially and temporally invariant systems.

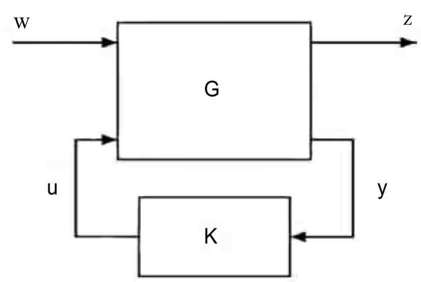


Figure 4.1: Standard Feedback Configuration

We assume that G is stable, and the transmission from w to z is denoted by T_{zw}. We have:

$$\begin{pmatrix} z \\ y \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} w \\ u \end{pmatrix} \tag{4.10}$$

All stabilizing spatio-temporal invariant controllers [84]

$$K = -Q(I - G_{22}Q)^{-1} \tag{4.11}$$

with Q stable, denote

$$Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}})) \tag{4.12}$$

The subspace $\mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ is defined as follows:
For $f \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$, for each $\theta \in [0, 2\pi)$, $f(e^{i\theta}, \cdot) \in \mathcal{L}^\infty(\mathbf{T})$, and for each $\lambda \in \bar{\mathcal{D}}$, $f(\cdot, \lambda) \in \mathcal{H}^\infty$.

In other words, all functions $f \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ can be viewed as $\mathcal{H}^\infty$ functions which take values in $\mathcal{L}^\infty(\bar{\mathcal{D}})$, i.e., $\mathcal{L}^\infty(\bar{\mathcal{D}})$ valued $\mathcal{H}^\infty$ functions. This interpretation is carried out to other subspaces in a similar vein.

With the parametrization (4.11), the H_2-control problem is formulated as:

$$\begin{aligned} \mu_2 &:= \inf_{K_{stabilizing}} \|T_{zw}\|_{\mathcal{H}^2} \\ &= \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_1 - T_2 Q\|_{\mathcal{H}^2}, \end{aligned} \tag{4.13}$$

where $T_2, T_1 \in \mathcal{H}^2(\mathcal{L}^2)$, that is, $\mathcal{L}^2$-valued functions belonging to $\mathcal{H}^2$.
For this problem to make sense, the system transform $G(z, \lambda)$ must be strictly proper in the frequency λ. This makes the subspace $\{T_2 Q : Q \in H^\infty(L^\infty(\bar{\mathcal{D}}))\}$ not closed. As a result, there may not exist a minimizer Q.

The space $\mathcal{H}^2(\mathcal{L}^2)$ can be viewed as the closure of the tensor Hilbert space $\mathcal{H}^2 \otimes \mathcal{L}^2$ in the norm (6.5). The inner product on $\mathcal{H}^2 \otimes \mathcal{L}^2$ is defined as for $f_1 \otimes g_1, f_2 \otimes g_2$:

$$\begin{gathered} f_i \in \mathcal{H}^2,\ g_i \in \mathcal{L}^2,\ i = 1, 2 : \\ \langle f_1 \otimes g_1,\ f_2 \otimes g_2 \rangle = \langle f_1,\ f_2 \rangle_{\mathcal{H}^2} \langle g_1,\ g_2 \rangle_{\mathcal{L}^2}, \end{gathered} \tag{4.14}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}^2}$ and $\langle \cdot, \cdot \rangle_{\mathcal{L}^2}$ are the standard inner products in $\mathcal{H}^2$ and $\mathcal{L}^2$, respectively.

The norm induced by the inner product (4.14) is

$$\alpha(f) = \sqrt{\langle f,\ f \rangle}, \quad f \in \mathcal{H}^2 \otimes \mathcal{L}^2. \tag{4.15}$$

The completion of the tensor space $\mathcal{H}^2 \otimes \mathcal{L}^2$ with respect to (4.15) will be denoted henceforth, by $\mathcal{H}^2 \otimes_\alpha \mathcal{L}^2$.
Then we have

$$\mathcal{H}^2(\mathcal{L}^2) = \mathcal{H}^2 \otimes_\alpha \mathcal{L}^2. \tag{4.16}$$

Therefore, these spaces will be used interchangeably.

A standard result from H^p-theory [100] asserts that if $f \in \mathcal{H}^2$, then $f(e^{jw}) \in \mathcal{L}^2$, that is, $\mathcal{H}^2$ may be viewed as a closed subspace of $\mathcal{L}^2$. For example,

$$L_2(-\infty, \infty) = L_2\,(-\infty, 0] \oplus L_2\,[0, \infty) \tag{4.17}$$

Each function f in $L_2(-\infty, \infty)$ has a unique decomposition $f = f_1 + f_2$ with $f_1 \in L_2(-\infty, 0]$ and $f_2 \in L_2[0, \infty)$:

$$\begin{aligned} f_1(t) &= f(t),\; f_2(t) = 0, \quad t \leq 0, \\ f_1(t) &= 0,\; f_2(t) = f(t), \quad t > 0. \end{aligned}$$

The function $f \to f_1$ from $L_2(-\infty, \infty)$ to $L_2\,(-\infty, 0]$ is an operator, the orthogonal projection of $L_2(-\infty, \infty)$ onto $L_2\,(-\infty, 0]$. It's easy to prove that its norm equals 1 [92].
Letting $H^{2\perp}$ be the orthogonal complement in $\mathcal{L}^2$, then

$$\mathcal{L}^2 = \mathcal{H}^2 \oplus H^{2\perp}, \tag{4.18}$$

that is, every $f \in \mathcal{L}^2$ can be written uniquely as:

$$f = f_1 + f_2, \tag{4.19}$$

with $f_1 \in \mathcal{H}^2$ and $f_2 \in H^{2\perp}$.

Definition 4.1. *The set of functions* $\{e^{jnw}\}_{n=-\infty}^{\infty}$ *is an orthogonal basis for* $\mathcal{L}^2$*, i.e.,* $\mathcal{L}^2 = \overline{span}\{e^{jnw}\}_{n=-\infty}^{\infty}, \forall f \in \mathcal{L}^2$

$$f(e^{jw}) = \sum_{n=-\infty}^{\infty} a_n e^{jnw}. \tag{4.20}$$

Similarly,

$$\begin{aligned} \mathcal{H}^2 &= \overline{span}\{e^{jnw}\}_{n=0}^{\infty}, \\ H^{2\perp} &= \overline{span}\{e^{jnw}\}_{n=-\infty}^{-1}, \end{aligned}$$

where if A is a set, $\bar{A}$ denotes its closure.

Every function $g \in \mathcal{H}^2$, and $h \in H^{2\perp}$ satisfy

$$g(e^{jw}) = \sum_{n=0}^{\infty} g_n e^{jnw} \tag{4.21}$$

$$h(e^{jw}) = \sum_{n=-\infty}^{-1} h_n e^{jnw} \tag{4.22}$$

Define the positive and negative Riesz projections from $\mathcal{L}^2$ into $\mathcal{H}^2$ and $H^{2\perp}$, respectively, as:

$$\begin{aligned} P_+ : \mathcal{L}^2 &\rightarrow \mathcal{H}^2 \\ f &\rightarrow P_+ f(e^{jw}) = \sum_{n=0}^{\infty} a_n e^{jnw} \end{aligned} \tag{4.23}$$

$$\begin{aligned} P_- : \mathcal{L}^2 &\rightarrow H^{2\perp} \\ f &\rightarrow P_- f(e^{jw}) = \sum_{n=-\infty}^{-1} a_n e^{jnw} \end{aligned} \tag{4.24}$$

where P_+ and P_- are orthogonal projections.

By a result in [101], if I denotes the identity operator in the space $\mathcal{L}^2$ (of functions of $e^{j\theta}$), then $P_+ \otimes_\alpha I$ is the orthogonal projection of $\mathcal{L}^2 \otimes_\alpha \mathcal{L}^2$ onto $\mathcal{H}^2 \otimes_\alpha \mathcal{L}^2$. Similarly, $P_- \otimes_\alpha I$ is the orthogonal projection onto $H^{2\perp} \otimes_\alpha \mathcal{L}^2$.

Assuming $T_1(e^{j\theta}, \lambda)$ is strictly proper in λ and for fixed θ, $T_2(e^{i\theta}, \cdot)$ as a function of λ is non-zero on the boundary of $\bar{\mathcal{D}}$, i.e., $\partial\bar{\mathcal{D}}$.

Then by using the *spatio-temporal inner-outer factorization*, we have:

$$T_2(e^{j\theta}, \lambda) = T_{2in}(e^{j\theta}, \lambda)T_{2out}(e^{j\theta}, \lambda) \tag{4.25}$$

where $T_{2in}(e^{j\theta}, \lambda)$ is an isometry and $T_{2out}(e^{j\theta}, \lambda)$ is outer invertible.

The optimal $\mathcal{H}^2-$control problem is then solved in the following theorem.

Theorem 4.2. *Under the assumptions above, the infimum in (4.13) is achieved, i.e.,*

$$\begin{aligned} \mu_2 &= \min_{Q\in\mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_{2in}^* T_1 - T_{2out}Q\|_{\mathcal{L}^2} && (4.26)\\ &= \|(P_- \otimes_\alpha I)(T_{2in}^* T_1)(e^{j\theta}, \lambda)\|_{\mathcal{L}^2} && (4.27)\end{aligned}$$

and the minumum is achieved by choosing any $Q \in H^\infty(L^\infty(\bar{\mathcal{D}}))$ such that

$$T_{2out}Q = (P_+ \otimes_\alpha I)(T_{2in}^* T_1) \tag{4.28}$$

or equivalently, $Q = \frac{(P_+\otimes_\alpha I)(T_{2in}^* T_1)}{T_{2out}}$.

Proof. Write

$$T_{2in}^* T_1 = (P_+ \otimes_\alpha I)(T_{2in}^* T_1) + (P_- \otimes_\alpha I)(T_{2in}^* T_1). \tag{4.29}$$

Since

$$(P_+ \otimes_\alpha I)(T_{2in}^* T_1) \in \mathcal{H}^2 \otimes_\alpha \mathcal{L}^2, \tag{4.30}$$

and

$$(P_- \otimes_\alpha I)(T_{2in}^* T_1) \in H^{2\perp} \otimes_\alpha \mathcal{L}^2. \tag{4.31}$$

By Pythagorean theorem, for any Q we have

$$\begin{aligned}
\|T_{2in}^* T_1 - T_{2out} Q\|_{\mathcal{L}^2}^2 &= \|(P_+ \otimes_\alpha I)(T_{2in}^* T_1) - T_{2out} Q\|_{\mathcal{L}^2}^2 \\
&+ \|(P_- \otimes_\alpha I)(T_{2in}^* T_1)\|_{\mathcal{L}^2}^2 \qquad (4.32) \\
&\geq \|(P_- \otimes_\alpha I)(T_{2in}^* T_1)\|_{\mathcal{L}^2}^2. \qquad (4.33)
\end{aligned}$$

Notice the equality (and therefore the minimum) is achieved by choosing Q satisfying:

$$\begin{aligned}
T_{2out} Q &= (P_+ \otimes_\alpha I)(T_{2in}^* T_1) \\
\text{or equivalently,} \\
Q &= \frac{(P_+ \otimes_\alpha I)(T_{2in}^* T_1)}{T_{2out}} \qquad (4.34)
\end{aligned}$$

and the theorem is proved. □□□

Remark 4.3. *The solution provided in Theorem 4.2 may not in general be distributed or decentralized. To derive optimal decentralized controllers, we adopt the framework proposed in [84], where subspaces of cone causal and funnel causal systems which remain invariant under the Youla parametrization were studied. In particular, the authors considered a relaxation problem to the original decentralized problem where the first N coefficients*

$$Q(\lambda, z) = Q_0(z) + Q_1(z)\lambda + \cdots \tag{4.35}$$

are constrained to correspond to band-operator Q_i with $(z_{i+1})-$diagonal for $i = 0, 1, \cdots, N-1$, where N is arbitrary [84]. This problem is interesting in its own right.

The authors in [84] show how it can be solved, and that it provides a lower bound to the original optimal decentralized control problem. The lower bound converges to the optimal solution as $N \to \infty$ [84]. In [85], the authors introduce a state-space framework to the relaxed problem, and provide a variety of numerical optimization algorithm to compute suboptimal controllers.

In the next section, we consider the original optimal decentralized control problem. In particular, we give an exact solution which relies on the fact that $\{z^n\}_{n=-\infty}^{\infty}$ is a basis for $\mathcal{L}^2$ and the principle of orthogonality using Riesz projections as in Theorem 4.2.

4.4 Design of $\mathcal{H}^2$ Optimal Decentralized Controllers

The framework considered in this section is borrowed from [84], where the particular structure of interest is when the spatio-temporal transform G_{22} yields the following form [84]:

$$G_{22}(z,\lambda) = \sum_{i=-\infty}^{\infty} g_i(\lambda) z^i \tag{4.36}$$

$$\text{with} \quad g_i(\lambda) = \lambda^{|i|}\tilde{g}_i(\lambda) \tag{4.37}$$

where as previously λ corresponds to the *temporal* transform variable, and z corresponds to the *spatial* two-sided transform.
The transfer functions $\tilde{g}_i(\lambda)$, $i = 1, 2, \cdots$ correspond to temporal causal systems [84]. The interpretation of this framework is that the input u_i to the ith system g_i affects the output y_j of the jth system g_j which is $|j-i|$ spatial location aways with a delay of $|j-i|$ time steps [84].

The controllers K we are looking for are assumed to have the same structure as G_{22}, that is,

$$
\begin{aligned}
k(z,\lambda) &= \sum_{i=-\infty}^{\infty} k_i(\lambda) z^i, && (4.38)\\
\text{with} \quad k_i(\lambda) &= \lambda^{|i|}\tilde{k}_i(\lambda), && (4.39)
\end{aligned}
$$

which means that the measurements of the jth location are made available at the ith system after $|j-i|$ time steps [84].

The following proposition asserts that the Youla parametrization transforms the decentralized constraints on K to convex constraints on the Youla parameter Q [84].

Proposition 4.4. *[84] All stabilizing controllers K with the structure (4.38)-(4.39) are given by*

$$
K = -Q(I - G_{22}Q)^{-1}, \tag{4.40}
$$

with Q stable given by

$$
Q(z,\lambda) = \sum_{i=-\infty}^{\infty} q_i(\lambda) z^i, \tag{4.41}
$$

where

$$
q_i(\lambda) = \lambda^{|i|}\tilde{q}_i(\lambda), \tag{4.42}
$$

where $\tilde{q}_i$ is stable.

The $\mathcal{H}^2$-decentralized optimal control problem can then be formulated as:

$$
\mu_{2d} \quad := \quad \inf_{K_{stabilizing} \text{ s.t.}(4.38)-(4.39) \text{ hold}} \|T_{zw}\|_{\mathcal{H}^2} \tag{4.43}
$$

$$= \inf_{Q \text{ stable s.t.}(6.16)-(6.17) \text{ hold}} \|T_1 - T_2 Q\|_{\mathcal{H}^2}, \tag{4.44}$$

where T_{zw} is defined in (5.1).

Again, using inner-outer factorization of T_2, we can write $T_2 = T_{2in}T_{2out}$ with $T_{2in}^* T_{2in} = I$. Then, (4.44) reduces to:

$$\mu_{2d} = \inf_{Q \text{ stable s.t.}(6.16)-(6.17) \text{ hold}} \|T_{2in}^* T_1 - T_{2out} Q\|_{\mathcal{L}^2} \tag{4.45}$$

Notice that the outer function T_{2out} admits the following expansions w.r.t. the basis $\{z^i\}_{-\infty}^{\infty}$, such that:

$$T_{2out}(z, \lambda) = \sum_{i=-\infty}^{\infty} \upsilon_i(\lambda) z^i,$$

with $\upsilon_i(\lambda)$ stable.

Therefore

$$T_{2out}(z, \lambda) Q(z, \lambda) = \sum_{i=-\infty}^{\infty} \tilde{\tilde{q}}_i(\lambda) z^i, \tag{4.46}$$

with $\tilde{\tilde{q}}_i(z)$ stable, and

$$\tilde{\tilde{q}}_i(\lambda) = \sum_j \lambda^{|j|} \tilde{q}_j(\lambda) \upsilon_{i-j}(\lambda).$$

Since $\{z^i\}_{i=-\infty}^{\infty}$ is an orthogonal basis of $\mathcal{L}^2$, $T_{2in}^* T_1$ can be written as:

$$T_{2in}^* T_1(z, \lambda) = \sum_{i=-\infty}^{\infty} \tilde{T}_i(\lambda) z^i \tag{4.47}$$

where

$$\tilde{T}_i(\lambda) \in \mathcal{L}^2$$

Substituting (4.46) and (4.47) into the decentralized optimization (4.45) yields:

$$\mu_{2d}^2 = \inf_{\tilde{\tilde{q}}_i(\lambda)\,\text{stable}\; -\infty<i<\infty} \| \sum_{i=-\infty}^{\infty} \tilde{T}_i(\lambda) z^i - \sum_{i=-\infty}^{\infty} \tilde{\tilde{q}}_i(\lambda) z^i \|_{\mathcal{L}^2}^2 \tag{4.48}$$

Using Parseval's identity, we get:

$$\mu_{2d}^2 = \inf_{\tilde{\tilde{q}}_i(\lambda)\,\text{stable}\, -\infty<i<\infty} \sum_{i=-\infty}^{\infty} \|\tilde{T}_i(\lambda) - \tilde{\tilde{q}}_i(\lambda)\|_{\mathcal{L}^2}^2. \tag{4.49}$$

Remark 4.5. *The optimal decentralized control problem for spatially invariant systems using this framework reduces to applying the orthogonality principle.*

The following theorem gives the **optimal solution**.

Theorem 4.6. *Under the assumptions above, the infimum in (4.48) is achieved, i.e.,*

$$\mu_{2d}^2 = \inf_{\tilde{\tilde{q}}_i(\lambda)\ \text{stable}} \sum_{i=-\infty}^{\infty} \|\tilde{T}_i(\lambda) - \tilde{\tilde{q}}_i(\lambda)\|_{\mathcal{L}^2}^2 \tag{4.50}$$

$$= \sum_{i=-\infty}^{\infty} \|P_{-}\tilde{T}_i(\lambda)\|_{\mathcal{L}^2}^2, \tag{4.51}$$

and the minimum is achieved by $\tilde{\tilde{q}}_i(\lambda)$ s.t.

$$\tilde{\tilde{q}}_i(\lambda) = P_{+}(\tilde{T}_i(\lambda)), \tag{4.52}$$

$$\textit{with} \quad i = \cdots, -2, -1, 0, 1, 2, \cdots.$$

Proof: First note that:

$$\mu_{2d}^2 \geq \sum_{i=-\infty}^{\infty} \inf_{\tilde{\tilde{q}}_i(\lambda)\, stable} \|\tilde{T}_i(\lambda) - \tilde{\tilde{q}}_i(\lambda)\|_{\mathcal{L}^2}^2 \tag{4.53}$$

By orthogonality for each i, for all $\tilde{\tilde{q}}_i(\lambda)$ we have

$$\begin{aligned}\|\tilde{T}_i(\lambda) - \tilde{\tilde{q}}_i(\lambda)\|^2_{\mathcal{L}^2} &= \|P_+(\tilde{T}_i(\lambda)) - \tilde{\tilde{q}}_i(\lambda)\|^2_{\mathcal{L}^2} + \|P_-(\tilde{T}_i(\lambda))\|^2_{\mathcal{L}^2} && (4.54)\\ &\geq \|P_-(\tilde{T}_i(\lambda))\|^2_{\mathcal{L}^2}. && (4.55)\end{aligned}$$

The equality and therefore the minimum is achieved by choosing $\tilde{\tilde{q}}_i(\lambda)$ satisfying:

$$\tilde{\tilde{q}}_i(\lambda) = P_+(\tilde{T}_i(\lambda)), \quad -\infty < i < \infty \tag{4.56}$$

and the theorem is proved. ■

The optimal decentralized Youla parameter is then

$$T_{2out}Q(z,\lambda) = \sum_{i=-\infty}^{\infty} \tilde{\tilde{q}}_i(z) z^i, \tag{4.57}$$

which implies that:

$$Q(z,\lambda) = \frac{\sum_{i=-\infty}^{\infty} P_+(\tilde{T}_i(\lambda))z^i}{T_{2out}(z,\lambda)}. \tag{4.58}$$

From (6.15), the corresponding optimal decentralized controller K is then given by:

$$\begin{aligned}K &= -Q(I - G_{22}Q)^{-1}\\ &= -\frac{\sum_{i=-\infty}^{\infty} P_+(\tilde{T}_i(\lambda))z^i}{T_{2out}(z,\lambda)} \times \left[(I - G_{22}(z,\lambda)\frac{\sum_{i=-\infty}^{\infty} P_+(\tilde{T}_i(\lambda))z^i}{T_{2out}(z,\lambda)}\right]^{-1}. && (4.59)\end{aligned}$$

4.5 Numerical Examples

In this section, the proposed approach is tested on the same numerical example as given in Voulgaris et al. [84]. As discussed in Section 4.4, the proposed orthogonal projection technique leads to a solution to the optimal H_2 decentralized control problem.

We consider the following spatio-temporal system which comes from the finite-difference discretization of a certain PDE.

$$\begin{aligned} y(i,k+1) - y(i,k) &= \frac{T}{\mathcal{L}^2} y(i+1,k) - 2y(i,k) + y(i-1,k) \\ &- \epsilon y(i,k) + u(i,k) \end{aligned} \tag{4.60}$$

Taking the appropriate transforms one obtains the transfer function, weighting function and stabilizing controller parameterization.

We want to compute a decentralized controller for optimal H_2 attenuation of an additive disturbance on the system output with weighting function

$$W(z,\lambda) = \frac{\lambda}{1 - (\frac{1}{8}z^{-1} + \frac{1}{4} + \frac{z}{8})\lambda\,.} \tag{4.61}$$

We assume $W(z;\lambda)$ to be asymptotically stable and yield the same structure as the plant itself.

As defined in [84], with $K(z,\lambda)$ and $Q(z,\lambda)$ of the prescribed form, the problem can be stated as

$$\inf_Q \|(1 - GQ)W\|_\infty = \inf_Q \|H - UQ\|_\infty, \tag{4.62}$$

where

$$\begin{aligned} H(z,\lambda) &= \frac{\lambda}{1 - r(z)\lambda}, \\ U(z,\lambda) &= \frac{T\lambda^2}{(1-\rho(z)\lambda)(1-r(z)\lambda)}, \end{aligned} \tag{4.63}$$

and

$$\begin{aligned}\rho(z) &= z^{-1}/6 + 1/3 + z/6, \\ r(z) &= z^{-1}/8 + 1/4 + z/8. \end{aligned} \tag{4.64}$$

An inner−outer factorization of $U(z, \lambda)$ yields

$$U_{in}(z, \lambda) = \lambda^2, \tag{4.65}$$

$$U_{out}(z, \lambda) = \frac{T}{(1 - \rho(z)\lambda)(1 - r(z)\lambda)\,.} \tag{4.66}$$

By plugging in the corresponding $T_{2in}^* T_1$ here in this example, we compute the negative Riesz projection and obtain the optimal solution as developed in *Theorem 4.6*.

Ideally, we want to have the whole set of functions $\{e^{jnw}\}_{n=-\infty}^{\infty}$ to be the orthogonal basis for $\mathcal{L}^2$. However, we have to pick a finite number N as the discretization resolution.

As shown in Figure. 4.2, we choose N varying from 2000 to 3000 with 100 increment for each run. As expected, better performance is achieved by assigning a more fine discretization. Moreover, we compare the results of the proposed work with those benchmarks reported in Paper I [84] and Paper II [85]. Clearly, we see our proposed distributed technique (red line) achieves better performance than both existing techniques (represented by black and blue lines). Finally, we also provide the best centralized results using proposed algorithm denoted by green line in the figure.

4.6 Summary

In this chapter, the optimal $\mathcal{H}^2$ control problem for spatially invariant systems was considered in the framework proposed in [84, 81]. In particular, the $\mathcal{H}^2$ optimal control problem was solved via the computation of an orthogonal projection of a

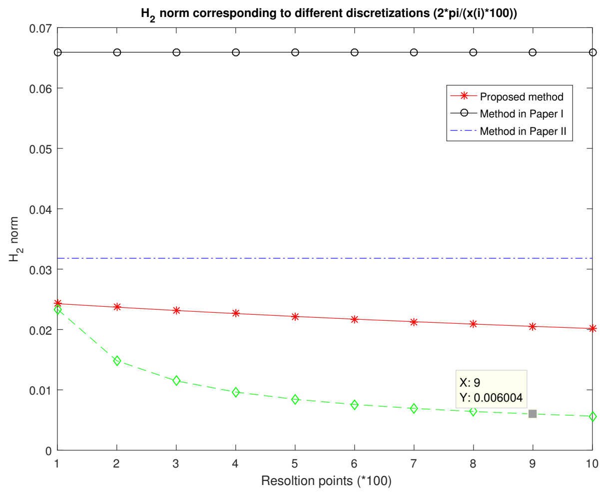


Figure 4.2: Comparison for optimal H^2 performance

tensor Hilbert space onto a particular subspace. The optimal $\mathcal{H}^2$ decentralized control problem was solved by computing an orthogonal projection on a class of Youla parameters with a decentralized structure. The latter uses Riesz projections after invoking a particular $\mathcal{L}^2$-basis.

Chapter 5

Optimal H^∞ Control Problem for Spatially Invariant Systems

We consider the problem of optimal distributed control of spatially invariant systems. We develop an input-output framework for problems of this class. Spatially invariant systems are viewed as multiplication operators from a particular Hilbert function space into itself in the Fourier domain. Optimal distributed performance is then posed as a distance minimization in a general L-infinity space from a vector function to a subspace with a mixed $\mathcal{L}^\infty$ and $\mathcal{H}^\infty$ space structure. In this framework, a generalized version of the Youla parametrization plays a central role. The duality structure of the problem is characterized by computing various dual and pre-dual spaces. The annihilator and pre-annihilator subspaces are also calculated for the dual and pre-dual problems. Furthermore, the latter is used to show the existence of optimal distributed controllers and dual extremal functions under certain conditions. Our approach is purely input-output and does not use any state space realization. We show that these spaces together with the pre-annihilator and annihilator subspaces can be realized explicitly as specific tensor spaces and subspaces, respectively. The tensor space formulation leads to a solution in terms of an operator given by a tensor product. Specifically, the optimal distributed control performance for spatially

invariant systems is equal to the operator induced norm of this operator. The results obtained in this chapter bridge the gap between control theory and the metric theory of tensor product spaces. The dual and pre-dual formulations lead to finite dimensional convex optimizations which approximate the optimal solution within desired accuracy. These optimizations can be solved using convex programming methods. This chapter is concluded with a numerical example.
Parts of the results in this chapter appeared in [98, 99].

5.1 Introduction

There has been resonant interest in analysis and synthesis of distributed coordination and control algorithms for spatially interconnected systems. For recent work on this class and some of the background for the present work, we refer the reader to [6, 7, 8, 9, 10], and the references therein. The basic idea for this spatially distributed problem is to perform distributed computations over a network to implicitly solve a global optimization problem.

A networked system is a collection of dynamic units that interact over an information exchange network. Such systems are ubiquitous in diverse areas of science and engineering [13]. There are many important problems that have been cast in the form of a large-scale finite-dimensional or an infinite-dimensional constraint optimization problem [14]. Such problems can range from physical, biological to mechanical and social systems [15, 16, 17, 18]. Distributed control has become a successful strategy to handle such design issues as coordinated control, formation control and synchronization of multi-agent systems [20, 21, 22, 23, 24, 25].

Even if the subsystems interact locally, the optimal controller will need global information to produce the feedback signal. Standard control design techniques are inadequate since most optimal control techniques cannot handle systems of very high dimensions and with a large number of inputs and outputs. A preferred alternative is to have control signals computed using only local communication among neighboring subsystems as motivated in [6] and [7]. Using an approach based on spatial Fourier transforms and operator theory, Bamieh *et al.* discussed the optimal control of linear spatially invariant systems with standard linear quadratic (LQ) criterion in [6]. As pointed out in [32], the above results are valid when system operator could generate a semigroup on $L_2(\mathbb{R}^n)$.

After the recent advances in communication technologies, the design of distributed controllers for physically interconnected systems has become an attractive and fruitful research direction [9, 41]. A body of literature has been worked out for the spatially distributed systems, where all signals are functions of both spatial and temporal variables. The linear matrix inequality (LMI) conditions for spatially interconnected systems consisting of homogeneous units are introduced in [9, 42]. Control synthesis results have employed consensus-based observer to guarantee leaderless synchronization of multiple identical linear dynamic systems under switching communication topologies[43]; neighbor-based observer to solve the synchronization problem for general linear time-invariant systems [44]; and individual-based observer with low-gain technique to synchronize a group of linear systems [45]. Synchronization of multiple heterogeneous linear systems has been studied in [8, 10], where the interconnection topology is represented by an arbitrary graph, using a conservative analysis-LMI. A similar problem is investigated under both fixed and switching communication topologies heterogeneous spatially distributed systems [29, 30].

In this chapter, we focus our investigation on spatially invariant systems. We show that the spatially invariant systems can be viewed as multiplication operators from a particular Hilbert function space into itself in the Fourier domain. A key distinctive feature of this chapter with respect to the existing literature, is that *we propose a new technique to pose the optimal distributed performance as a distance minimization in a general $\mathcal{L}^\infty$ space*, from a vector function to a subspace with a mixed $\mathcal{L}^\infty$ and $\mathcal{H}^\infty$ space structure via a spatial-temporal Youla parametrization [84].

We first derive the pre-dual and dual optimizations along the lines of [102] via tools from functional analysis. The duality structure of the problem is characterized by computing various dual and pre-dual spaces. The annihilator and pre-annihilator subspaces are also calculated for the dual and pre-dual problems. The latter is used to show the existence of optimal distributed controllers and dual extremal functions under certain conditions. Motivated by these findings, we continue the study of the Banach space duality structure. In particular, the duality structure of the problem is characterized in terms of tensor product spaces. This complements the above study where the dual and pre-dual formulations were in terms of abstract spaces. Here, we show that these spaces together with the pre-annihilator and annihilator subspaces can be realized explicitly as specific tensor spaces and subspaces, respectively. The tensor space formulation leads to a solution in terms of an operator given by a tensor product. Specifically, the optimal distributed control performance for spatially invariant systems is equal to the operator induced norm of this operator. The results obtained bridge the gap between control and the metric theory of tensor product spaces. In the sequel, we discuss how the dual and pre-dual formulations lead to finite dimensional convex optimizations which approximate the optimal solution within desired tolerance. These optimizations can be solved by convex programming methods. In addition, a numerical solution is given, where the optimal solution is approximated from both above and below via finite variable convex programming within *arbitrarily close*.

This chapter is organized as follows.

In Section 5.2, we explained that the dual and pre-dual formulations lead to finite dimensional convex optimizations which approximate the optimal solution within desired tolerance. It is followed by the operator theoretic approach discussed in Section 5.3. A discussion of a numerical solution is given in Section 5.4 to demonstrate that the reduced finite dimensional optimization problems can estimate the optimal solutions within desired accuracy. And a numerical example is given in Section 5.5 to show the utility of the method proposed in this chapter. Finally, we close this chapter with some concluding remarks in Section 5.6.

5.1.1 Problem Formulation

Based on the preliminary knowledge introduced in Sec. 3.3.1, we are concerned with a standard feedback configuration of Figure 5.1, where The plant G and controller K are spatially and temporally invariant systems, and the signals w, z, y and u are defined as follows: w = exogenous disturbances; z = output signals to be regulated; y = measured plant outputs; and u = control inputs to the plants.

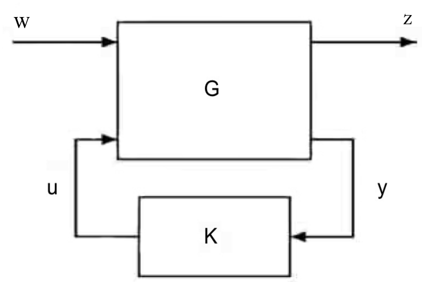


Figure 5.1: Standard Feedback Configuration

We assume that G is *stable*, and let T_{zw} denote the result closed-loop dynamics from w to z.
We have

$$\begin{pmatrix} z \\ y \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} w \\ u \end{pmatrix}. \tag{5.1}$$

All stabilizing spatio-temporal invariant controllers [84]

$$K = -Q(I - G_{22}Q)^{-1}, \tag{5.2}$$

with Q stable, denote

$$Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}})). \tag{5.3}$$

Definition 5.1. *The subspace $\mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ is defined as follows.*
We denote $f \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ if for each $\theta \in [0, 2\pi)$, $f(e^{i\theta}, \cdot) \in \mathcal{L}^\infty(\mathbf{T})$, and for each $\lambda \in \bar{\mathcal{D}}, f(\cdot, \lambda) \in \mathcal{H}^\infty$.

In other words, all functions $f \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ can be viewed as $\mathcal{H}^\infty$ functions which take values in $\mathcal{L}^\infty(\bar{\mathcal{D}})$, i.e., $\mathcal{L}^\infty(\bar{\mathcal{D}})$ valued $\mathcal{H}^\infty$ functions. This interpretation is carried out to other subspaces in a similar vein.

With the parametrization (5.2), optimal disturbance rejection can be formulated as:

$$\begin{aligned} p: \quad &= \inf_{K stabilizing} \sup_{\|w\|_2 \le 1} \|T_{zw} w\|_2 \\ &= \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_1 - T_2 Q\|_\infty \end{aligned} \tag{5.4}$$

Using a "spatio-temporal inner-outer factorization" [84], we have:

$$T_2(e^{j\theta}, \lambda) = T_{2in}(e^{j\theta}, \lambda) T_{2out}(e^{j\theta}, \lambda), \tag{5.5}$$

where $T_{2in}(e^{j\theta}, \lambda)$ is an isometry and $T_{2out}(e^{j\theta}, \lambda)$ is (temporally) causally invertible.

Therefore, the optimal performance index (5.4) can be written as:

$$\psi := \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_{2in}^* T_1 - T_{2out} Q\|_\infty, \tag{5.6}$$

where

$$T_{2in}^* := T_{2in}(z^{-1}, \lambda^{-1}), \tag{5.7}$$

after absorbing T_{2out} in Q and denote the product by Q by abuse of notation.
Then

$$\psi = \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_{2in}^* T_1 - Q\|_\infty. \tag{5.8}$$

Note that $T_{2in}^{*}T_1 \in \mathcal{L}^{\infty}(\mathbf{T} \times \bar{\mathcal{D}})$ such that (5.8) can be viewed as a distance minimization from the function $T_{2in}^{*}T_1$ to the subspace S of $\mathcal{L}^{\infty}(\mathbf{T} \times \bar{\mathcal{D}})$ defined by:

$$S := \mathcal{H}^{\infty}(\mathcal{L}^{\infty}(\bar{\mathcal{D}})). \tag{5.9}$$

In the next section, we will show the existence of optimal controllers using duality theory.

5.2 Pre-dual and dual Characterizations in abstract spaces

5.2.1 Pre-dual Characterization

Let $\mathbb{B}$ be a Banach space with norm $\|\cdot\|$. Its dual space denoted $\mathbb{B}^*$ is the space of bounded linear functionals defined on $\mathbb{B}$.

Definition 5.2. *Isometric isomorphism between Banach space is denoted by "$\simeq$" as in Definition 3.9. In particular, $\mathbb{B}^*$ is said to be the pre-dual space of $\mathbb{B}$ if $\mathbb{B}^* \simeq \mathbb{B}$.*

Definition 5.3. *For a subset M of $\mathbb{B}$, the* annihilator *of M in $\mathbb{B}^*$ is denoted $M^{\perp}$ and is defined by [103]:*

$$M^{\perp} := \{\Phi \in \mathbb{B}^* : \Phi(m) = 0, \quad \forall m \in M\} \tag{5.10}$$

In other words, $M^{\perp}$ is the set of bounded linear functionals on $\mathbb{B}$ which vanish on M. It is a sort of generation of orthogonal subspace in the Banach space setting.

Definition 5.4. *Similarly, if $\tilde{M}$ is a subset of $\mathbb{B}^*$, then the* pre-annihilator *of $\tilde{M}$ in $\mathbb{B}$ is denoted ${}^{\perp}\tilde{M}$, which is defined by [103]:*

$${}^{\perp}\tilde{M} := \{b \in \mathbb{B} : \psi(b) = 0, \quad \psi \in \tilde{M}\} \tag{5.11}$$

Remark 5.5. *Obviously, the pre-annihilator satisfies* $({}^{\perp}\tilde{M})^{\perp} \simeq M$.

Theorem 5.6. *[103] Following standard result of Banach space duality theory [103], the existence of a pre-annihilator implies that the following identity holds:*

$$\min_{\tilde{m}\in\tilde{M}} \|b - \tilde{m}\| = \sup_{\substack{\tilde{b}\in{}^{\perp}\tilde{M} \\ \|b\|\le 1}} | < b, \tilde{b} > | \tag{5.12}$$

where $< \cdot, \cdot >$ *denotes the duality product.*

It is readily seen that for problem (5.8):

$$\begin{aligned} \mathbb{B} = \mathcal{L}^{\infty}(\mathbf{T} \times \bar{\mathcal{D}}), \quad b = T_{2in}^{*}T_1 \in \mathcal{L}^{\infty}(\mathbf{T} \times \bar{\mathcal{D}}) \\ M = S = \mathcal{H}^{\infty}(\mathcal{L}^{\infty}(\bar{\mathcal{D}})) \end{aligned} \tag{5.13}$$

To apply the pre-duality result (5.12), we need to compute the pre-dual space of $\mathcal{L}^{\infty}(\mathbf{T} \times \bar{\mathcal{D}})$ and the pre-annihilator of S, ${}^{\perp}S$.

Let us first characterize the pre-dual space of $\mathcal{L}^{\infty}(\mathbf{T}\times\bar{\mathcal{D}})$. In order to do so, define the Banach space $L^1(\mathbf{T} \times \bar{\mathcal{D}})$ of measurable and absolutely integrable functions on $\mathbf{T} \times \bar{\mathcal{D}}$ under the L_1-norm for $f \in L^1(\mathbf{T} \times \bar{\mathcal{D}})$

$$\|f\|_1 := \int_{\bar{\mathcal{D}}} \int_{\mathbf{T}} |f(e^{i\theta}, \lambda)| d\theta d\lambda \tag{5.14}$$

To show that $\mathcal{L}^{\infty}(\mathbf{T}\times\bar{\mathcal{D}})$ is isometrically isomorphic to the dual space of $L^1(\mathbf{T}\times\bar{\mathcal{D}})$, let us introduce the concept of **tensor product of spaces**.

Let X, Y and Z be linear spaces over the same scalar field $\mathbb{K}(\mathbb{R}$ or $\mathbb{C})$. A function $\varphi : X \times Y \to Z$ is bilinear if $\varphi(x, \cdot) : Y \to Z$ is linear for each $x \in X$, and $\varphi(\cdot\,, y) : X \to Z$ for each $y \in Y$. The set of all bilinear functions from $X \times Y$ into Z is denoted by $B(X, Y; Z)$. If $Z = \mathbb{K}$, it is denoted simply by $B(X, Y)$.

For a linear space X, the space of all linear functionals on X is denoted by X'. For $x \in X, y \in Y$, the elementary tensor denoted $x \otimes y$ is the element of $B(X, Y)'$

defined by [103]:

$$(x \otimes y)(\varphi) = \varphi(x, y), \quad \forall \varphi \in B(X, Y)' \tag{5.15}$$

The tensor product $X \otimes Y$ is the linear span of all elementary tensors $\{x \otimes y : x \in X,\ y \in Y\}$.

Therefore, if $z \in X \otimes Y$, we then have:

$$z = \sum_{i=1}^{n} \lambda_i \, x_i \otimes y_i \tag{5.16}$$

for some certain $\{\lambda_1, \cdots, \lambda_n\}$ which are scalars, $\{x_1, \cdots, x_n\} \in X$, $\{y_1, \cdots, y_n\} \in Y$ and $n \in \mathbb{N}$ is arbitrary.

Definition 5.7. *For any $z \in X \otimes Y$, define the tensor norm [81]:*

$$\gamma(z) = \inf\{\sum_{i=1}^{n} \|x_i\|\|y_i\| : x_i \in X, y_i \in Y, z = \sum_{i=1}^{n} x_i \otimes y_i\}$$

In general, the space $X \otimes Y$ under the norm $\gamma(\cdot)$ is not complete. We will denote its completion by $X \otimes_\gamma Y$.

A result in [101] asserts that if Γ and $\widetilde{\Gamma}$ are σ- finite measure spaces, then

$$L^1(\Gamma \times \widetilde{\Gamma}) = L^1(\Gamma) \otimes_\gamma L^1(\widetilde{\Gamma}) \tag{5.17}$$

In our case, $\mathbf{T}$ and $\bar{\mathcal{D}}$ are finite measure spaces and therefore σ-finite measure spaces. Hence it follows that:

$$L^1(\mathbf{T} \times \bar{\mathcal{D}}) = L^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}})$$

And for $z \in L^1(\mathbf{T}) \otimes L^1(\bar{\mathcal{D}})$,

$$\gamma(z) = \inf\{\sum_{i=1}^{n} \|x_i(e^{i\theta})\|_1 \|y_i(\lambda)\|_1 : x_i(\cdot) \in L^1(\mathbf{T}),$$
$$y_i(\cdot) \in L^1(\bar{\mathcal{D}}),\ z = \sum_{i=1}^{n} x_i \otimes y_i\} \tag{5.18}$$

where

$$\|x_i(e^{i\theta})\|_1 = \int_{[0,2\pi)} |x_i(e^{i\theta})| d\theta \tag{5.19}$$

and

$$\|y_i(\lambda)\|_1 = \int_{\bar{\mathcal{D}}} |y_i(\lambda)| d\lambda \tag{5.20}$$

Remark 5.8. *The dual space of $L^1(\mathbf{T} \times \bar{\mathcal{D}})$, denoted $L^1(\mathbf{T} \times \bar{\mathcal{D}})^*$, can then be identified with $L^\infty(\mathbf{T} \times \bar{\mathcal{D}})$:*

$$L^1(\mathbf{T} \times \bar{\mathcal{D}})^* \simeq \left(L^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}})\right)^* \simeq L^\infty(\mathbf{T} \times \bar{\mathcal{D}}) \tag{5.21}$$

by the Steinhaus-Nikodym theorem*.*

That is for each $F \in L^\infty(\mathbf{T} \times \bar{\mathcal{D}})$, define $\varphi_F \in L^1(\mathbf{T} \times \bar{\mathcal{D}})^*$ such that:

$$\varphi_F(f, g) = \int_{\bar{\mathcal{D}}} \int_{[0,2\pi)} F(e^{i\theta}, \lambda) f(e^{i\theta}) g(\lambda) d\theta d\lambda \tag{5.22}$$

Expression (5.22) characterizes every bounded linear functionals on $L^1(\mathbf{T} \times \bar{\mathcal{D}})$.

Next, the pre-annihilator of $S = \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ in $L^\infty(\mathbf{T} \times \bar{\mathcal{D}})$ is computed. First note that the pre-annihilator of $H^\infty(\bar{\mathcal{D}})$ in $L^\infty(\mathbf{T})$ is:

$$H_0^1(\mathbf{T}) := \{f \in L^1(\mathbf{T}) : \hat{f}(n) = 0,\ \forall n \leq 0\}$$

where $\hat{f}(n)$ denotes the *n-th Fourier coefficients* of f.

That is $\forall g \in \mathcal{H}^\infty(\bar{\mathcal{D}})$, $\forall f \in H_0^1(\mathcal{D})$ we have:

$$\int_{[0,2\pi)} g(e^{i\theta}) f(e^{i\theta}) = 0 \tag{5.23}$$

To compute the pre-annihilator of $S = \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$, it suffices to notice that for each $f \in H_0^1(L^1(\mathbf{T}))$, i.e. $f(z,\lambda)$ with $z \in \mathcal{D}, \lambda \in \bar{\mathcal{D}}$ and for each fixed λ, $f(\cdot,\lambda) \in H_0^1$ and for each fixed $z \in \mathcal{D}$, $f(z,\cdot) \in L^1(\bar{\mathcal{D}})$.

Hence, for $\forall F \in H^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ and $\forall f \in H_0^1(L^1(\bar{\mathcal{D}}))$,

$$\int_{\bar{\mathcal{D}}} \underbrace{\int_{[0,2\pi)} F(e^{i\theta},\lambda) f(e^{i\theta},\lambda) d\theta}_{=0} \, d\lambda = 0 \tag{5.24}$$

Then the pre-annihilator of S is:

$$^\perp S = H_0^1(L^1(\bar{\mathcal{D}})) \tag{5.25}$$

The existence of a pre-dual space $L^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}})$ and a pre-annihilator $^\perp S$ implies the following theorem which is a standard result in Banach space duality theory relating the distance from a vector to a subspace and an extremal functional in the predual (*Theorem 2* in [103]).

Theorem 5.9. *There exists at least one optimal $Q_0 \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ achieving optimal performance μ in (5.4). Moreover the following identities hold:*

$$\begin{aligned} \mu &= \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_{2in}^* T_1 - Q\|_\infty = \|T_{2in}^* T_1 - Q_0\|_\infty \\ &= \sup_{\substack{F \in H_0^1(L^1(\mathbf{T})) \\ \|F\|_1 \le 1}} \int_{\bar{\mathcal{D}}} \int_{[0,2\pi)} T_{2in}^* T_1(e^{i\theta},\lambda) F(e^{i\theta},\lambda) d\theta d\lambda \end{aligned} \tag{5.26}$$

The optimal controller can then be computed by letting

$$Q = T_{2out}^{-1} Q_0$$

and

$$K = -Q(I - G_{22}Q_0)^{-1}$$

Note that the supremum in the pre-dual characterization is in general not attained. However, if we assume that $T_{2in}^* T_1$ is continuous on $\mathbf{T} \times \bar{\mathcal{D}}$, then it is possible to show that in fact the supremum is achieved. This is carried out in the next section.

5.2.2 Dual characterization

Let's first introduce the space of continuous functions on $\mathbf{T} \times \bar{\mathcal{D}}$, which is denoted $\mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$ under the sup-norm:

$$\sup_{\substack{\theta \in [0,2\pi) \\ \lambda \in \bar{\mathcal{D}}}} |G(e^{i\theta}, \lambda)| < \infty, \text{ for } G \in \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}}) \tag{5.27}$$

As mentioned before, in this section we make the following assumption.

Assumption 5.10.

$$T_{2in}^* T_1 \in \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}}), \tag{5.28}$$

that is $T_{2in}^* T_1$ *is continuous on* $\mathbf{T} \times \bar{\mathcal{D}}$.

Since $\mathbf{T}$ and $\bar{\mathcal{D}}$ are compact, the dual space of $\mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$, henceforth denoted $\mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})^*$, is isometrically isomorphic to the space of Borel measure $M(\mathbf{T} \times \bar{\mathcal{D}})$ on $\mathbf{T} \times \bar{\mathcal{D}}$ under the total variation norm for $\mu \in M(\mathbf{T} \times \bar{\mathcal{D}})$:

$$\|\mu\| = |\mu|(\mathbf{T} \times \bar{\mathcal{D}}), \tag{5.29}$$

where $|\mu|$ is the total variation of μ.

The isometric isomorphism is given by:
For $\mu \in M(\mathbf{T} \times \bar{\mathcal{D}})$ and $f \in \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$ letting

$$I_\mu(f) = \int_{[0,2\pi)\times\bar{\mathcal{D}}} f(e^{i\theta}, \lambda) d\mu(\theta, \lambda). \tag{5.30}$$

The map $\mu \to I_\mu$ is the isometric isomorphism from $M(\mathbf{T} \times \bar{\mathcal{D}})$ to $\mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$ [104].

Corollary 5.10.1. *Banach duality states that for a Banach space $\mathbb{B}$, and a subspace M of $\mathbb{B}$, we have*

$$\inf_{m \in M} \|b - m\| = \max_{\substack{\tilde{m} \in M^\perp \\ \|\tilde{m}\| \leq 1}} |\tilde{m}(b)| \tag{5.31}$$

where $M^\perp$ is the annihilator of M in B^*, the dual space of B, defined in (5.31).

Define a subspace of $\mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$ as follows:

$$S_c = S \bigcap \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}}) = \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}})) \bigcap \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}}) \tag{5.32}$$

In the following lemma we establish that the distance from $T_{2in}^* T_1 \in \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$ to S_c is the same as to S. Observe that with assumption (A_1), functions in S are continuous in the second variable λ.

Lemma 5.10.1.

$$\mu = \|T_{2in}^* T_1 - Q_0\|_\infty = \inf_{\tilde{Q} \in S_c} \|T_{2in}^* - \tilde{Q}\|_\infty \tag{5.33}$$

Proof: The proof can be found in the Appendix A.2. ■

Define the space A known as the disc algebra $A_0 = \mathcal{C}(\mathbf{T}) \cap \mathcal{H}^\infty$. So S_c can be written in the following form:

$$S_c = A_0(\mathcal{C}(\bar{\mathcal{D}})) \tag{5.34}$$

To compute the annihilator of S_c, $S_c^\perp$ in $M(\mathbf{T} \times \bar{\mathcal{D}})$ it suffices to notice that the annihilator of A_0 in $M(\mathbf{T})$ is $H_0^1(\mathbf{T})$, and use a similar argument as (**??**)

$$S_c^\perp = H_0^1(M(\bar{\mathcal{D}})) \tag{5.35}$$

Using the duality theory result, we deduce the following theorem:

Theorem 5.11.

$$\begin{aligned}
\mu &= \max_{\substack{\phi \in H_0^1(M(\bar{\mathcal{D}})) \\ \|\phi\| \le 1}} |\phi(T_{2in}^* T_1)| \\
&= \max_{\substack{\phi \in H_0^1(M(\bar{\mathcal{D}})) \\ \|\phi\| \le 1}} \left| \int_{[0,2\pi) \times \bar{\mathcal{D}}} T_{2in}^* T_1(e^{i\theta}, \lambda) d\phi(e^{i\theta}, \lambda) \right| \\
&= \int_{[0,2\pi) \times \bar{\mathcal{D}}} T_{2in}^* T_1(e^{i\theta}, \lambda) d\phi_0(e^{i\theta}, \lambda) \qquad (5.36)
\end{aligned}$$

where $\phi_0(\cdot, \cdot)$ is the dual extremal functional, and $\phi \in H_0^1(M(\bar{\mathcal{D}}))$ means for each fixed $\theta \in [0, 2\pi)$, $\phi(e^{i\theta}, \cdot)$ is a bounded Borel measure on $\bar{\mathcal{D}}$ and for each fixed $\lambda \in \bar{\mathcal{D}}$, $d\phi(\cdot, \lambda) = G(e^{i\theta})d\theta$ for some function $G(\cdot) \in H_0^1$.

Moreover, by *Lemma 5.10.1* under *Assumption 5.10*, the search of Q can be restricted to the subspace S_c. This will play an important role in finding a numerical solution as discussed in Section 5.4.

5.3 An Operator Theoretic Approach

In the sequel, we shall show that the pre-annihilator of S can be written as a tensor subspace. This will allow us to *separate* between θ and λ. We have seen that the pre-dual space can be characterized as the following tensor space,

$$L^1(\mathbf{T} \times \bar{\mathcal{D}}) \simeq L^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}}), \tag{5.37}$$

where $L^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}})$ is the closure in the γ-norm of the linear space $L^1(\mathbf{T}) \otimes L^1(\bar{\mathcal{D}})$ defined by:

$$\begin{aligned} L^1(\mathbf{T}) \otimes L^1(\bar{\mathcal{D}}) = \{F(e^{i\theta}, \lambda) = \sum_{i=1}^{n} f_i(e^{i\theta}) g_i(\lambda), \\ f_i \in L^1(\mathbf{T}),\ g_i \in L^1(\bar{\mathcal{D}}) \text{ and } n \text{ arbitrary integer.}\} \end{aligned} \tag{5.38}$$

The dual space of $L^1(\mathbf{T} \times \bar{\mathcal{D}})$ is given by $L^\infty(\mathbf{T} \times \bar{\mathcal{D}})$.

The pre-annihilator of $S = \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ in $L^\infty(\mathbf{T} \times \bar{\mathcal{D}})$ is characterized as

$$^\perp S = H_0^1(L^1(\bar{\mathcal{D}})). \tag{5.39}$$

We shall write $^\perp S$ as a particular tensor subspace. In order to do this, we define the following tensor:

$$H_0^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}}), \tag{5.40}$$

which is the closure in the γ-norm of the tensor space given below:

$$\begin{aligned} H_0^1(\mathbf{T}) \otimes L^1(\bar{\mathcal{D}}) &= \{F(e^{i\theta}, \lambda) \in L^1(\mathbf{T} \times \bar{\mathcal{D}}) : F(e^{i\theta}, \lambda) \\ &= \sum_{i=1}^{n} f_i(e^{i\theta}) g_i(\lambda), \end{aligned} \tag{5.41}$$

for $f_i \in H_0^1(\mathbf{T}), g_i \in L^1(\bar{\mathcal{D}})$ and n arbitrary integer$\}$. And where the γ-norm in this case is defined as:

$$\gamma(\sum_{i=1}^{n} f_i \otimes g_i) = \inf\{\sum_{i=1}^{n} \|f_i\|_1 \|g_i\|_1;$$

$$f_i \in H^1_0(\mathbf{T}),\ g_i \in L^1(\bar{\mathcal{D}})\}. \tag{5.42}$$

In the following Lemma we show that the pre-annihilator ${}^{\perp}S$ can be written as the tensor subspace (5.40).

Lemma 5.11.1. *The following isometric isomorphism holds:*

$$^{\perp}S = H^1_0(L^1(\bar{\mathcal{D}})) \simeq H^1_0(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}}). \tag{5.43}$$

Proof: The proof can be found in Appendix A.3. ■

A consequence of Lemma 5.11.1 is that Theorem 5.9 can be formulated as follows:

Theorem 5.12.

$$\begin{aligned}
\mu &= \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T^*_{2in}T_1 - Q\|_\infty \\
&= \|T^*_{2in}T_1 - Q_0\|_\infty \\
&= \sup_{\hat{S}} \int_{\bar{\mathcal{D}}} \int_{[0,2\pi)} T^*_{2in}T_1(e^{i\theta}, \lambda) F(e^{i\theta}, \lambda) d\theta d\lambda
\end{aligned}$$

$$= \sup_{S^-} \sum_{i=1}^{n} \int_{\bar{\mathcal{D}}} \int_{[0,2\pi)} T^*_{2in}T_1(e^{i\theta}, \lambda) x_i(\lambda) y_i(e^{i\theta}) d\theta d\lambda, \tag{5.44}$$

where the two sets are defined as:

$$\hat{S} := F \in H^1_0(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}}),\ \|F\|_\gamma \leq 1;$$

$$S^- := \sum_{i=1}^{n} x_i \otimes y_i \in H^1_0(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}}),\ \gamma(\sum_{i=1}^{n} x_i \otimes y_i) \leq 1.$$

Identity (5.44) can be used to give a solution in terms of operator theory as outlined in the following section.

5.3.1 Pre-dual Characterization

Each function $y_i, i = 1, \cdots, n$ in (5.44) belongs to $H_0^1(\mathbf{T})$.

By *F. Riesz Representation Theorem* [62], there is a factorization:

$$y_i(e^{i\theta}) = y_i^1(e^{i\theta})y_i^2(e^{i\theta}), \quad i = 1, \cdots, n. \tag{5.45}$$

where y_i^1 and y_i^2 are in $H^2(\mathbf{T})$ and $H_0^2(\mathbf{T})$ respectively.

Recall that $H^2(\mathbf{T})$ and $H_0^2(\mathbf{T})$ are the Hardy spaces of analytic and absolutely integrable functions in the unit disc.
In other words:

$$H^2(\mathbf{T}) = f(e^{i\theta}) \in L^2(\mathbf{T}) : \quad \hat{f}(m) = 0, \forall\, m < 0, \tag{5.46}$$

and

$$H_0^2(\mathbf{T}) = f(e^{i\theta}) \in L^2(\mathbf{T}) : \quad \hat{f}(n) = 0, \forall\, n \leq 0, \tag{5.47}$$

where $\hat{f}(n)$ denotes the *n-th Fourier coefficient* of f.

And $|y_i^1(e^{i\theta})|^2 = |y_i^2(e^{i\theta})|^2 = |y_i(e^{i\theta})|$ almost everywhere (a.e.).

From (5.44) we have

$$\begin{aligned} & \int_{\bar{\mathcal{D}}} \int_{[0,2\pi)} T_{2in}^* T_1(e^{i\theta}, \lambda) y_i^1(e^{i\theta}) y_i^2(e^{i\theta}) x_i(\lambda) d\theta d\lambda \\ = \quad & \int_{\bar{\mathcal{D}}} \int_{[0,2\pi)} (T_1(e^{i\theta}, \lambda) y_i^1(e^{i\theta}))(T_{2in}^*(e^{i\theta}, \lambda) y_i^2(e^{i\theta})) x_i d\theta d\lambda \end{aligned}$$

$$=: \; < T_1 y_i^1 x_i, T_{2in} \overline{y_i^2} > . \tag{5.48}$$

Note that $y_i^2 \in H^2(\mathbf{T})^\perp$, the orthogonal complement of $H^2(\mathbf{T})$ in $L^2(\mathbf{T})$. So for each fixed $\lambda \in \bar{\mathcal{D}}$,

$$\begin{gathered}(T_{2in}(\cdot, \lambda) H^2(\mathbf{T}))^\perp = \\ (H^2(\mathbf{T}) \ominus T_{2in}(\cdot, \lambda) H^2(\mathbf{T})) \oplus H^{2\perp}(\mathbf{T}) \quad ,\end{gathered} \tag{5.49}$$

where "$\oplus$" denotes the direct sum, and $(H^2(\mathbf{T}) \ominus T_{2in}(\cdot, \lambda) H^2(\mathbf{T}))$ is the orthogonal complement of $T_{2in}(\cdot, \lambda) H^2(\mathbf{T})$.

Remark 5.13. *Here orthogonality is understood to be with respect to the inner product of matrices.*

Let P denote the orthogonal projection in $L^2(\mathbf{T})$ with range $H^2(\mathbf{T}) \ominus T_{2in}(\cdot, \lambda) H^2(\mathbf{T})$. The function $T_{2in}(\cdot, \lambda) \overline{y_i^2} \in (T_{2in}(\cdot, \lambda) H^2(\mathbf{T}))^\perp$, where $\overline{y}$ denotes the complex conjugate of y.
Therefore

$$T_{2in}(\cdot, \lambda) \overline{y_i^2} - P T_{2in}(\cdot, \lambda) \overline{y_i^2} \in H^{2\perp}(\mathbf{T}). \tag{5.50}$$

It follows then:

$$< T_1 y_i^1 x_i, T_{2in} \overline{y_i^2} > = < T_1 y_i^1 x_i, (P T_{2in} \overline{y_i^2}) > . \tag{5.51}$$

Furthermore, for each $\lambda \in \bar{\mathcal{D}}$, the function

$$y_i^1 - P y_i^1 \in T_{2in}(\cdot, \lambda) H^2(\mathbf{T}), \tag{5.52}$$

and therefore

$$T_1(\cdot,\lambda)(y_i^1 - Py_i^1) \in T_{2in}(\cdot,\lambda)H^2(\mathbf{T}). \tag{5.53}$$

Hence,

$$\begin{aligned} < T_1 y_i^1 x_i, (PT_{2in}\overline{y_i^2}) > &= < T_1 P y_i^1 x_i, (PT_{2in}\overline{y_i^2}) > \\ &= < PT_1 P y_i^1 x_i, (T_{2in}\overline{y_i^2}) > \end{aligned}$$

Reporting in (5.44) yields:

$$\mu = \sup_{S^+} \sum_{i=1}^{n} < PT_1 P y_i^1 x_i, (T_{2in}\overline{y_i^2}) >, \tag{5.54}$$

where

$$\begin{aligned} S^+ := \{ \sum_{i=1}^{n} y_i^1 y_i^2 \otimes x_i \in H^2 H_0^2(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}}), \\ \inf \sum_{i=1}^{n} \|y_i^1\|_2 \|y_i^2\|_2 \|x_i\|_1 \leq 1 \}. \end{aligned}$$

Next, we define the following key operator:

$$\begin{aligned} \Xi &: H^2(\mathbf{T}) \ominus T_{2in} H^2(\mathbf{T}) \to H^2(\mathbf{T}) \ominus T_{2in} H^2(\mathbf{T}) \\ \Xi &:= PT_1 \mid_{H^2(\mathbf{T}) \ominus T_{2in} H^2(\mathbf{T})}, \end{aligned}$$

and the tensor operator

$$\begin{gathered} \Xi \otimes I : \\ (H^2(\mathbf{T}) \ominus T_{2in} H^2(\mathbf{T})) \otimes_\gamma L^1(\mathbf{T}) \longrightarrow \\ (H^2(\mathbf{T}) \ominus T_{2in}(\mathbf{T})) \otimes_\gamma L^1(\mathbf{T}). \end{gathered}$$

where $(H^2(\mathbf{T}) \ominus T_{2in}H^2(\mathbf{T})) \otimes_\gamma L^1(\mathbf{T})$ is the **completion** of $(H^2(\mathbf{T}) \ominus T_{2in}H^2(\mathbf{T})) \otimes L^1(\mathbf{T})$ in the γ-norm given by:

$$\gamma(\sum_{i=1}^{n} x_i \otimes y_i) = \inf\{\sum_{i=1}^{n} \|x_i\|_2 \|y_i\|_1, \\ x_i \in H^2(\mathbf{T}) \ominus T_{2in}H^2(\mathbf{T}),\ y_i \in L^1(\mathbf{T})\}. \tag{5.55}$$

Then $\Xi \otimes I$ is a bounded linear operator, and (5.54) is in fact its operator induced norm. That is:

$$\mu = \|\Xi \otimes I\|. \tag{5.56}$$

This fact is summarized in the following theorem.

Theorem 5.14. *There exists at least one optimal $Q_0 \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$ s.t.*

$$\begin{aligned} \mu &= \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_{2in}^* T_1 - Q\|_\infty \\ &= \|T_1 - T_{21}Q_0\|_\infty \\ &= \|\Xi \otimes I\|. \end{aligned} \tag{5.57}$$

Theorem 5.14 shows that optimal distributed performance is equal to the operator induced norm of the operator $\Xi \otimes I$. Ξ plays a central role in finding an explicit solution to our problem through the above theorem which quantifies the optimal performance.

5.3.2 Dual characterization using tensor spaces

Let us introduce the space of continuous functions on $\mathbf{T} \times \bar{\mathcal{D}}$, which is denoted $\mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$ under the sup-norm:

$$\|G\|_\infty = \sup_{\substack{\theta \in [0,2\pi) \\ \lambda \in \bar{\mathcal{D}}}} |G(e^{i\theta}, \lambda)|, \text{ for } G \in \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}}). \tag{5.58}$$

In Section 5.2.2, we provided the dual characterization based on *Assumption 5.10*, i.e.,

$T_{2in}^{*}T_1 \in \mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$, that is $T_{2in}^{*}T_1$ is continuous on $\mathbf{T} \times \bar{\mathcal{D}}$.

Here, $\mathcal{C}(\mathbf{T} \times \bar{\mathcal{D}})$ is viewed as a tensor space. This will allow us to separate between the independent variables θ and λ.

Definition 5.15. *Define the tensor space $\mathcal{C}(\mathbf{T}) \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}})$ as the closure of $\mathcal{C}(\mathbf{T}) \otimes \mathcal{C}(\bar{\mathcal{D}})$ in the λ-norm defined as:*

$$\begin{aligned} \lambda(\sum_{i=1}^{n} x_i \otimes y_i) = \sup\{\| \sum_{i=1}^{n} \phi(x_i) y_i \|_\infty \quad &: \quad \phi \in (\mathcal{C}(\mathbf{T}))^*, \\ \|\phi\| \quad &= \quad 1\}, \end{aligned} \tag{5.59}$$

where $x_i \in \mathcal{C}(\mathbf{T})$, $y_i \in \mathcal{C}(\bar{\mathcal{D}})$, and $(\mathcal{C}(\mathbf{T}))^*$ is the dual space of $\mathcal{C}(\mathbf{T})$.

Remark 5.16. *$(\mathcal{C}(\mathbf{T}))^*$ can be identified under an isometric isomorphism with the space of Borel measures $M(\mathbf{T})$ on the unit circle $\mathbf{T}$ under (bounded) total variation norm.*

For $\beta \in M(\mathbf{T})$, $\|\beta\| = |\beta|(\mathbf{T})$, where $|\beta|$ is the *total variation* of β.

The isometric isomorphism is given by:

For $\beta \in M(\mathbf{T})$ and $f \in \mathcal{C}(\mathbf{T})$

$$I_\beta(f) = \int_{[0,2\pi)} f(e^{i\theta}) d\beta(\theta). \tag{5.60}$$

With this identification, (5.59) can be written more explicitly as:

$$\lambda(\sum_{i=1}^{n} x_i \otimes y_i) = \sup_{\substack{\beta \in M(\mathbf{T}) \\ |\beta|=1}} \| \sum_{i=1}^{n} \int_{[0,2\pi)} x_i(e^{i\theta}) d\beta(\theta) y_i \|_\infty \tag{5.61}$$

It follows by a result in [101] that

$$\mathcal{C}(\mathbf{T}\times\bar{\mathcal{D}}) \simeq \mathcal{C}(\mathbf{T}) \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}}). \tag{5.62}$$

Thus the subspace S_c defined in (5.32) can be rewritten as

$$\begin{aligned} S_c &= S\bigcap \mathcal{C}(\mathbf{T}\times\bar{\mathcal{D}}) = \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))\bigcap \mathcal{C}(\mathbf{T}\times\bar{\mathcal{D}}) \\ &= \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))\bigcap \mathcal{C}(\mathbf{T}) \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}}). \end{aligned} \tag{5.63}$$

Since the dual spaces of $\mathcal{C}(\mathbf{T})$ and $\mathcal{C}(\bar{\mathcal{D}})$ are given by the spaces of Borel measures $M(\mathbf{T})$ and $M(\bar{\mathcal{D}})$, with total variation norms, respectively.
Considering $\mathcal{C}(\mathbf{T}\times\bar{\mathcal{D}}) \simeq \mathcal{C}(\mathbf{T}) \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}})$, the dual space of $\mathcal{C}(\mathbf{T}\times\bar{\mathcal{D}})$ follows that

$$\begin{aligned} (\mathcal{C}(\mathbf{T}) \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}}))^* &\simeq \mathcal{C}^*(\mathbf{T}) \otimes_\gamma \mathcal{C}^*(\bar{\mathcal{D}}) \\ &\simeq M(\mathbf{T}) \otimes_\gamma M(\bar{\mathcal{D}}), \end{aligned} \tag{5.64}$$

where

$$\gamma(\sum_{i=1}^n \phi_i \otimes \psi_i) = \inf\{\sum_{i=1}^n \|\phi_i\|\|\psi_i\| : \phi_i \in M(\mathbf{T}),\ \psi_i \in M(\bar{\mathcal{D}})\}.$$

The isometric isomorphism "$\simeq$" is given by:

$$f \in \mathcal{C}(\mathbf{T}) \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}}), \qquad f = \sum_{i=1}^n f_i \otimes g_i;$$

and

$$\mu \in M(\mathbf{T}) \otimes_\gamma M(\bar{\mathcal{D}}), \qquad \mu = \sum_{i=1}^n \phi_i \otimes \psi_i; \tag{5.65}$$

$$I(f) = \sum_{i=1}^{n} \int_{[0,2\pi)\times\bar{\mathcal{D}}} f_i(e^{i\theta}) g_i(\lambda) \phi_i(d\theta) \psi_i(d\lambda). \tag{5.66}$$

The subspace S_c can then be written as:

$$\begin{aligned} S_c &= \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}})) \bigcap [\mathcal{C}(\mathbf{T}) \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}})] \\ &= A_0 \otimes_\lambda \mathcal{C}(\bar{\mathcal{D}}). \end{aligned} \tag{5.67}$$

The annihilator of S_c, $S_c^\perp$ in $M(\mathbf{T}) \otimes_\gamma M(\bar{\mathcal{D}})$ can be determined explicitly as:

$$S_c^\perp = H_0^1(\mathbf{T}) \otimes_\gamma M(\bar{\mathcal{D}}), \tag{5.68}$$

since $H_0^1(\mathbf{T})$ is the annihilator of A_0 in $M(\mathbf{T})$.

By the *Radon-Nikodym Theorem* [104], each measure $\phi \in M(\bar{\mathcal{D}})$ can be written as:

$$d\phi(\lambda) = \varphi(\lambda) d\nu(\lambda), \tag{5.69}$$

where ν is the total variation of ϕ on $\bar{\mathcal{D}}$, and $\varphi \in L^1(d\nu)$.

Then, the γ-norm can be written as: for $\phi \in H_0^1(\mathbf{T}) \otimes M(\bar{\mathcal{D}})$,

$$\begin{aligned} d\phi(e^{i\theta}, \lambda) = \sum_{i=1}^{n} \psi_i(e^{i\theta}) \varphi_i(\lambda) d\theta d\nu_i(\lambda); \\ \psi_i \in H_0^1(\mathbf{T}), \quad \varphi_i \in L^1(d\nu_i), \end{aligned}$$

$$\gamma(\phi) = \inf\{\sum_{i=1}^{n} \|\psi_i\|_1 \int_{\bar{\mathcal{D}}} |\varphi_i(\lambda)| d\nu_i(\lambda),$$
$$\psi_i \in H_0^1(\mathbf{T}), \varphi_i \in L^1(d\nu_i)\}. \tag{5.70}$$

Using duality theory, we deduce the following theorem which formulates the dual problem in terms of tensor space.

Theorem 5.17.

$$\begin{aligned} \mu &= \max_{\substack{\phi \in H_0^1(\mathbf{T}) \otimes_\gamma M(\bar{\mathcal{D}}) \\ \gamma(\phi) \le 1}} \int_{[0,2\pi)\times\bar{\mathcal{D}}} |T_{2in}^* T_1(e^{i\theta}, \lambda) d\phi(e^{i\theta}, \lambda)| \\ &= \int_{[0,2\pi)\times\bar{\mathcal{D}}} T_{2in}^* T_1(e^{i\theta}, \lambda) d\phi_0(e^{i\theta}, \lambda), \end{aligned} \tag{5.71}$$

for some $\phi_0 \in H_0^1(\mathbf{T}) \otimes_\gamma M(\bar{\mathcal{D}}), \gamma(\phi_0) \le 1.$

Theorem 5.17 shows that to solve the dual extremal problem, the search for the maximization can be restricted to tensors of the form (5.65).

The pre-dual and dual formulation lead to numerical solutions based on finite variable convex programming.

5.4 Discussion of A Numerical Solution

The *infimum* in (5.33) is termed the primal optimization and corresponds to the following representation:

$$\mu = \inf_{Q \in S_c} \|T_{2in}^* T_1 - Q\|_\infty \tag{5.72}$$

if Q is restricted to the subspace P_{mn} consisting of polynomial in two variables of the form:

$$P_{mn}(z,\lambda) := \sum_{j=-m}^{m}\sum_{i=0}^{n}\alpha_{ij}z^i\lambda^j; \quad \alpha_{ij}\in\mathbb{R}, \tag{5.73}$$
$$\text{for} \quad |z|=1, \qquad |\lambda|\le 1$$

Remark 5.18. *Note that these polynomials are analytic in the first variable z for $|z|<1$ since Q is analytic in z for $|z|<1$.*

Then we get an **upper bound** for (5.72) that is for

$$\mu_{mn} := \inf_{Q\in P_{mn}} \|T_{2in}^*T_1 - Q\|_\infty \tag{5.74}$$

That is, $\mu_{mn}\ge\mu$ since $P_{mn}\subset S_c$, the infimum being taken over a smaller subspace. Since the polynomials P_{mn} are dense in $\mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))$, therefore we have:

$$\mu_{mn}\downarrow\mu \quad \text{as} \quad m,n\to\infty, \tag{5.75}$$

i.e., μ_{mn} converges to the optimal μ from above.

The optimization problem (5.74) is finite dimensional, since reduces to searching for the coefficients $\{\alpha_{ij}\}_{i=0,j=-m}^{n,m}$ that minimize $\|T_{2in}^*T_1 - Q\|_\infty$.

Now, we turn our attention to the dual problem which is:

$$\mu = \sup_{\substack{F\in H_0^1(L^1(\bar{\mathcal{D}}))\\ \|F\|_1\le 1}} \int_{\bar{\mathcal{D}}}\int_{[0,2\pi)} T_{2in}^*T_1(e^{i\theta},\lambda)F(e^{i\theta},\lambda)\,d\theta\, d\lambda \tag{5.76}$$

By restricting the search to polynomials $P_{k\ell}$ of two variables of the form:

$$P_{k\ell}(z,\lambda) := \sum_{j=-k}^{k}\sum_{i=1}^{\ell}\beta_{ij}z^i\lambda^j; \quad \beta_{ij}\in\mathbb{R},$$
$$\text{for} \quad |z|=1, \qquad |\lambda|\le 1 \tag{5.77}$$

with norm $|P_{k\ell}|_1 \leq 1$. Note in the sum over i we start from 1 since $P_{k\ell}(z,\lambda) \in H_0^1(L^1(\bar{\mathcal{D}}))$. We get the finite dimensional optimization:

$$\mu_{k\ell} := \sup_{\substack{P_{k\ell} \in \tilde{P}_{k\ell} \\ \|P_{k\ell}\|_1 \leq 1}} \int_{\bar{\mathcal{D}}} \int_{[0,2\pi)} T_{2in}^* T_1(e^{i\theta}, \lambda) P_{k\ell}(e^{i\theta}, \lambda) d\theta d\lambda \tag{5.78}$$

since the search in (5.78) is over the coefficients $\{\beta_{ij}\}_{i=1,j=-k}^{\ell,k}$.

Moreover, $P_{k\ell}$ is a subspace of $H_0^1(L^1(\bar{\mathcal{D}}))$, then $\mu_{k\ell} \leq \mu$, since the supremum is taken over a smaller set, i.e., we get a **lower bound** for μ.

Polynomials of the form (5.78) are dense in $H_0^1(L^1(\bar{\mathcal{D}}))$, therefore we have:

$$\mu_{k\ell} \uparrow \mu \quad \text{as} \quad k, \ell \to \infty \tag{5.79}$$

In other words, $\mu_{k\ell}$ converges to the optimal μ from below. Combining (5.75) and (5.79), we then have that:

$$\mu_{k\ell} \uparrow \mu \downarrow \mu_{mn} \ \text{ as } \ k, \ell, m, n \longrightarrow \infty \tag{5.80}$$

squeezing the optimum within desired accuracy by taking large enough k, ℓ, m and n.

Therefore the finite dimensional optimization (5.74) and (5.77) estimate μ within desired tolerance and compute the corresponding Q in P_{mn}, which in turn leads to the computation of distributed spatially invariant controllers K as close as desired to the optimal ones through the parametrization (5.2). Solving such problems are then applications of finite variable convex programming.

5.5 Numerical Examples

This section contains a numerical example that illustrates the utility of the method proposed in this chapter. As discussed in Section 5.4, the proposed duality theory leads to a dual pair of numerical solutions, which approach the optimal μ from opposite directions.

We take the same discrete time example as given in Voulgaris et al. [84]. Consider the following spatio-temporal system which comes from the finite-difference discretization of a certain PDE.

$$\begin{aligned} y(i,k+1) - y(i,k) &= \frac{T}{L^2} y(i+1,k) - 2y(i,k) + y(i-1,k) \\ &- \epsilon y(i,k) + u(i,k) \end{aligned} \tag{5.81}$$

Taking the appropriate transforms one obtains the transfer function, weighting function and stabilizing controller parameterization introduced in Section 3.3.

We want to compute a decentralized controller for optimal attenuation of an additive disturbance on the system output with weighting function

$$W(z,\lambda) = \frac{\lambda}{1 - (\frac{1}{8}z^{-1} + \frac{1}{4} + \frac{z}{8})\lambda\,.} \tag{5.82}$$

We assume $W(z;\lambda)$ to be asymptotically stable and yield the same structure as the plant itself.

As defined in [84], with $K(z,\lambda)$ and $Q(z,\lambda)$ of the prescribed form, the problem can be stated as

$$\inf_Q \|(1 - GQ)W\|_\infty = \inf_Q \|H - UQ\|_\infty, \tag{5.83}$$

where

$$H(z,\lambda) = \frac{\lambda}{1 - r(z)\lambda}, \qquad U(z,\lambda) = \frac{T\lambda^2}{(1-\rho(z)\lambda)(1-r(z)\lambda)}, \tag{5.84}$$

and

$$\begin{aligned}
\rho(z) &= z^{-1}/6 + 1/3 + z/6, \\
r(z) &= z^{-1}/8 + 1/4 + z/8. \qquad (5.85)
\end{aligned}$$

An inner−outer factorization of $U(z, \lambda)$ yields

$$U_{in}(z, \lambda) = \lambda^2, \qquad (5.86)$$

$$U_{out}(z, \lambda) = \frac{T}{(1 - \rho(z)\lambda)(1 - r(z)\lambda)\,.} \qquad (5.87)$$

The finite variable convex programs can be solved by any standard convex optimization algorithm [105, 106]. In this example, we used CVX, a free package for specifying and solving convex programs [107].

Solving their corresponding approximated convex problems (5.74), (5.77), both the lower and upper bounds can be obtained. Without loss of generality, we pick the 6-th order polynomial Q which holds the following expression

$$Q(z, \lambda) = \sum_{j=-3}^{3} \sum_{i=0}^{6} \alpha_{ij} z^i \lambda^j; \quad \alpha_{ij} \in \mathbb{R}. \qquad (5.88)$$

By observing all the possible combinations of the orders for z and λ, we notice that it should come with 49 coefficients, i.e. 49 different α_{ij}'s corresponding to various combinations of i, j's, respectively.

Figure 5.2 shows the upper and lower limits that we seek to calculate. Figure 5.2 compares the optimal disturbance rejections with coefficients being 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, respectively. It is worth mentioning that the x axis represents the change of the weighting function (i.e. we multiply the weighting function with these coefficients). The red line represents the upper bound and the blue line represents the lower bound. The black stars correspond to the difference between upper and lower

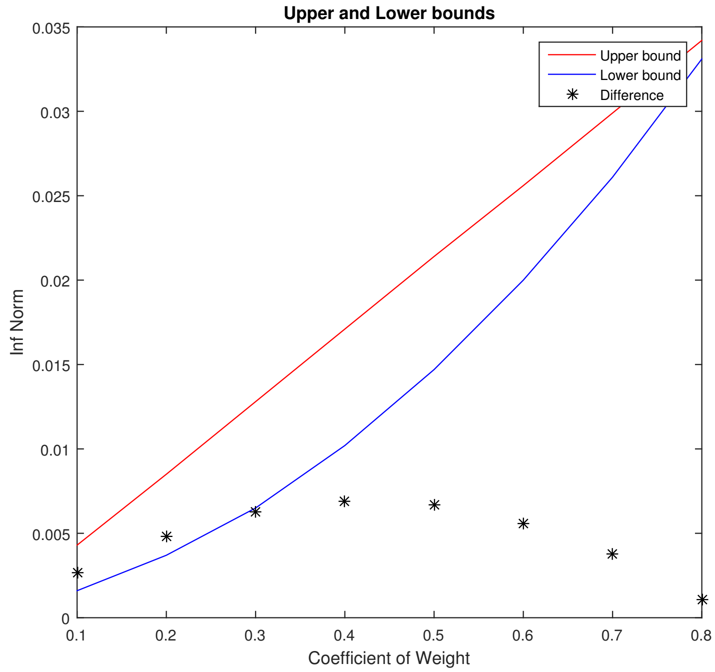


Figure 5.2: Upper and lower bounds

bounds at different coefficients. As expected, we observe the performance deteriorates as uncertainty increases.

As discussed before, by making the discretization more dense and increasing the order of the polynomial, we can make the band between lower and upper bounds more tight. Therefore, we can approximate the optimal solution within desired accuracy.

5.6 Summary

In this chapter, the duality structure of optimal H^∞ control of spatially invariant systems was characterized by computing the pre-dual and dual spaces after formulating the problem as a distance minimization. The pre-annihilator and annihilator subspaces were computed explicitly showing that an optimal distributed control exists. A dual extremal functional is also shown to exist.

Then the duality structure of the problem was characterized in terms of tensor product spaces. We showed that pre-dual and dual spaces together with the pre-annihilator and annihilator subspaces can be realized explicitly as specific tensor spaces and subspaces, respectively. The tensor space formulation lead to a solution in terms of a tensor product operator. Specifically, the optimal distributed control performance for spatially invariant systems was shown to be equal to the operator induced norm of this operator.

A discussion of a numerical solution is provided. Numerical solutions have also been given using finite variable convex programming methods based on the proposed technique.

Chapter 6

Mixed $\mathcal{L}^2/\mathcal{H}^\infty$ control problem Synthesis for Spatially Invariant Systems

It is well-known that $\mathcal{H}_\infty$ synthesis guarantees robust stability in the face of worst-case disturbance while $\mathcal{H}_2$ synthesis is more adapted to deal with transient performance. It is therefore natural to consider a mixed design framework that can integrate optimal transient performance and robustness in a single controller. This is the main motivation to develop a multi-objective design problem, i.e., the so-called mixed $\mathcal{L}^2/\mathcal{H}^\infty$ synthesis problem for spatially invariant systems. The mixed solution allows $\mathcal{L}^2/\mathcal{H}^\infty$ trade-offs to be made for a wide class of systems. In this chapter, Banach space duality theory developed in the previous chapters is used for the distributed mixed $\mathcal{L}^2/\mathcal{H}^\infty$ synthesis for spatially invariant systems.

We show how to minimize the nominal $\mathcal{H}_2$-norm performance in one channel while keeping bounds on the $\mathcal{H}_2$-norm or $\mathcal{H}_\infty$-norm performance (implying robust stability) in other channels.

6.1 Introduction

In our previous two chapters, we used the fact that spatially invariant systems can be viewed as multiplication operators from a particular Hilbert function space into itself in the Fourier domain. We have successfully posed the optimal distributed performance as a distance minimization in a general $\mathcal{L}^\infty$ space, from a vector function to a subspace with a mixed $\mathcal{L}^\infty$ and $\mathcal{H}^\infty$ space structure via a spatial-temporal Youla parametrization [99]. More recently, we have the $\mathcal{H}^2$ optimal control problem was solved via the computation of an orthogonal projection of a tensor Hilbert space onto a particular subspace [86]. The optimal $\mathcal{H}^2$ decentralized control problem was solved by computing an orthogonal projection on a class of Youla parameters with a decentralized structure. The latter uses Riesz projections after invoking a particular $\mathcal{L}^2$-basis. This chapter is a continuation of the work undertaken in previous Chapter 4. It combines the above two performance norms for spatially invariant systems.

The $\mathcal{L}^2/\mathcal{H}^\infty$ control means that the decision maker may construct a controller which can attenuate the external disturbance with mixed structure, as impulses with random input directions [61], i.e., $w_i(t) = \eta_i \delta_i(k)$, where i corresponds to the disturbance in the i-th subsystem and the disturbance w_i is finite energy in time.

In this chapter, we focus on minimizing the mixed $\mathcal{L}^2/\mathcal{H}^\infty$ norm of the spatial, temporal closed loop systems, respectively. Such a mixed norm is induced by the aforementioned disturbances $\{w_i\}$, which allows for more flexibility and accuracy of the desirable closed-loop behavior, in particular:

- The $\mathcal{H}_\infty$ performance is convenient to enforce robustness to model uncertainty and to express frequency-domain specifications such as bandwidth, low-frequency gain, and roll-off.
- The $\mathcal{H}_2$ performance is useful to handle stochastic aspects such as measurement noise and random disturbance which will be defined in Section 6.2.3.

We solve this problem by utilizing the orthogonal projection technique proposed in Chapter 4 (i.e. [86]). Once the $\mathcal{L}^2$ problem solved, the counterpart involving $\mathcal{H}^\infty$-norm could be achieved by following standard techniques in solving *model matching problem*.

The rest of the chapter is organized as follows. In Section 6.2, we introduce mathematical preliminaries for discrete spatio-temporal invariant systems. Section 6.3 provides the main result of the chapter, the optimal decentralized $\mathcal{H}^2$ control problem is solved through orthogonal projection and the $\mathcal{H}^\infty$ problem is solved following standard technique. A numerical example is provided in Section 6.4 to evaluate the proposed approach. Finally, some concluding remarks are drawn in Section **??**.

6.2 Preliminaries

6.2.1 Notation and Operator Theoretic Preliminaries

We use the following standard notation. Denote the closed (open) unit disk of the complex domain $\mathbb{C}$ by $\bar{\mathcal{D}} = \{z \in \mathbb{C},\ |z| \leq 1\}$ ($\mathcal{D} = \{z \in \mathbb{C},\ |z| < 1\}$) and unit circle by $\mathbf{T}$ or $\partial\mathcal{D}$. The closed space outside the unit disk is denoted by $\bar{\mathcal{D}}_+ = \{z \in \mathbb{C},\ |z| \geq 1\}$. The set of reals is denoted by $\mathbb{R}$ and the set of integers is denoted by $\mathbb{Z}$. The set of non-negative integers is denoted by $\mathbb{Z}^+$. Let the real double sequences $f = \{f_i(t)\}_{i=-\infty,t=0}^{i=\infty,t=\infty}$ denote the so-called spatiotemporal signals with a two-sided spatial support ($-\infty \leq i \leq \infty$) and a one-sided temporal support ($0 \leq t \leq \infty$). Moreover, let $\mathbf{T}$ denote the unit circle and $\bar{\mathcal{D}}(\mathcal{D})$ denote the closed (open) unit disc of the complex domain $\mathbb{C}$.

As is standard, $\mathcal{H}_2$ denotes the Hardy space of functions analytic in the unit disk $\mathcal{D}$, with square-summable power series. In general, for $1 \leq p < \infty$ the Hardy space H^p is defined as the space of all analytic functions f in the unit disk $\mathcal{D}$ for which the norm

$$\|f\|_p = \sup_{r<1}(\frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|^p d\theta)^{\frac{1}{p}} < \infty \tag{6.1}$$

The set $\mathcal{H}_2^{\perp}$ is the orthogonal complement of $\mathcal{H}_2$ in $\mathcal{L}_2$.

6.2.2 Discrete Spatio-Temporal Invariant Systems

Consider the standard feedback configuration of Figure 6.1, where w is the external disturbance, z is the controlled output, y is the measurement signal, and u is the control for all spatio-temporal sequences. The plant G and controller K are spatially and temporally invariant systems.

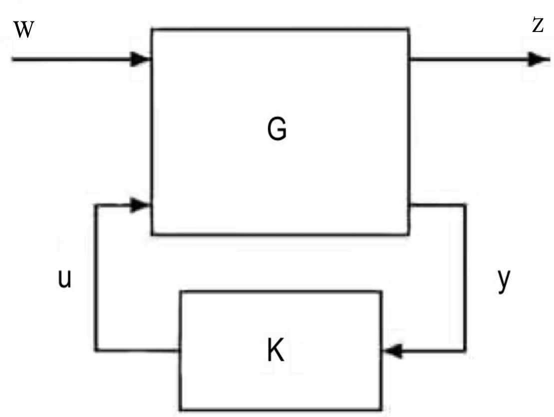


Figure 6.1: Standard Feedback Configuration under Mixed Norms

The system $G(z,\lambda)$ can be viewed as a multiplication operator on $\mathcal{L}_2(\mathbf{T},\bar{\mathcal{D}})$ where $\mathbf{T}$ is the unit circle and $\bar{\mathcal{D}}(\mathcal{D})$ is the closed (open) unit disk of the complex domain $\mathbb{C}$. If we assume that $G(z,\lambda)$ is stable, then [84, 81]:

$$\begin{aligned} G(z,\lambda):\ \mathcal{L}_2(\mathbf{T},\bar{\mathcal{D}}) &\longrightarrow \mathcal{L}_2(\mathbf{T},\bar{\mathcal{D}}), \\ u &\longrightarrow Gu = G(e^{i\theta},\lambda)u(e^{i\theta},\lambda), \end{aligned}$$

where $\theta \in [0, 2\pi),\ |\lambda| \leq 1$.

6.2.3 Disturbance

Now we introduce the disturbance signal for spatially invariant systems. Let $\hat{g}(\kappa, i) = (G \star K)(\kappa, j)$ denote the *Redheffer star product* of G and K [61]. Then let η_i represents some random noise with zero mean and covariance equal to the identity matrix,

$$y(t, i) = \sum_{j=-\infty}^{\infty} \sum_{\kappa=-\infty}^{\infty} \hat{g}(\kappa, j)\tilde{u}(t-\kappa)\eta_j \delta(i-j). \tag{6.2}$$

As mentioned in the introduction, the disturbance $w_i = \eta_j \delta(i-j)$, where $\mathbb{E}\eta^i \eta_i^T = I$ and $\mathbb{E}\eta_j \eta_i = 0$, $\forall j \neq i$.
Since $\delta(i-j)$ is a Dirac function, we get

$$y(t, i) = \sum_{\kappa=-\infty}^{\infty} \hat{g}(\kappa, i)\tilde{u}(t-\kappa)\eta_i. \tag{6.3}$$

6.2.4 Stability

Define the spatial ℓ_2-norm induced by the above disturbance can be defined as

$$\|G\|_2 = \left(\sum_{i=-\infty}^{\infty} |\hat{g}(t, i)|^2 \right)^{\frac{1}{2}}, \tag{6.4}$$

and the $\mathcal{H}^2$-norm of its transform $G(z, \lambda)$ is given by

$$\|G\|_{\mathcal{L}^2} = \left[\left(\frac{1}{2\pi}\right) \int_{\theta \in [0, 2\pi)} \left|G(e^{i\theta}, e^{iw})\right|^2 d\theta \right]^{\frac{1}{2}}. \tag{6.5}$$

By *Parseval's* theorem:

$$\|G\|_2 = \|G\|_{\mathcal{L}^2}. \tag{6.6}$$

The system G is said to be stable if its uniform norm

$$\|G\|_\infty = ess\sup_{\substack{0\le\theta<2\pi \\ \|\lambda\|\le 1}} |G(e^{i\theta}, \lambda)| < \infty. \tag{6.7}$$

We are looking for stabilizing controllers with the same structure as G_{22}. Thus we are imposing an implicit spatial-temporal structure on the controller K. In the next section, the optimal control problem for spatially invariant systems is formulated and solved.

6.3 Main Results

There exists an alternative solution applying the results from the aforementioned two sections. On the one hand, the $\mathcal{H}^2$-norm problem is considered with respect to the *spatial frequency* which could be solved using the result from the technique proposed in [86]. On the other hand, the $\mathcal{H}^\infty$ problem is considered with respect to the *temporal frequency* which can be solved using standard $\mathcal{H}^\infty$ control results.

6.3.1 Formulation

We first define the mixed norm used here

$$\psi = \inf_{Q\in\mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_{2in}^* T_1(e^{i\theta}, \lambda) - Q\|_{2,\infty}. \tag{6.8}$$

Recall the $\mathcal{H}^2$ norm of $G(z, \lambda)$ is given by

$$\|G\|_{\mathcal{L}^2} = \left[\left(\frac{1}{2\pi}\right)\int_{\theta\in[0,2\pi)} \left|G(e^{i\theta}, e^{iw})\right|^2 d\theta\right]^{\frac{1}{2}} \tag{6.9}$$

And the optimal $\mathcal{H}^\infty$ performance index can be written as:

$$\psi := \inf_{Q \in \mathcal{H}^\infty(\bar{\mathcal{D}})} \|T_{2in}^* T_1(e^{i\theta}, \lambda) - Q\|_\infty. \tag{6.10}$$

Therefore, the **mixed norm** could be defined as:

$$J_m := \|T_{2in}^* T_1(e^{i\theta}, \lambda) - Q\|_{2,\infty} = \\ \left[\frac{1}{2\pi}\int_0^{2\pi}\left(\operatorname{esssup}_{\|\lambda\|\leq 1}\left|T_{2in}^* T_1(e^{i\theta}, \lambda) - Q\right|\right)^2 d\theta\right]^{\frac{1}{2}}. \tag{6.11}$$

Based on the discussion in [86] and [98], the optimal mixed performance index can be written as:

$$\psi := \inf_{Q \in \mathcal{H}^\infty(\mathcal{L}^\infty(\bar{\mathcal{D}}))} \|T_{2in}^* T_1(e^{i\theta}, \lambda) - T_{2out} Q(e^{i\theta}, \lambda)\|_{2,\infty},$$

and

$$T_{2in}^* := T_{2in}(z^{-1}, \lambda^{-1}), \tag{6.12}$$

where as previously λ corresponds to the *temporal* transform variable, and z corresponds to the *spatial* two-sided transform.

Moreover, the optimal mixed performance is denoted as:

$$\begin{aligned} \psi &= \inf_{Q \in \mathcal{H}^\infty(\bar{\mathcal{D}})} \|T_{2in}^* T_1(e^{i\theta}, \lambda) - Q\|_{2,\infty} \\ &= \inf_{Q \in \mathcal{H}^\infty(\bar{\mathcal{D}})} J_m, \end{aligned} \tag{6.13}$$

where J_m is defined in (6.11)

6.3.2 Solution

Now we are ready to solve this problem by utilizing the orthogonal projection techniques proposed in Chapter 4. It should be remarked that, we may denote the spatial frequency $e^{i\theta}$ interchangeably with λ. We first recall some results about the $\mathcal{H}^2$ problem [86]. Since $\{z^i\}_{i=-\infty}^{\infty}$ is an orthogonal basis of $\mathcal{L}^2(\bar{\mathcal{D}})$, $T_{2in}^{*}T_1$ can be written as:

$$T_{2in}^{*}T_1(z,\lambda) = \sum_{i=-\infty}^{\infty} \tau_i(\lambda) z^i, \tag{6.14}$$

where

$$\tau_i(\lambda) \in \mathcal{L}^{\infty}(d\lambda).$$

Similarly, the outer function T_{2out} admits the following expansions w.r.t. the basis $\{z^i\}_{-\infty}^{\infty}$, such that:

$$T_{2out}(z,\lambda) = \sum_{i=-\infty}^{\infty} \chi_i(\lambda) z^i,$$

with $\chi_i(\lambda)$ stable, i.e., $\chi_i(\lambda) \in \mathcal{H}^{\infty}$.

Recall a result from [84, 108, 86] that, all stabilizing controllers K are given by

$$K = -Q(I - G_{22}Q)^{-1}, \tag{6.15}$$

with Q stable given by

$$Q(z,\lambda) = \sum_{i=-\infty}^{\infty} q_i(\lambda) z^i, \tag{6.16}$$

where

$$q_i(\lambda) = \lambda^{|i|}\tilde{q}_i(\lambda), \tag{6.17}$$

where $\tilde{q}_i$ is stable [84].

Therefore

$$T_{2out}(z,\lambda)Q(z,\lambda) = \sum_{i=-\infty}^{\infty} \xi_i(\lambda) z^i, \tag{6.18}$$

with $\xi_i(z)$ stable, and

$$\xi_i(\lambda) = \sum_j \lambda^{|j|} \tilde{q}_j(\lambda) \chi_{i-j}(\lambda).$$

Substituting (6.18) and (6.14) into the mixed performance index yields:

$$\begin{aligned} \psi^2 &= \inf_{\xi_i(\lambda)\in\mathcal{H}^\infty} \left\| \sum_{i=-\infty}^{\infty} \tau_i(\lambda) z^i - \sum_{i=-\infty}^{\infty} \xi_i(\lambda) z^i \right\|_{2,\infty}^2 \\ &= \inf_{\xi_i(\lambda)\in\mathcal{H}^\infty} \frac{1}{2\pi} \int_0^{2\pi} \operatorname{esssup}_{|\lambda|\leq 1} \Big| \sum_{i=-\infty}^{\infty} (\tau_i(\lambda) \\ &- \xi_i(\lambda)) e^{i\theta} \Big|^2 d\theta. \end{aligned} \tag{6.19}$$

Moreover, notice

$$\begin{aligned} & \| \sum_{i=-\infty}^{\infty} \tau_i(\lambda) z^i - \sum_{i=-\infty}^{\infty} \xi_i(\lambda) z^i \|_{2,\infty}^2 \\ = & \frac{1}{2\pi} \int_0^{2\pi} \operatorname{esssup}_{|\lambda|\leq 1} | \sum_{i=-\infty}^{\infty} (\tau_i(\lambda) - \xi_i(\lambda)) e^{i\theta} |^2 d\theta \\ = & \sum_{i=-\infty}^{\infty} \operatorname{esssup}_{|\lambda|\leq 1} |\tau_i(\lambda) - \xi_i(\lambda)|^2. \end{aligned} \tag{6.20}$$

Therefore, by definition, we have the mixed performance as:

$$\psi^2 = \inf_{\xi_i(\lambda)\in\mathcal{H}^\infty} \sum_{i=-\infty}^{\infty} \|\tau_i(\lambda) - \xi_i(\lambda)\|_\infty^2. \tag{6.21}$$

Remark 6.1. *Notice in (6.21), the first term $\tau_i(\lambda) \in \mathcal{L}^\infty$ and the second term $\xi_i(\lambda) \in \mathcal{H}^\infty$, therefore the optimal performance index ψ reduces to an infinite number of Nehari problems for local Youla parameters $\xi_i(\lambda)$.*
Moreover, each of the model matching problem can be solved using Hankel operators with symbols $\tau_i(\lambda) \in \mathcal{L}^\infty(d\lambda)$ by utilizing Nehari's theorem. This theorem establishes the existence of an optimal solution of an $\mathcal{L}^\infty$ minimization problem. Once the existence is established, the construction of the solution is immediate.

6.3.3 Nehari problem

Let $\varphi \in \mathcal{L}^\infty$, the multiplicative operator L_φ associated to φ on $\mathcal{L}^2$ is defined as:

$$(L_\varphi f)(z) = \varphi(z) f(z). \tag{6.22}$$

Definition 6.2. *[109] An infinite matrix is called a Hankel matrix if it has the form*

$$\Gamma = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \cdots \\ \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \cdots \\ \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \cdots \\ \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where $\alpha = \{\alpha_j\}_{j\geq 0}$ is a sequence of complex numbers. It should be noted that Hankel matrix is skew symmetric whose entries depend only on the sum of the coordinates.

Definition 6.3. *Let Γ_i be the Hankel operator associated to $\tau_i \in \mathcal{L}^\infty$, then for $i = \cdots, -1, 0, 1, \cdots,$*

$$\Gamma_i : \quad H^2 \;\to\; H^{2\perp}$$

$$\varphi \rightarrow \Gamma_i\varphi := (\mathbb{P}_-\tau_i)(\varphi), \tag{6.23}$$

where $\mathbb{P}_-$ is the *negative Riesz projection* (orthogonal projection from $\mathcal{H}^2$ onto $\mathcal{H}^{2\perp}$, clearly, $\mathbb{P}_- = I - \mathbb{P}_+$).

Then we define the operator *induced norm* of Γ as:

$$\begin{aligned} \|\Gamma_i\| : &= \sup_{\|\varphi\|_2 \leq 1} \|\Gamma_\varphi\|_2 \\ &= \|\mathbb{P}_-\tau_i\|. \end{aligned} \tag{6.24}$$

Then according to [110], we have the follwoing lemma,

Lemma 6.3.1.

$$\inf_{\xi_i(\lambda)\in\mathcal{H}^\infty} \|\tau_i(\lambda) - \xi_i(\lambda)\|_\infty = \|\mathbb{P}_-\tau_i\|. \tag{6.25}$$

Therefore, the optimal decentralized mixed performance index

$$\psi^2 = \sum_{i=-\infty}^{\infty} \|\mathbb{P}_-\tau_i\|^2. \tag{6.26}$$

If there exists a maximizing vector $\tilde{\varphi}_i$ for each Γ_i, more can be said about the optimal solution, namely,

Theorem 6.4. *If there exists a maximizing vector $\tilde{\varphi}_i \in \mathcal{H}^2$, i.e.,*

$$\|\Gamma_i\tilde{\varphi}_i\| = \|\Gamma_i\|\|\tilde{\varphi}_i\|_2, \tag{6.27}$$

then there exists a unique $\xi_i^*(\lambda) \in \mathcal{H}^\infty$, such that [110]

$$\inf_{\xi_i(\lambda)\in\mathcal{H}^\infty} \|\tau_i(\lambda) - \xi_i(\lambda)\|_\infty = \|\tau_i(\lambda) - \xi_i^*(\lambda)\|_\infty, \tag{6.28}$$

and the optimizing Youla parameter is given by:

$$\xi_i^*(\lambda) = \tau_i(\lambda) - \frac{\mathbb{P}_- \tau_i(\lambda)\tilde{\varphi}_i(\lambda)}{\tilde{\varphi}(\lambda)}, i = \cdots, -1, 0, 1, \cdots . \tag{6.29}$$

Remark 6.5. *Recall that by convention, we denoted $Q = T_{2out}Q$ by Q as* (**??**)*. Therefore to realize the optimal decentralized control law, we need one additional step following the result given in Theorem 6.4:*

$$\hat{\xi}_i^* = T_{2out}^* \xi_i^* . \tag{6.30}$$

Remark 6.6. *As discussed in [110], bounded operators may not have maximizing vectors on a Banach space. However, compact operators on Hilbert spaces do have maximizing vectors.*

Corollary 6.6.1. *[110] If Γ_i, for $i = \cdots, -1, 0, 1, \cdots$, is compact, then a maximizing vector always exists.*

In order to verify the existence and uniqueness of the $\mathcal{L}^\infty$ approximation problem, we shall establish the compactness criterion for Hankel operators.

Corollary 6.6.2. *Γ_i is compact if and only if $\tau_i(\lambda)$ can be decomposed into the following form:*

$$\tau_i(\lambda) = \hat{\tau}_i(\lambda) + h_i(\lambda), \tag{6.31}$$

where $\hat{\tau}_i$ is continuous on the unit circle, i.e., for $|\lambda| = 1$, and $h_i(\lambda) \in \mathcal{H}^\infty$.

Now we can obtain the compactness criterion for Hankel operators.

Theorem 6.7. *[109] With $\tau_i \in \mathcal{L}^\infty$, the following statements are equivalent:*

1. Γ_i *is compact for* $i = \cdots, -1, 0, 1, \cdots$.
2. $\tau_i \in \mathcal{H}^\infty + C$ *for* $i = \cdots, -1, 0, 1, \cdots$*;*

Theorem 6.7 guarantees existence and uniqueness of optimal Youla parameters.

6.3.4 Numerical algorithm

It should be noticed that only temporal frequency variable remains after applying the negative Riesz projection for each spatial frequency. In other words, we need to solve all the $\mathcal{H}^\infty$-norm problems regarding each negative spatial frequency basis.

The overall algorithm for computing the mixed norm is described in *Algorithm 1*. And one standard procedure to compute the $\mathcal{H}^\infty$-norm of Step 3 can be found in [61].

Algorithm 1 Mixed norm computation

Step 1:
    Initialize θ, λ, μ
Step 2:
    Negative projection for the optimal $\mathcal{H}^2$-norm to fix the spatial frequency θ.
Step 3:
for all Basis index $i = -\infty$ to ∞ **do**
    Optimal $\mathcal{H}^\infty$-norm μ_i with respect to the temporal frequency λ can be solved following the solution to standard *model matching problem*.
    $\mu = \mu + \mu_i$
end for
Step 4:
    Finish!

The Hankel operator has a matrix representation Γ as given in Definition 6.2:

$$\Gamma = \begin{bmatrix} \tau_{-1} & \tau_{-2} & \tau_{-3} & \tau_{-4} & \cdots \\ \tau_{-2} & \tau_{-3} & \tau_{-4} & \tau_{-5} & \cdots \\ \tau_{-3} & \tau_{-4} & \tau_{-5} & \tau_{-6} & \cdots \\ \tau_{-4} & \tau_{-5} & \tau_{-6} & \tau_{-7} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Remark 6.8. *In general, it is not trivial to solve the model matching problem, since the Hankel operators may turn out to be infinite dimensional though the compactness condition satisfied. For a special class of Hankel operators which are finite dimensional, we can apply the result for matrix operator using singular value decomposition (SVD) [61].*

Lemma 6.8.1. *Suppose matrix Γ has a SVD*

$$\begin{aligned} \Gamma &= U\Sigma V^* \\ \Sigma &= diag\{\sigma_1, \sigma_2, \cdots\} \\ U &= [u_1,\, u_2,\, \cdots] \\ V &= [v_1,\, v_2,\, \cdots] \end{aligned} \tag{6.32}$$

with $U^*U = UU^* = I$ and $V^*V = VV^* = I$. Then let σ_1 be the largest Hankel singular value, i.e., $\|\Gamma\| = \sigma_1$, and

$$(\tau_i - \xi_i^*)U = \sigma_1 V. \tag{6.33}$$

6.4 Numerical Example

In this section, we aim at applying the proposed control strategy in previous sections for a spatially invariant system given in Voulgaris et al. [84]. As discussed in Section 6.3, the proposed orthogonal projection and Hankel operator technique lead to an optimal decentralized control that minimizes the mixed $\mathcal{L}^2/\mathcal{H}^\infty$ performance for spatially invariant systems.

We consider the following spatio-temporal system which comes from the finite-difference discretization of a certain PDE [84].

$$\begin{aligned} y(i,k+1) - y(i,k) &= \frac{T}{L^2} y(i+1,k) - 2y(i,k) + y(i-1,k) \\ &- \epsilon y(i,k) + u(i,k) \end{aligned} \tag{6.34}$$

Taking the appropriate transforms one obtains the transfer function, weighting function and stabilizing controller parameterization.

We want to compute a decentralized controller for optimal H_2 attenuation of an additive disturbance on the system output with weighting function

$$W(z,\lambda) = \frac{\lambda}{1 - (\frac{1}{8}z^{-1} + \frac{1}{4} + \frac{z}{8})\lambda\,.} \tag{6.35}$$

We assume $W(z;\lambda)$ to be asymptotically stable and yield the same structure as the plant itself.

Then the mixed $\mathcal{L}^2/\mathcal{H}^\infty$ performance for this specific spatially invariant system can be stated as:

$$\inf_{Q\in\mathcal{H}^\infty} \|(1-GQ)W\|_{2,\infty} = \inf_{Q\in\mathcal{H}^\infty} \|H-UQ\|_{2,\infty}, \tag{6.36}$$

where

$$\begin{aligned} H(z,\lambda) &= \frac{\lambda}{1-r(z)\lambda}, \\ U(z,\lambda) &= \frac{T\lambda^2}{(1-\rho(z)\lambda)(1-r(z)\lambda)}, \end{aligned} \tag{6.37}$$

and

$$\begin{aligned} \rho(z) &= z^{-1}/6 + 1/3 + z/6, \\ r(z) &= z^{-1}/8 + 1/4 + z/8. \end{aligned} \tag{6.38}$$

An inner−outer factorization of $U(z,\lambda)$ yields

$$U_{in}(z,\lambda) = \lambda^2, \tag{6.39}$$

$$U_{out}(z,\lambda) = \frac{T}{(1-\rho(z)\lambda)(1-r(z)\lambda)\,.} \tag{6.40}$$

We can plug in the corresponding $T_{2in}^{*}T_1$ which comes from simple multiplication of $U_{in}^{*}(z,\lambda)$ and $H(z,\lambda)$ in this example. Hence this control problem fits in with our problem setting and our proposed control approach can be applied. We first compute the negative Riesz projection and then calculate the norm of the Hankel operators for each fixed spatial frequency as developed in *Theorem 6.4*.

It should be mentioned that, ideally, we want to have the whole set of functions $\{e^{jnw}\}_{n=-\infty}^{\infty}$ to be the orthogonal basis for L^2. However, we have to pick a finite number N as the discretization resolution which is chosen to be $200, 400, 600, 800, 1000, 1200$ and 1400. The simulation results utilizing the proposed algorithm are provided in Table 6.1. For each fixed N, σ_i's represent different singular values of the Hankel matrix Γ such that $\sigma_{max}(\sigma_{min})$ denote the maximum (minimum) Hankel singular values. Finally the mixed performance ψ is computed.

Table 6.1: Hankel norms and optimal mixed performance corresponding to different N.

N	σ_{max}	σ_{min}	ψ
200	0.6745	0.3125	0.4508
400	0.6708	0.3154	0.4519
600	0.6695	0.3164	0.4523
800	0.6688	0.3169	0.4525
1000	0.6684	0.3171	0.4527
1200	0.6681	0.3173	0.4528
1400	0.6679	0.3175	0.4529

6.5 Summary

This chapter considers the mixed $\mathcal{L}^2/\mathcal{H}^\infty$ synthesis problem for spatially invariant systems. The mixed solution provides optimal solution for minimizing a mixed $\mathcal{H}_2$

and $\mathcal{H}_\infty$ norm for a wide class of systems. In particular, the optimal decentralized $\mathcal{H}^2$ control problem is solved through orthogonal projection, and the $\mathcal{H}^\infty$ problem is solved using an infinite number of Hankel operators.

Chapter 7

Decentralized Stochastic Control: A Stochastic Maximum Principle via Malliavin Calculus Approach

Recently, another theoretical mechanism for decentralized stochastic control was proposed in [46]. A general model of decentralized stochastic control called partial history sharing information structure is presented. In this model, at each step the controller of each individual agent shares part of their observation and control history with each neighboring agent.

The optimal control problem at the coordinator level is shown to be a *partially observable Markov decision process* (POMDP). Instead of the POMDP, we consider a controlled Itô-Lévy process where the 'Markovian' property does not hold any more. It should be noted that this is a more complicated situation than the case where the standard stochastic maximum principle would fail. Therefore, we may need to apply a Malliavin calculus approach to derive a maximum principle, where the adjoint processes are explicitly expressed by the parameters and the states of the system.

When investigating this problem, we obtained two conclusions. First, the literature pertaining the Malliavin calculus is rather difficult, with proofs containing very little detail. Second, the investigations of the applications of the Malliavin calculus to the stochastic optimal control have been dominated by mathematicians, and mostly been constrained to the context of the Black-Scholes model. The aim of this chapter (together with Appendix B) is to apply results in Malliavin calculus to more realistic control problems including various hot topics in control society including Markov Decision Process, Team Decision Problems, and so on. Again, the fundamental motivation for this chapter is to provide an accessible sketch to bridge the gap between pure mathematic Malliavin calculus and practical stochastic optimal control, especially the hot topic - stochastic decentralized control problem.

7.1 Introduction

7.1.1 Markov decision process

Markov decision process (MDP) models have proven to be useful in a variety of sequential planning applications where it is crucial to account for uncertainty in the process. The partially observable MDP model (POMDP) generalizes the MDP model to allow for even more forms of uncertainty to be accounted for in the process. A POMDP is really just an MDP; we have a set of states, a set of actions, transitions and immediate rewards. The actions' effects on the state in a POMDP is exactly the same as in an MDP. The only difference is in whether or not we can observe the current state of the process. In a POMDP we add a set of observations to the model. So instead of directly observing the current state, the state gives us an observation which provides a hint about what state it is in. The observations can be probabilistic; so we need to also specify an observation function. This observation function simply tells us the probability of each observation for each state in the model. Fortunately, it turns out that simply maintaining a probability distribution over all of the states provides us with the same information as if we maintained the complete history.

7.1.2 Stochastic Maximum principle

The stochastic maximum principle is another important extension of the Pontryagin maximum principle for systems subject to randomness. In the stochastic case, there are basically different approaches based on the assumptions used to derive the (stochastic) minimum principle. Specifically, [111] uses spike variations and Neustadts variational principle, [112] uses Girsanovs measure transformation for nondegenerate controlled diffusion processes, while [113] utilizes the martingale representation to derive the adjoint equation. An excellent account of the stochastic minimum principle is found in [114] which also includes an extensive list of references.

In applications of control theory, there are many problems in the physical sciences and engineering where systems are modeled by stochastic differential equations driven by controls which are also stochastic processes with a specific information structure, such as full information or partial information. Mathematically, information structures are modeled via the minimal sigma algebra generated by the available information process. And it is this process that the controller uses to generate control actions. For full information problems in which the information structure is Markovian, one often employs Bellmans principle of optimality to construct what is known as the HJB (Hamilton-Jacob-Bellman) equation, a nonlinear PDE defined on the state space of the system under consideration. This equation describes the evolution of the value function which is used to construct the state feedback control law provided this function is at least once differentiable with respect to the state variable. This however requires solving the HJB equation which may have a viscosity solution but is not sufficiently smooth [114]. For non-Markovian controlled diffusion processes with general information structures, the HJB equation does not apply. For information structures which correspond to full information or partial information, the stochastic minimum principle is often employed.

7.1.3 Malliavin Calculus

The Malliavin calculus (also known as stochastic calculus of variation) is a differential calculus for functions (i.e. random variables) defined on a space with a Gaussian measure (usually some version of the Wiener space).

The Malliavin calculus is a method originally developed for proving smoothness of $p(t, x, y)$ in the variable y, where $p(t, x, y)$ is the transition density of a process associated to an operator with smooth coefficients [115]. It is an infinite-dimensional differential calculus, whose operators act on functionals of general Gaussian processes [116]. Initiated by Paul Malliavin [117], the theory is based on an integration by parts formula in an infinite-dimensional space. As mentioned in [118], it was first applied to study smoothness of solutions to partial differential operators. Then after that, many other applications have been developed, for example to stochastic differential equations and stochastic integrals.

7.2 Problem Formulation

If the control strategy for the future is fixed as a function of future beliefs, then the current belief is a sufficient statistic for predicting future costs under any choice of the current action.

Based on the information commonly known to all the controllers, a "**fictious**" **coordinator** was created in order to reformulate the decentralized problem as an equivalent centralized problem from the perspective of a **coordinator**. It should be remarked that the key technique to bypass the information limitation faced by the traditional centralized control technique is to involve only partial information that are shared by connected agents. Hence, an unsolvable large scale centralized problem is successfully decoupled into solvable local hierarchical coordinators.

Then, the optimal control problem at the coordinator level is shown to be a *partially observable Markov decision process* (POMDP) which can be solved by existing techniques specialized for POMDP [119, 120].

A review of algorithms to solve POMDPs can be found in [120] and references therein. The goal of MDP is to derive a mapping from states to actions, which represents the best actions to take for each state, for a given horizon length. Specifically, the coordinator first chooses a decision strategy to minimize a total cost (by solving the POMDP problem) knows the common information; then based on this supervisory control decision, the coordinator selects prescriptions that map each controllers local information to its control actions.

This assumes that the next state depends only upon the current state (and action). There are situations where the effects of an action might depend not only on the current state, but upon the last few states. The MDP model will not model these situations directly. The Markov assumption made by the MDP model is that the next state is solely determined by the current state (and current action).

As mentioned before, the main breakthrough in this mechanism is to apply the results from POMDP. However, it would be interesting to study if we could relax the "Markov" assumptions in the problem formulation. So, we could formulate a similar decentralized control problem as [46] where common information is available for the coordinator. Instead of the POMDP, we will face a controlled Itô-Lévy process since the 'Markov' property does not hold any more.

The newly developed stochastic maximum principle [121] may be recalled to solve the problem in the coordinator level. Furthermore, it could be treated as a mean-field type stochastic control problem discussed in [122]. It is promising to apply these two novel coordination algorithms since in both of the two problems, the dynamics is governed by a controlled Itô-Lévy process and the information available to the controller is possibly less than the overall information. All the system coefficients and the objective performance functional are allowed to be random, possibly non-Markovian [122].

It should be noted that this is a more complicated situation than the case where standard stochastic maximum principle would fail. Therefore, we may need to apply a Malliavin calculus approach to derive a maximum principle, where the adjoint processes are explicitly expressed by the parameters and states of the system.

7.2.1 Why Care about Malliavin Calculus?

- By relaxing the Markovian assumption from [46], we can develop a Decentralized Stochastic maximum principle via Malliavin calculus based on a common information mechanism.

  In [46], a general model of decentralized stochastic control called partial history sharing information structure is presented. In this model, at each step the controller of each individual agent shares part of their observation and control history with each neighboring agent. The optimal control problem at the coordinator level is shown to be a POMDP which can be solved by existing techniques specialized for POMDP. Instead of the POMDP, we will face a controlled Itô-Lévy process if no 'Markov' property is assumed.
- This maximum principle based on Malliavin Calculus is actually much more powerful than the traditional backward SDE method in the literature, such as in [114].
- Bass in [115] claim that

  There are two main approaches, one using the *Girsanov transformation* [123] and the other using the *Ornstein-Uhlenbeck* operator. Both are interesting and both are useful.

  Provided the above statement is true, all the derivations using *Girsanov transformation* (for example, the team games framework by Charalambos [124, 125, 126]) retain one alternative way to get proved by the Malliavin Calculus. Furthermore, this will not only proves the same result, but also relax the critical assumption "**convexity**" from the whole framework.

We recall the basic definition and properties of Malliavin Calculus in Appendix B.

7.3 Decentralized Stochastic Control

Following [46], we know that common information can be exploited to convert the decentralized optimization problem into a centralized optimization problem involving a coordinator. In this section, we briefly explain how to convert the decentralized problem into a centralized stochastic control problem (in particular, a POMDP), identify the structure of optimal control strategies, and provide a dynamic program like decomposition for the decentralized problem.

For two random variables (or random vectors) X and Ytaking values in $\mathcal{X}$ and $\mathcal{Y}$. Consider a system with two controllers. The system operates in discrete time for a horizon T. Let $X_t \in \mathcal{X}_t$ denote the state of system at time t, $U_t^i \in \mathcal{U}_t^i$ denote the control action of controller i at time t, and $\mathbf{U}_t$ denote the vector $(U_t^1, \cdots, U_t^n)$. The dynamic system with finite horizon T will evolve according to:

$$X_{t+1} = f_t(X_t, \mathbf{U}_t, W_t^0),\ t = 0, 1, 2, \cdots, T, \tag{7.1}$$

where $\{W_t^0\}_{t=1}^T$ is a sequence of independent and identically distributed (i.i.d.) random variables with probability distribution Q_W^0.

Each controller of DM^i makes its own **local observation** $Y_t^i \in \mathcal{Y}_t^i$

$$Y_t^i = h_t^i(X_t, W_t^i), \tag{7.2}$$

where $\{W_t^i, t = 0, 1, \cdots, T\}$, $i = 1, 2, \cdots T$, represents primitive random variables with known statistics.

Moreover, each controller stores a subset M_t^i of its past local observations and its past actions in a **local memory**

$$M_t^i \subset \{Y_{0:t-1}^i, U_{0:t-1}^i\}. \tag{7.3}$$

Notice that, at $t = 1$, the local memory is empty, i.e., $M_0^i = \emptyset$.
In addition to local memory, each controller has access to a **shared memory**. The contents C_t of the shared memory at time t are a subset of the past local observations and control actions of all controllers:

$$C_t \subset \{\mathbf{Y}_{1:t-1}, \mathbf{U}_{1:t-1}\}, \tag{7.4}$$

where $\mathbf{Y}_t$ and $\mathbf{U}_t$ denote the vectors $(Y_t^1, \cdots, Y_t^n)$ and $(U_t^1, \cdots, U_t^n)$, respectively.
Controller i chooses action U_t^i as a function of the total data (Y_t^i, M_t^i, C_t) available to it. Specifically, for every controller i,

$$U_t^i = g_t^i(Y_t^i, M_t^i, C_t), \tag{7.5}$$

where g_t^i is the **control law** of controller i.
The collection $\mathbf{g}^{1:n} = (\mathbf{g}^1, \cdots, \mathbf{g}^n)$ is called the **control strategy** of the system.
At time t, the system incurs a cost $\ell(X_t, \mathbf{U}_t)$. The performance of the control strategy of the system is measured by the expected total cost

$$J(\mathbf{g}^{1:n}) := \mathbb{E}^{\mathbf{g}^{1:n}} \left[\sum_{t=1}^{T} \ell(X_t, \mathbf{U}_t) \right]. \tag{7.6}$$

The **optimization problem** is to find a control strategy $\mathbf{g}^{1:n}$ for the system that minimizes the expected total cost given by (7.6).

For simplicity, assume we have **two Decision Makers** (DMs), DM^1 and DM^2. All the discussions can be easily generalized to cases with $n \geq 2$.
Since two DMs may access different information, **key difficulties** involved are [46]:

- Since costs depend on system state and DMs actions, any prediction of future costs must involve:
  - a belief on the system state;
  - some means of predicting other DMs' actions.
- Different controllers have different information, hence the beliefs formed by each DM and their predictions of future costs are not consistent.
- Even with fixed strategies (g^1, g^2) a DM can not exactly predict the other DMs action.

Therefore, to address the above challenges we aim to use common knowledge/information. First, beliefs based on common knowledge are consistent among DMs. Secondly, though DM^1 cannot know U_t^2, it knows exactly mapping from Y_t^2, M_t^2 to U_t^2 using a given realization c_t of common knowledge.

$$\begin{aligned} U_t^2 &= g_t^2(Y_t^2, M_t^2, c_t) \\ U_t^2 &= g_t^2(\cdot, \cdot, c_t) = \gamma_t^2(\cdot, \cdot) \end{aligned} \tag{7.7}$$

where γ_t^2 denotes the partial decision rule for the given realization c_t of common knowledge [46].

7.3.1 Common Knowledge 4-step Scheme

In general, to design decentralized control based on common knowledge can be divided into the following main steps [46]:

1. Introduce a problem with a fictitious coordinator
   - Coordinators beliefs are based on common knowledge
   - Coordinator selects partial decision rules (prescriptions) based on common knowledge
2. Establish equivalence between the original problem and the problem with the coordinator

3. Establish structural results/sufficient statistic and a sequential decomposition for the problem with the coordinator
4. Use the equivalence (Step 2) to conclude a structural result and a sequential decomposition for the original problem

In the following, we will explain each step one by one.

Step 1 - Fictitious coordinator

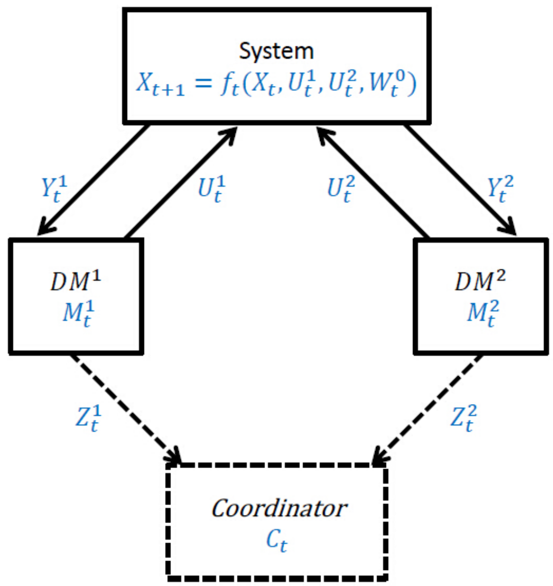


Figure 7.1: Fictitious coordinator [2]

Step 2 - Establish equivalence

This is an important step to show that the total expected cost of the coordinated system is the same as the original strategy without adding the fictious coordinator.

Step 3 - The coordinator's problem

From the coordinator's point of view, the original system and the DMs together can be viewed as the coordinator's environment. As shown in Figure 7.4, we can consider

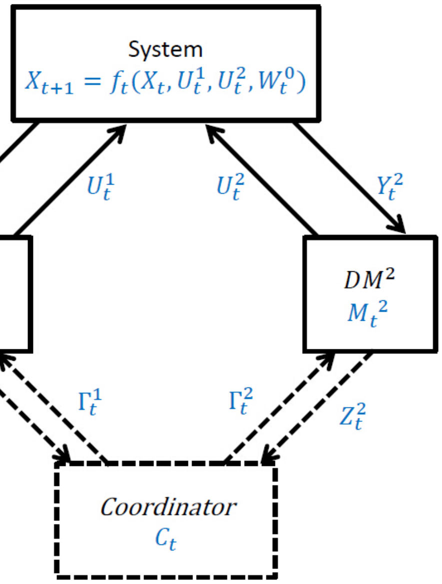


Figure 7.2: Fictitious coordinator with decision signals [2]

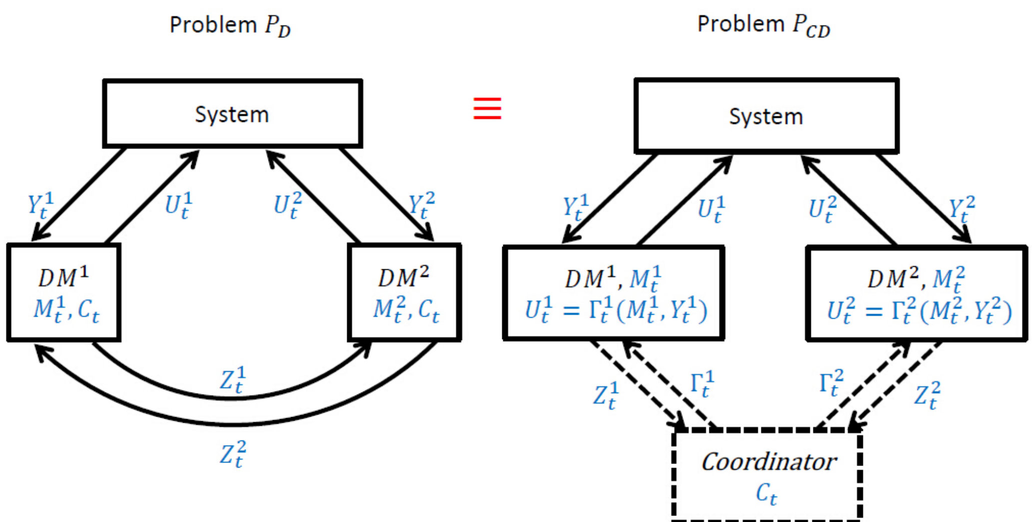


Figure 7.3: Equivalence of the two problems [2]

everything inside the red box as a "black box" to the coordinator. Then, at time t, the outputs of the "black box", Z_t^1, Z_t^2 work as observations for the coordinator; similarly, the decisions from coordinator, Γ_t^1, Γ_t^2 work as inputs to the "black box".

Let S_t denote the state of coordinator,

$$S_t = (X_t, M_t^1, M_t^2, Y_t^1, Y_t^2). \tag{7.8}$$

Therefore, the system dynamic of the coordinator follows:

$$\begin{aligned} S_{t+1} &= \hat{f}_t(S_t, \Gamma_t^1, \Gamma_t^2, \omega_t) && (7.9) \\ Z_t^i &= \hat{h}_t^i(S_t, \Gamma_t^i),\ i = 1, 2, && (7.10) \end{aligned}$$

where $\hat{f}_t$ describes state functions, $\hat{h}_t^1, \hat{h}_t^2$ represent output functions, and ω_t corresponds to system noise.

If we assume the Markov property holds here, we have

$$\mathbb{P}(S_{t+1}, Z_t^1, Z_t^2 | S_{0:t}, Z_{0:t-1}^1, Z_{0:t-1}^2, \Gamma_t^1, \Gamma_t^2) = \mathbb{P}(S_{t+1}, Z_t^1, Z_t^2 | S_t, \Gamma_t^1, \Gamma_t^2). \tag{7.11}$$

By equivalence with the original problem, there exist functions $\hat{\ell}_t$, $t = 0, 1, \cdots, T$ such that

$$\ell_t(X_t, U_t^1, U_t^2) = \hat{\ell}_t(S_t, \Gamma_t^1, \Gamma_t^2), \text{ for all}\, t = 0, 1, \cdots, T. \tag{7.12}$$

Hence the coordinator's problem (P_{CD}) can be rewritten as

$$\min_{\psi} J_{CD,T}^{\psi} := \mathbb{E}^{\psi}\{\sum_{t=0}^{T} \hat{\ell}_t(S_t, \Gamma_t^1, \Gamma_t^2)\} \tag{7.13}$$

subject to

$$\begin{aligned} S_{t+1} &= \hat{f}_t(S_t, \Gamma_t^1, \Gamma_t^2, \omega_t) \\ Z_t^i &= \hat{h}_t^i(S_t, \Gamma_t^i),\ i = 1, 2, \\ \Gamma_t^i &= \psi_t^i(C_t, \Gamma_{0:t-1}^1, \Gamma_{0:t-1}^2),\ i = 1, 2. && (7.14) \end{aligned}$$

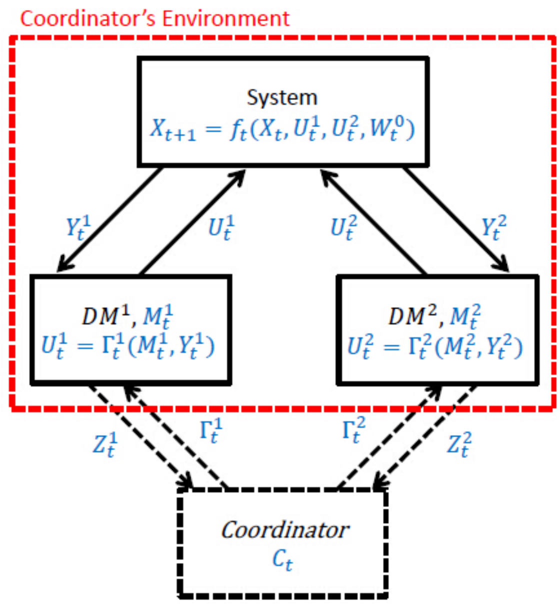


Figure 7.4: The coordinators problem [2]

Step 4 - From coordinator to the original problem

Theorem 7.1. *[46] For problem P_D there exist optimal policies of the DMs of the form*

$$U_t^i = g_t^i(Y_t^i, M_t^i, \Pi_t),\ i = 1, 2, \tag{7.15}$$

$$\Pi_t = P(X_t, M_t^1, M_t^2, Y_t^1, Y_t^2 | C_t, \Gamma^1_{0:t-1}, \Gamma^2_{0:t-1}),\ for\, t = 0, 1, \cdots, T, \tag{7.16}$$

where Π_t is the conditional distribution on X_t, Y_t, M_t given C_t.

As defined in [46], we call Π_t the *common information* state.

7.4 Main Results

Based on the scheme presented in Sec. 7.3.1, we can reformulate the decentralized problem as an equivalent centralized problem from the perspective of a coordinator.

The coordinator knows the common information and selects prescriptions that map each controllers local information to its control actions [46]. The optimal control problem for the coordinated system is further assumed to be POMDP, i.e., the state process at the coordinator follows a Markov Chain Process, which means the next state depends only on the current state.

This "Memorylessness" assumption limits the application situations dramatically, and degrade the performance since the coordinator does not utilize all the available historical information. Instead of considering the special MDP, we propose to compute the optimal decision based on a much more general stochastic maximum principle.

Remark 7.2. *Actually, we are not the first one to pursue solutions for decentralized stochastic decision systems. For instance, Charalambos and Ahmed derived team optimality and person-by person optimality conditions for distributed stochastic differential systems with different information structures in [124, 125, 126] and several following papers. In [124], given a fixed probability space* $\Big(\Omega, \mathbb{F}, \{\mathbb{F}_{0,t} : t \in [0,T]\}, \mathbb{P}\Big)$*, each subsystem* i *has its own state space* $\mathbb{R}^{n_i}$*, action space* $\mathbb{A}^i \subset \mathbb{R}^{d_i}$*, an exogenous noise space* $\mathbb{W}^i \triangleq \mathbb{R}^{m_i}$*, and an initial state* $x^i(0) = x^i_0$*. A decentralized stochastic system is formulated as:*

$$\begin{aligned} dx^i(t) =& f^i(t, x^i(t), u^i_t)dt + \sigma^i(t, x^i(t), u^i_t)dW^i(t) + \sum_{j=1, j\neq i}^{N} f^{ij}(t, x^j(t), u^j_t)dt \\ &+ \sum_{j=1, j\neq i}^{N} \sigma^{ij}(t, x^j(t), u^j_t)dW^j(t), \quad x^i(0) = x^i_0, \;\; t \in (0,T], \;\; i \in \mathbb{Z}_N. \end{aligned} \tag{7.17}$$

On the product space $(\mathbb{X}^{(N)}, \mathbb{A}^{(N)}, \mathbb{W}^{(N)})$*, where* $\mathbb{X}^{(N)} \triangleq \times_{i=1}^N \mathbb{R}^{n_i}$*,* $\mathbb{A}^{(N)} \triangleq \times_{i=1}^N \mathbb{A}^i$*,* $\mathbb{W}^{(N)} \triangleq \times_{i=1}^N \mathbb{R}^{m_i}$*. In contrast to their framework where coupled stochastic differential equations of Itô type are introduced from the beginning, we only formulate the coordinator-level control problem as a mean-field type stochastic control problem where the dynamics is governed by a controlled Itô-Lévy process. Therefore, our proposed algorithm yields a nice balance between complexity and optimality.*

The coordinator-level optimal decision problem is considered as a partially observed optimal control for forward stochastic systems which are driven by Brownian motions and an independent Poisson random measure with a feature that the cost functional is of mean-field type [122]. The key difference with [46] is that all the system coefficients and the objective performance functionals are allowed to be random, possibly non-Markovian. Malliavin calculus is employed to derive a maximum principle for the optimal control. It should be mentioned that mean-field type stochastic control problem, whose state equation is related to a kind of McKeanVlasov equation (refer to, for example, [127, 128]).

Remark 7.3. *Though we usually have multiple states in the coordinator-level control problem, we limit our discussion to the one-dimensional case for simplicity and readability (especially the notation complexity from multivariate Lévy process). Our solution can be easily generalized to the higher dimension $\mathbb{R}^N$ case.*

Coordinated Stochastic Systems

Suppose the state process $S(t) = S^{(u)}(t, \omega); t \geq 0, \omega \in \Omega$ is a controlled Itô-Lévy process in one-dimensional $\mathbb{R}$, the system dynamic of the coordinator (7.9) can be rewritten as

$$\begin{aligned} dS(t) &= b(t, S(t), u(t), \omega)dt + \sigma(t, S(t), u(t), \omega)dB_t \\ &+ \int_{\mathbb{R}_0} \theta(t, S(t^{-1}), u(t^{-1}), z, \omega)\tilde{N}(dt, dz); \qquad (7.18) \\ S(0) &= s \in \mathbb{R} \qquad (7.19) \end{aligned}$$

where $\mathbb{R}_0 = \mathbb{R} \setminus \{0\}, B(t) = B(t, \omega)$ and $\eta(t) = \eta(t, \omega)$, given by

$$\eta(t) = \int_0^t \int_{\mathbb{R}_0} z\tilde{N}(d\xi, dz);\ t \geq 0,\ \omega \in \Omega, \qquad (7.20)$$

are a 1D Brownian motion and an independent pure jump Lévy martingale, respectively, on a given filtered probability space $(\omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$.

Thus

$$\tilde{N}(dt, dz) := N(dt, dz) - \nu(dz)dt \tag{7.21}$$

Cost functional

Let $\mathcal{A} = \mathcal{A}_\epsilon$ denote a given family of controls, contained in the set of ϵ_t-adapted càdlàg controls $u(\cdot)$ such that (7.18) has a unique strong solution up to time T. Consider the system (7.18), given $u \in \mathcal{A}_\epsilon$, we define the cost functional or performance criterion by

$$J(u) = \mathbb{E}\left[\int_0^T f(t, S(t), \mathbb{E}\left[f_0(S(t))\right], u(t), \omega)dt + g\left(S(T), \mathbb{E}\left[g_0(S(T))\right], \omega\right)\right] \tag{7.22}$$

where $\mathbb{E} = \mathbb{E}_p$ denotes expectation with respect to P, $f_0 : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are given functions such that $\mathbb{E}\left[|f_0(S(t))|\right] < \infty$ for all t and $\mathbb{E}\left[|g_0(S(T))|\right] < \infty$, and $f : [0, T] \times \mathbb{R} \times \mathbb{R} \times U \times \Omega \to \mathbb{R}$ and $g : \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}$ are given $\mathcal{F}_t$-adapted processes with (7.22) finite.

Then the coordinator-level control problem is the following:

Problem 7.4. *Find minimum cost function $\Phi_\epsilon \in \mathbb{R}$ and optimal control $u^* \in \mathcal{A}_\epsilon$ (if it exists) such that*

$$\Phi_\epsilon = \inf_{u \in \mathcal{A}_\epsilon} J(u) = J(u^*) \tag{7.23}$$

Remark 7.5. *It should be mentioned that, the integration term inside (7.22) is called the running cost, while the second term represents the terminal cost. Moreover, the cost functional (7.22) involves the mean of functions of the state variable. As mentioned in [122], this extra mean term breaks the nice property - time consistency, which leads to the failure of dynamic programming approach.*

Instead of dynamic programming approach, there are attempts to solve this problem via maximum principle. For example, [129] developed a maximum principle for mean-field SDE's with the adjoint processes defined by backward SDE's (BSDE's). It is well-known that these BSDE's are difficult to solve. On the other hand, the duality involved via Malliavin derivative enables us to derive an explicit form for the adjoint process.

We follow [122] to investigate the use of Malliavin calculus to derive mean-field stochastic maximum principle.

7.4.1 Stochastic Maximum Principle

In order to obtain the solution to Problem 7.4, we use the results from [122], which rely on the following assumptions.

Remark 7.6. *We denote the states $S(t)$ in the coordinator level by $X(t)$ in the following.*

Assumption 7.7. *The following functions are all continuously differentiable (C^1) with respect to the arguments (if depending on them), where $x \in \mathbb{R}_0, x_0 \in \mathbb{R}$ and $u \in \mathcal{U}$ for each $t \in [0,T]$.*

- $b(t,x,u,\omega) : [0,T] \times \mathbb{R} \times U \times \Omega \to \mathbb{R}$
- $\sigma(t,x,u,\omega) : [0,T] \times \mathbb{R} \times U \times \Omega \to \mathbb{R}$
- $\theta(t,x,u,z,\omega) : [0,T] \times \mathbb{R} \times U \times \mathbb{R}_0 \times \Omega \to \mathbb{R}$
- $f(t,x,x_0.u,\omega) : [0,T] \times \mathbb{R} \times \mathbb{R} \times U \times \Omega \to \mathbb{R}$
- $g(x,x_0,\omega) : \mathbb{R} \times \mathbb{R}\Omega \to \mathbb{R}$
- $f_0(x_0) : \mathbb{R} \to \mathbb{R}$
- $g_0(x_0) : \mathbb{R} \to \mathbb{R}$

Assumption 7.8. *For all $t, r \in (0,T), t \leq r$, and all bounded ϵ_t-measurable random variables $\alpha = \alpha(\omega)$ the control*

$$\beta_\alpha(s) = \alpha(\omega)\chi_{[t,r]}(s);\ s \in [0,T] \tag{7.24}$$

belongs to $\mathcal{A}_\epsilon$.

Assumption 7.9. *For all $u, \beta \in \mathcal{A}_\epsilon$ with β bounded, there exists $\delta > 0$ such that*

$$u + y\beta \in \mathcal{A}_\epsilon \quad \textit{for all} \quad y \in (-\delta, \delta). \tag{7.25}$$

Furthermore, if we define

$$\begin{aligned} \tilde{f}\left(t, X(t), \mathbb{E}\left[f_0(X(t))\right], u(t)\right) := &\frac{\partial f}{\partial x}\left(t, X(t), \mathbb{E}\left[f_0(X(t))\right], u(t)\right) \\ &+ \mathbb{E}\left[\frac{\partial f}{\partial x_0}\left(t, X(t), \mathbb{E}\left[f_0(X(t))\right], u(t)\right)\right] f_0'(X(t)), \end{aligned} \tag{7.26}$$

$$\begin{aligned} \tilde{g}\left(X(t), \mathbb{E}\left[g_0(X(t))\right]\right) := &\frac{\partial g}{\partial x}\left(X(t), \mathbb{E}\left[g_0(X(t))\right]\right) \\ &+ \mathbb{E}\left[\frac{\partial g}{\partial x_0}\left(X(T), \mathbb{E}\left[g_0(X(T))\right]\right)\right] g_0'(X(T)), \end{aligned} \tag{7.27}$$

then, the family

$$\begin{aligned} \Big\{ &\tilde{f}\left(t, X^{u+y\beta}(t), \mathbb{E}\left[X^{u+y\beta}(t)\right], u(t) + y\beta(t)\right) \frac{d}{dy} X^{u+y\beta}(t) \\ &\frac{\partial f}{\partial u}\left(t, X^{u+y\beta}(t), u(t) + y\beta(t)\right)\beta(t)\Big\}_{y \in (-\delta, \delta)} \end{aligned} \tag{7.28}$$

is $\lambda \times P$-uniformly integrable and the family

$$\left\{ \tilde{g}\left(X^{u+y\beta}(T), \mathbb{E}\left[g_0\left(X^{u+y\beta}(T)\right)\right]\right) \frac{d}{dy} X^{u+y\beta}(T) \right\}_{y \in (-\delta, \delta)} \tag{7.29}$$

is $P-$uniformly integrable.

Assumption 7.10. *For all $u, \beta \in \mathcal{A}_\epsilon$ with β bounded, the process $Y(t) = Y^{(\beta)}(t) = \frac{d}{dy} X^{(u+y\beta)}(t)|_{y=0}$ exists and satisfies the equation*

$$
\begin{aligned}
dY(t) = Y(t^-) \Big[& \frac{\partial b}{\partial x}(t, X(t), u(t))\, dt + \frac{\partial \sigma}{\partial x}(t, X(t), u(t))\, dB(t) \\
& + \int_{\mathbb{R}_0} \frac{\partial \theta}{\partial x}\left(t, X(t^-), u(t^-), z\right) \tilde{N}(dt, dz) \Big] \\
& + \beta(t^-) \Big[\frac{\partial b}{\partial u}(t, X(t), u(t))\, dt + \frac{\partial \sigma}{\partial u}(t, X(t), u(t))\, dB(t) \\
& + \int_{\mathbb{R}_0} \frac{\partial \theta}{\partial u}\left(t, X(t^-), u(t^-), z\right) \tilde{N}(dt, dz) \Big]
\end{aligned}
\tag{7.30}
$$

Assumption 7.11. *For all $u \in \mathcal{A}_\epsilon$, with definitions (7.26) and (7.27), the following processes all exist for $0 \leq t \leq s \leq T$, $z \in \mathbb{R}_0$:*

- $K(t) := \tilde{g}\left(X(T), \mathbb{E}\left[g_0(X(T))\right]\right) + \int_t^T \tilde{f}\left(s, X(s), \mathbb{E}\left[f_0(X(s))\right], u(s)\right) ds$
- $D_t K(t) := D_t \tilde{g}\left(X(T), \mathbb{E}\left[g_0(X(T))\right]\right) + \int_t^T D_t \tilde{f}\left(s, X(s), \mathbb{E}\left[f_0(X(s))\right], u(s)\right) ds$
- $D_{t,z} K(t) := D_{t,z} \tilde{g}\left(g_0(X(T))\right) + \int_t^T D_{t,z} \tilde{f}\left(s, X(s), \mathbb{E}\left[f_0(X(s))\right], u(s)\right) ds$
- $$
\begin{aligned}
H_0(s, x, u) := & K(s) b(s, x, u) + D_s K(s) \sigma(s, x, u) \\
& + \int_{\mathbb{R}_0} D_{s,z} K(s) \theta(s, x, u, z) \nu(dz)
\end{aligned}
$$
- $$
\begin{aligned}
G(t, s) := \exp\Bigg(& \int_t^s \left\{ \frac{\partial b}{\partial x}(r, X(r), u(r), \omega) - \frac{1}{2} \left(\frac{\partial \sigma}{\partial x}\right)^2 (r, X(r), u(r), \omega) \right\} dr \\
& \int_t^s \frac{\partial \sigma}{\partial x}(r, X(r), u(r), \omega) dB(r) \\
& + \int_t^s \int_{\mathbb{R}_0} \left\{ ln \left(1 + \frac{\partial \theta}{\partial x}(r, X(r^-), u(r^-), z, \omega) \right) \right. \\
& \left. - \frac{\partial \theta}{\partial x}(r, X(r), u(r), z, \omega) \right\} \nu(dz) dr \\
& + \int_t^s \int_{\mathbb{R}_0} ln \left(1 + \frac{\partial \theta}{\partial x}(r, X(r^-), u(r^-), z, \omega) \right) \tilde{N}(dr, dz) \Bigg),
\end{aligned}
$$

- $p(t) := K(t) + \int_t^T \frac{\partial H_0}{\partial x}(s, X(s), u(s))G(t,s)ds$
- $q(t) := D_t p(t),$ *and* $r(t,z) := D_{t,z}p(t)$

To derive the stochastic maximum principle, we first define a *general stochastic Hamiltonian.*

Definition 7.12. *The general stochastic Hamiltonian is the process*

$$H(t,x,u,\omega) : [0,T] \times \mathbb{R} \times U \times \Omega \to \mathbb{R} \tag{7.31}$$

defined by

$$\begin{aligned} H(t,x,u,\omega) &= f(t,x,\mathbb{E}\left[f_0(X(t))\right],u,\omega) + p(t)b(t,x,u,\omega) + q(t)\sigma(t,x,u,\omega) \\ &+ \int_{\mathbb{R}_0}^{max} r(t,z)\theta(t,x,u,z,\omega)\nu(dz). \end{aligned} \tag{7.32}$$

Theorem 7.13. *[122] $u^* \in \mathcal{A}_\epsilon$ is a critical point for $J(u)$, in the sense that*

$$\frac{d}{dy}J(u^* + y\beta)|_{y=0} = 0 \, \textit{for all bounded} \, \beta \in \mathcal{A}_\epsilon. \tag{7.33}$$

if and only if the following equation holds

$$\mathbb{E}\left[\frac{\partial \hat{H}}{\partial u}\left(t, \hat{X}, \hat{u}\right)|\epsilon_t\right] = 0 \quad \textit{for almost all} \quad t, \omega, \tag{7.34}$$

where

$$\begin{aligned} \hat{X}(t) &= X^{(\hat{u})}(t), \\ \hat{H}(t,\hat{X}(t),u) &= f(t,\hat{X},\mathbb{E}\left[f_0(\hat{X}(t))\right],u) + \hat{p}(t)b(t,\hat{X}(t),u) + \hat{q}(t)\sigma(t,\hat{X}(t),u) \\ &+ \int_{\mathbb{R}_0} \hat{r}(t,z)\theta(t,\hat{X}(t),u,z)\nu(dz), \end{aligned} \tag{7.35}$$

with

$$\begin{aligned}
\hat{p}(t) &= \hat{K}(t), + \int_t^T \frac{\partial H_0}{\partial x}(s, \hat{X}(s), \hat{u}(s))\hat{G}(t,s)ds, \\
\hat{q} &:= D_t \hat{p}(t), \textit{ and} \qquad (7.36) \\
\hat{r}(t,z) &:= D_{t,z}\hat{p}(t), \qquad (7.37)
\end{aligned}$$

where

$$\begin{aligned}
\hat{G}(t,s) &= exp\Bigg(\int_t^s \left\{\frac{\partial b}{\partial x}(r, \hat{X}(r), u(r), \omega) - \frac{1}{2}\left(\frac{\partial \sigma}{\partial x}\right)^2 (r, \hat{X}(r), u(r), \omega)\right\} dr \\
&+ \int_t^s \frac{\partial \sigma}{\partial x}(r, \hat{X}(r), u(r), \omega) dB(r) \\
&+ \int_t^s \int_{\mathbb{R}_0} \left\{ ln\left(1 + \frac{\partial \theta}{\partial x}(r, \hat{X}(r), u(r), z, \omega)\right) - \frac{\partial \theta}{\partial x}(r, \hat{X}(r), u(r), z, \omega)\right\}\nu(dz)dr \\
&+ \int_t^s \int_{\mathbb{R}_0} ln\left(1 + \frac{\partial \theta}{\partial x}(r, \hat{X}(r^-), u(r^-), z, \omega)\right)\tilde{N}(dr, dz)\Bigg) \qquad (7.38)
\end{aligned}$$

and

$$\hat{K}(t) = K^{\hat{u}}(t) = \tilde{g}(\hat{X}(T), \mathbb{E}\left[f_0(\hat{X}(T))\right]) + \int_t^T \tilde{f}(s, \hat{X}(s), \mathbb{E}\left[f_0(\hat{X}(t))\right], \hat{u}(s))ds.$$

Proof. Detailed proof can be found in [122]. □□□

It should be mentioned that, the power technique from Malliavin Calculus, integration by parts (duality formula), were used several times in the proof. To emphasis this nice property, we recall the formula although we have it in the appendix.

Definition 7.14. *Integration by parts*
Suppose $u(t)$ is $\mathcal{F}_t-$adapted with $\mathbb{E}\left[\int_0^T u^2(t)dt\right] < \infty$ and let $F \in \mathcal{D}_{1,2}$ (Appendix B.5).
Then

$$\mathbb{E}\left[F\int_0^T u(t)dB(t)\right] = \mathbb{E}\left[\int_0^T u(t)D_t F dt\right] \qquad (7.39)$$

Similarly, a Malliavin derivative in the pure jump martingale case together with the duality formula for Lévy process are provided in [122].

It's also worth mentioning that, *Girsanov transformation* [123] was recalled several times inside the proof as we can see from the expressions of the last two bullets in Assumption 7.11 and (7.36)-(7.37).

7.5 Applications of the Generalized Model

As discussed in [46], many decentralized stochastic control problems can be solved by the proposed common information approach. Depend on information sharing structure, these problems can be divided into three main categories:

1. Controllers with identical information
2. Coupled subsystems with control sharing information structure
3. Broadcast information structure

Controllers with identical information

Roughly speaking, the first category represents the case when all controllers only make the common observation (i.e., they share identical information without any local observation or local memory). This is not an interesting scenario except explaining traditional centralized control is a special case our general framework.

Broadcast information structure

The third category works by assigning one lead/central node which keeps broadcasting without any local observations. Mathematically, the detailed information flow is given in Figure 7.5 (assume node 1 is the *central node*, and all the other nodes are called *peripheral nodes* [46]).

State update for central node: $X_{t+1}^1 = f_t^1(X_t^1, U_t^1, N_t^1)$

State update for peripheral nodes ($i \geq 2$): $X_{t+1}^i = f_t^i(X_t^i, X_t^1, U_t^i, U_t^1, N_t^i)$

Common observation update: $Y_t^{com} = X_t^1$

Local observation for peripheral nodes ($i \geq 2$): $Y_t^i = X_t^i$
Central node keeps broadcasting but **no** local observation!

Shared memory $C_t = Y_{1:t-1}^{com} = X_{1:t-1}^1$

Figure 7.5: Information flow for Broadcasting information structure

State of the system at time t is: $X_t = (X_t^1, X_t^2, \cdots, X_t^n, X_t^*)$,
where X_t^i, $i = 1, \cdots, n$ corresponds to the
lcoal state of subsystem i, and X_t^* is a global state.

State update for global state: $X_{t+1}^* = f_t^*(X_t^*, U_t, N_t^0)$;
State update for local state: $X_{t+1}^i = f_t^i(X_t^i, X_t^*, U_t, N_t^i)$.
N_t^0 and N_t^i are mutually independent i.i.d. noise processes.

Common observation: $Y_t^{com} = X_t^*$;
Local observation: $Y_t^i = X_t^i$.

Shared memory $C_t = \{X_{1:t-1}^*, U_{1:t-1}\}$.

Controller i sends $Z_t^i = \{X_t^*, U_t^i\}$ to the shared memory .

Figure 7.6: Coupled subsystems with control sharing information structure

Coupled subsystems with control sharing information structure

The flow chart of this information sharing structure is given in Figure 7.6, which relates to a huge number of real applications for distributed control. To name a few, the distributed control problems for multi-zone buildings/connected and automated vehicles both fall into this category.

7.5.1 A toy example in multi-zone buildings

If the Heating, Ventilating, and Air Conditioning (HVAC) unit is not over-designed to condition the whole building area, there exists a tradeoff between the comfort of the buildings occupants and the available heating/cooling power at critical load hours.

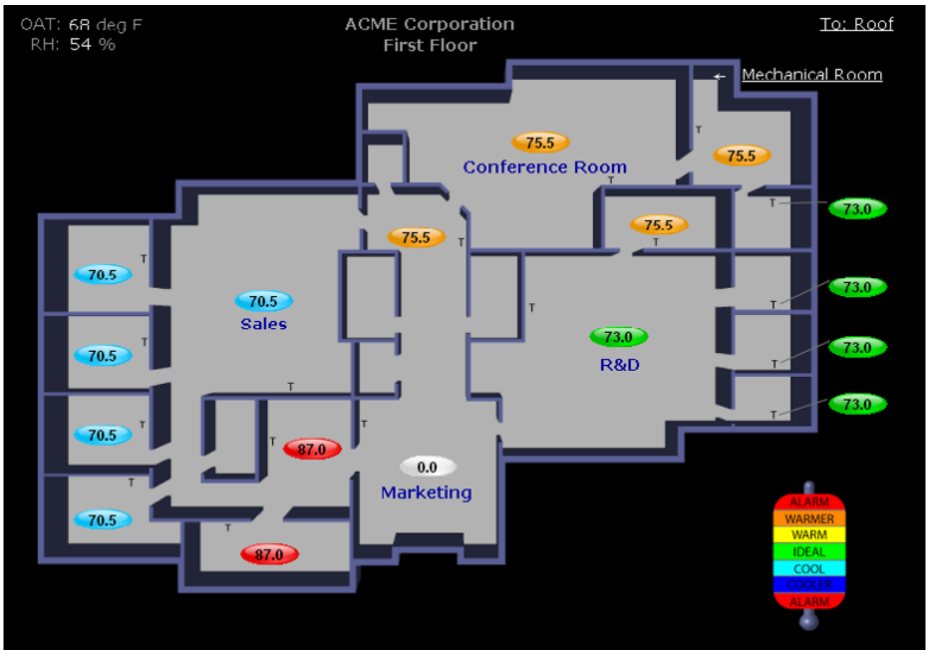


Figure 7.7: Floor plan of one sample Carrier commercial HVAC system

As we can see from Figure 7.7, it is impossible to reach the required comfort temperature despite the fact that all the available power is used to cool the rooms. However, the controller aims at minimizing the variance of the error across the rooms. Therefore it is desired to dispatch cooling/heating resources among different zones in a optimal coordinated way.

Using the proposed framework, we consider each zone (marked with different colors for the temperatures) as a agent. Each zone is assumed to share its temperature and control actions with its direct neighboring zones by sending the shared information to a fictious coordinator.

Then, the main task of the coordinator is to optimally dispatch cooling/heating resources (summation of control actions from connected agents under this specific coordinator) to each zone in order to minimize the differences across zones. It should be mentioned that, the coordinator only send prescription signals Γ_i to each agent instead of directly controlling them in this mechanism. In this specific example, Γ_i will be interpreted as *temperature setting points* for each zone. Once the temperature setting points for each zone is assigned, the local controller would take charge to decide whether turn on/off the local AC unit. The intuition of this method is pretty clear since cooling resources are more likely to be dispatched to the hottest zones first in order to reach a mean-field objective.

7.6 Summary

This chapter proposes to solve the Non-Markovian decentralized stochastic control problem by stochastic maximum principle via Malliavin Calculus. It is promising to obtain a complete solution to this problem since neither '**Markovian**' nor '**Convexity**' are required in the framework. Some necessary backgrounds for Malliavin calculus and Stochastic Maximum Principle are provided either in this chapter or Appendix B. Basic ideas about *Polynomial Chaos* are given in Appendix B.9.

Chapter 8

Conclusions and Future Directions

8.1 Summary

In this dissertation, we studied the general area of complex networked systems that consist of interconnected components and usually operate in uncertain environments and with incomplete information. We first developed unifying stochastic control approach for improving energy efficiency while meeting probabilistic constraints. We then proposed optimal distributed control for spatially invariant systems.

In order to address the energy optimization problems in building systems and virtualized web servers, we have developed the constrained quadratic control of room temperature on a dynamic building climate model. We introduced a quadratic cost function in terms of temperature errors and control inputs, which were subject to several constraints on the room temperature and control input. In particular, we only considered the case where we assume that the disturbance is *Gaussian* and the problem was formulated to minimize the expected cost subject to a linear constraint on control input and a *probabilistic constraint* on the state. We have also proposed an efficient algorithm to reduce the probabilistic constraint to a hard constraint on the control input exactly [47]. The problems are formulated as semidefinite optimization problems which may be solved through semidefinite programming (SDP) for the optimal solutions efficiently.

Then, we have studied the problem of optimal distributed control of spatially invariant systems. We have developed an input-output framework for problems of this class. Spatially invariant systems were viewed as multiplication operators from a particular Hilbert function space into itself in the Fourier domain. Optimal performance was posed as a distance minimization in a general $\mathcal{L}^\infty$ space from a vector function to a subspace with a mixed $\mathcal{L}^\infty$ and $\mathcal{H}^\infty$ space structure. Moreover, the H^2 optimal control problem was solved via the computation of an orthogonal projection of a tensor Hilbert space onto a particular subspace. The optimal H^2 decentralized control problem was also solved by computing an orthogonal projection on a class of Youla parameters with a decentralized structure. Then we investigated minimizing the mixed $\mathcal{L}^2/\mathcal{H}^\infty$ norm of the spatial, temporal closed loop systems, respectively. Such a mixed norm was induced by the aforementioned disturbances $\{w_i\}$, which allowed for more flexible and accurate specification of the desirable closed-loop behavior We have obtained solution to this problem by utilizing the orthogonal projection techniques.

Finally, we turn our attention to a stochastic version of the distributed control problem. Instead of using the solution from Markov Decision Process, we consider a controlled Itô-Lévy process where the 'Markovian' property does not hold any more. It should be noted that this is a more complicated situation than the case where standard stochastic maximum principle would fail. Therefore, we need to apply a Malliavin calculus approach to derive a maximum principle, where the adjoint processes are explicitly expressed by the parameters and the states of the system.

8.2 Directions for Future Research

So far, the results of our exploration on simple-structured optimal distributed controllers are encouraging. These results, however, are only form the tip of the iceberg in distributed control, and much remains to be done.

The following problems need to be further considered and addressed in the future research work, which are potentially the building blocks of a systematic approach to distributed (decentralized) stochastic control.

- **A stochastic approach to distributed building climate control**

  In our previous work, it is assumed that all the probabilistic constraints and stochastic disturbance are only associated with states which are three different temperatures in the model. A more practical consideration is towards the generalization of control scheme to the case of whole buildings, which leads to increased complexity for both the models and control algorithms. Given a deployed building with multi-zones, a challenging issue is how to schedule HVAC controllers so that the total run cost the system can be minimized or aggregating HVAC load for ancillary services? In case the state space model cannot provide accurate characteristics for the model when only on/off control actions are allowed for the HVAC, how to solve the stochastic control problem using SDE models with regime switching? Mean-variance regime switching technique has not received much attention in the control society and I would like to apply this innovative solution to improve many aspects of building operation.

- **Extension to distributed $\mathcal{H}^\infty$ control for spatially invariant systems**

  The $\mathcal{H}^\infty$ control problem we solved in Chapter 5 is actually a centralized performance. It is interesting to investigate how to generalize the solution to a decentralized structure as for the $\mathcal{H}^2$ problem.

- **Application to true physical interconnected systems**

  As discussed in [46] and also in Section 7.5, this common information approach unifies various ad-hoc approaches taken in the literature, it is possible to compare our proposed technique with existing literature. More importantly, if we can compute the exact building example proposed in Section 7.5.1 using both our technique or the MDP approach [46], we will be able justify how "memory" affect the performance at the coordinator level. Furthermore, we can also investigate the computation cost since solving POMDP is known to be computationally costly, while our technique provides an explicit solution which can be computed much easier. This mean-field stochastic maximum principle for Itô-Lévy process using Malliavin calculus may work as a powerful approach to solve various stochastic control problems in power systems, especially when uncertainty and jumps are involved.

Bibliography

[1] J. M. Fowler and R. D'Andrea, "A formation flight experiment for research in control of interconnected systems." xv, 45

[2] D. Teneketzis, "On the role of common knowledge in decision-making over networks," IMA Workshop on Distributed Control and Decision Making over Networks, 2015. xvi, 124, 125, 127

[3] Energy Information Administration, "Commercial buildings energy consumption survey 2009," Tech. Rep., 2009. 2, 3, 14

[4] K. Law, K. Dauber, and X. Pan, "Energy impact of commercial building controls and performance diagnostics: Market characterization, energy impact of building faults and energy savings potential," Technical Report TIAX Reference, Tech. Rep., 2005. 2, 14

[5] S. Katipamula, R. M. Underhill, J. K. Goddard, D. Taasevigen, M. Piette, J. Granderson, R. E. Brown, S. M. Lanzisera, and T. Kuruganti, "Small-and medium-sized commercial building monitoring and controls needs: A scoping study," *Pacific Northwest National Laboratory (PNNL), Richland, WA (US), Tech. Rep*, 2012. 2, 14

[6] B. Bamieh, F. Paganini, and M. A. Dahleh, "Distributed control of spatially invariant systems," *IEEE Trans. Autom. Contr.*, vol. 47, no. 7, pp. 1091–1107, 2002. 3, 5, 6, 49, 50, 69, 70

[7] N. Motee and A. Jadbabaie, "Optimal control of spatially distributed systems," *IEEE Trans. Autom. Contr.*, vol. 53, no. 7, pp. 1616–1629, 2008. 3, 5, 49, 69, 70

[8] C. Langbort, R. S. Chandra, and R. D'Andrea, "Distributed control design for systems interconnected over an arbitrary graph," *IEEE Trans. Autom. Contr.*, vol. 49, no. 9, pp. 1502–1519, 2004. 3, 6, 7, 49, 69, 70

[9] R. D'Andrea and G. E. Dullerud, "Distributed control design for spatially interconnected systems," *IEEE Trans. Autom. Contr.*, vol. 48, no. 9, pp. 1478–1495, 2003. 3, 7, 49, 69, 70

[10] G. E. Dullerud and R. D'Andrea, "Distributed control of heterogeneous systems," *IEEE Trans. Autom. Contr.*, vol. 49, no. 12, pp. 2113–2128, 2004. 3, 6, 7, 49, 69, 70

[11] S. Zhuravlev, S. Blagodurov, and A. Fedorova, "Addressing shared resource contention in multicore processors via scheduling," in *ACM SIGARCH Computer Architecture News*, vol. 38, no. 1. ACM, 2010, pp. 129–142. 4, 16

[12] R. Singh, U. Sharma, E. Cecchet, and P. Shenoy, "Autonomic mix-aware provisioning for non-stationary data center workloads," in *Proceedings of the 7th international conference on Autonomic computing*. ACM, 2010, pp. 21–30. 4, 16

[13] A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt, "Controllability of multi-agent systems from a graph-theoretic perspective," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 162–186, 2009. 4, 69

[14] N. Motee and A. Jadbabaie, "Distributed multi-parametric quadratic programming," *IEEE Trans. Autom. Contr.*, vol. 54, no. 10, pp. 2279–2289, 2009. 4, 69

[15] J. Cortés, S. Martínez, and F. Bullo, "Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions," *IEEE Trans. Autom. Contr.*, vol. 51, no. 8, pp. 1289–1298, 2006. 4, 69

[16] H. G. Tanner, A. Jadbabaie, and G. J. Pappas, "Flocking in fixed and switching networks," *IEEE Trans. Autom. Contr.*, vol. 52, no. 5, pp. 863–868, 2007. 4, 69

[17] N. Chopra and M. W. Spong, "On exponential synchronization of kuramoto oscillators," *IEEE Trans. Autom. Contr.*, vol. 54, no. 2, pp. 353–357, 2009. 4, 69

[18] H. Bai, M. Arcak, and J. Wen, *Cooperative control design: a systematic, passivity-based approach.* Springer, 2011, vol. 89. 4, 69

[19] W. Feng, A. J. Salgado, C. Wang, and S. M. Wise, "Preconditioned Steepest Descent Methods for some Nonlinear Elliptic Equations Involving p-Laplacian Terms," *ArXiv preprint arXiv:1607.01475*, 2016. 4

[20] P. Wieland, J.-S. Kim, and F. Allgöwer, "On topology and dynamics of consensus among linear high-order agents," *International Journal of Systems Science*, vol. 42, no. 10, pp. 1831–1842, 2011. 4, 69

[21] S. E. Tuna, "Conditions for synchronizability in arrays of coupled linear systems," *IEEE Trans. Autom. Contr.*, vol. 54, no. 10, pp. 2416–2420, 2009. 4, 69

[22] A. Jadbabaie, N. Motee, and M. Barahona, "On the stability of the kuramoto model of coupled nonlinear oscillators," in *Proc. IEEE Conf. on Decision and Control*, vol. 5, 2004, pp. 4296–4301. 4, 69

[23] F. Dorfler, M. Chertkov, and F. Bullo, "Synchronization assessment in power networks and coupled oscillators," in *Proc. IEEE Conf. on Decision and Control*, 2012, pp. 4998–5003. 4, 69

[24] W. Ren, "Synchronization of coupled harmonic oscillators with local interaction," *Automatica*, vol. 44, no. 12, pp. 3195–3200, 2008. 4, 69

[25] M. Fardad, F. Lin, and M. Jovanovic, "Design of optimal sparse interconnection graphs for synchronization of oscillator networks," *IEEE Trans. Autom. Contr.*, vol. PP, no. 99, pp. 1–1, 2014. 4, 44, 69

[26] M. Gwerder and J. Tödtli, "Predictive control for integrated room automation," in *8th REHVA World Congress Clima*, 2005. 5, 14, 22, 24, 25

[27] A. Sarwar, P. G. Voulgaris, and S. M. Salapaka, "On ℓ_∞ and ℓ_2 robustness of spatially invariant systems," *International Journal of Robust and Nonlinear Control*, vol. 20, no. 6, pp. 607–622, 2010. 5, 38

[28] Butkovskiĭ, *Distributed control systems.* 5, 38

[29] J. Lunze, "Synchronization of heterogeneous agents," *IEEE Trans. Autom. Contr.*, vol. 57, no. 11, pp. 2885–2890, 2012. 6, 7, 70

[30] L. Alvergue, A. Pandey, G. Gu, and X. Chen, "Output consensus control for heterogeneous multi-agent systems," in *Proc. IEEE Conf. on Decision and Control*, vol. 1, 2013, pp. 1502–1507. 6, 7, 70

[31] B. L. Jones and E. C. Kerrigan, "When is the discretization of a spatially distributed system good enough for control?" *Automatica*, vol. 46, no. 9, pp. 1462–1468, 2010. 6, 50

[32] R. F. Curtain, "Comments on" on optimal control of spatially distributed systems"," *IEEE Trans. Autom. Contr.*, vol. 54, no. 6, p. 1423, 2009. 6, 49, 50, 70

[33] B. S. Institution, *Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality,*

thermal environment, lighting and acoustics. European committee for Standardization, 2007. 6, 15

[34] A. Parisio, D. Varagnolo, D. Risberg, G. Pattarello, M. Molinari, and K. H. Johansson, "Randomized model predictive control for hvac systems," in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings.* ACM, 2013, pp. 1–8. 6, 15

[35] F. Oldewurtel, A. Parisio, C. N. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth, "Energy efficient building climate control using stochastic model predictive control and weather predictions," in *American control conference (ACC), 2010.* IEEE, 2010, pp. 5100–5105. 6, 15, 18, 19, 22, 24, 25

[36] F. Oldewurtel, C. N. Jones, and M. Morari, "A tractable approximation of chance constrained stochastic mpc based on affine disturbance feedback," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on.* IEEE, 2008, pp. 4731–4736. 6, 15, 18, 19

[37] Y. Wang, X. Wang, M. Chen, and X. Zhu, "Partic: Power-aware response time control for virtualized web servers," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 22, no. 2, pp. 323–336, 2011. 7, 31

[38] X. Wang and Y. Wang, "Coordinating power control and performance management for virtualized server clusters," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 22, no. 2, pp. 245–259, 2011. 7

[39] D. Kusic, J. O. Kephart, J. E. Hanson, N. Kandasamy, and G. Jiang, "Power and performance management of virtualized computing environments via lookahead control," *Cluster computing*, vol. 12, no. 1, pp. 1–15, 2009. 7

[40] R. Nathuji, P. England, P. Sharma, and A. Singh, "Feedback driven qos-aware power budgeting for virtualized servers," in *Fourth International Workshop*

on Feedback Control Implementation and Design in Computing Systems and Networks (FeBID), 2009. 7

[41] O. Demir and J. Lunze, "A decomposition approach to decentralized and distributed control of spatially interconnected systems," in *Proc. of the 18th IFAC World Congress, Milano, Italy*, 2011. 7, 49, 70

[42] R. D'Andrea, "A linear matrix inequality approach to decentralized control of distributed parameter systems," in *Proc. IEEE American Control Conf.*, vol. 3, 1998, pp. 1350–1354. 7, 70

[43] L. Scardovi and R. Sepulchre, "Synchronization in networks of identical linear systems," *Automatica*, vol. 45, no. 11, pp. 2557–2562, 2009. 7, 70

[44] Z. Li, Z. Duan, G. Chen, and L. Huang, "Consensus of multiagent systems and synchronization of complex networks: a unified viewpoint," *IEEE Trans. Circuits and Systems I: Regular Papers,*, vol. 57, no. 1, pp. 213–224, 2010. 7, 49, 70

[45] J. H. Seo, H. Shim, and J. Back, "Consensus of high-order linear systems using dynamic output feedback compensator: low gain approach," *Automatica*, vol. 45, no. 11, pp. 2659–2664, 2009. 7, 49, 70

[46] A. Nayyar, A. Mahajan, and D. Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," *Automatic Control, IEEE Transactions on*, vol. 58, no. 7, pp. 1644–1658, 2013. 8, 115, 119, 120, 121, 122, 123, 127, 128, 129, 136, 142, 143

[47] D. Bertsimas and D. B. Brown, "Constrained stochastic lqc: a tractable approach," *Automatic Control, IEEE Transactions on*, vol. 52, no. 10, pp. 1826–1841, 2007. 9, 17, 18, 19, 20, 21, 22, 140

[48] X. Shi, J. Dong, S. M. Djouadi, Y. Wang, X. Ma, and Y. Feng, "Power-efficient resource management for co-located virtualized web servers: A

stochastic control approach," in *Green Computing Conference (IGCC), 2014 International*. IEEE, 2014, pp. 1–9. 13

[49] X. Shi, J. Dong, S. M. Djouadi, Y. Feng, X. Ma, and Y. Wang, "Papmsc: Power-aware performance management approach for virtualized web servers via stochastic control," *Journal of Grid Computing*, pp. 1–21, 2015. 13

[50] X. Ma, J. Dong, S. M. Djouadi, J. J. Nutaro, and T. Kuruganti, "Stochastic control of energy efficient buildings: A semidefinite programming approach," in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2015, pp. 780–785. 13

[51] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica *et al.*, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010. 14

[52] X. Shi, J. Dong, S. M. Djouadi, Y. Wang, X. Ma, and Y. Feng, "Power-efficient resource management for co-located virtualized web servers: A stochastic control approach," in *Green Computing Conference (IGCC), 2014 International*. IEEE, 2014, pp. 1–9. 17

[53] Z. Jin and C. Huang, "Dfig voltage control based on dynamically adjusted control gains," *Journal of Power and Energy Engineering*, vol. 2, no. 08, p. 45, 2014. 17

[54] C. Huang, F. Li, T. Ding, Z. Jin, and X. Ma, "Second-order cone programming-based optimal control strategy for wind energy conversion systems over complete operating regions," *Sustainable Energy, IEEE Transactions on*, vol. 6, no. 1, pp. 263–271, 2015. 17

[55] C. Huang, F. Li, and Z. Jin, "Maximum power point tracking strategy for large-scale wind generation systems considering wind turbine dynamics," *Industrial Electronics, IEEE Transactions on*, vol. 62, no. 4, pp. 2530–2539, 2015. 17

[56] T. Ding, K. Sun, C. Huang, Z. Bie, and F. Li, "Mixed-integer linear programming-based splitting strategies for power system islanding operation considering network connectivity." 17

[57] C. Huang, F. Li, T. Ding, Y. Jiang, J. Guo, and Y. Liu, "A bounded model of the communication delay for system integrity protection schemes," *IEEE Transactions on Power Delivery*, vol. PP, no. 99, pp. 1–1, 2016. 17

[58] J. Vargas and J. Parise, "Simulation in transient regime of a heat pump with closed-loop and on-off control," *International Journal of Refrigeration*, vol. 18, no. 4, pp. 235–243, 1995. 27

[59] M. Arlitt and T. Jin, "A workload characterization study of the 1998 world cup web site," *Network, IEEE*, vol. 14, no. 3, pp. 30–37, 2000. 31

[60] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback control theory*. Courier Corporation, 2013. 37

[61] K. Zhou, J. C. Doyle, K. Glover *et al.*, *Robust and optimal control*. Prentice hall New Jersey, 1996, vol. 40. 39, 42, 43, 99, 102, 110, 111

[62] K. Hoffman, *Banach spaces of analytic functions*. Courier Corporation, 2007. 39, 84

[63] C. Foias, H. Ozbay, A. R. Tannenbaum, and M. Thoma, *Robust control of infinite dimensional systems: frequency domain methods*. Springer-Verlag New York, Inc., 1996. 39

[64] R. A. Martínez-Avendaño and P. Rosenthal, *An introduction to operators on the Hardy-Hilbert space*. Springer Science & Business Media, 2007, vol. 237. 40, 41

[65] R. L. Raffard, C. J. Tomlin, and S. P. Boyd, "Distributed optimization for cooperative agents: Application to formation flight," in *Proc. IEEE Conf. on Decision and Control*, vol. 3. IEEE, 2004, pp. 2453–2459. 43

[66] F. Borrelli, T. Keviczky, and G. J. Balas, "Collision-free uav formation flight using decentralized optimization and invariant sets," in *Proc. IEEE Conf. on Decision and Control*, vol. 1. IEEE, 2004, pp. 1099–1104. 43

[67] H. Raza and P. Ioannou, "Vehicle following control design for automated highway systems," *Control Systems, IEEE*, vol. 16, no. 6, pp. 43–60, 1996. 43

[68] M. L. El-Sayed and P. Krishnaprasad, "Homogeneous interconnected systems: An example," *IEEE Trans. Autom. Contr.*, vol. 26, no. 4, pp. 894–901, 1981. 43

[69] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Contr.*, vol. 48, no. 6, pp. 988–1001, 2003. 43

[70] D. L. Laughlin, M. Morari, and R. D. Braatz, "Robust performance of cross-directional basis-weight control in paper machines," *Automatica*, vol. 29, no. 6, pp. 1395–1410, 1993. 43

[71] S. Mijanovic, G. Stewart, G. Dumont, and M. Davies, "$\mathcal{H}_\infty$ robustification of a paper machine cross-directional control system," in *Proc. IEEE American Control Conf.*, vol. 3. IEEE, 2001, pp. 2203–2209. 43

[72] S. S. Stankovic, "Stochastic inclusion principle applied to decentralized automatic generation control," *International Journal of Control*, vol. 72, no. 3, pp. 276–288, 1999. 43

[73] I. Kamwa, R. Grondin, and Y. Hébert, "Wide-area measurement based stabilizing control of large power systems-a decentralized/hierarchical approach," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 136–153, 2001. 43

[74] F. Dörfler, J. Simpson-Porco, and F. Bullo, "Breaking the hierarchy: Distributed control & economic optimality in microgrids," *arXiv preprint arXiv:1401.1767*, 2014. 43

[75] M. Napoli and B. Bamieh, "Modeling and observer design for an array of electrostatically actuated microcantilevers," in *Proc. IEEE Conf. on Decision and Control*, vol. 5. IEEE, 2001, pp. 4274–4279. 43

[76] R. D'Andrea and G. E. Dullerud, "Distributed control design for spatially interconnected systems," *IEEE Trans. Autom. Contr.*, vol. 48, no. 9, pp. 1478–1495, 2003. 43

[77] E. Taylor Michael, "Partial differential equations i. basic theory," *Applied Mathematical Sciences*, vol. 115. 43

[78] R. W. Brockett and J. L. Willems, "Discretized partial differential equations: examples of control systems defined on modules," *Automatica*, vol. 10, no. 5, pp. 507–515, 1974. 43

[79] E. Sontag, "Linear systems over commutative rings. a survey," *Ricerche di Automatica*, vol. 7, no. 1, pp. 1–34, 1976. 43

[80] F. Paganini and B. Bamieh, "Decentralization properties of optimal distributed controllers," in *Proc. IEEE Conf. on Decision and Control*, vol. 2. IEEE, 1998, pp. 1877–1882. 43

[81] P. Voulgaris, G. Bianchini, and B. Bamieh, "Optimal decentralized controllers for spatially invariant systems," in *Proc. IEEE Conf. on Decision and Control*, vol. 4. IEEE, 2000, pp. 3763–3768. 43, 44, 45, 46, 51, 52, 53, 66, 76, 101

[82] B. Bamieh, F. Paganini, M. Dahleh *et al.*, "Distributed control of spatially invariant systems," *IEEE Trans. Autom. Contr.*, vol. 47, no. 7, pp. 1091–1107, 2002. 43

[83] M. Rinehart, M. Roozbehani, and M. A. Dahleh, "$\mathcal{H}_2$ performance bounds for voltage regulation on a spatially-invariant dc power grid," in *Proc. IEEE American Control Conf.*. IEEE, 2011, pp. 3911–3917. 44

[84] P. G. Voulgaris, G. Bianchini, and B. Bamieh, "Optimal $\mathcal{H}_2$ controllers for spatially invariant systems with delayed communication requirements," *Systems & control letters*, vol. 50, no. 5, pp. 347–361, 2003. 44, 45, 46, 50, 51, 52, 53, 54, 59, 60, 61, 64, 65, 66, 70, 72, 73, 94, 101, 105, 106, 111

[85] M. Fardad and M. R. Jovanović, "Design of optimal controllers for spatially invariant systems with finite communication speed," *Automatica*, vol. 47, no. 5, pp. 880–889, 2011. 44, 50, 60, 66

[86] S. M. Djouadi and J. Dong, "Operator theoretic approach to the optimal distributed control problem for spatially invariant systems," in *Proc. IEEE Conf. Decision and Control*, 2015. 48, 99, 100, 103, 104, 105

[87] W. Ren, R. W. Beard, and E. M. Atkins, "A survey of consensus problems in multi-agent coordination," in *Proc. IEEE American Control Conf.* IEEE, 2005, pp. 1859–1864. 49

[88] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Autom. Contr.*, vol. 49, no. 9, pp. 1465–1476, 2004. 49

[89] H. Zhang, F. L. Lewis, and A. Das, "Optimal design for synchronization of cooperative systems: state feedback, observer and output feedback," *IEEE Trans. Autom. Control*, vol. 56, no. 8, pp. 1948–1952, 2011. 49

[90] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, pp. 857–881, 1962. 49

[91] V. D. Blondel and J. N. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol. 36, no. 9, pp. 1249–1274, 2000. 49

[92] B. A. Francis, M. Thoma, and A. Wyner, *A course in H_∞ control theory*. Springer-verlag New York, 1987, vol. 88. 49, 56

[93] D. C. Youla, H. Jabr, and J. J. Bongiorno, "Modern wiener-hopf design of optimal controllers–part ii: The multivariable case," *IEEE Trans. Autom. Control*, vol. 21, no. 3, pp. 319–338, 1976. 50

[94] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *IEEE Trans. Autom. Contr.*, vol. 51, no. 2, pp. 274–286, 2006. 50

[95] J.-H. Kim and S. Lall, "Explicit solutions to separable problems in optimal cooperative control," *IEEE Trans. Autom. Contr.*, vol. 60, no. 5, pp. 1304–1319, 2015. 50

[96] B. Bamieh and P. G. Voulgaris, "A convex characterization of distributed control problems in spatially invariant systems with communication constraints," *Systems & Control Letters*, vol. 54, no. 6, pp. 575–583, 2005. 50

[97] M. Rotkowitz and S. Lall, "A characterization of convex problems in decentralized control," *IEEE Trans. Automa. Control*, vol. 51, no. 2, pp. 274–286, Feb 2006. 50

[98] S. M. Djouadi and J. Dong, "Duality of the optimal distributed control for spatially invariant systems," in *Proc. IEEE American Control Conf.*, 2014. 51, 69, 104

[99] ——, "Operator theoretic approach to the optimal distributed control problem for spatially invariant systems," in *Proc. IEEE American Control Conf.*, 2015. 51, 69, 99

[100] P. L. Duren, *Theory of H^p spaces.* Dover, 2000. 56

[101] A. Defant and K. Floret, *Tensor norms and operator ideals.* Elsevier, 1992, vol. 176. 57, 76, 89

[102] S. M. Djouadi, "Operator theoretic approach to the optimal two-disk problem," *IEEE Trans. Autom. Contr.*, vol. 49, no. 10, pp. 1607–1622, 2004. 71

[103] D. G. Luenberger, *Optimization by vector space methods.* John Wiley & Sons, 1968. 74, 75, 76, 78

[104] G. B. Folland and G. Folland, *Real analysis: modern techniques and their applications.* Wiley New York, 1999, vol. 361. 80, 90

[105] N. Z. Shor, *Minimization methods for non-differentiable functions.* Springer-Verlag New York, Inc., 1985. 95

[106] S. P. Boyd, C. H. Barratt, S. P. Boyd, and S. P. Boyd, *Linear controller design: limits of performance.* Prentice Hall Englewood Cliffs, NJ, 1991. 95

[107] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, mar 2014. 95

[108] M. Vidyasagar, *Control system synthesis: a factorization approach.* Morgan & Claypool Publishers, 2011. 105

[109] V. Peller, *Hankel operators and their applications.* Springer Science & Business Media, 2012. 107, 109

[110] N. Young, *An introduction to Hilbert space.* Cambridge university press, 1988. 108, 109

[111] H. Kushner, "Necessary conditions for continuous parameter stochastic optimization problems," *SIAM Journal on Control*, vol. 10, no. 3, pp. 550–565, 1972. 117

[112] U. G. Haussmann, *A stochastic maximum principle for optimal control of diffusions*. John Wiley & Sons, Inc., 1986. 117

[113] J.-M. Bismut, "An introductory approach to duality in optimal stochastic control," *SIAM review*, vol. 20, no. 1, pp. 62–78, 1978. 117

[114] J. Yong and X. Y. Zhou, *Stochastic controls: Hamiltonian systems and HJB equations*. Springer Science & Business Media, 1999, vol. 43. 117, 120

[115] R. F. Bass, *Diffusions and elliptic operators*. Springer Science & Business Media, 1998. 118, 120

[116] I. Nourdin and G. Peccati, *Normal approximations with Malliavin calculus: from Stein's method to universality*. Cambridge University Press, 2012, vol. 192. 118

[117] P. Malliavin, "Stochastic calculus of variation and hypoelliptic operators," in *Proc. Intern. Symp. SDE Kyoto 1976*. Kinokuniya, 1978, pp. 195–263. 118

[118] S. Janson, *Gaussian hilbert spaces*. Cambridge university press, 1997, vol. 129. 118

[119] P. Whittle, *Optimization over time*. John Wiley & Sons, Inc., 1982. 119

[120] H. Zhang, "Partially observable markov decision processes: A geometric technique and analysis," *Operations Research*, vol. 58, no. 1, pp. 214–228, 2010. 119

[121] T. Meyer-Brandis, X. Zhou, and B. Oksendal, "A stochastic maximum principle via malliavin calculus," *arXiv preprint arXiv:0911.3720*, 2009. 119

[122] T. Meyer-Brandis, B. Øksendal, and X. Y. Zhou, "A mean-field stochastic maximum principle via malliavin calculus," *Stochastics an International Journal of Probability and Stochastic Processes*, vol. 84, no. 5-6, pp. 643–666, 2012. 119, 129, 130, 131, 134, 135, 136

[123] I. V. Girsanov, "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures," *Theory of Probability & Its Applications*, vol. 5, no. 3, pp. 285–301, 1960. 120, 136

[124] C. D. Charalambous and N. U. Ahmed, "Centralized versus decentralized team games of distributed stochastic differential decision systems with noiseless information structures-part i: Thoery," *arXiv preprint arXiv:1302.3416*, 2013. 120, 128

[125] ——, "Centralized versus decentralized team games of distributed stochastic differential decision systems with noiseless information structures-part ii: Applications," *arXiv preprint arXiv:1302.3416*, 2013. 120, 128

[126] ——, "Team games optimality conditions of distributed stochastic differential decision systems with decentralized noisy information structures," *arXiv preprint arXiv:1304.3246*, 2013. 120, 128

[127] M. Kac, "Foundations of kinetic theory," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1955, no. 3, 1954, pp. 171–197. 129

[128] H. P. McKean, "A class of markov processes associated with nonlinear parabolic equations," *Proceedings of the National Academy of Sciences*, vol. 56, no. 6, pp. 1907–1911, 1966. 129

[129] D. Andersson and B. Djehiche, "A maximum principle for sdes of mean-field type," *Applied Mathematics & Optimization*, vol. 63, no. 3, pp. 341–356, 2011. 130

[130] I. Karatzas and S. E. Shreve, "Brownian motion and stochastic calculus, vol. 113 of," *Graduate texts in mathematics*, 1991. 168

[131] A. Sokol, "Stochastic analysis and mathematical finance," Ph.D. dissertation, University of Copenhagen, 2008. 169

[132] D. Nualart, *The Malliavin calculus and related topics.* Springer, 2006, vol. 1995. 169

[133] D. Xiu, *Numerical methods for stochastic computations: a spectral method approach.* Princeton University Press, 2010. 178

Appendix

Appendix A

Proofs

A.1 Proof of Proposition 2.3

Proof: The original system can be written in terms of error dynamics, at time k,

$$e_k = \tilde{\mathbf{A}}_{k-1} e_0 + \tilde{\mathbf{B}}_{k-1} \mathbf{u} + \tilde{\mathbf{C}}_{k-1} \omega + L_{k-1}$$

where $L_{k-1} = A_d^{k-1} x_r - x_r$ and from (2.14), x_k, ω_k are 3×1 and u_k is a scalar so that

$$\begin{aligned} \tilde{\mathbf{A}}_{k-1} &= A_d^k \\ \tilde{\mathbf{B}}_{k-1} &= [A_d^{k-1} B_d \quad \cdots \quad B_d \ \mathbf{0}_{3\times(N-k)}] \\ \tilde{\mathbf{C}}_{k-1} &= [A_d^{k-1} C_d \quad \cdots \quad C_d \ \mathbf{0}_{3\times 3(N-k)}] \end{aligned} \tag{A.1}$$

Then, after some manipulations, the error state term in cost function becomes:

$$\begin{aligned} e_k^T Q e_k &= e_0^T \tilde{\mathbf{A}}_{k-1}^T Q \tilde{\mathbf{A}}_{k-1} e_0 + 2 e_0^T \tilde{\mathbf{A}}_{k-1}^T Q (\tilde{\mathbf{B}}_{k-1} \mathbf{u} + \tilde{\mathbf{C}}_{k-1} \omega) \\ &+ \mathbf{u}^T \tilde{\mathbf{B}}_{k-1}^T Q \tilde{\mathbf{A}}_{k-1} \mathbf{u} + \omega^T \tilde{\mathbf{C}}_{k-1}^T Q \tilde{\mathbf{C}}_{k-1} \omega \\ &+ 2 \mathbf{u}^T \tilde{\mathbf{B}}_{k-1}^T Q \tilde{\mathbf{C}}_{k-1} \omega + 2 L_{k-1}^T Q (\tilde{\mathbf{A}}_{k-1} e_0 \\ &+ \tilde{\mathbf{B}}_{k-1} \mathbf{u} + \tilde{\mathbf{C}}_{k-1} \omega) + L_{k-1}^T Q L_{k-1} \end{aligned} \tag{A.2}$$

Thus, we reach the formula of the cost function stated above with

$$
\begin{aligned}
\mathbf{A} &= \sum_{k=1}^{N} \tilde{\mathbf{A}}_{k-1}^{T} Q \tilde{\mathbf{A}}_{k-1} \\
\mathbf{B} &= \mathbf{diag}(R, \cdots, R) + \sum_{k=1}^{N} \tilde{\mathbf{B}}_{k-1}^{T} Q \tilde{\mathbf{B}}_{k-1} \\
\mathbf{C} &= \sum_{k=1}^{N} \tilde{\mathbf{C}}_{k-1}^{T} Q \tilde{\mathbf{C}}_{k-1}, \ \mathbf{D} = \sum_{k=1}^{N} \tilde{\mathbf{B}}_{k-1}^{T} Q \tilde{\mathbf{C}}_{k-1}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{c} &= \left(\sum_{k=1}^{N} \tilde{\mathbf{C}}_{k-1}^{T} Q \tilde{\mathbf{A}}_{k-1} \right) e_0 + \sum_{k=1}^{N} \tilde{\mathbf{C}}_{k-1}^{T} Q L_{k-1} \\
\mathbf{a} &= \sum_{k=1}^{N} \tilde{\mathbf{A}}_{k-1}^{T} Q L_{k-1}, \ \hat{l} = \sum_{k=1}^{N} L_{k-1}^{T} Q L_{k-1} \\
\mathbf{b} &= \left(\sum_{k=1}^{N} \tilde{\mathbf{B}}_{k-1}^{T} Q \tilde{\mathbf{A}}_{k-1} \right) e_0 + \sum_{k=1}^{N} \tilde{\mathbf{B}}_{k-1}^{T} Q L_{k-1}
\end{aligned}
\tag{A.3}
$$

■

A.2 Proof of Lemma 5.10.1

This appendix is devoted to the proof of Lemma 5.10.1.

Proof: First notice that since $S_c \subset S$, then

$$\mu \leq \inf_{\tilde{Q} \in S_c} \|T_{2in}^* - \tilde{Q}\|_\infty$$

since the infimum in the definition of ψ is taken over a larger subspace.

For the reverse inequality, let $F := T_{2in}^* T_1$ and $0 < r < 1$, call:

$$\begin{aligned} F_r(e^{i\theta}, \lambda) &:= F(re^{i\theta}, \lambda) \\ Q_{or}(e^{i\theta}, \lambda) &:= Q_o(re^{i\theta}, \lambda) \end{aligned} \tag{A.4}$$

Then,

$$\begin{aligned} \|F - Q_{or}\|_\infty &= \|F - F_r + F_r - Q_{or}\|_\infty \\ &\leq \|F_r - Q_{or}\|_\infty + \|F - F_r\|_\infty \end{aligned} \tag{A.5}$$

Now, note that $(F_r - Q_{or})$ is bounded above by $\|F - Q_o\|_\infty$, since in the latter norm the supremum is taken over a larger set, that is, for $g(e^{i\theta,\lambda}) \in \mathcal{H}^\infty(\mathcal{L}^\infty \mathbf{T})$, and $\|g_r\|_\infty$ is a non-decreasing function of r for $r \in [0, 1]$. Therefore,

$$\|F - Q_{or}\|_\infty \leq \|F - Q_o\|_\infty + \|F - F_r\|_\infty \tag{A.6}$$

Considering $T_{2in} T_1$ is continuous on $\mathbf{T} \times \bar{\mathcal{D}}$, and the definition that $F := T_{2in}^* T_1$, then F is also continuous. Therefore, $\forall \epsilon > 0$, there exists $0 < r < 1$ such that

$$\|F - F_r\|_\infty < \epsilon \tag{A.7}$$

and

$$\|F - Q_{or}\|_{\mathcal{D}} \leq \mu + \epsilon \tag{A.8}$$

Q_{or} being in S_c implies:

$$\inf_{\tilde{Q} \in S_c} \|F - \tilde{Q}\|_{\infty} \leq \mu + \epsilon \tag{A.9}$$

Since ϵ is arbitrary, therefore:

$$\inf_{\tilde{Q} \in S_c} \|F - \tilde{Q}\|_{\infty} \leq \mu \tag{A.10}$$

so the lemma holds, completing the proof. ■

A.3 Proof of Lemma 5.11.1

We shall now focus on the proof of Lemma 5.11.1.

Proof: : For

$$z = \sum_{i=1}^{n} x_i \otimes y_i \in H_0^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}}), \tag{A.11}$$

where $z(e^{i\theta}, \lambda) = \sum_{i=1}^{n} x_i(\lambda) y_i(e^{i\theta})$.

Associate to $z(\cdot,\cdot)$ the following function:

$$\begin{aligned} \psi_z : \bar{\mathcal{D}} &\rightarrow H_0^1(\mathbf{T}), \\ \psi_z(\lambda) &= \sum_{i=1}^{n} x_i(\lambda) y_i(e^{i\theta}). \end{aligned}$$

The function $x_i(\lambda)$ can be approximated in $L^1(\bar{\mathcal{D}})$ by simple functions as closely as desired.

Moreover,

$$\begin{aligned} \|\psi_z\|_1 &= \int_{\bar{\mathcal{D}}} \|\psi_z(\lambda) d\lambda\| = \int \| \sum_{i=1}^{n} x_i(\lambda) y_i \| ds \\ &\leq \int_{\bar{\mathcal{D}}} \sum_{i=1}^{n} |x_i(\lambda)| \|y_i\|_1 d\lambda \\ &= \sum_{i=1}^{n} \int_{\bar{\mathcal{D}}} |x_i(\lambda)| \|y_i\|_1 d\lambda \\ &= \sum_{i=1}^{n} \|x_i\|_1 \|y_i\|_1 \end{aligned} \tag{A.12}$$

So $\psi_z \in L^1(\bar{\mathcal{D}}) \otimes_\gamma H_0^1(\mathbf{T})$.

Now taking the infimum over all representation of z, we get

$$\|\psi_z\|_1 \leq \gamma(z).$$

The linear map $z \to \psi_z$ has therefore a continuous extension to $H_0^1(\mathbf{T}) \otimes_\gamma L^1(\bar{\mathcal{D}})$.

To show that $\|F_z\| = \gamma(z)$. Assume that F_z is a simple function in $z = \sum_{i=1}^n x_i \otimes y_i$. We can choose x_i and y_i to be simple functions, such that $x_i(\lambda)x_j(\lambda) = 0$ for $i \neq j$ and $\sum_{i=1}^n x_i(\lambda) = 1$.
Then

$$\begin{aligned}
\|F_z\| &= \int_{\bar{\mathcal{D}}} \| \sum_{i=1}^n x_i(\lambda) y_i(e^{i\theta}) \|_1 d\lambda \\
&= \sum_{i=1}^n \int_{\bar{\mathcal{D}}} |x_i(\lambda)| \|y_i\|_1 d\lambda \\
&= \sum_{i=1}^n \|x_i\| \|y_i\|_1 \\
&\geq \gamma(z).
\end{aligned}$$

Since simple functions are dense in $H_0^1(L^1(\bar{\mathcal{D}}))$, and the Lemma holds, completing the proof. ■

Appendix B

The Malliavin Calculus

B.1 First taste of Malliavin Calculus

The two main challenging parts to understand this new calculus are to understand "what" is Malliavin calculus and "why" we need it? Instead of directly answering these two questions in an abstract way, we can relate Malliavin calculus to the same kind of questions when we were first exposed to *measure theory (Lebesgue Integral)* and *Soblev spaces*. The most intuitive ideas that help us understand these concepts would lead us smoothly to answer the two questions for Malliavin calculus here.

Intuitive answers to the questions:

- **What is Lebesgue Integral, why we need it and measure theory?**
- **What is Soblev space, why we need it?**

Definition B.1. *[130] The probability measure P_* on $(C[0,\infty), \mathcal{B}(C[0,\infty)))$, under which the coordinate mapping process $W_t(\omega) \triangleq \omega(t)$, $0 \le t < \infty$, is a standard, one-dimensional Brownian motion, is called Wiener measure.*

Remark B.2. *[130] A standard, one-dimensional Brownian motion defined on any probability space can be thought of as a random variable with values in $C[0,\infty)$; regarded this way, Brownian motion includes the Wiener measure on*

$(C[0,\infty), \mathcal{B}(C[0,\infty)))$. *For this reason, we call* $(C[0,\infty), \mathcal{B}(C[0,\infty)), P_*)$, *where* P_* *is Wiener measure, the canonical probability space for Brownian motion.*

We will first give an one-dimensional toy example for Malliavin Calculus, which works as the recipe of the whole paper. Then we will follow [131] to define the Malliavin derivative slightly differently than in Nualart [132], initially using coordinates of the Brownian motion instead of stochastic integrals with deterministic integrands. We then extend the operator by the usual closedness argument. Next, we prove the chain rule and the integration-by-parts formula. We use the generalized chain rule to show that our definition of the Malliavin derivative coincides with that in Nualart [132]. Finally, we develop the Hilbert space theory of the Malliavin derivative and use it to obtain a chain rule for Lipschitz transformations.

B.2 One-dimensional toy example

The aim of this section is to introduce the realm of Malliavin operators, by focusing on the one-dimensional case only. In particular, we are going to define derivative, divergence and Ornstein-Uhlenbeck operators acting on random variables of the type $F = f(N)$, where f is a deterministic function and $N \sim \mathcal{N}(0,1)$ has a standard Gaussian distribution. Actually, one-dimensional Malliavin operators coincide with familiar objects of functional analysis, which are more accessible to us.

Definition B.3. *Define* $\mathcal{S}$ *a* ***smooth function*** *which denotes the set of* $\mathcal{C}^\infty$*-functions* $f : \mathbb{R} \to \mathbb{R}$ *such that* f *and all its derivatives have at most polynomial growth.*

Remark B.4. *All the following operators will be defined on domains that can be obtained as the closure of* $\mathcal{S}$ *with respect to an appropriate norm.*

Proposition B.5. *The monomials* $x^n : n = 0, 1, 2, \cdots$ *generate a dense subspace of* $L^q(\gamma)$ *for every* $q \in [1,\infty)$*.In particular, for any* $q \in [1,\infty)$ *the space* $\mathcal{S}$ *is a dense subset of* $L^q(\gamma)$.

Proof. Using standard *Hahn-Banach theory.* □□□

Remark B.6. *In general, our aim would be to develop a canonical representation for element of $\mathcal{S}$, develop results to yield flexible representations of elements in $\mathcal{S}$ and to prove that $\mathcal{S}$ is an algebra which is dense in $\mathcal{L}^p(\mathcal{F}_T)$, $p \geq 1$.*

Fix $f \in \mathcal{S}$ for every $p = 1, 2, \cdots$, we write $f^{(p)}$ or, equivalently, $D^p f$ to indicate the pth derivative of f. Note that the mapping $f \mapsto D^p f$ is an operator from $\mathcal{S}$ into itself.

Lemma B.6.1. *The operator $D^p : \mathcal{S} \subset L^q(\gamma) \to L^q(\gamma)$ is closable for every $q \in [1, \infty)$ and every integer $p \geq 1$.*

Fix $q \in [1, \infty)$ and an integer $p \geq 1$. We set $\mathbb{D}^{p,q}$ to be the closure of $\mathcal{S}$ with respect to the norm

$$\|f\|_{\mathbb{D}^{p,q}} = \left(\int_{\mathbb{R}} |f(x)|^q d\gamma(x) + \int_{\mathbb{R}} |f^{(1)}(x)|^q d\gamma(x) + \cdots + \int_{\mathbb{R}} |f^{(p)}(x)|^q d\gamma(x) \right)^{\frac{1}{q}}. \quad \text{(B.1)}$$

Remark B.7. *Equivalently, $\mathbb{D}^{p,q}$ is the Banach space of all functions in $L^q(\gamma)$ whose derivatives up to the order p in the sense of distributions also belong to $L^q(\gamma)$.*

Definition B.8. *We denote by Domδ^p the subset of $L^2(\gamma)$ composed of those functions g such that there exists $c > 0$ satisfying the property that, for all $f \in \mathcal{S}$ (or, equivalently, for all $f \in \mathbb{D}^{p,2}$),*

$$\left| \int_{\mathbb{R}} f^{(p)}(x) g(x) d\gamma(x) \right| \leq c \sqrt{\int_{\mathbb{R}} f^2(x) d\gamma(x)} \quad \text{(B.2)}$$

Fix $g \in \mathrm{Dom}\delta^p$. Since condition (B.2) holds, the linear operator $f \mapsto \int_r e f^{(p)}(x) g(x) d\gamma(x)$ is continuous from $\mathcal{S}$, equipped with the $L^2(\gamma)-$norm, into $\mathbb{R}$. Thus, we can use Riesz representation theorem to extend this operator to a linear operator from $L^2(\gamma)-$ into $\mathbb{R}$. There exists a unique element $\delta^p g$ in $L^2(\gamma)$, such that

$$\int_{\mathbb{R}} f^{(p)}(x) g(x) d\gamma(x) = \int_{\mathbb{R}} f(x) \delta^p g(x) d\gamma(x) \quad \text{for all } f \in \mathcal{S}. \quad \text{(B.3)}$$

Then we can get the integrtaion by parts formula:

$$\int_{\mathbb{R}} f'(x)g(x)d\gamma(x) = \int_{\mathbb{R}} xf(x)g(x)d\gamma(x) - \int_{\mathbb{R}} f(x)g'(x)d\gamma(x). \tag{B.4}$$

B.3 Malliavin derivative

Define probability space $(\Omega, \mathcal{F}, P)$ with a one-dimensional Wiener process W on $[0, T]$, where the variables to be differentiated are in some suitable sense related to W. First consider our probability space $C_0[0, T]$, endowed with Wiener measure P. Notice that $C_0[0, T]$ is a Banach space under the uniform (sup) norm, where W can be considered as an identity mapping, W is a Wiener process on $[0, T]$.

Therefore, we want to identify a differentiability concept for variables of the form $\omega \mapsto W_t(\omega)$ which can be generalized to the setting of an abstract probability space with a Wiener process. Then we consider a mapping X from $C_0[0, T]$ to $\mathbb{R}$.

Following Gâteaux directional derivatives: The derivative in direction $h \in C_0[0, T]$ at ω is the element $D_h X(\omega) \in \mathbb{R}$ such that $\lim_{\epsilon \to 0} \frac{X(\omega+\epsilon h) - X(\omega)}{\epsilon} = D_h X(\omega)$.
Then for $h \in C_0[0, T]$, the derivative of W_t in direction h is

$$D_h W_t(\omega) = \lim_{\epsilon \to 0} \frac{X(\omega + \epsilon h) - X(\omega)}{\epsilon} = \lim_{\epsilon \to 0} \frac{\omega_t + \epsilon h_t - \omega_t}{\epsilon} = h_t \tag{B.5}$$

Consider $h_t = \int_0^t g(s)ds$ for some $g \in \mathcal{L}^1[0, T]$. Then

$$D_h W_t(\omega) = h_t = \int_0^T 1_{[0,t]}(s)g(s)ds. \tag{B.6}$$

So the Gâteaux derivative for any $\omega \in C_0[0, T]$ is actually characterized by the mapping $1_{[0,T]} \mapsto \mathbb{R}$. Thus, $1_{[0,t]}$ can be considered as a kind of derivative of W_t.

Using the same idea, we can generalize this to general mappings $C_0[0,T] \to \mathbb{R}$ as:

$$f : [0,T] \to \mathbb{R}, \text{ for } \quad h \in C_0[0,T] \quad \text{with} \quad h(t) = \int_0^t g(s)ds, \tag{B.7}$$

$$D_h X(\omega) = \int_0^T f(s)g(s)ds \tag{B.8}$$

where f is the derivative of X, and we define $D_{\mathbb{F}}X = f$.

Assume chain rule holds for D:

If f is continuously differentiable with sufficient growth conditions, namely polynomial growth of the mapping itself and its partial derivatives, we would like to have:

$$Df(W_{t_1}, \cdots W_{t_n}) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(W_{t_1}, \cdots, W_{t_n}) DW_{t_k}, \tag{B.9}$$

where $DW_{t_k} = 1_{[0,t_k]}$. Hence, to study the derivative of f is sufficient to study the operator D.

Remark B.9. *We wanna specify the immediate domain of D and consider its properties.*

- *The derivative given above is a stochastic function from $[0,T]$ to $\mathbb{R}$.*
- *With sufficient growth conditions on f, it is an element of $\mathcal{L}^p([0,T] \times \Omega), p \geq 1$.*

B.4 The Space $\mathcal{S}$ and $\mathcal{L}^p(\Pi)$

A multi-index of order n is an element $\alpha \in \mathbb{N}_0^n$. The degree of the multi-index is $|\alpha| = \sum_{k=1}^{n} \alpha_k$. A polynomial in n variables of degree k is a map:

$$p : \mathbb{R}^n \mapsto \mathbb{R}, \, p(x) = \sum_{|\alpha| \leq k} a_\alpha x^\alpha. \tag{B.10}$$

the sum in the above runs over all multi-indicies α with $|\alpha \leq k|$, with $x^\alpha = \Pi_{i=1}^n x^{\alpha_i}$. The space of polynomials of degree k in any number of variables is denoted β_k.

We introduce some smooth notations:

- $C^1(\mathbb{R}^n)$ are the mappings $f : \mathbb{R}^n \mapsto \mathbb{R}$ which are continuously differentiable.
- $C^\infty(\mathbb{R}^n)$ are the mappings $f : \mathbb{R}^n \mapsto \mathbb{R}$ which are differentiable infinitely often.
- $C_p^\infty(\mathbb{R}^n)$ are the elements $f \in C^\infty(\mathbb{R}^n)$ such that f and its partial derivatives are dominated by polynomials
- $C_c^\infty(\mathbb{R}^n)$ are the elements of $C^\infty(\mathbb{R})$ with compact support.

We need to justify a little about the $C_p^\infty(\mathbb{R}^n)$, and make a connection with the well-known Soblev space knowledge.

Lemma B.9.1. *Clearly, we know*

$$C_c^\infty(\mathbb{R}^n) \subset C_p^\infty(\mathbb{R}^n) \subset C^\infty(\mathbb{R}^n) \subset C^1(\mathbb{R}^n). \tag{B.11}$$

Moreover, if we consider the standard regularity properties of boundary Ω, we see $C_p^\infty(\mathbb{R}^n)$ is the weakest one among them, i.e.,
The uniform C^m -regularity condition ($m \geq 2$).
$\Rightarrow$ *the strong local Lipschitz condition*
$\Rightarrow$ *the uniform cone condition*
$\Rightarrow$ *the segment condition*
$\Rightarrow$ ***the polynomial condition***

Definition B.10. *If $t \in [0,T]^n$, we write $W_t = (W_{t_1}, \cdots, W_{t_n})$. By $\mathcal{S}$, we denote the space of variables $f(W_t)$, where $f \in C_p^\infty(\mathbb{R}^n)$ and $t \in [0,T]^n$.*

Then we want be able to extend the coordinates which F depends on, and also to reorder the coordinates using Lemma B.10.1.

Lemma B.10.1. *Let $t \in [0,T]^n$, $s \in [0,T]^m$ and $F \in \mathcal{S}$ with $F = f(W_t)$. Assume that $t_1, \cdots, t_n \subset s_1, \cdots, s_m$. There exists $g \in C_p^\infty(\mathbb{R}^m)$ such that $F = g(W_s)$.*

Corollary B.10.1. *Let $F \in \mathcal{S}$. There exists $t \in [0,T]^n$ such that $0 < t_1 < \cdots < t_n$ and $f \in C_p^\infty(\mathbb{R}^n)$ such that $F = f(W_t)$.*

This corollary states that any element of $\mathcal{S}$ has a representation where we can assume that all the coordinates of the Wiener process in the element are different, positive and ordered. This observation will make our lives a good deal easier in the following. Our next result shows that any element of $\mathcal{S}$ can be written as a transformation of an n-dimensional standard normal variable.

Lemma B.10.2. *The space $\mathcal{S}$ is an algebra, and $\mathcal{S} \subseteq \mathcal{L}^p(\Omega)$ for all $p \geq 1$.*

Corollary B.10.2. *The space $\mathcal{S}$ is an algebra, and $\mathcal{S} \subseteq \mathcal{L}^p(\mathcal{F}_T)$ for any $p \geq 1$.*

We have defined the Malliavin derivative D on $\mathcal{S}$, we will investigate its basic properties and extend it to a larger space.

Remark B.11. *Even dense sets can be quite slim, and we need to extend the operator D to a larger space before it can be of any actual use. Then, we would like to show D is closable. We can then extend it by taking the closure. (Closed Graph Theory)*

Theorem B.12. *The operator D is* ***closable*** *from $\mathcal{S}$ to $\mathcal{L}^p(\Pi)$.*

Then for any $p \geq 1$, a linear operator $D : \mathbb{D}_{1,p} \to \mathcal{L}^p(\Pi)$ with the following two properties:

- For $F \in \mathcal{S}$ with $F = f(W_t), DF = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(W_t)\mathbb{1}_{[0,t_k]}$.
- When considering $\mathbb{D}_{1,p}$ under the norm $\|\cdot\|_p$, D is closed.

B.5 The Malliavin Derivative on $\mathbb{D}_{1,2}$

We will consider the properties of the Malliavin derivative as operator $D : \mathbb{D}_{1,2} \mapsto \mathcal{L}^2(\Pi)$.

Theorem B.13. *Let $F = (F_1, \cdots, F_n)$, where $F_1, \cdots, F_n \in \mathbb{D}_{1,2}$, and let $\varphi \in C^1(\mathbb{R}^n)$. Assume $\varphi(F) \in \mathcal{L}^2(\mathcal{F}_T)$ and $\sum_{k=1}^{n} \frac{\partial \varphi}{\partial x_k}(F)DF_k$.*

Corollary B.13.1. *Let $F, G \in \mathbb{D}_{1,2}$. If the integrability conditions $FG \in \mathcal{L}^2(\mathcal{F}_T)$ and $(DF)G + F(DG) \in \mathcal{L}^2(\Pi)$ hold, $FG \in \mathbb{D}_{1,2}$ and $D(FG) = (DF)G + G(DG)$.*

The alternative integration-by-parts formula that's easy to apply in reality:

Corollary B.13.2. *Let $F, G \in \mathcal{D}_{1,2}$ with $FG \in \mathcal{L}^2(\mathcal{F}_T)$ and $F(DG) + G(DF) \in \mathcal{L}^2(\Pi)$. Then*

$$E(G < DF, h >_{[0,T]}) = E(FG(\theta h)) - EF < DG, h >_{[0,T]} \tag{B.12}$$

and another similar formula as

Corollary B.13.3. *Let $F \in \mathcal{D}_{1,2}$ and $G \in \mathcal{S}_h$. Then*

$$E(G < DF, h >_{[0,T]}) = E(FG(\theta h)) - EF < DG, h >_{[0,T]} . \tag{B.13}$$

B.6 Skorohod Integral (Divergence Operator)

The Skorohod integral δ is defined as the adjoint operator of the Malliavin derivative D. This is a linear operator defined on a dense subspace $\mathbb{S}_{1,2}$ of $\mathcal{L}^2(\Pi)$ mapping into $\mathcal{L}^2(\mathcal{F}_T)$, characterized by the duality relationship

$$\langle F, \delta u \rangle_{\mathcal{F}_T} = \langle DF, u \rangle_{\Pi} \tag{B.14}$$

Basically, as the derivative maps $L^2(\Omega)$ to $L^2(\Omega; H)$, its adjoint will be an operator on $L^2(\Omega; H)$ taking values in $L^2(\Omega)$.

Definition B.14. *The divergence operator, sometimes also called the Skorohod integral. The domain $D(\delta)$ consists of those $u \in L^2(\Omega; H)$ such that there exists $X \in L^2(\Omega)$ with*

$$\mathbb{E} < DY, u >_H = \mathbb{E}(Y \cdot X) \tag{B.15}$$

for all $Y \in \mathbb{D}^{1,2}$. Since $\mathbb{D}^{1,2}$ is dense in $L^2(\Omega)$, there is at most one such element X. We write $\delta(u) := X$.

Remark B.15. *Note that an element u of $L^2(\Omega; H)$ belongs to the domain $D(\delta)$ if and only if there exists a constant $c \geq 0$ such that*

$$|\mathbb{E} < DY, u > | \leq c(\mathbb{E}|Y|^2)^{\frac{1}{2}} \tag{B.16}$$

for all $X \in \mathbb{D}^{1,2}$.

Indeed, the map $\varphi(Y) := \mathbb{E} < DY, h >$ is a linear functional on $\mathbb{D}^{1,2}$ which is bounded with respect to the 2-norm. It thus has a unique bounded extension to all of $L^2(\Omega)$. By the Riesz-Fischer theorem, this extension is of the form $Y \mapsto \mathbb{E}(YX)$ for a certain $X \in L^2(\Omega)$.

Example B.16. *Let $u \in \mathcal{L}(H)$, say $u = \sum_{j=1}^{n} X_j h_j$. Then $X \in D(\delta)$ and*

$$\delta(u) = \sum_{j=1}^{n} X_j W(h_j) - \sum_{j=1}^{n} < DX_j, h_j > . \tag{B.17}$$

by

$$\mathbb{E} < DY, u >= \mathbb{E}X < DY, h >= \mathbb{E}Y(XW(h) - < X, h >) \tag{B.18}$$

for all $Y \in \mathbb{D}^{1,2}$.

We can factor out a scalar random variable in divergence.

Proposition B.17.

$$\delta(Xu) = X\delta(u) - < DX, u > . \tag{B.19}$$

We can also show both the Malliavin derivative and Skorohod integral are *local operators*.

Definition B.18. *An operator T defined on a space of random variables local is $X = 0$ a.e. on a set $A \in \Sigma$ implies that also $TX = 0$ e.e. on A.*

B.7 The Ornstein-Uhlenbeck Semigroup

Definition B.19. *Let J_n denote the projection on $L^2(\Omega)$ onto the n-th Wiener chaos. The Ornstein-Uhlenbeck Semigroup is the one-parameter semigroup $(T(t))_{t\geq 0}$ defined by*

$$T(t)X : \sum_{n=0}^{\infty} e^{-nt} J_n X. \tag{B.20}$$

It turns out that the *Ornstein-Uhlenbeck Semigroup* is a strongly continuous semigroup of self-adjoint operators.

Then we finally have the relation between operators L, D and δ.

Theorem B.20. *We have $L = -\delta D$, i.e. $X \in D(L)$ if and only if $X \in \mathbb{D}^{1,2}$ and $DX \in D(\delta)$; in that case $\delta(DX) = -LX$.*

B.8 Itô's Integral and the Clark-Ocone Formula

We study properties of the Malliavin derivative and the divergence operators in the white noise setting.

Proposition B.21. *Let $u \in L^2_{\mathcal{F}}((0,\tau) \times \Omega)$ and define $X := \int_0^\tau u(s) dB_s$. Then $X \in \mathbb{D}^{1,2}$ if and only if $u \in \mathbb{L}^{1,2}$. In that case, $t \mapsto D_t u(s)$ belongs to $L^2_{\mathbb{F}}$ and for $t \in (0, \tau)$ we have*

$$D_t X = u(t) + \int_t^\tau D_t u(s) dB_s \tag{B.21}$$

almost surely.

By Ito

$$X = \mathbb{E}X + \int_0^\tau u(t) dB_t \tag{B.22}$$

How to compute the process u given X. For random variables $X \in \mathcal{D}^{1,2}$ we have the following result, called the *Clark-Ocone formula.*

Theorem B.22. *Let $T = [0, \tau]$ and set as usual $B_t := W(\mathbb{1}_{(0,t]})$ and $\mathcal{F}_t = \sigma(B_s : s \leq t)$. Then for $X \in \mathbb{D}^{1,2}$ we have*

$$X = \mathbb{E}X + \int_0^\tau \mathbb{E}(D_t X | \mathcal{F}_t) dB_t, \tag{B.23}$$

which means $u(t) = \mathbb{E}(D_t X | \mathcal{F}_t)$.

B.9 Polynomial Chaos

We first review the basics of orthogonal polynomials, which play a central role in modern approximation theory. more in-depth discussions can be found in many standard books such as [133].

A general polynomial of degree n takes the form

$$Q_n(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad a_n \neq 0. \tag{B.24}$$

where a_n is the coefficient of the polynomial.
A system of polynomials Q_n is an orthogonal system of polynomials with respect to some real positive measure γ if the following orthogonality relations hold:

$$\int_S Q_n(x) Q_m(x) d\gamma(x) = \varsigma_n \delta_{mn} \tag{B.25}$$

where $\delta_{mn} = 0$ if $m \neq n$ and $\delta_{mn} = 1$ if $m = n$.

Let $A = \mathcal{B}(\mathbb{R})$, and $\gamma(A) = \int_A \alpha(t) dt$. So assume space $\mathcal{L}^2(\mathbb{R}, \gamma)$.

Let $H_n(x)$ denote the Hermite polynomial with

$$H_0 = 1, \quad H_1 = x,$$
$$H_{n+1}(x) - x H_n(x) + n H_{n-1}(x) = 0, \tag{B.26}$$

where $\{H_n(x)\}_{n=0}^{\infty}$ is the orthogonal basis of $\mathcal{L}^2(\mathbb{R},\gamma)$.

Moreover, we wanna show it's actually an orthonormal basis.

$$\begin{aligned}\int_{\mathbb{R}} H_n(x)d\gamma(x) &= 0, \text{ for } n \geq 1; &\text{(B.27)}\\ \int_{\mathbb{R}} (x^2-1)d\gamma(x) &= \int_{\mathbb{R}} x^2 d\gamma(x) - 1 = \mathbb{E}(\xi^2) - 1. &\text{(B.28)}\end{aligned}$$

Then

$$< H_n, H_m >= n!\delta_{nm}. \tag{B.29}$$

So $h_n = \{\frac{1}{\sqrt{n!}}H_n\}_{n=1}^{\infty}$ is orthonormal basis in $\mathcal{L}^2(\mathbb{R},\gamma)$.

Therefore, $\forall f(x) \in \mathcal{L}^2(\mathbb{R},\gamma)$,

$$\begin{aligned}f(x) &= \sum_{n=0}^{\infty} \alpha_n h_n(x), &\text{(B.30)}\\ \alpha_n &= < f(x), h_n(x) >= \mathbb{E}\left[f(\xi)h_n(\xi)\right]. &\text{(B.31)}\\ f(\xi) &= \sum_{n=0}^{\infty} \alpha_n h_n(\xi) &\text{(B.32)}\\ \Rightarrow f(t,\xi) &= \sum_{n=0}^{\infty} \alpha_n(t) h_n(\xi) &\text{(B.33)}\end{aligned}$$

After defining the derivative operator, we came to a point to ask what more benefits can we get from the Malliavin Calculus. An interesting connection to polynomial chaos. Consider functions on the real axis $\mathbb{R} = (-\infty, \infty)$ equipped with the Gaussian measure

$$\mu(dx) = \rho(x)dx, \quad \rho(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \tag{B.34}$$

where dx is the Lebesgue measure.

We denote the space of square integrable functions with the Gaussian measure μ as

$$L^2(\mathbb{R}, \mu) = \{f(x); \int_{-\infty}^{+\infty} f(x)^2 \mu(dx) < \infty\}. \tag{B.35}$$

Then the expectation

$$\int_{\mathbb{R}} f(x) d(\mu x) := \mathbb{E} f(N) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-\frac{x^2}{2}} dx \tag{B.36}$$

The inner product on this space is defined as

$$(f, g)_\mu = \int_{-\infty}^{\infty} f(x) g(x) \mu(dx) = \int_{-\infty}^{\infty} f(x) g(x) \rho(x) dx. \tag{B.37}$$

Suppose ξ is a standard Gaussian random variable with distribution $N(0,1)$, then

$$(f, g)_\mu = E[f(\xi) g(\xi)], \tag{B.38}$$

where E denotes the expectation operator.

Hermite polynomial $\{h_n\}_{n=0}^{\infty}$ gives a complete orthonormal basis, then

$$f(x) = \sum_{n=0}^{\infty} a_n h_n(x). \tag{B.39}$$

Example B.23. *(Fractional Brownian motion) A fractional Brownian motion is a Gaussian process with covariance function*

$$c_H(t, s) = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}) \tag{B.40}$$

where $H \in (0,1)$ is the so-called Hurst parameter. The choice $H = \frac{1}{2}$ yields $c_{\frac{1}{2}}(t, s) = \min\{t, s\}$ which is the covariance function of Brownian motion.

Definition B.24. *(Isonormal Gaussian processes)*
Let H be a real, separable Hilbert space with inner product $< \cdot, \cdot >$. An H-Isonormal Gaussian process is a family $W = W(h) : h \in H$ of real-valued random variables, defined on a common probability space $(\Omega, \Sigma, \mathbb{P})$, such that $W(h)$ is a Gaussian random variable for all $h \in H$ and, for $h, g \in H$, we have $\mathbb{E}\left(W(h)W(g)\right) =< h, g >$.

Then we have

$$< \mathbb{1}_{(0,t]}, \mathbb{1}_{(0,s]} >_H = c_H(t, s). \tag{B.41}$$

Hence, Isonormal Gaussian process W gives rise to a fractional Brownian motion with Hurst parameter H.

Definition B.25. *(Hermite Polynomials)*
For $n \in \mathbb{N}_0$, the n-th Hermite polynomial H_n is defined by $H_0 \equiv 1$, $H_{-1}(x) = 0$ and

$$H_n(x) := (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n}(e^{-\frac{x^2}{2}}) \tag{B.42}$$

for $n \geq 1$.

For example, consider the function $F(t, x) := exp(tx - \frac{t^2}{2})$, then the Hermite polynomial are the coefficients in the power series expansion of F with respect to t. Indeed, we have

$$\begin{aligned} F(t,x) &= exp(\frac{x^2}{2} - \frac{1}{2}(x - t^2)) \\ &= e^{\frac{x^2}{2}} \sum_{n=0}^{\infty} \frac{t^n}{n!} \frac{d^n}{dt^n} e^{-\frac{(x-t)^2}{2}} |_{t=0} \\ &= \sum_{n=0}^{\infty} t^n \frac{(-1)^n}{n!} e^{\frac{x^2}{2}} \frac{d^n}{dz^n} e^{-\frac{z^2}{2}} |_{z=x} \\ &= \sum_{n=0}^{\infty} t^n H_n(x)/n! \end{aligned} \tag{B.43}$$

Some properties of the Hermite polynomials

Lemma B.25.1. *For $n \geq 1$, we have:*

- $H'_n(x) = nH_{n-1}(x)$
- $H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$
- *The n-th Hermite polynomial H_n is a polynomial of degree n.*

Lemma B.25.2. *It holds that for $h, g \in \mathcal{L}^2[0,T]$ with $\|h\|_2 = \|g\|_2 = 1$,*

$$\mathbb{E}H_n(\theta h)H_m(\theta g) = \begin{cases} 0 & if \quad n \neq m \\ n! < h, g >^n & if \quad n = m \end{cases} \tag{B.44}$$

Then we use the orthogonal basis $\mathcal{H}_n$ to decompose $\mathcal{L}^2(\mathcal{F}_T)$ and $\mathcal{L}^2(\Pi)$ given by

$$\mathcal{L}^2(\mathcal{F}_T) = \oplus_{n=0}^{\infty}\mathcal{H}_n \quad and \quad \mathcal{L}^2(\Pi) = \oplus_{n=0}^{\infty}\mathcal{H}_n(\Pi). \tag{B.45}$$

Then insert Gaussian random variables into polynomials.

Definition B.26. *Let W be an H-isonormal Gaussian process. The n-th Wiener chaos $\mathcal{H}_n$ is the closure in $L^2(\Omega, \Sigma, \mathbb{P})$ of the linear span of the set $\{H_n(W(h)) : h \in H, \|h\| = 1\}$.*

Appendix C

List of Publications

Journal Papers

1. **J. Dong**, S. M. Djouadi and T. Kuruganti, "Optimal Production Control for Home Energy Management in a Micro Grid," (in preparation).
2. **J. Dong** and S. M. Djouadi, "Optimal Distributed Mixed $\mathcal{L}^2/\mathcal{H}^\infty$ control Synthesis for Spatially Invariant Systems," *IEEE Transactions on Automatic Control* (in preparation).
3. S. M. Djouadi and **J. Dong**, "On the Distributed Control of Spatially Invariant Systems," *IEEE Transactions on Automatic Control* (in preparation).
4. S. M. Djouadi and **J. Dong**, "Operator theoretic approach to the optimal distributed control problem for spatially invariant systems," *IEEE Transactions on Automatic Control* (in preparation).
5. I. Sharma, **J. Dong**, A. Andreas, M. Street, J. Ostrowski, T. Kuruganti and R. Jackson, "Energy Management System for Smart Operations of a Residential Building", *Energy and Buildings*, 2016 (Accepted).
6. X. Shi, **J. Dong**, S. M. Djouadi, Y. Wang, X. Ma and Y. Feng, "PAPMSC: Power-Aware Performance Management for Web Applications on Virtualized Servers via Stochastic Control," *Journal of Grid Computing*, pp. 1-21, 2015.

7. **J. Dong**, X. Ma, S. M. Djouadi, H. Li, and Y. Liu, "Frequency Prediction of Power Systems in FNET based on State Space Approach and Uncertain Basis Functions," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2602-2612, 2014.
8. A. Wu, **J. Dong** and G. Duan, "Robust H-infinity estimation for linear time-delay systems: An improved LMI approach", *International Journal of Control, Automation and Systems*, vol. 7, no. 4, pp. 668-673, 2009.
9. A. Wu, G. Duan, **J. Dong** and Y. Fu, "Design of proportional-integral observers for discrete-time descriptor linear systems", *IET Control Theory & Applications*, vol. 3, no. 1, pp. 79-87, 2009.

Conference Proceedings

1. **J. Dong**, S. M. Djouadi, H. Li and P. T. Kuruganti, "Short-term Solar Irradiation prediction based on uncertain basis function," in *Proc. IEEE International Conference on Smart Grid Communications (SmartGridComm), 2016* (Submitted).
2. **J. Dong** and S. M. Djouadi, "Distributed Mixed $\mathcal{L}^2/\mathcal{H}^\infty$ control problem Synthesis for Spatially Invariant Systems," *IEEE Conference on Decision and Control (CDC)*, 2016 (Submitted).
3. **J. Dong**, A. Malikopoulos, S. M. Djouadi and T. Kuruganti, "Application of Optimal Production Control theory for Home Energy Management in a Micro Grid," in *Proc. IEEE American Control Conference (ACC), 2016*.
4. **J. Dong**, C. Winstead, S. M. Djouadi, J. Nutaro and T. Kuruganti, "Occupancy-based Optimal Stochastic Control for HVAC Systems in Energy Efficient Buildings," *4th International High Performance Buildings Conference at Purdue), 2016*.
5. M. Mahmoudi, **J. Dong**, K. Tomsovic and S. M. Djouadi, "Application of Distributed Control to Mitigate Disturbance Propagations in Large Power

Networks," in *North American Power Symposium (NAPS), 2015*, vol., no., pp.1-6, 4-6 Oct. 2015.

6. X. Ma, **J. Dong**, S. M. Djouadi, J. Nutaro and P. T. Kuruganti, "Stochastic Control of Energy Efficient Buildings: A Semidefinite Programming Approach", in *Proc. IEEE SmartGridComm*, 2015.
7. S. M. Djouadi and **J. Dong**, "On the Distributed Control of Spatially Invariant Systems", in *IEEE Conference on Decision and Control (CDC)*, 2015.
8. S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska and **J. Dong**, "Finite Energy and Bounded Actuator Attacks on Cyber-Physical Systems", in *Proc. IEEE European Control Conference (ECC)*, 2015.
9. S. M. Djouadi and **J. Dong**, "Operator theoretic approach to the optimal distributed control problem for spatially invariant systems", in *Proc. IEEE American Control Conference (ACC)*, pp. 2613-2618, 2015.
10. X. Shi, **J. Dong**, S. M. Djouadi, Y. Wang, X. Ma and Y. Feng, "Power-efficient resource management for co-located virtualized web servers: A stochastic control approach," in *Proc. IEEE International Green Computing Conference (IGCC)*, pp. 1-9, 2014.
11. S. M. Djouadi and **J. Dong**, "Duality of the optimal distributed control for spatially invariant systems," in *Proc. IEEE American Control Conference (ACC)*, pp. 2214-2219, 2014.
12. S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska and **J. Dong**, "Finite Energy and Bounded Attacks on Control System Sensor Signals," in *Proc. IEEE American Control Conference (ACC)*, pp. 1716-1722, 2014.
13. **J. Dong**, S. M. Djouadi, J. Nutaro and P. T. Kuruganti "Secure Control Systems with Application to Cyber-Physical Systems", in *ACM 9th Annual Cyber and Information Security Research (CISR) Conference*, pp. 9-12, 2014.
14. **J. Dong**, X. Ma, S. M. Djouadi, H. Li and P. T. Kuruganti, "Real-time prediction of power system frequency in FNET: A state space approach,"

in *Proc. IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 109-114, 2013.

15. S. Sahyoun, **J. Dong** and S. M. Djouadi. "Reduced Order Modeling for Fluid Flows Based on Nonlinear Balanced Truncation," in *Proc. IEEE American Control Conference (ACC)*, pp. 1284-1289, 2013.

16. S. Sahyoun, **J. Dong** and S. M. Djouadi. "Reduced order modeling for fluid flows subject to quadratic type nonlinearities," in *Proc. IEEE 51st Conference on Decision and Control (CDC)*, pp. 961-966, 2012.

Vita

Jin Dong was born on the 8th of February 1988 in a beautiful city named Yangzhou which is a prefecture-level city in central Jiangsu, China. He grew up with his grandparents and parents and had a happy childhood in the exquisite, vigorous ancient city. After completing his high school education in Yangzhou High School of Jiangsu Province, in September 2006, he went to Harbin Institute of Technology (HIT) in Heilongjiang province and received his B.S. from Harbin Institute of Technology, Harbin, China, in 2010. He is mentored by Dr. Seddik. M. Djouadi towards the Ph.D. degree in the Department of Electrical Engineering and Computer Science (EECS) at the University of Tennessee, Knoxville. He was the recipient of Min Kao Fellowship from 2010 to 2013. Jin will join the Whole-Building and Community Integration (WBCI) Group at Oak Ridge National Laboratory as a R&D staff member.