CLOSED-LOOP ANALYSIS AND FEEDBACK DESIGN IN THE PRESENCE OF LIMITED INFORMATION

BY

DAPENG LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mechanical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Naira Hovakimyan, Chair
Professor P.R. Kumar
Professor Geir Dullerud
Professor Prashant Mehta

ABSTRACT

Recent progress in communication technologies and their use in feedback control systems motivate to look deeper into the interplay of control and communication in the closed-loop feedback architecture. Among several research directions on this topic, a great deal of attention has been given to the fundamental limitations in the presence communication constraints. Entropy rate inequalities corresponding to the information flux in a typical causal closed loop have been derived towards obtaining a Bode-like integral formula.

This work extends the discrete-time result to continuous-time systems. The main challenge in this extension is that Kolmogorov's entropy rate equality, which is fundamental to the derivation of the result in discrete-time case, does not hold for continuous-time systems. Mutual information rate instead of entropy rate is used to represent the information flow in the closed-loop, and a limiting relationship due to Pinsker towards obtaining the mutual information rate between two continuous time processes from their discretized sequence is used to derive the Bode-like formula. The results are further extended to switched systems and a Bode integral formula is obtained under the assumption that the switching sequence is an ergodic Markov chain. To enable simplified calculation of the resulting lower bound, some Lie algebraic conditions are developed.

Besides analysis results, this dissertation also includes joint control/communication design for closed-loop stability and performance. We consider the stabilization problem within Linear Quadratic Regulator framework, where a control gain is chosen to minimize a linear quadratic cost functional while subject to the input power constraint imposed by an additive Gaussian channel which closes the loop. Also focused on Gaussian channel, the channel noise attenuation problem is addressed, by using H-infinity/H2 methodology. Similar feedback optimal estimation problem is solved by using Kalman filtering theory.

ACKNOWLEDGMENTS

I would like to thank Prof. Naira Hovakimyan for her outstanding advisorship over these years from Virginia Tech to University of Illinois. Her appreciation of mathematical rigor as well as the deep understanding of various engineering problems has been infused into my blood. Prof. Hovaikmyan's uncompromising insistence on the academic righteousness and excellency also has equipped me with a strong mind ready for future challenges. She is the one who best understands and appreciates my work, gives me a strong support and motivates my further scientific exploration.

I also would like to thank Prof. P.R. Kumar, Geir Dullerud and Prashant Mehta for serving as my Ph.D. committee members. They have been nice and helpful during the whole course of my Ph.D. study. Their comments on my work have given me many opportunities to retrospect and learn more.

My gratitude also goes to Prof. Chengyu Cao, whose engineering insight helped me with solutions to some meaningful problems. I am also indebted to Prof.Tryphon T. Georgiou for his input to our work on adaptive control and information theory.

I would like to express my gratitude to my fellow group members, they are Xiaofeng Wang, Ronald Choe,Enric Xargay, Hui Sun, Zhiyuan Li, Vervoorst, Jan, Evegeny Karisov, Justin Vanness, Mohammad Naghnaeian and David Caballero.

To my parents Xinmin Li and Meili Huang, you are the most important part in my life. The dissertation would not have been possible without your selfless love. My sisters Yan Li and Xiaoli Li have also been giving me persistent support over the years. Thank you all, my family members!

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Control theory explores the feedback structures and uses them to design feedback controllers to achieve desired closed-loop behaviors. Information theory, which was developed slightly later than control theory, deals with information compression and transmission with or without loss. These two seemingly distinct disciplines, however, are deeply related. In fact, their intrinsic relationship has been exploited ever since their inception. Wiener, one of the founding fathers of control theory, succinctly defined cybernetics as "*the study of communication and control in the animal and the machine*" [1], where the role of communication of information was explicitly pointed out. On the other hand, in [2] Shannon made the following comment regarding the possible usefulness of feedback in reliable communications "*. . . can be pursued further and is related to a duality between past and future and the notions of control and knowledge. Thus we may have knowledge of the past and cannot control it; we may control the future but have no knowledge of it.*".

Recently, a renewed interest of studying the relationship between the two subjects has been stimulated by the need for understanding and developing new technologies that merge control, communication and computation, [3]. For example, when multiple actuators and sensors are present in a complex control system in a distributed fashion, where wired networks are being replaced by wireless networks, the communication among the elements cannot be simply ignored. A set of nontrivial questions can be therefore formulated related to the communication limitations. A basic one is: under certain information patterns, what is the lower bound for the channel capacity to guarantee the closed-loop stability. In addition, more questions can be raised if the performance and robustness of the closed loops are also of interest. Results can be also developed on the information theory& communication side. Though feedback is not able to increase the capacity of communication channels significantly [4], it significantly simplifies the coding schemes with stronger reliability guarantees. With feedback being cheaply and reliably implemented, recent research holds a great promise for improved performance in modern communication systems. Rather than benefiting control/communication design, the unification of information theory and control theory enables a fresh perspective on complex and highly connected systems, which are ubiquitous in biological and social networks, [5].

In this dissertation, the main focus is on:

- *Obtaining Bode-type fundamental limitation results for continuous-time as well as discrete-time stochastic switched plants by using information theoretic machineries;*
- *Control and feedback estimation design in the presence of communication limitations for real-time as well as stationary closed-loop systems.*

1.1 Chapter 2 Bode's integral in with limited information

1.1.1 Problem Formulation

We consider the following closed loop in the presence of disturbance. Under

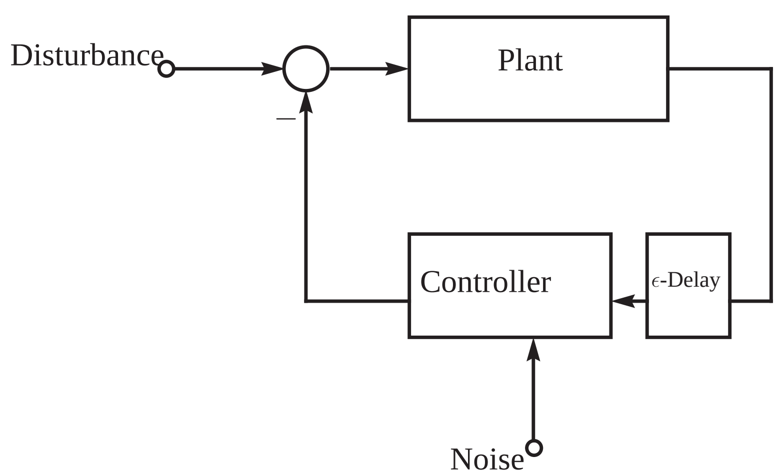

Figure 1.1: A Feedback Closed Loop with Disturbance

the assumption that both the plant and the controller are linear time-invariant and the loop-transfer function $L(s)$ has relative degree at least 1, a log integral is obtained [6]:

$$\frac{1}{2\pi}\int_0^\infty \log|S(j\omega)| = \sum_i p_i\,,$$

where S is the sensitivity transfer function, and p_i represents the open-loop unstable eigenvalues.

However, when the linearity and the deterministic nature of the system dynamics are removed, such a relationship may fail to hold. Therefore, the objective of this research is to establish a similar relationship in a general setting, where information theoretic quantities like entropy and mutual information are expected to play a major role.

1.1.2 Literature Review

Most of the previous results on the intersection of control theory and information theory are derived for discrete-time dynamical systems. In this chapter, we investigate continuous-time systems for the following reasons. First, a large number of real-life systems are continuous-time in nature, and therefore it is of interest to develop the corresponding continuous-time tools for closed-loop analysis. Second, although digital channels dominate almost all communication systems, some continuous-time models such as continuous-time Additive Gaussian White Noise (AWGN) channels attract significant attention because of their theoretical simplicity [7, 8]. From technique perspective of view, the continuous-time case imposes challenges for both control theory and information theory. As for control, except for the classical Bode's result and its extensions [9], where Bode's integral formulae for continuous-time and discrete-time are bridged by Poisson's integral formula, there is no similar mathematical tool available yet for the general setting. As for information theory, we point out that the results in [10] and [11], together with several others [12–14], rely heavily upon the following entropy rate equality originated by Kolmogorov [15]:

$$\bar{h}(\xi) = \log(2\pi\sqrt{e}) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_\xi(\lambda)d\lambda\,, \tag{1.1}$$

where ξ is a discrete-time stationary process, $\bar{h}$ stands for the entropy rate, and f_ξ is the spectral density function of ξ. This formula, however, is only applicable to discrete-time processes, and its continuous-time extension *has to be derived otherwise* [16]. However, no such extension has been carried out since Kolmogrov's comment because of the undesirable behavior of differential entropy rate for continuous-time processes.

1.1.3 Main Contribution

In this chapter, we attempt to use tools from information theory to analyze performance limitations for continuous-time systems with stochastic disturbances. We first derive the mutual information rate inequality by assuming causality of the closed-loop system. A Bode-type formula is then obtained to address the fundamental limitation of the stabilization problem in frequency domain. The techniques utilized here are different from discrete-time case in that: 1. Mutual information rate instead of entropy rate is adopted to represent the information flow in a closed-loop; 2. To get the Bode-type integral, we use the result from [17], which helps to circumvent Kolmogorov's formula (1.1). To get insight into the resulting Bode's integral, we employ tools from complex analysis to identify an extra term of performance limitation induced by the controller/channel noise. We also quantify the negative portion of the Bode's integral and relate it to closed-loop communication constraint. Finally we apply this framework to communication–control interconnection to study the relationship between the channel capacity and the stability of the closed-loop systems.

1.2 Chapter 3: Bode's Integral For Stochastic Switched Systems

1.2.1 Problem Formulation

We consider the closed loop Fig. 1.2, where the plant is switching among finite modes.

The objective is to derive a Bode-type formula by using information theory. The statistical properties of the switching signal contribute significantly to the closed-loop performance and need to be quantified explicitly.

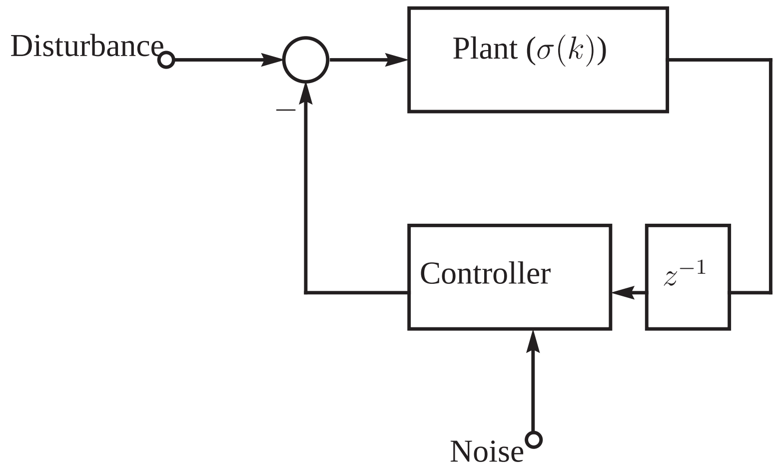

Figure 1.2: A Feedback Closed Loop with Disturbance and Plant Switching

1.2.2 Literature Review

While switched control systems have been studied from various perspectives [18], it is still not clear how to characterize their fundamental limitations within an appropriate framework. The problem becomes especially challenging, when such closed loops are further subject to communication constraints. A notable effort was made in [19], where the authors consider the stabilization problems and derive the lower bound of the required data-rate.

In economics, typical dynamic programming problems in macroeconomics are considered with a mutual information type of constraint, which is regarded as an appropriate model of *rational inattention* [20]. Rational inattention is the lack of infinite capability of receiving and passing information for economic entities, individuals and firms. The limited information processing capability contributes to many aspects of economic fluctuations. For policy makers, rational inattention is an especially important factor, when curial monetary policies are craft. To evaluate the consequence of the different policies, a recently developed frequency domain approach in terms of a Bode's integral, is appealing for its simplicity and novelty [21].

1.2.3 Main contribution

In this chapter, we extend the framework from [10] to closed loops with stochastic switched plants. We address the problem by using an information theoretic framework towards obtaining a Bode integral formula, under the assumptions that the switching sequence is an ergodic Markov chain. We first derive a closed-loop information conservation law by using information theoretic arguments similar to [22] and [10]. Then, under some stationarity assumption, a Bode integral-like theorem is obtained, characterizing a lower bound on the performance limitations. To enable the simplified calculation of the resulting lower bound, some Lie algebraic conditions are developed.

To demonstrate the usefulness of the theoretical result, we propose two different examples. The first one is NCS with random packet dropouts, which has been widely used in control literature to model typical computer network protocols, such as TCP and UDP [23]. We develop a Bode integral to show that the degree of instability of the plants determines the lower bound of the performance limitation.

The second potential illustration is in the field of macroeconomics, where feedback is used to generate optimal policies with respect to certain criteria. We apply Bode's integral to propose a simple frequency domain method for optimal monetary policy evaluation under a regime of switching economy. Furthermore, we extend the method to enable visualization of the impact of individual's limited information processing capability on the policy design limits. The content of the chapter is reported in [24].

1.3 Chapter 4: Continuous Time Linear Quadratic Design

1.3.1 Problem Formulation

In this chapter, we consider the control design problem with limited information. More specifically, we formulate the problem in the Linear Quadratic Regulation framework, where the state-control minimizes a infinite quadratic functional, while subject to the power constraint imposed by an additive Gaussian channel in the closed loop.

1.3.2 Literature Review

In most of the previous work, plants and communication channels are modeled as discrete-time systems, since discrete-time models well fit the digital communication channels. Nevertheless, it is still worth investigating the continuous-time systems, since many plants to be controlled are continuous-time in nature. Furthermore, as pointed out in [25], a number of communication channels in practice could be conveniently modeled as continuous-time additive Gaussian channels (AGC). Some recent effort has been made towards this direction, among which [25] has provided if and only if conditions for observability and stabilizability of LTI systems over a class of Gaussian channels. Reference [26] proposes a method of obtaining a tight upper bound on SNR based on H_2 control type argument.

The communication constrained LQG problems have also been addressed in [27] in discrete time, where the communication channel is modeled as a finite rate quantization. For the case of additive Gaussian channels, a simple scalar case was considered in [28].

1.3.3 Main Contribution

This chapter is to investigate the continuous-time linear quadratic regulator control problem over an additive white Gaussian noise (AWGN) channel with input power constraint. A new framework based on stochastic differential equations(SDE) is established to address both the plant and the channel dynamics, which are introduced by the noise of the channel with some randomness. Within the framework, an LMI convex optimization problem is proposed to calculate the controller parameters.

1.4 Chapter 5: Noise Attenuation Over Additive Gaussian Channels

1.4.1 Problem Formulation

While Shannon's theory solves the information transmission problem with arbitrary accuracy (probability of error), the communication channels in control systems may not share the same feature because the accuracy of reconstruction of messages needs a certain amount of time, which is not tolerable for control systems, especially when certain performances need to be achieved timely. It is then reasonable to assume that the channel noise propagates into the systems, and a controller should be able to cope with the disturbance noise. In this chapter, we consider a state feedback control problem with input power containt for the channel input.

1.4.2 Main Contribution

In this chapter we propose a new control design strategy to address the stabilization and the noise attenuation problems in AWGN channels. The solution turns out to fit into the mixed $\mathcal{H}_\infty/\mathcal{H}_2$ framework. The design approach is based on linear matrix inequalities (LMI). The LMI solution gives more computational efficiency, and it also avails a possibility of dealing with multiple-input-multiple-output (MIMO) channels.

1.5 Chapter 6: Optimal State Estimation Over Gaussian Channels with Noiseless Feedback

1.5.1 Problem Formulation

The scheme is depicted in Fig. 1.3 where the transmitter has access to the time-history of the channel output via a noiseless feedback.

A transmitter and an estimator need to be designed to estimate the state of a possibly unstable linear dynamics, while achieving mean square optimality.

1.5.2 Literature Review

Gaussian channel and its variants have been one of the central topics in information and communication theory for their capability of capturing several important aspects of real-life communication systems. To consider the relationship between control and communication, Gaussian channels are also a popular choice.

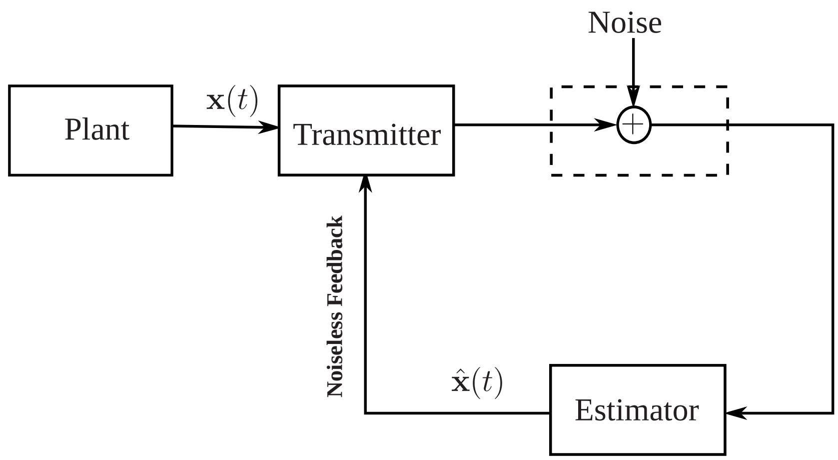

Figure 1.3: State Estimation via Noiseless Feedback

Ref. [29] has captured the relation between the state (output) feedback stabilization of a linear time-invariant (LTI) system and the signal-to-noise ratio (SNR) constraint of the channel for both continuous-time and discrete-time cases; [27] and [30] have considered the linear quadratic Gaussian framework to derive the data-rate bound and provide a fairly complete scheme for design of the encoder, the controller and the decoder. In [10], Gaussianity plays an important role in obtaining the Bode's integrals in terms of log integral of relevant power spectral densities in the closed loop.

The state estimation under communication limitations has been investigated for its close relationship with controls as well as its own importance. References [31]and [32] tried to fit the problem into the framework developed in [10] and [11] with the hope to use the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control theory in this context. In a more general setting, feedback has long been used to improve the performance of the communication systems in terms of better convergence rate of the error probability. In the discrete-time setting in case of additive white gaussian noise (AWGN) channel, inspired by Robbins-Monro stochastic iterative root seeking algorithm from [33], S-K feedback coding is presented [34]. A large number of results followed this seminal work along with various of extensions. Recently, this classical result caught much attention from control community, starting from [12], which linked the optimal estimation with optimal encoding/decoding, with a fundamental observation unifying control, estimation and communication (see also [35]). Another similar development from the information theory perspective is reported in [22], where colored gaussian channel with the capacity of coding is discussed in a fairly general setting. The continuous-time version of S-K scheme is presented in [36], where the derivation heavily relies on the stochastic calculus and optimal filtering theory.

1.5.3 Main Contribution

The objective of this chapter is to solve the continuous-time optimal estimation problem in the presence of an AWGN channel with an input power constraint. The contribution of the chapter is three-fold:

- It establishes a framework to analyze some important quantities in a stable closed loop, such as minimal mean-square error (MMSE) and channel capacity (or signal to noise ratio), where stationarity is not assumed;
- Based on this framework, we not only recover the existing relation between channel capacity and the open-loop instability in stable closed loops, but also provide a tighter bound to guarantee an exponentially decaying mean square of estimation error.
- The detailed procedure and algorithms are provided for the transmitter and estimator design, together with the rigorous proof of optimality.

CHAPTER 2

BODE-LIKE INTEGRAL FOR CONTINUOUS-TIME CLOSED-LOOP SYSTEMS IN THE PRESENCE OF LIMITED INFORMATION

The chapter is organized as follows. In Section 2.1 we introduce the closed-loop feedback configuration and some basic definitions and facts from information theory and the theory of stochastic processes. Section 2.2 studies a general feedback scheme, within which we develop a mutual information inequality and a Bode-type integral formula. Section 2.3 further explores the relation of Bode's integral with the information transmission rate of the closed loop, while Section 2.4 carries out the in-depth analysis of the the Bode-type integral by using complex integration techniques. The paper is concluded in Section 2.6. We note that Sections 2.4, 2.3 and 2.5 are developed in somewhat parallel manner, and the reader should not be surprised to find forward cross-referencing among these sections.

2.1 Preliminaries

Notation:

- $\mathbb{R}$ denotes the field of real numbers; $\mathbb{C}$ stands for complex plane; $\mathbb{C}^-$ and $\mathbb{C}^+$ stand for the left half and right half of $\mathbb{C}$ respectively.

- Random variables defined in appropriate probability spaces are represented using boldface letters, such as $\mathbf{x}$, $\mathbf{y}$. If not otherwise stated, the random variables take values in $\mathbb{R}$ throughout the chapter.

- If $\mathbf{x}(k), k \in \mathbb{N}^+$, is a discrete time stochastic process, we denote its segment $\{\mathbf{x}(k)\}_{k=l}^{u}$ by $\mathbf{x}_l^u$, and use $\mathbf{x}_0^n := \mathbf{x}^n$ for simplicity.

- Consider a continuous time stochastic process $\mathbf{x}(t), t \in \mathbb{R}^+$. A sample path on an interval $[t_1, t_2),\ 0 \leq t_1 < t_2 \leq +\infty$, is indicated as $\mathbf{x}_{t_1}^{t_2}$. We also denote $\mathbf{x}_0^t := \mathbf{x}^t$ for simplicity.

- $\mathbf{x}^{(h)}$ is the discrete-time process obtained from sampling of $\mathbf{x}(t)$ on $t \in [t_1, t_2)$ with an interval $h > 0$. We denote $\mathbf{x}_i^{(h)} = \mathbf{x}^{(h)}(i) := \mathbf{x}(t_1 + ih),\ \ i = 0, 1, ...$

- The probability density (if it exists) of a random variable $\mathbf{x}$ is represented as $p_{\mathbf{x}}$.

- $\mathbf{E}[\cdot]$ is the expectation operator of a random variable.

- $(\cdot)^+ = \max\{\cdot, 0\}$ and $(\cdot)^- = \min\{\cdot, 0\}$.

- $\Re(\cdot)$ gives the real part of a complex number.

- $\lambda_i(\cdot)$ gives the eigenvalues of a square matrix.

- $Re(\cdot; z)$ gives the residue of a analytical function about $z \in \mathbb{C}$.

In this section, several basic definitions and related facts from information theory and stochastic processes are introduced. We rely on [4] an [37] as main references.

2.1.1 Entropy, Mutual Information and Related Facts

In this subsection, we introduce some elementary definitions and results from information theory, most of which are taken from [4].

Definition 2.1.1 (*Differential Entropy*). The *differential entropy* of a continuous random variable $\mathbf{x}$ with density $p_{\mathbf{x}}$ is defined as

$$h(\mathbf{x}) := -\mathbf{E}[\log p_{\mathbf{x}}] = -\int_{\mathbb{S}} p_{\mathbf{x}} \log p_{\mathbf{x}} d\mathbf{x}\,, \tag{2.1}$$

where $\mathbb{S}$ is an abstract space where the random variable $\mathbf{x}$ is defined.

Definition 2.1.2 (*Conditional Entropy*). If there are two random variables $\mathbf{x}$ and $\mathbf{y}$, the conditional entropy $h(\mathbf{x}|\mathbf{y})$ is defined as

$$h(\mathbf{x}|\mathbf{y}) := -\int_{\mathbb{S}^2} p_{\mathbf{xy}} \log p_{\mathbf{x}|\mathbf{y}} d\mathbf{x} d\mathbf{y} \tag{2.2}$$

Definition 2.1.3 (*Joint Entropy*). The entropy of the random vector $\mathbf{x}^n := \mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n$, comprised of random variables with density $p_{\mathbf{x}^n}$, is defined as

$$h(\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n) := -\mathbf{E}[\log p_{\mathbf{x}^n}] = -\int_{\mathbb{S}^n} p_{\mathbf{x}^n} \log p_{\mathbf{x}^n} d\mathbf{x}^n \tag{2.3}$$

Definition 2.1.4 (*Mutual Information*). The mutual information between the two random variables $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$I(\mathbf{x};\mathbf{y}) := -\mathbf{E}_{\mathbf{xy}}\left[\log \frac{p_{\mathbf{xy}}}{p_{\mathbf{x}}p_{\mathbf{y}}}\right] = -\int_{\mathbb{S}^2} p_{\mathbf{xy}} \log \frac{p_{\mathbf{xy}}}{p_{\mathbf{x}}p_{\mathbf{y}}} d\mathbf{x}d\mathbf{y} \tag{2.4}$$

Definition 2.1.5 (*Conditional Mutual Information*). The mutual information between the two random variables $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\begin{aligned} I(\mathbf{x};\mathbf{y}|\mathbf{z}) &:= -\mathbf{E}_{\mathbf{xyz}}\left[\log \frac{p_{\mathbf{xy}|\mathbf{z}}}{p_{\mathbf{x}|\mathbf{z}}p_{\mathbf{y}|\mathbf{z}}}\right] \\ &= -\int_{\mathbb{S}^3} p_{\mathbf{xyz}} \log \frac{p_{\mathbf{xy}|\mathbf{z}}}{p_{\mathbf{x}|\mathbf{z}}p_{\mathbf{y}|\mathbf{z}}} d\mathbf{x}d\mathbf{y}d\mathbf{z} \end{aligned} \tag{2.5}$$

Definition 2.1.6 (*Joint Mutual Information*). The joint mutual information between n dimensional vectors $\mathbf{x}^n := \mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n$ and $\mathbf{y}^n := \mathbf{y}_0, \mathbf{y}_1, ..., \mathbf{y}_n$ is defined as

$$\begin{aligned} I(\mathbf{x}^n;\mathbf{y}^n) &= -\mathbf{E}_{\mathbf{x}^n\mathbf{y}^n}\left[\log \frac{p_{\mathbf{x}^n\mathbf{y}^n}}{p_{\mathbf{x}^n}p_{\mathbf{y}^n}}\right] \\ &= -\int_{\mathbb{S}^{2n}} p_{\mathbf{x}^n\mathbf{y}^n} \log \frac{p_{\mathbf{x}^n\mathbf{y}^n}}{p_{\mathbf{x}^n}p_{\mathbf{y}^n}} d\mathbf{x}^n d\mathbf{y}^n \end{aligned} \tag{2.6}$$

Definition 2.1.7. [Entropy Rate] The entropy rate of $\mathbf{x}$ is defined as

$$\bar{h}(\mathbf{x}) := \lim_{n\to\infty} \frac{h(\mathbf{x}^n)}{n+1}\,, \tag{2.7}$$

given the existence of the limit.

Definition 2.1.8 (*Mutual Information Rate*). The mutual information rate of two stochastic processes is defined as

$$\bar{I}(\mathbf{x};\mathbf{y}) := \lim_{n\to\infty} \frac{I(\mathbf{x}^n;\mathbf{y}^n)}{n+1}\,, \tag{2.8}$$

given the existence of the limit.

To consider the information between two continuous-time stochastic processes we introduce the following definition.

Definition 2.1.9 (*Mutual Information of Continuous Processes*). The mutual information between two stochastic processes $\mathbf{x}$ and $\mathbf{y}$ on time interval $[s,t)$, $0 \leq s \leq t < \infty$, is defined as

$$I(\mathbf{x}_s^t; \mathbf{y}_s^t) := \int \log \frac{dP_{\mathbf{x}_s^t, \mathbf{y}_s^t}}{dP_{\mathbf{x}_s^t} \times dP_{\mathbf{y}_s^t}} dP_{\mathbf{x}_s^t, \mathbf{y}_s^t}\,, \tag{2.9}$$

where $P_{\mathbf{x}_s^t}$, $P_{\mathbf{y}_s^t}$ and $P_{\mathbf{x}_s^t, \mathbf{y}_s^t}$ are the probability measures, induced by random objects $\mathbf{x}_s^t$, $\mathbf{y}_s^t$ and $(\mathbf{x}_s^t, \mathbf{y}_s^t)$ respectively, and $\frac{dP_{\mathbf{x}_s^t, \mathbf{y}_s^t}}{dP_{\mathbf{x}_s^t} \times dP_{\mathbf{y}_s^t}}$ is the Radon-Nikodym derivative, given that $P_{\mathbf{x}_s^t, \mathbf{y}_s^t}$ is absolutely continuous with respect to the product measure $P_{\mathbf{x}_s^t} \times P_{\mathbf{y}_s^t}$.

Similar to Definition 2.1.8, we define the *information rate* for continuous-time processes.

Definition 2.1.10 (*Information Rate*). The information rate is given by

$$\bar{I}(\mathbf{x}; \mathbf{y}) := \lim_{T \to \infty} \frac{I(\mathbf{x}^T; \mathbf{y}^T)}{T}\,, \tag{2.10}$$

given the existence of the limit.

In (2.10), $\bar{I}$ could be viewed as the rate of mutual information for reliable transmission through any communication channel ($\mathbf{x}$ as input and $\mathbf{y}$ as output or vice versa).

Remark 2.1.11. It is worth mentioning that, according to convention, we avoid the notion of differential entropy $h(\cdot)$ for a segment of a continuous time process, because h can be infinite for certain processes, as shown in the following example.

Example 2.1.12. Let $\mathbf{w}(t)\,, t \in \mathbb{R}^+$, be a zero-mean white Gaussian noise process with unit variance. It is straightforward to see that $\mathbf{w}$ is an individually and identically distributed (i.i.d) process in continuous time. We take $N+1$ samples over the interval $[0,1)$ denoted as $\mathbf{w}_0, \mathbf{w}_1, ..., \mathbf{w}_N$. It is straightforward to see that $h(\hat{\mathbf{w}}_0, \hat{\mathbf{w}}_1, ...\hat{\mathbf{w}}_N) = \frac{N+1}{2} \log 2\pi e$, and from the fact that $\mathbf{w}_1, ..., \mathbf{w}_N$ is a function of $\mathbf{w}_0^1$ we have

$$h(\mathbf{w}_0^1) \geq \lim_{N \to \infty} h(\hat{\mathbf{w}}_0, \hat{\mathbf{w}}_1, ..., \hat{\mathbf{w}}_N) = \infty\,.$$

Therefore the counterpart of definition (2.1.7) in continuous time does not exist.

The next lemma gives the opportunity to represent continuous time mutual information as the limit of its discretized version.

Lemma 2.1.13. Consider separable stochastic processes $\mathbf{x}$ and $\mathbf{y}$. The mutual information between $\mathbf{x}_s^t$ and $\mathbf{y}_s^t$, $0 \leq s < t < \infty$, can be obtained as

$$\begin{aligned} & I(\mathbf{x}_s^t; \mathbf{y}_s^t) = \lim_{n \to \infty} I(\mathbf{x}_0^{(\delta(n))}, ..., \mathbf{x}_n^{(\delta(n))}; \mathbf{y}_0^{(\delta(n))}, ..., \mathbf{y}_n^{(\delta(n))})\,, \\ & \mathbf{x}_i^{(\delta(n))} = \mathbf{x}(s + i\delta(n))\,, i = 0, 1, ... \end{aligned} \tag{2.11}$$

for any fixed s and t with $\delta(n) = \frac{t-s}{n+1}$.

The proof of this lemma is given in 2.7.

This lemma is used successfully in [38] to connect discrete-time results with continuous-time ones regarding the channel sensitivity. The inherent sampling type of argument in the lemma *permits the general information measures to inherent many of its properties from the simpler discrete-time case* [39]. It will also serve as an important tool to obtain the main result. A list of useful properties of entropy and mutual information are given here, and are frequently used in the upcoming arguments.

(P1) *Symmetry and nonnegativity*:

$$I(\mathbf{x}; \mathbf{y}) = I(\mathbf{y}; \mathbf{x}) = h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \geq 0\,.$$

(P2) *Kolmogorov equality*:

$$I(\mathbf{x}; (\mathbf{y}, \mathbf{z})) = I(\mathbf{x}; \mathbf{z}) + I(\mathbf{x}; \mathbf{y}|\mathbf{z})$$

(P3) *Data processing inequality*:

$$I(\mathbf{x}; \mathbf{y}|\mathbf{z}) \geq I(\mathbf{x}; g(\mathbf{y})|\mathbf{z})$$

The equality holds, if $g(\cdot)$ is invertible.

(P4) *Invariance of mutual information (entropy)*

$$I(\mathbf{x}; \mathbf{y}|\mathbf{z}) = I(\mathbf{x} + g(\mathbf{z}); \mathbf{y}|\mathbf{z})\,, h(\mathbf{x}|\mathbf{z}) = h(\mathbf{x} + g(\mathbf{z})|\mathbf{z}),$$

where $g(\cdot)$ is a function.

(P5) *Chain rule*:

$$h(\mathbf{x}^n|\mathbf{y}) = \sum_{k=1}^{n} h(\mathbf{x}_k|\mathbf{y}, \mathbf{x}^{k-1})$$

(P6) *Maximum entropy*: Consider $\mathbf{x} \in \mathbb{R}^m$ and the covariance matrix given by $V := \mathbf{E}[\mathbf{x}\mathbf{x}^\top]$. Then we have

$$h(\mathbf{x}) \le h(\bar{\mathbf{x}}) = \frac{1}{2}\log((2\pi e)^m \det V)\,,$$

where $\bar{\mathbf{x}}$ is a Gaussian process with the same covariance as $\mathbf{x}$. Equality holds, if $\mathbf{x}$ is Gaussian.

2.1.2 Spectral Analysis of Stationary Stochastic Processes

Here we introduce some results related to the spectral theory of stationary processes.

Definition 2.1.14 (*Wide Sense Stationary Process*). A zero-mean continuous-time stochastic process $\mathbf{x}(t) \in \mathbb{R}^n$, $t \ge 0$, is stationary, if for all $t \ge 0$ its covariance function, defined by

$$R_{\mathbf{x}}(\tau) = \mathbf{E}[\mathbf{x}(t+\tau)\mathbf{x}^\top(t)], \quad \tau \in \mathbb{R}\,, \tag{2.12}$$

is independent of t. Throughout this chapter, *wide sense stationary* is abbreviated as *stationary* for convenience.

The spectral decomposition of the covariance function $R_{\mathbf{x}}(t)$ is defined via Fourier transform:

$$f_{\mathbf{x}}(\omega) = \int_0^\infty e^{-it\omega} R_{\mathbf{x}}(t)dt\,, \tag{2.13}$$

and the function $f_{\mathbf{x}}(\cdot)$ is called *power spectral density (PSD)* of $\mathbf{x}$. The stationary process $\mathbf{x}$ admits a *spectral factorization*, if

$$f_{\mathbf{x}}(\omega) = \phi_{\mathbf{x}}(-j\omega)\phi_{\mathbf{x}}(j\omega)\,,$$

for some function $\phi_{\mathbf{x}}(\cdot)$. The following lemma from [40] shows that a rational PSD always admits a rational spectral factorization.

Lemma 2.1.15. If $f_{\mathbf{x}}(\omega)$ is rational, then there exits a minimum phase and asymptotically stable LTI system $\phi_{\mathbf{x}}(s)$, such that

$$f_{\mathbf{x}}(\omega) = \phi_{\mathbf{x}}(-j\omega)\phi_{\mathbf{x}}(j\omega)$$

There are various ways to find $\phi_{\mathbf{x}}$; the reader is referred to [41] for an extensive overview.

Definition 2.1.16 (*Markov Process*). A continuous-time stochastic process $\mathbf{x}(t),\ t \in \mathbb{R}^+$, is called a Markov process, if

$$P(\mathbf{x}(t) \in A|\mathbf{x}(u), u \leq s) = P(\mathbf{x}(t) \in A|\mathbf{x}(s)) \tag{2.14}$$

holds for every $s < t$ and every measurable set $A \subset \mathbb{S}$, where $P(\mathbf{x}_t \in A|\mathbf{x}_u, u \leq s)$ denotes the conditional probability of $\{\mathbf{x}_t \in A\}$, given the knowledge of $\mathbf{x}_u, u \leq s$.

While more general definitions of Markov processes can be found in many standard stochastic process texts, we adopt this simple one to avoid complex notations requiring more background from the reader. We define class $\mathbb{F}$ functions as follows [42].

Definition 2.1.17 (*Class $\mathbb{F}$ function*).

$$\mathbb{F} = \{l : l(\omega) = p(\omega)(1 - \varphi(\omega)), l(\omega) \in \mathbb{C}, \omega \in \mathbb{R}\}\,, \tag{2.15}$$

where $p(\cdot)$ is rational and $\varphi(\cdot)$ is a measurable function, such that $0 \leq \varphi < 1$ for all $\omega \in \mathbb{R}$ and $\int_{\mathbb{R}} |\log(1 - \varphi(\omega))|d\omega < \infty$.

It is obvious that all rational functions are in $\mathbb{F}$.

The following lemma is taken from [17], which gives a lower bound on the mutual information rate of two continuous-time Gaussian stationary processes.

Lemma 2.1.18. Suppose that two one-dimensional continuous-time processes $\mathbf{x}$ and $\mathbf{y}$ form a stationary Gaussian process $(\mathbf{x}, \mathbf{y})$. Then

$$\bar{I}(\mathbf{x}, \mathbf{y}) \geq -\frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 - \frac{|f_{\mathbf{xy}}(\omega)|^2}{f_{\mathbf{x}}(\omega)f_{\mathbf{y}}(\omega)}\right) d\omega\,. \tag{2.16}$$

The equality holds, if $f_{\mathbf{x}}$ or $f_{\mathbf{y}}$ belong to the class $\mathbb{F}$.

2.1.3 Closed-Loop System

Throughout the chapter we consider the feedback configuration depicted in Fig. 2.1.

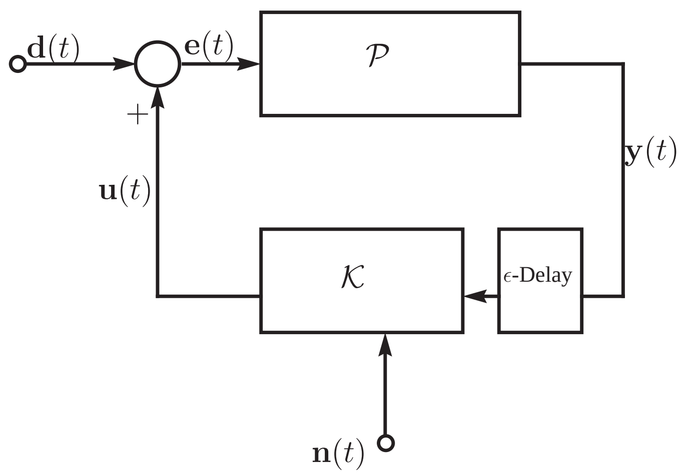

Figure 2.1: Basic Feedback Scheme

Several assumptions are made:

- The plant $\mathcal{P}$ is modeled by the following stochastic differential equation

$$\begin{aligned} \dot{\mathbf{x}}(t) &= A\mathbf{x}(t) + B\mathbf{e}(t)\,, \quad \mathbf{x}(0) = \mathbf{x}_0\,, \\ \mathbf{y}(t) &= C\mathbf{x}(t)\,. \end{aligned} \tag{2.17}$$

  Here $\mathbf{x}(t) \in \mathbb{R}^n$, and $\mathbf{x}_0$ is assumed to have finite differential entropy or $|h(\mathbf{x}_0)| < \infty$.

- An arbitrary small time-delay $\epsilon > 0$ is imposed on the output signal $\mathbf{y}$.

- The disturbance $\mathbf{d}(t)$ is a Markov process, and $\mathbf{n}(t)$ is a stochastic process that models the controller noise. We assume that $\mathbf{d}(t)$, $\mathbf{n}(t)$ and $\mathbf{x}_0$ are mutually independent.

- The controller $\mathcal{K}$ is given as a deterministic causal map such that

$$\mathcal{K} : (\mathbf{y}_0^{t-\epsilon}, \mathbf{n}_0^t) \mapsto \mathbf{u}(t)\,.$$

Definition 2.1.19 (*Sensitivity-like Function*). A sensitivity-like function of the closed loop is defined as

$$S_{\mathbf{d},\mathbf{e}}(\omega) = \sqrt{\frac{f_{\mathbf{e}}(\omega)}{f_{\mathbf{d}}(\omega)}}\,, \tag{2.18}$$

where $\mathbf{e}$ and $\mathbf{d}$ are stationary and stationarily correlated.

Remark 2.1.20. The function $S_{\mathbf{d},\mathbf{e}}(\omega)$ is the stochastic analogue of the sensitivity function $|S(j\omega)|$ in Bode's original work [43].

Throughout, we adopt the following stability definition.

Definition 2.1.21 (*Mean-square Stability*). The closed loop given in Fig. 2.1 is said to be mean-square stable, if

$$\sup_{t\geq 0} \mathbf{E}[\mathbf{x}^{\top}(t)\mathbf{x}(t)] < \infty\,. \tag{2.19}$$

2.2 Information Conservation Law and Extension of Bode's Integral Formula

As it has been revealed in [10], causality plays a central role in obtaining a Bode-type formula for a discrete-time feedback loop with stochastic disturbance. Bearing this observation in mind, we then obtain a set of mutual information rate inequalities resulting directly from the feedback structure and causality of the closed loop shown in Fig 2.1. In turn, an analogue of Bode's theorem is obtained by assuming certain stationarity and Markov properties for the disturbance signal.

To start with, we introduce the following Lemma, where the sum of all the unstable eigenvalues (or the degree of instability) of the open loop state matrix A is upper bounded by the mutual information rate between the initial value $\mathbf{x}_0$ and the error signal $\mathbf{e}$.

Lemma 2.2.1. If the closed-loop system in Fig. 2.1 is stable, then the following inequality holds

$$\bar{I}(\mathbf{x}_0;\mathbf{e}) \geq \sum_i \Re(\lambda_i(A))^+\,, \tag{2.20}$$

where $\Re(\lambda_i(A))^+ := \max\{0, \Re(\lambda_i(A))\}$.

Proof. If A is Hurwitz, then $\sum_i \Re(\lambda_i(A))^+ = 0$ and (2.20) trivially holds. In case A is not Hurwitz, it is obvious that there exists a nonsingular matrix $G \in \mathbb{R}^{n\times n}$ such that

$$G^{-1}AG = \begin{bmatrix} A_s & 0 \\ 0 & A_u \end{bmatrix}, \tag{2.21}$$

where A_s and A_u stand for the Jordan blocks with stable and unstable eigenvalues respectively. Accordingly, the state $\mathbf{x}(t)$ can be represented as $\mathbf{x}(t) = G[\mathbf{x}_s^\top(t), \mathbf{x}_u^\top(t)]^\top$, where $\mathbf{x}_s$ and $\mathbf{x}_u$ indicate the stable and unstable sub-state vectors respectively. We then consider the following unstable dynamics:

$$\dot{\mathbf{x}}_u(t) = A_u\mathbf{x}_u(t) + B_u\mathbf{e}(t)\,, \tag{2.22}$$

where B_u stands for the submatrix of BG^{-1} corresponding to A_u. The solution to (2.22) is written as

$$\begin{aligned}
\mathbf{x}_u(t) &= \exp(A_u t)\mathbf{x}_u(0) + \int_0^t \exp(A_u(t-\tau))b_u\mathbf{e}(\tau)d\tau \\
&= \exp(A_u t)\left(\mathbf{x}_u(0) + \int_0^t \exp(-A_u\tau)b_u\mathbf{e}(\tau)d\tau\right) \\
&= \exp(A_u t)(\mathbf{x}_u(0) + \hat{\mathbf{x}}_u(t)) \quad \forall\, t > 0\,,
\end{aligned} \tag{2.23}$$

where we have defined

$$\hat{\mathbf{x}}_u(t) := \int_0^t \exp(-A_u\tau)b_u\mathbf{e}(\tau)d\tau\,.$$

The condition in (3.2) implies that for all t

$$\begin{aligned}
+\infty > M > \log\mathbf{E}\left(\det(\mathbf{x}_u(t)\mathbf{x}_u^\top(t))\right) &= 2t\log\left(\det(\exp(A_u))\right) \\
&+ \log\mathbf{E}\left(\det(\mathbf{x}_u(0)+\hat{\mathbf{x}}_u(t))(\mathbf{x}_u(0)+\hat{\mathbf{x}}_u(t))^\top\right)
\end{aligned} \tag{2.24}$$

for some $M \in \mathbb{R}^+$. On the other hand,

$$
\begin{aligned}
I(\mathbf{x}_0; \mathbf{e}^t) &\overset{(a)}{\geq} I(\mathbf{x}_u(0); \mathbf{e}^t) \\
&\overset{(b)}{\geq} I(\mathbf{x}_u(0); \hat{\mathbf{x}}_u(t)) \\
&\overset{(c)}{=} h(\mathbf{x}_u(0)) - h(\mathbf{x}_u(0)|\hat{\mathbf{x}}_u(t)) \\
&\overset{(d)}{=} h(\mathbf{x}_u(0)) - h(\mathbf{x}_u(0) + \hat{\mathbf{x}}_u(t)|\hat{\mathbf{x}}_u(t)) \\
&\overset{(e)}{\geq} h(\mathbf{x}_u(0)) - h(\mathbf{x}_u(0) + \hat{\mathbf{x}}(t)) \\
&\overset{(f)}{\geq} h(\mathbf{x}_u(0)) - \log(2\pi e)^n \\
&- \log\left(\det \mathbf{E}\left[(\mathbf{x}_u(0) + \hat{\mathbf{x}}_u(t))(\mathbf{x}_u(0) + \hat{\mathbf{x}}_u(t))^\top\right]\right) \,.
\end{aligned} \tag{2.25}
$$

Here, (a) follows from (P3) since $\mathbf{x}_u$ is a function of $\mathbf{x}$; (b) follows from (P3) since $\hat{\mathbf{x}}_u$ is a function of $\mathbf{e}^t$; (c) follows from (P1); (d) follows from (P4); (e) follows from (P1) and (f) is from (P6).

In what follows, we combine (2.24) and (2.25) to obtain

$$
\begin{aligned}
\frac{I(\mathbf{x}_0; \mathbf{e}^t)}{t} \geq \frac{h(\mathbf{x}_u(0))}{t} - \frac{n \log(2\pi e)}{2t} \\
- \frac{M}{2t} + \log\left(\det(\exp(A_u))\right)
\end{aligned} \tag{2.26}
$$

Note that

$$
\log\left(\det(\exp(A_u))\right) = \sum_i \lambda_i(A_u) = \sum_i \Re(\lambda_i(A))^+ \,, \tag{2.27}
$$

and taking the limit on both sides of (2.26), as $t \to \infty$, we obtain (2.20). $\square$

The following Lemma is a consequence of closed-loop causality. It will be used in subsequent derivations.

Lemma 2.2.2. Consider the feedback loop in Fig. 2.1, with all signals sampled with the given δ interval, $0 < \delta \leq \epsilon$. The following identity holds:

$$
I(\mathbf{d}^{(\delta)}(i); [\mathbf{u}^{(\delta)}]^i, \mathbf{x}_0 | [\mathbf{d}^{(\delta)}]^{i-1}) = 0 \,, \quad \forall\, i \geq 1. \tag{2.28}
$$

Proof.

$$
\begin{aligned}
& I(\mathbf{d}^{(\delta)}(i);[\mathbf{u}^{(\delta)}]^i,\mathbf{x}_0|[\mathbf{d}^{(\delta)}]^{i-1}) \\
& \overset{(a)}{\leq} I(\mathbf{d}^{(\delta)}(i);\mathbf{u}^{\delta i},\mathbf{u}^{(\delta)}(i),\mathbf{x}_0|[\mathbf{d}^{(\delta)}]^{i-1}) \\
& \overset{(b)}{\leq} I(\mathbf{d}^{(\delta)}(i);\mathbf{y}^{\delta i-\epsilon},\mathbf{n}^{\delta i}|[\mathbf{d}^{(\delta)}]^{i-1}) \\
& \overset{(c)}{\leq} I(\mathbf{d}^{(\delta)}(i);\mathbf{d}^{\delta i-\epsilon},\mathbf{x}_0,\mathbf{n}^{\delta i}|[\mathbf{d}^{(\delta)}]^{i-1}) \\
& \overset{(d)}{=} I(\mathbf{d}^{(\delta)}(i);\mathbf{d}^{\delta i-\epsilon},\mathbf{x}_0,\mathbf{n}^{\delta i},[\mathbf{d}^{(\delta)}]^{i-1}) - I(\mathbf{d}^{(\delta)}(i);[\mathbf{d}^{(\delta)}]^{i-1}) \\
& \overset{(e)}{=} I(\mathbf{d}^{(\delta)}(i);\mathbf{d}^{\delta i-\epsilon},[\mathbf{d}^{(\delta)}]^{i-1}) - I(\mathbf{d}^{(\delta)}(i);[\mathbf{d}^{(\delta)}]^{i-1}) \\
& \overset{(f)}{=} I(\mathbf{d}^{(\delta)}(i);\mathbf{d}^{(\delta)}(i-1)) - I(\mathbf{d}^{(\delta)}(i);\mathbf{d}^{(\delta)}(i-1)) \\
& = 0
\end{aligned}
\tag{2.29}
$$

Here, (a) follows from (P3), since $[\mathbf{u}^{(\delta)}]^i$ is a function of $(\mathbf{u}^{\delta i},\mathbf{u}(\delta i))$; (b) also follows from (P3), since $(\mathbf{u}^{\delta i},\mathbf{u}(\delta i))$ is a function of $\mathbf{y}^{\delta i-\epsilon}$ and $\mathbf{n}^{\delta i}$; (c) also follows from (P3), since $\mathbf{y}^{\delta i-\epsilon}$ is a function of $\mathbf{d}^{\delta i-\epsilon}$, $\mathbf{x}_0$ and $\mathbf{n}^{\delta i}$; (d) follows from (P2); (e) follows from the assumption that $\mathbf{n}$, $\mathbf{x}_0$ and $\mathbf{d}$ are mutually independent; (f) follows from Markov property of $\mathbf{d}$. □

We are ready to state the main theorem regarding closed loop causality.

Theorem 2.2.3. Consider the closed loop shown in Fig. 2.1. The following inequality holds:

$$
I(\mathbf{e}^t;\mathbf{u}^t) \geq I(\mathbf{d}^t;\mathbf{u}^t) + I(\mathbf{x}_0;\mathbf{e}^t)\,, \quad \forall\, t \in \mathbb{R}^+. \tag{2.30}
$$

Proof. Given $t > 0$, we take $k+1$ samples of each of the signals $\mathbf{e}$, $\mathbf{d}$ and $\mathbf{u}$ over $[0,t)$, by sampling the interval $\delta(k) > 0$ to get the discretized signals $\{\mathbf{e}^{(\delta(k))}(i) : 1 \leq i \leq k\}$, $\{\mathbf{d}^{(\delta(k))}(i) : 1 \leq i \leq k\}$ and $\{\mathbf{u}^{(\delta(k))}(i) : 1 \leq i \leq k\}$ respectively. Notice also that $(k+1)\delta(k) = t$.

We expand the following mutual information by Kolmogrov's formula (P4) for any $1 \leq i \leq k$:

$$
\begin{aligned}
-I(&\mathbf{d}^{(\delta(k))}(i);\mathbf{x}_0,[\mathbf{u}^{(\delta(k))}]^i|[\mathbf{d}^{(\delta(k))}]^{i-1}) \\
&= I(\mathbf{d}^{(\delta(k))}(i);[\mathbf{d}^{(\delta(k))}]^{i-1}) - I(\mathbf{d}^{(\delta(k))}(i);[\mathbf{d}^{(\delta(k))}]^{i-1},\mathbf{x}_0,[\mathbf{u}^{(\delta(k))}]^i) \\
&\overset{(a)}{=} h(\mathbf{d}^{(\delta(k))}(i)|[\mathbf{d}^{(\delta(k))}]^{i-1},\mathbf{x}_0,[\mathbf{u}^{(\delta(k))}]^i) - h(\mathbf{d}^{(\delta(k))}(i)|[\mathbf{d}^{(\delta(k))}]^{i-1}) \\
&\overset{(b)}{=} h(\mathbf{d}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1},\mathbf{x}_0,[\mathbf{u}^{(\delta(k))}]^i) - h(\mathbf{d}^{(\delta(k))}(i)|[\mathbf{d}^{(\delta(k))}]^{i-1})
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(c)}{=} h(\mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}, \mathbf{x}_0, [\mathbf{u}^{(\delta(k))}]^{i}) - h(\mathbf{d}^{(\delta(k))}(i)|[\mathbf{d}^{(\delta(k))}]^{i-1}) \\
&\overset{(d)}{=} h(\mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}) - I(\mathbf{x}_0; \mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}) \\
&\quad - I([\mathbf{u}^{(\delta(k))}]^{i}; \mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}, \mathbf{x}_0) - h(\mathbf{d}^{(\delta(k))}(i)|[\mathbf{d}^{(\delta(k))}]^{i-1}),
\end{aligned}
\tag{2.31}
$$

where (a) follows from (P1), (b) from the fact that $[\mathbf{e}^{(\delta(k))}]^{i-1} = [\mathbf{d}^{(\delta(k))}]^{i-1} + [\mathbf{u}^{(\delta(k))}]^{i-1}$ and therefore the map $([\mathbf{d}^{(\delta(k))}]^{i-1}, \mathbf{x}_0, [\mathbf{u}^{(\delta(k))}]^{i}) \mapsto ([\mathbf{e}^{(\delta(k))}]^{i-1}, \mathbf{x}_0, [\mathbf{u}^{(\delta(k))}]^{i})$ is invertible, (c) from (P4) since $\mathbf{e}^{(\delta(k))}(i) = \mathbf{d}^{(\delta(k))}(i) + \mathbf{u}^{(\delta(k))}(i)$, and (e) is from (P4).

On the other hand, Lemma 2.2.2 claims that

$$
I(\mathbf{d}^{(\delta(k))}(i); \mathbf{x}_0, [\mathbf{u}^{(\delta(k))}]^{i}|[\mathbf{d}^{(\delta(k))}]^{i-1}) = 0 \tag{2.32}
$$

Summing up $-I(\mathbf{d}^{(\delta(k))}(i); \mathbf{x}_0, [\mathbf{u}^{(\delta(k))}]^{i}|[\mathbf{d}^{(\delta(k))}]^{i-1})$ from 1 to k, $\forall\ k \geq 1$, and considering (2.31), we have

$$
\begin{aligned}
0 &= \sum_{i=1}^{k} I(\mathbf{d}^{(\delta(k))}(i); \mathbf{x}_0, [\mathbf{u}^{(\delta(k))}]^{i}|[\mathbf{d}^{(\delta(k))}]^{i-1}) \\
&\overset{(a)}{=} h([\mathbf{e}^{(\delta(k))}]^{k}) - I(\mathbf{x}_0; [\mathbf{e}^{(\delta(k))}]^{k}) - h([\mathbf{d}^{(\delta(k))}]^{k}) \\
&- \sum_{i=1}^{k} I([\mathbf{u}^{(\delta(k))}]^{i}; \mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}, \mathbf{x}_0) \\
&\overset{(b)}{=} h([\mathbf{e}^{(\delta(k))}]^{k}) - h([\mathbf{e}^{(\delta(k))}]^{k}|[\mathbf{u}^{(\delta(k))}]^{k} + h([\mathbf{d}^{(\delta(k))}]^{k}|[\mathbf{u}^{(\delta(k))}]^{k}) \\
&- I(\mathbf{x}_0; [\mathbf{e}^{(\delta(k))}]^{k}) - h([\mathbf{d}^{(\delta(k))}]^{k}) \\
&\quad - \sum_{i=1}^{k} I([\mathbf{u}^{(\delta(k))}]^{i}; \mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}, \mathbf{x}_0) \\
&\overset{(c)}{=} I([\mathbf{e}^{(\delta(k))}]^{k}; [\mathbf{u}^{(\delta(k))}]^{k}) - I(\mathbf{x}_0; [\mathbf{e}^{(\delta(k))}]^{k}) \\
&- \sum_{i=1}^{k} I([\mathbf{u}^{(\delta(k))}]^{i}; \mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}, \mathbf{x}_0) \\
&- I([\mathbf{d}^{(\delta(k))}]^{k}; [\mathbf{u}^{(\delta(k))}]^{k}) \\
&\overset{(d)}{\leq} I([\mathbf{e}^{(\delta(k))}]^{k}; [\mathbf{u}^{(\delta(k))}]^{k}) - I(\mathbf{x}_0; [\mathbf{e}^{(\delta(k))}]^{k}) \\
&- I([\mathbf{d}^{(\delta(k))}]^{k}; [\mathbf{u}^{(\delta(k))}]^{k})
\end{aligned}
\tag{2.33}
$$

Here (a) follows from (P5), (b) follows from (P4) since $h([\mathbf{e}^{(\delta(k))}]^{k}|[\mathbf{u}^{(\delta(k))}]^{k}) =$

$h([\mathbf{d}^{(\delta(k))}]^k|[\mathbf{u}^{(\delta(k))}]^k)$, (c) follows from (P1) and (d) follows from the non-negativeness of mutual information.

Taking the limit as $k \to \infty$, we have $\delta(k) \to 0$, which consequently implies that

$$0 \leq I(\mathbf{e}^t; \mathbf{u}^t) - I(\mathbf{d}^t; \mathbf{u}^t) - I(\mathbf{x}_0; \mathbf{e}^t) \,. \tag{2.34}$$

The inequality in (2.30) follows. □

Remark 2.2.4. The quantity $\sum_{i=1}^k I([\mathbf{u}^{(\delta(k))}]^i; \mathbf{e}^{(\delta(k))}(i)|[\mathbf{e}^{(\delta(k))}]^{i-1}, \mathbf{x}_0)$ in the equation (b) of (2.33) has been defined in [44] as *directed information* from $[\mathbf{u}^{(\delta(k))}]^k$ to $[\mathbf{e}^{(\delta(k))}]^k$ conditioned by $\mathbf{x}_0$, and is denoted as $I([\mathbf{u}^{(\delta(k))}]^k \to [\mathbf{e}^{(\delta(k))}]^k|\mathbf{x}_0)$. One can define the continuous-time version of directed information by letting $k \to \infty$. A preliminary exploration of continuous-time directed information and its relation with optimal estimation theory has been reported recently in [45].

An inequality for information rate is readily obtained by dividing both sides of (2.30) by t and letting t go to infinity (assuming that the limit exists). It is summarized in the following corollary.

Corollary 2.2.5. Given the closed loop system in Fig. 2.1, we have

$$\bar{I}(\mathbf{e}; \mathbf{u}) - \bar{I}(\mathbf{d}; \mathbf{u}) \geq \bar{I}(\mathbf{x}_0; \mathbf{e}) \tag{2.35}$$

The subsequent Theorem incorporates the mean square stability of the closed loop with the information rate inequality (2.35). Some stationarity assumptions are further enforced to derive a Bode-like formula. The details are summarized in the following theorem.

Theorem 2.2.6 (*Bode-Like Formula*). Suppose the closed-loop system shown in Fig. 2.1 is mean-square stable. Then

$$\bar{I}(\mathbf{e}; \mathbf{u}) \geq \bar{I}(\mathbf{d}; \mathbf{u}) + \sum_i \Re(\lambda_i(A))^+ \,. \tag{2.36}$$

Furthermore, if $(\mathbf{d}, \mathbf{u})$ and $(\mathbf{u}, \mathbf{e})$ form stationary processes and $f_\mathbf{u} \in \mathbb{F}$ and $\mathbf{d}$ is a stationary Gaussian Markov process, then

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right) d\omega \geq \sum_i \Re(\lambda_i(A))^+ \,. \tag{2.37}$$

Proof. The inequality in (2.36) directly follows from (2.20) and (2.35). To obtain (3.2.12), first we have

$$\begin{aligned}
&I(\mathbf{e}^t;\mathbf{u}^t) - I(\mathbf{d}^t;\mathbf{u}^t)\\
&\stackrel{(a)}{=} \lim_{k\to\infty}\{I([\mathbf{e}^{(\delta(k))}]^k;[\mathbf{u}^{(\delta(k))}]^k) - I([\mathbf{d}^{(\delta(k))}]^k;[\mathbf{u}^{(\delta(k))}]^k)\}\\
&\stackrel{(b)}{=} \lim_{k\to\infty}\{h([\mathbf{e}^{(\delta(k))}]^k) - h([\mathbf{d}^{(\delta(k))}]^k)\}\\
&\stackrel{(c)}{\leq} \lim_{k\ \to\infty}\{h([\bar{\mathbf{e}}^{(\delta(k))}]^k) - h([\mathbf{d}^{(\delta(k))}]^k)\}\\
&\stackrel{(d)}{=} \lim_{k\to\infty}\{I([\bar{\mathbf{e}}^{(\delta(k))}]^k;[\bar{\mathbf{u}}^{(\delta(k))}]^k) - I([\mathbf{d}^{(\delta(k))}]^k;[\bar{\mathbf{u}}^{(\delta(k))}]^k)\}\\
&\stackrel{(e)}{=} I(\bar{\mathbf{e}}^t;\bar{\mathbf{u}}^t) - I(\mathbf{d}^t;\bar{\mathbf{u}}^t)\,,
\end{aligned} \tag{2.38}$$

where $(\bar{\mathbf{e}},\bar{\mathbf{u}})$ stands for the Gaussian stationary process with the same covariance as $(\mathbf{e},\mathbf{u})$. Here (a) follows from Lemma 2.1.13; (b) follows from (P1); (c) follows from (P6); (d) follows from (P1), and we use the fact that $h([\bar{\mathbf{e}}^{(\delta(k))}]^k|[\bar{\mathbf{u}}^{(\delta(k))}]^k) = h([\bar{\mathbf{d}}^{(\delta(k))}]^k|[\bar{\mathbf{u}}^{(\delta(k))}]^k), \forall k \in \mathbb{N}^+$; (e) follows from Lemma 2.1.13. Then it is straightforward to show that

$$\bar{I}(\mathbf{e};\mathbf{u}) - \bar{I}(\mathbf{d};\mathbf{u}) \leq \bar{I}(\bar{\mathbf{e}};\bar{\mathbf{u}}) - \bar{I}(\mathbf{d};\bar{\mathbf{u}}) \tag{2.39}$$

Since $f_{\mathbf{u}} \in \mathbb{F}$, Lemma 2.1.18 implies

$$\begin{aligned}
&\bar{I}(\bar{\mathbf{e}};\bar{\mathbf{u}}) - \bar{I}(\mathbf{d};\bar{\mathbf{u}})\\
&= -\frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(1-\frac{f_{\mathbf{eu}}(\omega)f_{\mathbf{ue}}(\omega)}{f_{\mathbf{e}}(\omega)f_{\mathbf{u}}(\omega)}\right)d\omega + \frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(1-\frac{f_{\mathbf{du}}(\omega)f_{\mathbf{ud}}(\omega)}{f_{\mathbf{d}}(\omega)f_{\mathbf{u}}(\omega)}\right)d\omega\\
&= \frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(\frac{f_{\mathbf{e}}(\omega)}{f_{\mathbf{d}}(\omega)}\cdot\frac{f_{\mathbf{d}}(\omega)f_{\mathbf{u}}(\omega)-f_{\mathbf{du}}(\omega)f_{\mathbf{ud}}(\omega)}{f_{\mathbf{e}}(\omega)f_{\mathbf{u}}(\omega)-f_{\mathbf{eu}}(\omega)f_{\mathbf{ue}}(\omega)}\right)d\omega\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right)d\omega\,.
\end{aligned} \tag{2.40}$$

Here we have used the fact

$$\frac{f_{\mathbf{d}}(\omega)f_{\mathbf{u}}(\omega)-f_{\mathbf{du}}(\omega)f_{\mathbf{ud}}(\omega)}{f_{\mathbf{e}}(\omega)f_{\mathbf{u}}(\omega)-f_{\mathbf{eu}}(\omega)f_{\mathbf{ue}}(\omega)} = 1.$$

Indeed, since $\mathbf{d} = \mathbf{e} + \mathbf{u}$, then

$$
\begin{aligned}
f_{\mathbf{d}}(\omega) &= \int_{-\infty}^{\infty} e^{-it\omega} R_{\mathbf{e}+\mathbf{u}}(\tau) d\tau \\
&= \int_{-\infty}^{\infty} e^{-it\omega} (R_{\mathbf{e}}(\tau) + f_{\mathbf{e},\mathbf{u}}(\tau) + f_{\mathbf{u},\mathbf{e}}(-\tau) + R_{\mathbf{u}}(\tau)) d\tau \\
&= f_{\mathbf{e}}(\omega) + f_{\mathbf{eu}}(\omega) + f_{\mathbf{ue}}(\omega) + f_{\mathbf{u}}(\omega) \, ,
\end{aligned}
\tag{2.41}
$$

and

$$
\begin{aligned}
f_{\mathbf{du}} &= \int_{-\infty}^{\infty} e^{-it\omega} R_{\mathbf{e}+\mathbf{u},\mathbf{u}}(\tau) d\tau \\
&= \int_{-\infty}^{\infty} e^{-it\omega} (R_{\mathbf{e},\mathbf{u}} + R_{\mathbf{u}})(\tau) d\tau \\
&= f_{\mathbf{eu}}(\omega) + f_{\mathbf{u}}(\omega) \, .
\end{aligned}
\tag{2.42}
$$

Hence, (2.41) and (2.42) give

$$
\begin{aligned}
&\frac{f_{\mathbf{d}}(\omega) f_{\mathbf{u}}(\omega) - f_{\mathbf{du}}(\omega) f_{\mathbf{ud}}(\omega)}{f_{\mathbf{e}}(\omega) f_{\mathbf{u}}(\omega) - f_{\mathbf{eu}}(\omega) f_{\mathbf{ue}}(\omega)} \\
&= \frac{(f_{\mathbf{e}} + f_{\mathbf{eu}} + f_{\mathbf{ue}} + f_{\mathbf{u}}) f_{\mathbf{u}} - (f_{\mathbf{u}} + f_{\mathbf{eu}})(f_{\mathbf{u}} + f_{\mathbf{ue}})}{f_{\mathbf{e}} f_{\mathbf{u}} - f_{\mathbf{eu}} f_{\mathbf{ue}}} \\
&= 1 \, .
\end{aligned}
$$

The proof is complete. □

Remark 2.2.7. The equation (3.2.12) is formally identical to the inequality version of Bode's integral developed in the classical case [9], where a time delay is introduced to make the residual of $\log |S(s)|$ vanish at infinity for strictly proper plants. The same type of time delay in the course of our derivation is introduced to ensure closed-loop causality, so that the sequential relations among the signals residing in Fig. 2.1 are revealed by using information theoretical machineries.

Remark 2.2.8. We have hinged on *stationary* closed loops for the derivation of Bode's integral formula (3.2.12) from the information conservation law in (2.35) for simplicity. Nonetheless, the similar argument can be also extended to *asymptotically stationary* cases with minor modification.

2.3 Negative Component of Bode's Integral

In the section, we investigate the lower bound of $\bar{I}(\mathbf{d};\mathbf{u})$, with additional assumptions that $\mathbf{d}$ and $\mathbf{e}$ are mutually wide sense stationary and $\mathbf{d}$ is Gaussian. As shown in the subsequent result, the lower bound of $\bar{I}(\mathbf{d};\mathbf{u})$ is obtained as the negative portion of the Bode's integral obtained in the previous section.

The following theorem summarizes the main result

Theorem 2.3.1. Consider the feedback closed loop given in Fig 2.1, where $\mathbf{d}$ and $\mathbf{e}$ are mutually wide-sense stationary and $\mathbf{d}$ is a Gaussian Markov process. If $f_{\mathbf{u}}(\omega)$ is bounded away from zero, then the following inequality holds

$$\bar{I}(\mathbf{d};\mathbf{u}) \geq -\frac{1}{2\pi}\int_{-\infty}^{\infty} \left(\log S_{\mathbf{d},\mathbf{e}}(\omega)\right)^{-} d\omega \tag{2.43}$$

Proof. To begin with, we consider the following Wiener predictor

$$L(j\omega) = \frac{f_{\mathbf{d},\mathbf{u}}(\omega)}{f_{\mathbf{u}}(\omega)} e^{j\omega\epsilon} ,$$

which represents the minimal mean square error prediction of $\mathbf{d}$, given the observation of the entire time history of $\mathbf{u}$ with the time delay ϵ. To obtain a causal prediction of $\mathbf{d}(t)$ by using the possibly noncausal $L(j\omega)$, we define the following predictor:

$$\hat{\mathbf{d}}(t) = L(s)\lfloor \mathbf{u}(t) \rfloor_t ,$$

where $\lfloor \cdot \rfloor_t$ stands for the truncation operator.

The above Wiener predictor is now used to lower bound the quantity $\bar{I}(\mathbf{u};\mathbf{d})$. First, the process $\mathbf{d}(\tau), 0 \leq \tau < t$ is sampled with interval $\delta(k) = \frac{t}{k+1}$, leading to

$$
\begin{aligned}
& I([\mathbf{d}^{(\delta(k))}]^k;\mathbf{u}^{t-\epsilon}) \\
& \overset{(a)}{\geq} I([\mathbf{d}^{(\delta(k))}]^k;\hat{\mathbf{d}}^t) \\
& \overset{(b)}{\geq} I([\mathbf{d}^{(\delta(k))}]^k;[\hat{\mathbf{d}}^{(\delta(k))}]^k) \\
& \overset{(c)}{=} h([\mathbf{d}^{(\delta(k))}]^k) - h([\mathbf{d}^{(\delta(k))}]^k|[\hat{\mathbf{d}}^{(\delta(k))}]^k) \\
& \overset{(d)}{\geq} h([\mathbf{d}^{(\delta(k))}]^k) - h([\tilde{\mathbf{d}}^{(\delta(k))}]^k) \\
& \overset{(e)}{=} h([\mathbf{d}^{(\delta(k))}]^k) - h([\mathbf{d}^{(\delta(k))}]^k|[\hat{\mathbf{d}}^{(\delta(k))}]^k) + h([\tilde{\mathbf{d}}^{(\delta(k))}]^k|[\hat{\mathbf{d}}^{(\delta(k))}]^k) - h([\tilde{\mathbf{d}}^{(\delta(k))}]^k) \\
& = I([\mathbf{d}^{(\delta(k))}]^k;[\hat{\mathbf{d}}^{(\delta(k))}]^k) - I([\tilde{\mathbf{d}}^{(\delta(k))}]^k;[\hat{\mathbf{d}}^{(\delta(k))}]^k)\,,
\end{aligned}
$$

where $\tilde{\mathbf{d}} := \mathbf{d} - \hat{\mathbf{d}}$. Here (a) follows from (P3), since $\hat{\mathbf{d}}^t$ is a function of $\mathbf{u}^{t-\epsilon}$; (b) follows from (P3), since $[\hat{\mathbf{d}}^{(\delta(k))}]^k$ is a function $\hat{\mathbf{d}}^t$; (c) follows from P1; (d) follows from the fact that conditioning reduces entropy; (e) follows from $h([\mathbf{d}^{(\delta(k))}]^k|[\hat{\mathbf{d}}^{(\delta(k))}]^k) = h([\tilde{\mathbf{d}}^{(\delta(k))}]^k|[\hat{\mathbf{d}}^{(\delta(k))}]^k)$.

By applying Lemma 2.1.13, we have

$$
I(\mathbf{d}^t;\mathbf{u}^{t-\epsilon}) \geq I(\mathbf{d}^t;\hat{\mathbf{d}}^t) - I(\tilde{\mathbf{d}}^t;\hat{\mathbf{d}}^t)\,,
$$

which in turn gives the limiting case

$$
\bar{I}(\mathbf{d};\mathbf{u}) \geq \bar{I}(\mathbf{d};\hat{\mathbf{d}}) - \bar{I}(\tilde{\mathbf{d}};\hat{\mathbf{d}})\,. \tag{2.44}
$$

Note that $\mathbf{d}$ and $\hat{\mathbf{d}}$ are Gaussian and stationarily correlated and $f_{\mathbf{d}} \in \mathbb{F}$, and from Lemma 2.1.13 we have

$$
\begin{aligned}
& \bar{I}(\mathbf{d};\hat{\mathbf{d}}) - \bar{I}(\tilde{\mathbf{d}};\hat{\mathbf{d}}) \\
& = -\frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 - \frac{f_{\mathbf{d}\hat{\mathbf{d}}}(\omega) f_{\hat{\mathbf{d}}\mathbf{d}}(\omega)}{f_{\mathbf{d}}(\omega) f_{\hat{\mathbf{d}}}(\omega)}\right) d\omega + \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 - \frac{f_{\tilde{\mathbf{d}}\hat{\mathbf{d}}}(\omega) f_{\hat{\mathbf{d}}\tilde{\mathbf{d}}}(\omega)}{f_{\tilde{\mathbf{d}}}(\omega) f_{\hat{\mathbf{d}}}(\omega)}\right) d\omega \\
& = \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(\frac{f_{\mathbf{d}}(\omega)}{f_{\tilde{\mathbf{d}}}(\omega)}\right) d\omega \\
& = \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(\frac{f_{\mathbf{d}}(\omega)}{f_{\mathbf{d}}(\omega) - |L(j\omega)|^2 f_{\mathbf{u}}(\omega)}\right) d\omega\,,
\end{aligned}
$$

where we have used the fact

$$
f_{\hat{\mathbf{d}}}(\omega) = f_{\tilde{\mathbf{d}}}(\omega) = f_{\mathbf{d}}(\omega) - |L(j\omega)|^2 f_{\mathbf{u}}(\omega).
$$

We then note that

$$|L(j\omega)| = \frac{|f_{\mathbf{du}}(\omega)|}{|f_{\mathbf{u}}(\omega)|} \geq \frac{\Re(f_{\mathbf{du}}(\omega))}{f_{\mathbf{u}}(\omega)} = \frac{f_{\mathbf{du}}(\omega) + f_{\mathbf{ud}}(\omega)}{2f_{\mathbf{u}}(\omega)} = \frac{f_{\mathbf{e}}(\omega) - f_{\mathbf{d}}(\omega) - f_{\mathbf{u}}(\omega)}{2f_{\mathbf{u}}(\omega)}$$

Therefore (2.44) is further written as

$$\begin{aligned} &\bar{I}(\mathbf{d},\mathbf{u}) \geq \\ &\frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(\frac{4f_{\mathbf{d}}f_{\mathbf{u}}}{-f_{\mathbf{d}}^2 - f_{\mathbf{e}}^2 - f_{\mathbf{u}}^2 + 2f_{\mathbf{d}}f_{\mathbf{e}} + 2f_{\mathbf{d}}f_{\mathbf{u}} + 2f_{\mathbf{u}}f_{\mathbf{e}}}\right) d\omega \end{aligned} \tag{2.45}$$

Taking the maximum value of the right hand side of (2.45), we have

$$\begin{aligned} &\sup_{f_u>0} \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(\frac{4f_{\mathbf{d}}f_{\mathbf{u}}}{-f_{\mathbf{d}}^2 - f_{\mathbf{e}}^2 - f_{\mathbf{u}}^2 + 2f_{\mathbf{d}}f_{\mathbf{e}} + 2f_{\mathbf{d}}f_{\mathbf{u}} + 2f_{\mathbf{u}}f_{\mathbf{e}}}\right) d\omega \\ &= -\frac{1}{2\pi}\int_{-\infty}^{\infty} (\log S_{\mathbf{d},\mathbf{e}})^{-} \, d\omega \end{aligned} \qquad \square$$

The relation in (2.46) follows from the fact that (2.45) holds also for all $f_{\mathbf{u}}(\omega) > 0$.

Once the inequality (2.46) is obtained, we can employ the inequality (3.2.18) later in Section 2.5 to obtain the following theorem.

Theorem 2.3.2. Consider the closed loop shown in Fig. 2.1, where $\mathbf{e}$ and $\mathbf{d}$ are assumed jointly stationary, with $\mathbf{d}$ being a Gaussian Markov process. If the closed loop is mean square stable then the following holds:

$$-\frac{1}{2\pi}\int_{-\infty}^{\infty} (\log S_{\mathbf{d},\mathbf{e}}(\omega))^{-} \, d\omega \leq \bar{I}((\mathbf{x}(0),\mathbf{d});\mathbf{u}) - \sum_i \Re(\lambda_i(A))^{+}\,, \tag{2.46}$$

Remark 2.3.3. The upcoming discussion in Section 2.5 will show that $\bar{I}((\mathbf{x}(0),\mathbf{d});\mathbf{u})$ represents the total information flow in the closed loop. Therefore, the inequality in (2.46) implies that the negative portion of the Bode integral (where $S_{\mathbf{d},\mathbf{e}}(\omega) < 1$) is determined by both the degree of open-loop instability and the information rate transmitted through the closed loop. It can be clearly observed from (2.46) that if $\bar{I}((\mathbf{x}(0),\mathbf{d});\mathbf{u}) = \sum_i \Re(\lambda_i(A))^{+}$, then the $S_{\mathbf{d},\mathbf{e}}(\omega) \geq 1$ for all ω. Moreover, the same observation shows that, to achieve a desirable shaping of the sensitivity function, one needs a larger information transmission rate to allow for a less constraint on the negative part of $\log S_{\mathbf{d},\mathbf{e}}(\omega)$.

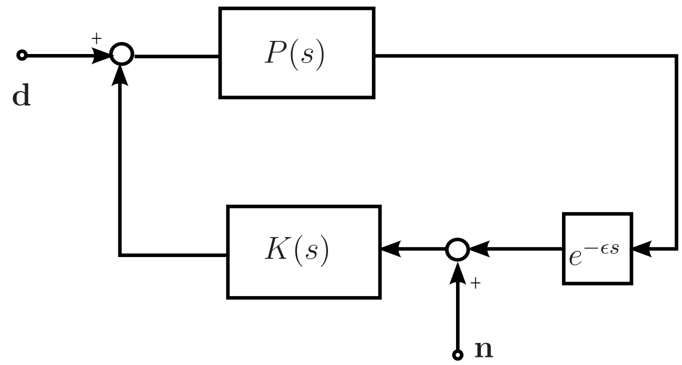

Figure 2.2: Linear Stochastic Closed Loop

2.4 Achievable Lower Bound of Bode's Integral for LTI Systems

This section is devoted to further investigation of the the tightness of the resulting Bode's integral. As is has been shown in (3.2.12), the sum of the unstable poles serves as a lower bound on the log-integral of the sensitivity function; however, the conservativeness of this *inequality* remains unclear. One can intuitively conclude that the controller noise $\mathbf{n}$ contributes to the increase of $\frac{1}{2\pi}\int_{-\infty}^{\infty} \log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right) d\omega$ by making $\mathbf{e}$ noisier within some frequency range. Detailed analysis of this issue is given subsequently, where the controller and the plant are given by LTI systems.

We now specialize the problem to the closed-loop configuration, shown in Fig. 2.2, where $P(s)$ is strictly proper and minimum phase, and the unstable poles are denoted as $\{p_1, p_2, ..., p_N\}$. In addition, we choose a proper stable stabilizing controller $K(s)$. The controller noise $\mathbf{n}(t)$ is a stationary (possibly colored) Gaussian process with zero mean; the disturbance signal $\mathbf{d}$ is a stationary Gaussian Markov process. A candidate $\mathbf{d}$ can be expressed as the following Itô integral, also known as Ornstein-Uhlenbeck Brownian motion.

$$\mathbf{d}(t) = b\int_0^t e^{-a(t-u)} dW_u \, ,$$

where $a > 0$ and $b \neq 0$ are real numbers, W_t is a standard Wiener process. The initial conditions for both $P(s)$ and $K(s)$ are set to 0.

Note that closed loop is stable (with sufficiently small $\epsilon > 0$) and that $\mathbf{d}$ and $\mathbf{n}$

are independent. We have

$$f_{\mathbf{e}}(\omega) = \frac{f_d(\omega)}{|1 - P(j\omega)K(j\omega)e^{-j\omega\epsilon}|^2} + \frac{|K(j\omega)|^2 f_{\mathbf{n}}(\omega)}{|1 - P(j\omega)K(j\omega)e^{-j\omega\epsilon}|^2} .$$

Subsequently, the sensitivity function is obtained as

$$S_{\mathbf{d},\mathbf{e}}(\omega) = \sqrt{\frac{f_{\mathbf{e}}(\omega)}{f_{\mathbf{d}}(\omega)}} = \frac{\sqrt{1 + |K(j\omega)|^2 \frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}}}{|1 - P(j\omega)K(j\omega)e^{-j\omega\epsilon}|} \tag{2.47}$$

Next, we prove the following theorem regarding the log-integral of sensitivity.

Theorem 2.4.1. Consider the closed loop shown in Fig 2.2. The following equality holds

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \log S_{\mathbf{d},\mathbf{e}}(\omega) d\omega = \sum_i \Re(\lambda_i(A))^+ + \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 + |K(j\omega)|^2 \frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}\right) d\omega , \tag{2.48}$$

Proof. By using (2.47), we have

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} \log S_{\mathbf{d},\mathbf{e}}(\omega) d\omega =& \frac{1}{2\pi}\int_{-\infty}^{\infty} \log\left(\frac{1}{|1 - P(j\omega)K(j\omega)e^{-j\omega\epsilon}|}\right) d\omega + \\ & \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 + |K(j\omega)|^2 \frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}\right) d\omega .\end{aligned}$$

Notice that $1/(1 - K(s)P(s))$ is stable and proper. Then we employ the same argument as in the proof of Theorem 3.1.4 in [9] to obtain

$$\begin{aligned}&\frac{1}{2\pi}\int_{-\infty}^{\infty} \log\left(\frac{1}{|1 - P(j\omega)K(j\omega)e^{-j\omega\epsilon}|}\right) d\omega \\ &= \frac{1}{2\pi j}\oint_{\mathscr{C}} \log\left(\frac{1}{|1 - P(s)K(s)e^{-s\epsilon}|}\right) ds \\ &= p_1 + ... + p_N = \sum_i \Re(\lambda_i(A))^+ .\end{aligned}$$

Here $\mathscr{C}$ denotes the right half plane closed contour, which has a sufficiently large radius and circumvents all the unstable poles of $P(s)$ [9]. The same integration can also be calculated by a simplified methodology developed in [46]. The proof is complete. □

The positive term $\kappa := \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 + |K(j\omega)|^2 \frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}\right) d\omega$ in (2.48) presents an additional performance limitation, on the top of the sum of the unstable poles. In order to gain some insight, we now illustrate the significance of this term from different perspectives.

- Although it is not easy to quantify κ in general (yet a special case is given later in Lemma 2.4.2 to calculate κ explicitly), we can roughly estimate its value by observing the magnitudes of $f_{\mathbf{d}}(\omega)$, $f_{\mathbf{n}}(\omega)$ and $K(j\omega)$. It becomes evident that, both a lower noise-to-disturbance ratio $f_{\mathbf{n}}(\omega)/f_{\mathbf{d}}(\omega)$ and a smaller controller magnitude $|K(j\omega)|$ lead to a less restrictive limitation on the closed loop.

- From information theoretical point of view, the expression of κ reminds of the mutual information rate of a continuous-time additive Gaussian channel [4]. For the non-feedback additive Gaussian channel shown in Fig. 2.3, the input/ouput mutual information can be calculated by Lemma 2.1.18 as

$$\begin{aligned}\bar{I}(\mathbf{v};\mathbf{z}) &= -\frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(1-\frac{|f_{\mathbf{vz}}(\omega)|^2}{f_{\mathbf{v}}(\omega)f_{\mathbf{z}}(\omega)}\right)d\omega\\ &= -\frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(1-\frac{|f_{\mathbf{v}}(\omega)|^2}{f_{\mathbf{v}}(\omega)(f_{\mathbf{v}}(\omega)+f_{\mathbf{d}}(\omega))}\right)d\omega\\ &= \frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(1+|K(j\omega)|^2\frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}\right)d\omega = \kappa\,.\end{aligned}$$

  The above interpretation of κ shows that the extra amount of performance limitation is induced by the mutual information rate between the propagated controller noise $\mathbf{v}$ and the observation $\mathbf{z}$. To reduce the mutual information rate, one can reduce the uncertainty of the channel source $\mathbf{v}$, which can be done by either lowering the magnitude of $K(s)$, or denoising the controller noise $\mathbf{n}$.

- κ can be also related to the famous H_∞ entropy [47]. Suppose there exits a proper transfer function $M(s)$ such that

$$\frac{1}{2}\frac{|K(j\omega)|^2\frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}}{1+|K(j\omega)|^2\frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}} = M(-j\omega)M(j\omega).$$

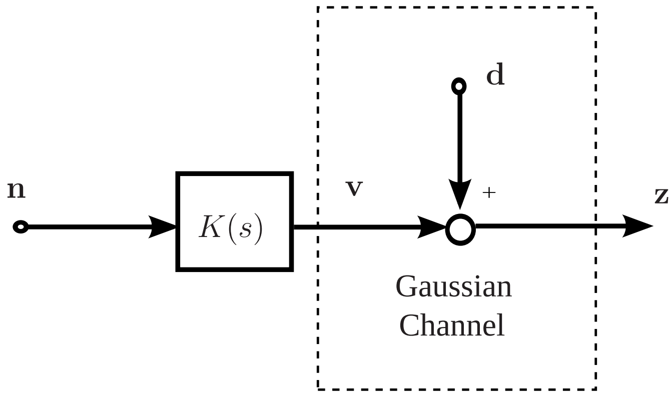

Figure 2.3: Additive Gaussian channel

Then the above relation leads to

$$\kappa = \frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 + |K(j\omega)|^2 \frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}\right) d\omega$$

$$= -\frac{\gamma^2}{2\pi}\int_{-\infty}^{\infty} \log\left(1 - \gamma^{-2} M(-j\omega)M(j\omega)\right) d\omega \,,$$

which is exactly the expression of the H_∞ entropy of $M(s)$ with disturbance rejection level $\gamma = 1/\sqrt{2}$. It has been shown that the minimal H_∞ entropy controller is equivalent to a suboptimal H_∞ controller ($\|M\|_{H_\infty} \leq \gamma$) [47]. Therefore the above observation actually proposes a way to minimize κ by resorting to various H_∞ methodologies for the design of $K(s)$. While the detailed development along this direction is not given here, the readers are encouraged to look into this interesting problem as it provides a potential link between H_∞ theory and information theory.

Next we will show that, under some mild assumptions, κ can be obtained explicitly, where we assume that $f_{\mathbf{d}}(\omega)$ and $f_{\mathbf{n}}(\omega)$ are rational and admit the following spectral factorizations:

$$f_{\mathbf{d}}(\omega) = \phi_{\mathbf{d}}(-j\omega)\phi_{\mathbf{d}}(j\omega), \qquad f_{\mathbf{n}}(\omega) = \phi_{\mathbf{n}}(-j\omega)\phi_{\mathbf{n}}(j\omega)\,.$$

Lemma 2.4.2. Assume that $K(s)\frac{\phi_{\mathbf{n}}(s)}{\phi_{\mathbf{d}}(s)}$ admits a minimal realization $(A_k, b_k, c_k^\top, d_k)$ with A_k being Hurwitz. Moreover, assume that there exists a matrix $Q > 0$ solving the following algebraic Riccati equation (ARE):

$$A_k^\top Q + QA_k - \frac{1}{1+d_k^2} Qb_k b_k^\top Q + \frac{1}{1+d_k^2} c_k c_k^\top = 0\,, \tag{2.49}$$

and ensuring that

$$A_k - \frac{1}{1+d_k^2} b_k b_k^\top Q \text{ is Hurwitz.} \tag{2.50}$$

Then

$$\frac{1}{4\pi}\int_{-\infty}^{\infty} \log\left(1 + |K(j\omega)|^2 \frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}\right) d\omega = \frac{1}{\sqrt{1+d_k^2}} b_k^\top Q b_k + \frac{d_k}{\sqrt{1+d_k^2}} c_k^\top b_k$$

Proof. We will first obtain the following spectral factorization:

$$1 + |K(j\omega)|^2 \frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)} = H(-j\omega)H(j\omega)\,,$$

where $H(s) = -\frac{1}{\sqrt{1+d_k^2}}(b_k^\top Q + d_k c_k^\top)(s\mathbb{I} - A_k)^{-1} b_k + \sqrt{1+d_k^2}$. Indeed, it can be verified that

$$\begin{aligned}
&H(-s)H(s) \\
&= \left(\frac{1}{\sqrt{1+d_k^2}} b_k^\top (s\mathbb{I} + A_k^\top)^{-1}(Qb_k + d_k c_k) + \sqrt{1+d_k^2}\right) \times \\
&\qquad \left(\frac{-1}{\sqrt{1+d_k^2}}(b_k^\top Q + d_k c_k^\top)(s\mathbb{I} - A_k)^{-1} b_k + \sqrt{1+d_k^2}\right) \\
&= b_k^\top(-s\mathbb{I} - A_k^\top)^{-1} c_k c_k^\top (s\mathbb{I} - A_k^\top)^{-1} b_k \\
&+ d_k c_k^\top (s\mathbb{I} - A_k^\top)^{-1} b_k + d_k c_k^\top(-s\mathbb{I} - A_k^\top)^{-1} b_k + 1 + d_k^2 \\
&= K(-s)\frac{\phi_{\mathbf{n}}(-s)}{\phi_{\mathbf{d}}(-s)} K(s) \frac{\phi_{\mathbf{n}}(s)}{\phi_{\mathbf{d}}(s)} + 1\,.
\end{aligned}$$

Next, note that both $K(s)\frac{\phi_{\mathbf{n}}(s)}{\phi_{\mathbf{d}}(s)}$ and $1/K(s)\frac{\phi_{\mathbf{n}}(s)}{\phi_{\mathbf{d}}(s)}$ are, as a consequence of (2.49) and (2.50), analytic on the right half plane. Hence, we have

$$\begin{aligned}&\frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(1+|K(j\omega)|^2\frac{f_{\mathbf{n}}(\omega)}{f_{\mathbf{d}}(\omega)}\right)d\omega\\&=\frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(H(-j\omega)H(j\omega)\right)d\omega\\&=\frac{1}{4\pi}\int_{-\infty}^{\infty}\log\left(|H(-j\omega)|^2\right)d\omega\\&=\Re\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\log\left(H(-j\omega)\right)d\omega\right)\\&=\Re\left(\frac{1}{2\pi j}\oint_{\mathscr{D}}\log\left(H(-s)\right)ds\right),\end{aligned}$$

where $\mathscr{D}$ denotes a contour encompassing from $-j\infty$ to $j\infty$ and enclosing $\mathbb{C}^+$. The value of the integration along the contour can then be evaluated by using the residue of $\log\left(H(-s)\right)$ about $s=\infty$, which is calculated as

$$\begin{aligned}Res(\log\left(H(-s)\right);\infty)&=-\lim_{s\to\infty}s(H(-s)-H(\infty))\\&=\frac{1}{\sqrt{1+d_k^2}}b_k^\top Qb_k+\frac{d_k}{\sqrt{1+d_k^2}}c_k^\top b_k.\end{aligned}$$

Residue theorem in turn yields

$$\Re\left(\frac{1}{2\pi j}\oint_{\mathscr{D}}\log\left(H(-s)\right)ds\right)=\frac{1}{\sqrt{1+d_k^2}}b_k^\top Qb_k+\frac{d_k}{\sqrt{1+d_k^2}}c_k^\top b_k\,.$$

The proof is complete. □

In summary, the following theorem holds.

Theorem 2.4.3. Consider the closed loop shown in Fig. 2.3, and assume that $K(s)\frac{\phi_{\mathbf{n}}(s)}{\phi_{\mathbf{d}}(s)}$ admits a minimal realization $(A_k, b_k, c_k^\top, d_k)$ and A_k is Hurwitz, and $Q>0$ is the unique solution to the ARE in (2.49) and satisfies (2.50). Then

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\log S_{\mathbf{d},\mathbf{e}}(\omega)d\omega=\sum_i\Re(\lambda_i(A))^+ +\frac{1}{\sqrt{1+d_k^2}}b_k^\top Qb_k+\frac{d_k}{\sqrt{1+d_k^2}}c_k^\top b_k\,. \tag{2.51}$$

Remark 2.4.4. The condition that $K(s)\frac{\phi_{\mathbf{n}}(s)}{\phi_{\mathbf{d}}(s)}$ needs to be proper does not impose a significant restriction on the class of closed loops, for which we can derive the same calculations as in Theorem 2.4.3, as one can always choose stabilizing $K(s)$ with higher relative degree, rendering $K(s)\frac{\phi_{\mathbf{n}}(s)}{\phi_{\mathbf{d}}(s)}$ proper.

2.5 Information Rate Inequality & Control with Communication Constraints

Another information rate inequality regarding the closed-loop stability based on the framework in Section 2.2 is obtained in this section. By using it, we investigate the stabilization problem, where the communication channel is modeled as a continuous-time Gaussian channel with certain Signal-to-Noise Ratio (SNR) level constraint.

The following lemma provides a lower bound for the mutual information rate $\bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u})$, which accounts for total information rate flow in the loop. Further insight into $\bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u})$ is provided later in Remark 2.5.2.

Lemma 2.5.1. Consider the closed-loop system shown in Fig. 2.1. We have the following inequality:

$$\bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u}) \geq \bar{I}(\mathbf{x}_0; \mathbf{e}) + \bar{I}(\mathbf{d}; \mathbf{u})\,. \tag{2.52}$$

Proof. Using Kolmogorov's formula (P2), we have

$$I((\mathbf{x}_0, \mathbf{d}^t); \mathbf{u}^t) = I(\mathbf{x}_0; \mathbf{u}^t|\mathbf{d}^t) + I(\mathbf{u}^t; \mathbf{d}^t)\,, \tag{2.53}$$

where $t \in \mathbb{R}^+$ is arbitrary time instance. We can lower bound $I((\mathbf{x}_0, \mathbf{d}^t); \mathbf{u}^t)$ as

$$\begin{aligned}
&I((\mathbf{x}_0, \mathbf{d}^t); \mathbf{u}^t) \\
&\overset{(a)}{=} I(\mathbf{x}_0; \mathbf{e}^t|\mathbf{d}^t) + I(\mathbf{u}^t; \mathbf{d}^t) \\
&\overset{(b)}{=} I(\mathbf{x}_0; \mathbf{e}^t) - I(\mathbf{x}_0; \mathbf{d}^t) + I(\mathbf{x}_0; \mathbf{d}^t|\mathbf{e}^t) + I(\mathbf{u}^t; \mathbf{d}^t) \\
&\overset{(c)}{=} I(\mathbf{x}_0; \mathbf{e}^t) + I(\mathbf{x}_0; \mathbf{d}^t|\mathbf{e}^t) + I(\mathbf{u}^t; \mathbf{d}^t) \\
&\overset{(d)}{\geq} I(\mathbf{x}_0; \mathbf{e}^t) + I(\mathbf{u}^t; \mathbf{d}^t)\,.
\end{aligned} \tag{2.54}$$

Here (a) follows from the fact that $I(\mathbf{x}_0; \mathbf{u}^t|\mathbf{d}^t) = I(\mathbf{x}_0; \mathbf{u}^t{+}\mathbf{d}^t|\mathbf{d}^t) = I(\mathbf{x}_0; \mathbf{e}^t|\mathbf{d}^t)$; (b) follows from (P2); (c) follows from the independence of $\mathbf{d}$ and $\mathbf{x}_0$; and (d) follows from the fact that $I(\mathbf{x}_0; \mathbf{d}^t|\mathbf{e}^t) \geq 0$. We have obtained the following inequality:

$$I((\mathbf{x}_0, \mathbf{d}^t); \mathbf{u}^t) \geq I(\mathbf{x}_0; \mathbf{e}^t) + I(\mathbf{u}^t; \mathbf{d}^t)\,. \tag{2.55}$$

The conclusion is readily obtained by dividing the terms on both sides of (3.9) by t and taking the limit as $t \to \infty$. □

Remark 2.5.2. To illustrate the importance of $\bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u})$, we consider the block diagrams shown in Fig. 2.4, which recast the closed loop in Fig. 2.1 into a typical analog communication scheme with feedback [48]. The "Message" to be transmitted is composed of the two independent sources $\mathbf{x}_0$ and $\mathbf{d}(t)$, and $\mathbf{u}(t)$ is the channel output. We can also identify the "Transmitter" and "Channel" in this "communication system" accordingly, though, in our current setup, they do not function the same way as their names suggest. It turns out to be clear that $\bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u})$ represents the input/ouput information rate, and therefore Lemma 3.2.18 indicates that the total information flow of the closed loop is bounded from below by the contributions of the initial value and the disturbance.

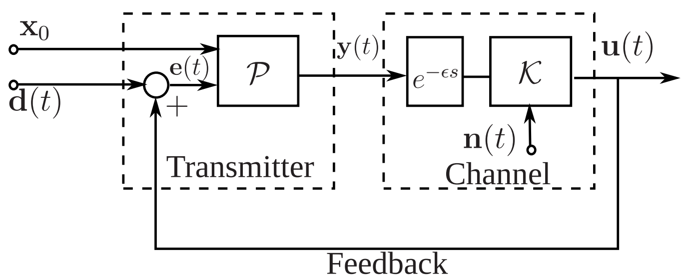

Figure 2.4: Closed loop configuration from the communication perspective

We can then define the *feedback capacity* of the closed loop in Fig. 2.1 as

$$\mathcal{C}_f := \sup_{\mathbf{x}_0, \mathbf{d}} \bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u})\,.$$

Notice that the discrete-time and non-causal version of the feedback capacity has been introduced in several chapters, as [22] and [30].

To take the closed-loop stability into consideration, we further elaborate the inequality (3.2.18) to get the following theorem.

Theorem 2.5.3. If the closed-loop system shown in Fig. 2.1 with feedback capacity $\mathcal{C}_f$ is mean-square stable, then

$$\bar{I}(\mathbf{u}; \mathbf{d}) \leq \mathcal{C}_f - \sum_i \Re(\lambda_i(A))^+\,. \tag{2.56}$$

Example: Stabilization with Gaussian Channel Constraint

Next we focus on the continuous time additive white Gaussian noise (AWGN) channel with input power constraint. This particular type of a communication channel, rooted in Shannon's celebrated work [7], has been intensively studied for its theoretical and practical significance in various chapters, [8] [49] and [50]. To consider the Gaussian channel in a feedback loop, we adopt the same scheme as in [29], which is shown in Fig. 2.5. Here, $\mathcal{P}$ is the same LTI system as in (3.1) and $\mathbf{y}(t) = \mathbf{x}(t)$; $K \in \mathbb{R}^{1\times n}$ is the control gain matrix; $\mathbf{u}(t)$ is the channel input with power constraint $\mathbf{E}[\mathbf{u}^2(t)] \leq \mathscr{P}, \;\; \forall\, t \geq 0$, for some power level $\mathscr{P} > 0$; $\mathbf{d}(t)$ is a Gaussian white noise process with SDF $f_{\mathbf{d}} \equiv \Phi > 0$.

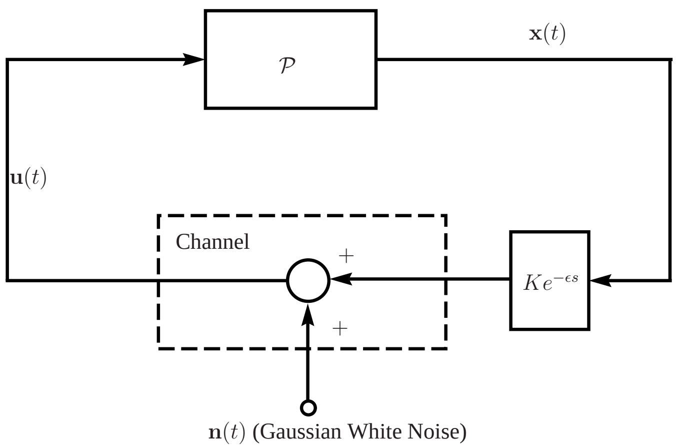

Figure 2.5: Feedback control in the presence of a Gaussian channel

The channel capacity $\mathcal{C}$ can be obtained by the following formula [37]:

$$\mathcal{C} = \frac{\mathscr{P}}{2\Phi}\,. \tag{2.57}$$

Regarding the closed-loop system stability, we have the following theorem.

Theorem 2.5.4. If the closed-loop system shown in Fig. 2.5 is mean-square stable, then the following relationship holds:

$$\frac{\mathscr{P}}{2\Phi} \geq \sum_i \Re(\lambda_i(A))^+\,. \tag{2.58}$$

Proof. Note that $\mathbf{d} \equiv 0$ and the fact that feedback does not change the capacity of memoryless white Gaussian additive channels imply

$$\frac{\mathscr{P}}{2\Phi} = \mathcal{C} = \mathcal{C}_f = \sup_{\mathbf{x}_0} \bar{I}(\mathbf{x}_0; \mathbf{u}) \,.$$

Therefore (2.56) is reduced to

$$\frac{\mathscr{P}}{2\Phi} = \mathcal{C} \geq \bar{I}(\mathbf{x}_0; \mathbf{u}) \geq \sum_i \Re(\lambda_i(A))^+ \,. \tag{2.59}$$

The proof is complete. □

Remark 2.5.5. This result provides a sufficient condition to solve *Problem 1* in [29]. A similar condition is also obtained in [25], where the authors have used the result from [49] on mutual information rate of a Gaussian channel. Different from [29], the method used here is purely information theory-based, and may be applied to more general systems rather than LTI.

2.6 Conclusion

In this chapter we investigated the continuous-time information conservation laws in a causal closed loop feedback setting as an extension from the well established discrete-time case. For the purpose of this extension, we resort to mutual information rate rather than differential entropy rate, whose behavior is not desirable in the continuous-time setting. As a result of the aforementioned conservation laws, a Bode-type integral formula is obtained, for which we have used mutual information integral inequalities instead of the widely used Kolmogorov's formula. We also pursue an in-depth investigation into the resulting Bode integral in terms its tightness and its relation with communication constraints. These conservation laws have also shown the ability of handling particular problems such as control with limited information.

2.7 Proofs

We first introduce an alternative definition of mutual information between two random variables [17].

Definition 2.7.1. Let ξ and η be random variables assuming values in the measurable spaces $(X, \mathscr{F}_x)$ and $(Y, \mathscr{F}_y)$ respectively. The mutual information between $\mathbf{x}$ and $\mathbf{y}$ is given as

$$I(\mathbf{x}, \mathbf{y}) = \sup \sum_{i,j} P_{\xi_n \eta_n}(E_i \times F_j) \log \frac{P_{\xi_n \eta_n}(E_i \times F_j)}{P_{\xi_n}(E_i) P_{\eta_n}(F_j)} \,,$$

where the supremum is taken over all partitions $\{E_i\}$ of X and $\{F_j\}$ of Y.

To prove Lemma 2.1.13, we need the following proposition:

Proposition 2.7.2. Let $\xi_n, n = 1, 2, ...$ and $\eta_n, n = 1, 2, ...$ be random variables. Then, if (ξ_n, η_n) converges to (ξ, η) in distribution, we have

$$I(\xi; \eta) \leq \lim_{n \to \infty} I(\xi_n; \eta_n)$$

Proof. By converging in distribution, we have

$$\lim_{n \to \infty} P_{\xi_n \eta_n}(A) = P_{\xi \eta}(A), \forall A \in \mathscr{F}_x \times \mathscr{F}_y \,.$$

Therefore, for any fixed partition $\{E_i\}$ and $\{F_i\}$ of X and Y, which satisfy $P_\xi(E_i) \neq 0$ and $P_\eta(F_j) \neq 0$, we have

$$\begin{aligned} & \lim_{n \to \infty} \sum_{i,j} P_{\xi_n \eta_n}(E_i \times F_j) \log \frac{P_{\xi_n \eta_n}(E_i \times F_j)}{P_{\xi_n}(E_i) P_{\eta_n}(F_j)} \\ = & \sum_{i,j} P_{\xi \eta}(E_i \times F_j) \log \frac{P_{\xi \eta}(E_i \times F_j)}{P_\xi(E_i) P_\eta(F_j)} \end{aligned}$$

Considering Definition 2.7.1, the following relation can be obtained

$$\begin{aligned} \lim_{n \to \infty} I(\xi_n; \eta_n) &= \lim_{n \to \infty} \sup \sum_{i,j} P_{\xi_n \eta_n}(E_i \times F_j) \log \frac{P_{\xi_n \eta_n}(E_i \times F_j)}{P_{\xi_n}(E_i) P_{\eta_n}(F_j)} \,. \\ &\geq \sup \sum_{i,j} P_{\xi \eta}(E_i \times F_j) \log \frac{P_{\xi \eta}(E_i \times F_j)}{P_\xi(E_i) P_\eta(F_j)} = I(\xi; \eta) \,. \end{aligned}$$

□

Proof of Lemma 2.1.13.[1]

[1]The proof is inspired by a private communication with Dr. V. Prelov

Proof. The first step is to establish the following inequality:

$$I(\mathbf{x}_s^t;\mathbf{y}_s^t) \le \lim_{n\to\infty} I(\mathbf{x}_0^{(\delta(n))},...,\mathbf{x}_n^{(\delta(n))};\mathbf{y}_0^{(\delta(n))},...,\mathbf{y}_n^{(\delta(n))})\,. \tag{2.60}$$

We define

$$\bar{\mathbf{x}}_n := \sum_{i=0}^{n} \mathbf{x}_i^{(\delta(n))}\chi_{[s+\delta(n)i)}\ \ \bar{\mathbf{y}}_n := \sum_{i=0}^{n} \mathbf{y}_i^{(\delta(n))}\chi_{[s+\delta(n)i)}\,,$$

where χ is the characteristic function. Since $\mathbf{x}$ and $\mathbf{y}$ are separable, we can always find the versions of $\mathbf{x}$ and $\mathbf{y}$ such that the joint distribution of them can be arbitrarily approximated by the corresponding discrete-time processes with countable samplings. Therefore the convergence in distribution is implied.

The inequality (2.60) is followed by applying Proposition 2.7.2

$$I(\mathbf{x}_s^t;\mathbf{y}_s^t) \le \lim_{n\to\infty} I(\bar{\mathbf{x}}_n;\bar{\mathbf{y}}_n) = \lim_{n\to\infty} I(\mathbf{x}_0^{(\delta(n))},...,\mathbf{x}_n^{(\delta(n))};\mathbf{y}_0^{(\delta(n))},...,\mathbf{y}_n^{(\delta(n))})\,.$$

On the other hand, the following relation is immediately obtained by (P1)

$$I(\mathbf{x}_s^t;\mathbf{y}_s^t) \ge I(\mathbf{x}_0^{(\delta(n))},...,\mathbf{x}_n^{(\delta(n))};\mathbf{y}_0^{(\delta(n))},...,\mathbf{y}_n^{(\delta(n))})\,, \forall n \ge 1\,. \tag{2.61}$$

The proof is completed by combining (2.60) and (2.61).

CHAPTER 3

BODE'S INTEGRAL FOR STOCHASTIC SWITCHED SYSTEMS

The chapter is organized as follows. In Section 3.1 we introduce the closed-loop feedback configuration and some basic definitions and facts from information theory and the theory of stochastic processes. Section 3.2 studies a general feedback scheme, within which we develop a mutual information inequality and a Bode-type integral formula. Section 3.3 applies Bode's integral to NCS, while Section 3.4 carries out the analysis of its application to macroeconomics. The chapter is concluded in Section 3.5.

3.1 Preliminaries & Problem Formulation

Notation:

- $\mathbb{R}$ denotes the field of real numbers; $\mathbb{C}$ stands for complex plane; $\mathbb{C}^-$ and $\mathbb{C}^+$ stand for the left half and right half of $\mathbb{C}$ respectively.
- Random variables defined in appropriate probability spaces are represented using boldface letters, such as $\mathbf{x}$, $\mathbf{y}$. If not otherwise stated, the random variables take values in $\mathbb{R}$ throughout the chapter.
- If $\mathbf{x}(k), k \in \mathbb{N}^+$, is a discrete time stochastic process, we denote its segment $\{\mathbf{x}(k)\}_{k=l}^{u}$ by $\mathbf{x}_l^u$, and use $\mathbf{x}_0^n := \mathbf{x}^n$ for simplicity.
- $\mathbf{E}[\cdot]$ is the expectation operator of a random variable.
- $(\cdot)^+ = \max\{\cdot, 0\}$ and $(\cdot)^- = \min\{\cdot, 0\}$.
- $\Re(\cdot)$ gives the real part of a complex number.
- $\lambda_j(\cdot)$ gives the eigenvalues of a square matrix.

- $h(\cdot)$ stands for (differential) entropy and $I(\cdot;\cdot|\cdot)$ for conditioned mutual information; $\bar{h}$ and $\bar{I}$ stand for the entropy rate and mutual information rate respectively.
- When A is a finite set, $|A|$ gives the number of elements in A.
- $sp\{\cdot\}$ denotes the spectrum of an operator.

A list of useful properties of entropy and mutual information are given here, and are frequently used in the upcoming arguments.

(P1) *Symmetry and nonnegativity*:

$$I(\mathbf{x};\mathbf{y}) = I(\mathbf{y};\mathbf{x}) = h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \geq 0\,.$$

(P2) *Kolmogorov equality*:

$$I(\mathbf{x};(\mathbf{y},\mathbf{z})) = I(\mathbf{x};\mathbf{z}) + I(\mathbf{x};\mathbf{y}|\mathbf{z})$$

(P3) *Data processing inequality*:

$$I(\mathbf{x};\mathbf{y}|\mathbf{z}) \geq I(\mathbf{x};g(\mathbf{y})|\mathbf{z})$$

The equality holds, if $g(\cdot)$ is invertible.

(P4) *Invariance of mutual information (entropy)*

$$I(\mathbf{x};\mathbf{y}|\mathbf{z}) = I(\mathbf{x}+g(\mathbf{z});\mathbf{y}|\mathbf{z})\,, h(\mathbf{x}|\mathbf{z}) = h(\mathbf{x}+g(\mathbf{z})|\mathbf{z}),$$

where $g(\cdot)$ is a function.

(P5) *Chain rule*:

$$h(\mathbf{x}^n|\mathbf{y}) = \sum_{k=1}^{n} h(\mathbf{x}_k|\mathbf{y},\mathbf{x}^{k-1})$$

(P6) *Maximum entropy*: Consider $\mathbf{x} \in \mathbb{R}^m$ and the covariance matrix given by $V := \mathbf{E}[\mathbf{x}\mathbf{x}^\top]$. Then we have

$$h(\mathbf{x}) \leq h(\bar{\mathbf{x}}) = \frac{1}{2}\log((2\pi e)^m \det V)\,,$$

where $\bar{\mathbf{x}}$ is a Gaussian process with the same covariance as $\mathbf{x}$. Equality holds, if $\mathbf{x}$ is Gaussian.

Throughout the paper we consider the feedback configuration depicted in Fig. 3.1.

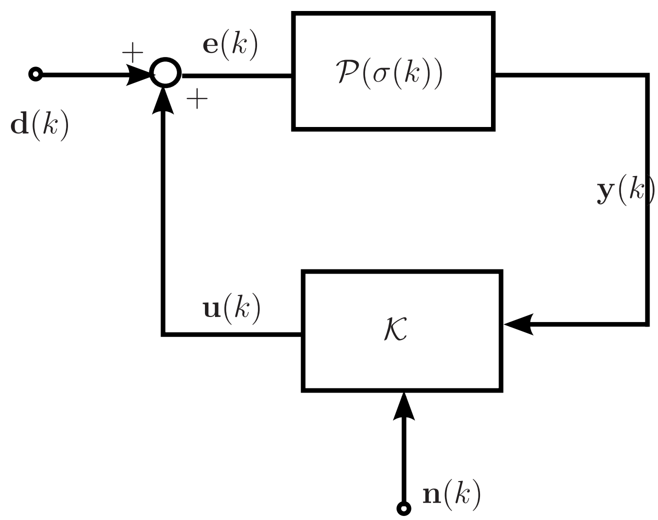

Figure 3.1: Basic Feedback Scheme

Several assumptions are made:

- The plant $\mathcal{P}$ is modeled by the following stochastic difference equation

$$\begin{aligned} \mathbf{x}(k+1) &= A(\boldsymbol{\sigma}(k))\mathbf{x}(k) + B(\boldsymbol{\sigma}(k))\mathbf{e}(k)\,, \quad \mathbf{x}(0) = \mathbf{x}_0\,, \\ \mathbf{y}(k) &= C(\boldsymbol{\sigma}(k))\mathbf{x}(k)\,, \quad k = 0, 1, 2.... \end{aligned} \tag{3.1}$$

  Here $\mathbf{x}(k) \in \mathbb{R}^m$, and $\mathbf{x}_0$ is assumed to have finite differential entropy or $h(\mathbf{x}_0) < \infty$, and $\boldsymbol{\sigma}(k) \in \{1, 2, ..., N\} =: \mathcal{N}$ is a finite state ergodic Markov process given by

$$P(\boldsymbol{\sigma}(k+1) = j|\boldsymbol{\sigma}(k) = i) := p_{ij} \geq 0\,,$$

  where p_{ij} is named as transition probability from state i to j, and $\sum_j p_{ij} = 1$ for all $i \in \mathcal{N}$. The stationary distribution of the Markov chain $\boldsymbol{\sigma}$, denoted as $\boldsymbol{\pi} = [\pi_1, ..., \pi_{|\mathcal{N}|}]$, is obtained by solving

$$\boldsymbol{\pi}^\top [p_{ij}]_{i,j\in\mathcal{N}} = \boldsymbol{\pi}^\top\,, \text{ and } \quad [1, .., 1]\boldsymbol{\pi} = 1\,.$$

- The disturbance $\mathbf{d}(k)$ is a stochastic process, and $\mathbf{n}(k)$ is a stochastic process that models the controller noise. We assume that $\boldsymbol{\sigma}(k)$, $\mathbf{d}(k)$, $\mathbf{n}(k)$ and $\mathbf{x}_0$ are mutually independent.

- The controller $\mathcal{K}$ is given as a deterministic causal map such that

$$\mathcal{K} : (k, \mathbf{y}^{k-1}, \mathbf{n}^k) \mapsto \mathbf{u}(k)\,.$$

Definition 3.1.1 (*Wide Sense Stationary Process*). A zero-mean stochastic process $\mathbf{x}(k) \in \mathbb{R}^n$, $t \geq 0$, is stationary, if for all $k \geq 0$ its covariance function, defined by

$$R_{\mathbf{x}}(l) = \mathbf{E}[\mathbf{x}(k+l)\mathbf{x}^\top(k)], \quad l \in \mathbb{N}^+\,,$$

is independent of l. Throughout this chapter, *wide sense stationary* is abbreviated as *stationary* for convenience.

Definition 3.1.2. The spectral density of a stationary process $\mathbf{v}$ is given as the following Fourier transform

$$f_{\mathbf{v}}(\omega) = \frac{1}{2\pi}\sum_{k=0}^{\infty} R_{\mathbf{v}}(k)e^{-j\omega k}$$

Definition 3.1.3 (*Sensitivity-like Function*). A sensitivity-like function of the closed loop is defined as

$$S_{\mathbf{d},\mathbf{e}}(\omega) = \sqrt{\frac{f_{\mathbf{e}}(\omega)}{f_{\mathbf{d}}(\omega)}}\,,$$

where $\mathbf{e}$ and $\mathbf{d}$ are stationary and stationarily correlated.

Remark 3.1.4. The function $S_{\mathbf{d},\mathbf{e}}(\omega)$ is the stochastic analogue of the sensitivity function $|S(j\omega)|$ in Bode's original work [43].

Throughout, we adopt the following stability definition.

Definition 3.1.5 (*Mean-square Stability*). The closed loop given in Fig. 3.1 is said to be mean-square stable, if

$$\sup_{k\geq 0} \mathbf{E}[\mathbf{x}^\top(k)\mathbf{x}(k)] < \infty\,. \tag{3.2}$$

Definition 3.1.6 (*Lie Algebra*). A Lie algebra is denoted as

$$\mathfrak{g} := \{A(n) : n \in \mathcal{N}\}_{LA} ,$$

which is generated by the matrices $A(n), n \in \mathcal{N}$, with respect to the standard Lie bracket

$$[A(1), A(2)] := A(1)A(2) - A(2)A(1) .$$

We say that the Lie algebra $\mathfrak{g}$ is *solvable* if the following derived series

$$\mathfrak{g} > [\mathfrak{g}, \mathfrak{g}] > [[\mathfrak{g}, \mathfrak{g}], [\mathfrak{g}, \mathfrak{g}]] > ...$$

becomes 0 eventually, where "$>$" denotes the relation of sub-algebra.

Theorem 3.1.7. [Simultaneous triangularization] The matrices $\{A(n) : n \in \mathcal{N}\}$ can be simultaneously triangularized by some linear operator $T \in \mathbb{C}^{m\times m}$, if and only if the Lie algebra $\mathfrak{g}$ is *solvable*.

3.2 Bode-like Integral Discrete Time Case

In this section we develop the information conservation law of the closed loop system depicted in Fig. 3.1. In turn, an analogue of Bode's formula is obtained with stationarity assumption.

3.2.1 Information conservation law

The following lemma is introduced to characterize the closed loop causality.

Lemma 3.2.1.

$$I(\mathbf{d}(i); (\mathbf{u}^i, \mathbf{x}_0, \boldsymbol{\sigma}^i)|\mathbf{d}^{i-1}) = 0 , \quad \forall\, i \geq 1. \tag{3.3}$$

Proof.

$$\begin{aligned}
&I(\mathbf{d}(i);(\mathbf{u}^i,\mathrm{x}_0,\boldsymbol{\sigma}^i)|\mathbf{d}^{i-1})\\
&\overset{(a)}{\leq} I(\mathbf{d}(i);(\mathbf{u}^i,\mathbf{n}^i,\boldsymbol{\sigma}^i,\mathrm{x}_0)|\mathbf{d}^{i-1})\\
&\overset{(b)}{\leq} I(\mathbf{d}(i);(\mathbf{d}^{i-1},\mathbf{n}^i,\boldsymbol{\sigma}^i,\mathrm{x}_0)|\mathbf{d}^{i-1})\\
&\overset{(c)}{=} I(\mathbf{d}(i);(\mathbf{n}^i,\boldsymbol{\sigma}^i,\mathrm{x}_0)|\mathbf{d}^{i-1})\\
&\overset{(d)}{=} 0
\end{aligned}$$

Here, (a) follows from (P3); (b) also follows from (P3), since $\mathbf{u}^i$ is a function of $(\mathbf{d}^{i-1},\mathbf{n}^i,\boldsymbol{\sigma}^i,\mathrm{x}_0)$; (c) follows from (P4), and (d) is implied because $\mathbf{n}$, $\boldsymbol{\sigma}$, x_0 and $\mathbf{d}$ are mutually independent. □

In what follows we use the result from Lemma 3.2.1 to achieve an equality, revealing a key relationship among signals residing in 3.1.

Lemma 3.2.2. Consider the closed loop in Fig. 3.1. The following inequality holds

$$h(\mathbf{e}^k)=h(\mathbf{d}^k)+I((\mathrm{x}_0,\boldsymbol{\sigma}^k);\mathbf{e}^k)+\sum_{i=1}^{k} I(\mathbf{u}^i;\mathbf{e}(i)|\mathbf{e}^{i-1},\mathrm{x}_0,\boldsymbol{\sigma}^k) \tag{3.4}$$

Proof. We break down the equality (3.3) by

$$\begin{aligned}
0&=I(\mathbf{d}(i);(\mathbf{u}^i,\mathrm{x}_0,\boldsymbol{\sigma}^i)|\mathbf{d}^{i-1})\\
&\overset{(a)}{=} I(\mathbf{d}(i);\mathbf{u}^i,\mathrm{x}_0,\boldsymbol{\sigma}^i,\mathbf{d}^{i-1})-I(\mathbf{d}(i);\mathbf{d}^{i-1})\\
&\overset{(b)}{=} I(\mathbf{d}(i);\mathbf{u}^i,\mathrm{x}_0,\boldsymbol{\sigma}^i,\mathbf{e}^{i-1})-I(\mathbf{d}(i);\mathbf{d}^{i-1})\\
&\overset{(c)}{=} -h(\mathbf{d}(i)|\mathbf{u}^i,\mathrm{x}_0,\boldsymbol{\sigma}^i,\mathbf{e}^{i-1})+h(\mathbf{d}(i)|\mathbf{d}^{i-1})\\
&\overset{(d)}{=} -h(\mathbf{e}(i)|\mathbf{u}^i,\mathrm{x}_0,\boldsymbol{\sigma}^i,\mathbf{e}^{i-1})+h(\mathbf{d}(i)|\mathbf{d}^{i-1})\\
&\overset{(e)}{=} -h(\mathbf{e}(i)|\mathbf{e}^{i-1})+I((\mathrm{x}_0,\boldsymbol{\sigma}^i));\mathbf{e}(i)|\mathbf{e}^{i-1})+\\
&\qquad I(\mathbf{u}^i;\mathbf{e}(i)|\mathbf{e}^{i-1},\mathrm{x}_0,\boldsymbol{\sigma}^i)+h(\mathbf{d}(i)|\mathbf{d}^{i-1})\,.
\end{aligned}$$

Here (a) follows from (P3), (b) follows from the fact that $\mathbf{e}^{i-1}=\mathbf{u}^{i-1}+\mathbf{d}^{i-1}$, (c) follows from (P1), (d) follows from (P4) and (f) from (P5). Summing up the above equality from 1 to k and using (P5), we have (3.4). □

Remark 3.2.3. The term $\sum_{i=1}^{k} I(\mathbf{u}^i; \mathbf{e}(i)|\mathbf{e}^{i-1}, \mathbf{x}_0, \boldsymbol{\sigma}^k)$ is alternatively represented as the directed information from $\mathbf{u}$ to $\mathbf{e}$ conditioned by $(\mathbf{x}_0, \boldsymbol{\sigma}^k)$ [44].

Theorem 3.2.4. Consider the closed loop shown in Fig. 3.1. The following entropy rate inequality holds

$$\bar{h}(\mathbf{e}) \geq \bar{h}(\mathbf{d}) + \bar{I}((\mathbf{x}_0, \boldsymbol{\sigma}); \mathbf{e})\,. \tag{3.5}$$

Proof. Considering the nonegativeness of the mutual information, from (3.4) we have

$$h(\mathbf{e}^k) \geq h(\mathbf{d}^k) + I((\mathbf{x}_0, \boldsymbol{\sigma}^k); \mathbf{e}^k)\,.$$

The proof is completed by dividing both sides of the above equality by k and letting $k \to \infty$. $\square$

Remark 3.2.5. The inequality in (3.5) has been derived in both information theory and control theory literature in different setups and with different generalities. Here we only assume causality of the closed loop.

3.2.2 Evaluating an important information rate

As it can be seen in (3.5), the mutual information rate $\bar{I}((\mathbf{x}_0, \boldsymbol{\sigma}); \mathbf{e})$ plays an important role in the conservation law. In this subsection we establish some nontrivial lower bounds for $\bar{I}((\mathbf{x}_0, \boldsymbol{\sigma}); \mathbf{e})$ assuming some algebraic conditions.

Theorem 3.2.6. Consider the closed loop in Fig. 3.1. The following inequality holds.

$$\bar{I}((\mathbf{x}_0, \boldsymbol{\sigma}); \mathbf{e}) \geq \liminf_{k\to\infty} \frac{1}{k}\mathbf{E} \sum_{j} \Re\left(\log \lambda_j\left(F_k\right)\right)^+ \,, \tag{3.6}$$

where $F_k := A(\boldsymbol{\sigma}(k))A(\boldsymbol{\sigma}(k-1))\cdots A(\boldsymbol{\sigma}(0))$.

Proof. We first consider the dynamics of the plant

$$\mathbf{x}(k+1) = \mathbf{x}(k)A(\boldsymbol{\sigma}(k)) + B(\boldsymbol{\sigma}(k))\mathbf{e}(k)\,,$$

which can be solved as

$$\begin{aligned}\mathbf{x}(k+1) &= \left(\prod_{i=0}^{k} A(\boldsymbol{\sigma}(i))\right)\mathbf{x}_0 + \sum_{i=0}^{k}\left(\prod_{l=i}^{k} A(\boldsymbol{\sigma}(l))\right)B(\boldsymbol{\sigma}(i))\mathbf{e}(i)\\ &= F_k(\mathbf{x}_0 - \hat{\mathbf{x}}_0(k+1)),\end{aligned}$$

where

$$\hat{\mathbf{x}}_0(k+1) := \left(\prod_{i=0}^{k} A(\boldsymbol{\sigma}(i))\right)^{-1} \sum_{i=0}^{k} \left(\prod_{l=i}^{k} A(\boldsymbol{\sigma}(l))\right) B(\boldsymbol{\sigma}(i))\mathbf{e}(i)\,.$$

F_k can be decomposed into the following form by a linear transformation T_k:

$$T_k^{-1} F_k T_k = \left[\begin{array}{cc} F_{ku} & 0 \\ 0 & F_{ks} \end{array}\right],$$

where F_{ku} is unstable and F_{ks} is stable. The same linear transformation can be applied to $\mathbf{x}_0$ and $\hat{\mathbf{x}}_0$ to have

$$T_k \mathbf{x}_0 = \left[\begin{array}{c} \mathbf{x}_{u0} \\ \mathbf{x}_{s0} \end{array}\right] \quad \text{and } T_k \hat{\mathbf{x}}_0 = \left[\begin{array}{c} \hat{\mathbf{x}}_{u0} \\ \hat{\mathbf{x}}_{s0} \end{array}\right].$$

We then establish the lower bound of $I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k)$ as follows

$$\begin{aligned}
& I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k) \\
& \stackrel{(a)}{=} I(\mathbf{x}_0; \boldsymbol{\sigma}^k) + I(\mathbf{x}_0; \mathbf{e}^k | \boldsymbol{\sigma}^k) \\
& \stackrel{(b)}{=} I(\mathbf{x}_0; \mathbf{e}^k | \boldsymbol{\sigma}^k) \\
& \stackrel{(c)}{=} I(\mathbf{x}_{u0}, \mathbf{x}_{s0}; \mathbf{e}^k, \boldsymbol{\sigma}^k) \\
& \stackrel{(d)}{=} h(\mathbf{x}_{u0}, \mathbf{x}_{s0}) - h(\mathbf{x}_{s0} | \mathbf{e}^k, \boldsymbol{\sigma}^k) - h(\mathbf{x}_{u0} | \mathbf{x}_{s0}, \mathbf{e}^k, \boldsymbol{\sigma}^k) \\
& \stackrel{(e)}{\geq} h(\mathbf{x}_{u0}, \mathbf{x}_{s0}) - h(\mathbf{x}_{s0}) - h(\mathbf{x}_{u0} | \mathbf{x}_{s0}, \mathbf{e}^k, \boldsymbol{\sigma}^k)\,.
\end{aligned}$$

Here (a) follows from P4, (b) follows from P1 and the fact that $\mathbf{x}_0$ and $\boldsymbol{\sigma}$ are independent (and therefore $I(\mathbf{x}_0; \boldsymbol{\sigma}^k) = h(\mathbf{x}_0)$), (c) follows from P3, (d) follows from P1 and (e) follows from the fact that $h(\mathbf{x}_{s0} | \mathbf{e}^k, \boldsymbol{\sigma}^k) \leq h(\mathbf{x}_{s0})$.

To evaluate the term $h(\mathbf{x}_{u0} | \mathbf{x}_{s0}, \mathbf{e}^k, \boldsymbol{\sigma}^k)$, we note that

$$\begin{aligned}
& h(\mathbf{x}_{u0} | \mathbf{x}_{s0}, \mathbf{e}^k, \boldsymbol{\sigma}^k) \\
& = h(\mathbf{x}_{u0} - \hat{\mathbf{x}}_{u0} | \mathbf{x}_{s0}, \mathbf{e}^k, \boldsymbol{\sigma}^k) \\
& \leq h(\mathbf{x}_{u0} - \hat{\mathbf{x}}_{u0}) \\
& \leq \log(2\pi e)^l - \log \mathbf{E} |\det F_{ku}| + \log \mathbf{E} \det \mathbf{x}_{u0}(k) \mathbf{x}_{u0}^\top(k) \\
& \leq \log(2\pi e)^l - \mathbf{E} \log |\det F_{ku}| + \log \mathbf{E} \det \mathbf{x}_{u0}(k) \mathbf{x}_{u0}^\top(k),
\end{aligned}$$

where l is the dimension of $\mathbf{x}_{u0}$ and the last inequality follows form Jensen's inequality.

Therefore we have

$$I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k) \geq -\log(2\pi e)^l + \mathbf{E} \log |\det F_{ku}| \\ - \log \mathbf{E} \det \mathbf{x}_u(k)\mathbf{x}_u^\top(k) \,.$$

Note that the stability of the closed loop system implies that $\mathbf{E} \det \mathbf{x}_u(k)\mathbf{x}_u^\top(k) < \infty \,, \forall k$. Then we have

$$\begin{aligned} \bar{I}(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k) &\geq \liminf_{k\to\infty} \frac{1}{k} \mathbf{E} \log |\det F_{ku}| \\ &= \liminf_{k\to\infty} \frac{1}{k} \mathbf{E} \sum_j \Re \left(\log \lambda_j \left(F_k\right)\right)^+ \,. \end{aligned}$$

□

Remark 3.2.7. The right hand side of (3.6) is actually a Lyapunov exponent for the dynamic system (3.1). For a complete treatment of Lyapunov exponents for stochastic switching systems, please refer to [51].

To overcome the difficulty of obtaining $\liminf_{k\to\infty} \frac{1}{k} \mathbf{E} \sum_j \Re \left(\log \lambda_j \left(F_k\right)\right)^+$ by using Lyapunov exponent method, we exploit the algebraic structure of the matrices $A(n), n \in \mathcal{N}$. From Theorem 3.1.7 we know that the solvability of $\mathfrak{g}$ implies that $\{A(n)\}\,, n \in \mathcal{N}$, can be simultaneously triangularizable by some linear transformation $T \in \mathbb{C}^{m\times m}$:

$$T^{-1}A(n)T = \begin{bmatrix} \lambda_1^{(n)} & \star & \star \\ 0 & \ddots & \star \\ 0 & 0 & \lambda_m^{(n)} \end{bmatrix} , \forall n \in \mathcal{N} \,. \tag{3.7}$$

Now we divide the index set $\{1, ..., m\}$ into two distinct sets $\mathcal{M}_u$ and $\mathcal{M}_s$, defined by

$$\mathcal{M}_u := \left\{ j : \prod_{n\in\mathcal{N}} |\lambda_j^{(n)}|^{\pi_n} > 1 \,, j = 1, 2, ..., m \right\} ,$$

$$\mathcal{M}_s := \{1, ..., m\} \setminus \mathcal{M}_u \,.$$

Corollary 3.2.8. Suppose that the Lie algebra $\mathfrak{g}$ is *solvable*. Then we have

$$\bar{I}((\mathbf{x}_0, \boldsymbol{\sigma}); \mathbf{e}) \geq \sum_{n\in\mathcal{N}} \sum_{j\in\mathcal{M}_u} \pi_n \log |\lambda_j^{(n)}|$$

Proof. We start with a mutually disjoint partition of the index set $\{1, 2, ..., \boldsymbol{\sigma}(k)\}$, given by

$$\{1, 2, ..., \boldsymbol{\sigma}(k)\} = \bigcup_{n \in \mathcal{N}} \mathcal{K}_n \,,$$

where $\mathcal{K}_n := \{i : \boldsymbol{\sigma}(i) = n, i = 1, 2, .., k\}$. Then we claim that the eigenvalues of F_k take the form $\lambda_j(F_k) = \prod_{n \in \mathcal{N}} \prod_{j=1}^{m} \left(\lambda_j^{(n)}\right)^{|K_n|}$, where $\lambda_j^{(n)}$ is the diagonal entry from (3.7). Indeed it is easy to see that $T^{-1}F_kT = T^{-1}A(\boldsymbol{\sigma}(k))TT^{-1}A(\boldsymbol{\sigma}(k-1))T \cdots T^{-1}A(\boldsymbol{\sigma}(0))T$ is a triangular matrix for all k. Further, the jth diagonal entry of $T^{-1}F_kT$ can be calculated as

$$\lambda_j(F_k) = \prod_{i=0}^{k} \lambda_j^{(\boldsymbol{\sigma}(i))} = \prod_{n \in \mathcal{N}} \left(\lambda_j^{(n)}\right)^{|K_n|}$$

Using the above relation and Fatou's Lemma we have

$$\begin{aligned}
&\liminf_{k\to\infty} \frac{1}{k} \mathbf{E} \sum_j \Re \left(\log \lambda_j \left(F_k\right)\right)^+ \\
&= \liminf_{k\to\infty} \mathbf{E} \frac{1}{k} \sum_j \Re \left(\log \lambda_j(F_k)\right)^+ \\
&\geq \mathbf{E} \liminf_{k\to\infty} \frac{1}{k} \sum_j \Re \left(\log \lambda_j(F_k)\right)^+ .
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\liminf_{k\to\infty} \frac{1}{k} \sum_j \Re \left(\log \lambda_j(F_k)\right)^+ \\
&= \lim_{k\to\infty} \frac{1}{k} \sum_j \Re \left(\log \lambda_j(F_k)\right)^+ \\
&= \sum_j \Re \left(\sum_n \pi_n \log \lambda_j^{(n)} \right)^+ \\
&= \sum_{n \in \mathcal{N}} \sum_{j \in \mathcal{M}_u} \pi_n \log |\lambda_j^{(n)}| \,,
\end{aligned}$$

where the second equality follows from ergodicity of $\boldsymbol{\sigma}(k)$. $\square$

Remark 3.2.9. As explained in [18], this modern algebraic approach, though mathematically appealing, shows a significant drawback for its lack of robustness, i.e. even a very small perturbation of system parameters can violate the solvability condition. One may conduct perturbation analysis to relax the algebraic structure requirement, though it is not trivial in general.

Here we propose yet another way to determine the Lyapunov exponent $\liminf_{k\to\infty}\frac{1}{k}\mathbf{E}\sum_i \Re\left(\log\lambda_i\left(F_k\right)\right)^+$ by using operator semigroup theory. To start with, we consider the semigroup generated by matrices $\{A(n)\,, n\in\mathcal{N}\}$ with respect to the matrix multiplication. The following lemma from [52] gives a sufficient condition for the permutability of the spectra of the product of the operators.

Theorem 3.2.10. If for all $n_1, n_2, n_3 \in \mathcal{N}$,

$$sp(A(n_1)A(n_2)A(n_3)) = sp(A(n_2)A(n_1)A(n_3))\,, \tag{3.8}$$

then for any sequence $A(n_1), ..., A(n_k)\,,\ \ n_1, ..., n_k \in \mathcal{N}$, the following identity holds for any permutation τ of $\{n_1, ..., n_k\}$

$$sp\left\{\prod_{i}^{k} A(n_i)\right\} = sp\left\{\prod_{\tau(i)}^{\tau(k)} A(n_{\tau(i)})\right\}.$$

The following corollary is now straightforward to prove.

Corollary 3.2.11. Suppose that the condition in (3.8) is satisfied. Then we have

$$\bar{I}((\mathbf{x}_0,\boldsymbol{\sigma});\mathbf{e}) \geq \sum_j \Re\left(\log\lambda_j\left(\prod_{n\in\mathcal{N}} A(n)^{\pi_n}\right)\right)^+ .$$

Proof. Theorem 3.2.10 implies that

$$sp(F_k) = sp\left\{\prod_{n\in\mathcal{N}} A^{|\mathcal{K}_n|}(n)\right\} = sp\left\{\prod_{n\in\mathcal{N}} A^{|\mathcal{K}_{\tau(n)}|}(\tau(n))\right\}$$

$$= \{\hat{\lambda}_1^{(k)}, \ldots, \hat{\lambda}_m^{(k)}\}$$

for any permutation $\tau(\cdot)$, where $\hat{\lambda}_j^{(k)} = \prod_{n\in\mathcal{N}}\left(\lambda_j^{(k)}\right)^{|K_n|}$. Following the same argument in the proof of Corollary 3.2.8, we have

$$\liminf_{k\to\infty}\frac{1}{k}\mathbf{E}\sum_j \Re\left(\log\lambda_j\left(F_k\right)\right)^+ \geq$$
$$\mathbf{E}\liminf_{k\to\infty}\frac{1}{k}\sum_j \Re\left(\log\hat{\lambda}_j^{(k)}\right)^+$$

and

$$
\begin{aligned}
&\liminf_{k\to\infty} \frac{1}{k} \sum_j \Re \left(\log \hat{\lambda}_j^{(k)}\right)^+ \\
&= \lim_{k\to\infty} \frac{1}{k} \sum_j \Re \left(\log \hat{\lambda}_j^{(k)}\right)^+ \\
&= \sum_j \Re \left(\sum_n \pi_n \log \lambda_j^{(n)}\right)^+ \\
&= \sum_j \Re \left(\log \lambda_j \left(\prod_{n\in\mathcal{N}} A(n)^{\pi_n}\right)\right)^+ .
\end{aligned}
$$

The theorem is proved. □

3.2.3 Bode's Integral

Theorem 3.2.12. Consider the closed loop in Fig. 3.1. If $\mathbf{d}$ and $\mathbf{e}$ form Gaussian stationary processes, then

$$
\frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right) d\omega \geq \liminf_{k\to\infty} \mathbf{E} \frac{1}{k} \sum_i \Re \left(\log \lambda_i \left(F_k\right)\right)^+ .
$$

Proof. This result is evident by considering the following relation, followed by Kolmogrov's formula [15]

$$
\bar{h}(\mathbf{d}) = \log(2\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log f_{\mathbf{d}}(\omega) d\omega \,,
$$

$$
\bar{h}(\mathbf{e}) = \log(2\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log f_{\mathbf{e}}(\omega) d\omega \,,
$$

together with Theorem 3.2.6. □

Since we have obtained various lower bounds for $\bar{I}(\mathbf{x}_0, \mathbf{d}, \boldsymbol{\sigma}; \mathbf{e})$ in the previous subsection, the following corollaries can be readily obtained.

Corollary 3.2.13. Consider the closed loop in Fig. 3.1. If $\mathbf{d}$ and $\mathbf{e}$ form Gaussian stationary processes, then

$$
\frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right) d\omega \geq \log \prod_{n\in\mathcal{N}} \left|\det A(n)\right|^{\pi_n} .
$$

Corollary 3.2.14. Consider the closed loop in Fig. 3.1. If $\mathbf{d}$ and $\mathbf{e}$ form Gaussian stationary processes, and the Lie algebra $\mathfrak{g}$ is *solvable*, then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right)d\omega \geq \sum_j \Re\left(\log\lambda_j\left(\prod_{n\in\mathcal{N}} A(n)^{\pi_n}\right)\right)^+ .$$

Corollary 3.2.15. Consider the closed loop in Fig. 3.1. If $\mathbf{d}$ and $\mathbf{e}$ form Gaussian stationary processes, and the condition in (3.8) is satisfied, then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right)d\omega \geq \sum_{n\in\mathcal{N}}\sum_{j\in\mathcal{M}_u}\pi_n\log|\lambda_j^{(n)}| .$$

Remark 3.2.16. Similar to its deterministic counterpart, Bode's integral in this stochastic setting also captures the performance limitation of a closed loop in frequency domain. The lower bound of the achievable performance is inherent from its open loop plant instability.

Remark 3.2.17. Though it is hard to determine whether the closed loop in Fig. 3.1 is stationary in general, some results for LTI systems can be found in [53] and [54].

3.2.4 Data Rate Inequality

Another inequality, resulting from the closed loop causality, is developed here. The following lemma provides a lower bound for the mutual information rate $\bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u})$, which accounts for total information rate flow in the loop. Further insight into $\bar{I}((\mathbf{x}_0, \mathbf{d}); \mathbf{u})$ can be found in [10] and [55].

Lemma 3.2.18. Consider the closed-loop system shown in Fig. 3.1. We have the following inequality:

$$\bar{I}((\mathbf{x}_0, \mathbf{d}, \boldsymbol{\sigma}); \mathbf{u}) \geq \bar{I}(\mathbf{x}_0, \boldsymbol{\sigma}; \mathbf{e}) + \bar{I}(\mathbf{d}; \mathbf{u}) .$$

Proof. Using Kolmogorov's formula (P2), we have

$$I((\mathbf{x}_0, \mathbf{d}^k, \boldsymbol{\sigma}^k); \mathbf{u}^k) = I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{u}^k|\mathbf{d}^k) + I(\mathbf{u}^k; \mathbf{d}^k) ,$$

where $k \in \mathbb{N}^+$ is an arbitrary time instance. We can lower bound $I((\mathbf{x}_0, \mathbf{d}^k); \mathbf{u}^k)$ as

$$
\begin{aligned}
& I((\mathbf{x}_0, \boldsymbol{\sigma}^k, \mathbf{d}^k); \mathbf{u}^k) \\
& \overset{(a)}{=} I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k | \mathbf{d}^k) + I(\mathbf{u}^k; \mathbf{d}^k) \\
& \overset{(b)}{=} I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k) - I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{d}^k) + \\
& \qquad\qquad I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{d}^k | \mathbf{e}^k) + I(\mathbf{u}^k; \mathbf{d}^k) \\
& \overset{(c)}{=} I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k) + I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{d}^k | \mathbf{e}^k) + I(\mathbf{u}^k; \mathbf{d}^k) \\
& \overset{(d)}{\geq} I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k) + I(\mathbf{u}^k; \mathbf{d}^k) \,.
\end{aligned}
$$

Here (a) follows from the fact that $I(\mathbf{x}_0; \mathbf{u}^k | \mathbf{d}^k) = I(\mathbf{x}_0; \mathbf{u}^k + \mathbf{d}^k | \mathbf{d}^k) = I(\mathbf{x}_0; \mathbf{e}^k | \mathbf{d}^k)$; (b) follows from (P2); (c) follows from the independence of $\mathbf{d}$ and $\mathbf{x}_0$; and (d) follows from the fact that $I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{d}^k | \mathbf{e}^k) \geq 0$. We have obtained the following inequality:

$$
I((\mathbf{x}_0, \mathbf{d}^k, \boldsymbol{\sigma}^k); \mathbf{u}^k) \geq I(\mathbf{x}_0, \boldsymbol{\sigma}^k; \mathbf{e}^k) + I(\mathbf{u}^k; \mathbf{d}^k) \,. \tag{3.9}
$$

The conclusion is readily obtained by dividing the terms on both sides of (3.9) by k and taking the limit as $k \to \infty$. $\square$

3.3 Networked Control Systems with Random Packet Dropouts

In this section, we apply the framework from the previous section to examine the performance limitation problems in the networked control systems (NCS). To be specific, we only consider the control systems with a lossy communication channel placed between the sensor and the controller, which has been studied in various chapters [56] [57] [58]. In this chapter we adopt a structure similar to [57], shown in Fig. 3.2, where an erasure channel is employed to model a packet dropout.

The packet dropouts are compensated for by an output of an LTI system, which has to be designed. The controller can be represented by any causal map from $\mathbf{y}_0^k$ to $u(k)$. The sequence of *ON's* and *OFF's* of the erasure channel is modeled as a

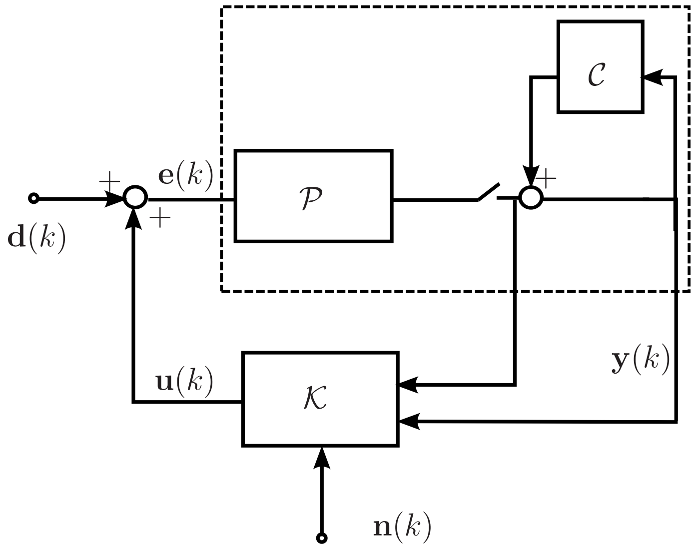

Figure 3.2: A networked control system

two-state Markov chain with transition probability matrix

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}, \quad 0 \le p, q \le 1\,.$$

One can calculate the stationary distribution as $\boldsymbol{\pi} = \left[\frac{q}{p+q}, \frac{p}{p+q}\right]$. Let the state space realization of the plant and the channel compensator be

$$\left[\begin{array}{c|c} A & B \\ \hline C & 0 \end{array}\right] \text{ and } \left[\begin{array}{c|c} A_c & B_c \\ \hline C_c & 0 \end{array}\right]$$

respectively. We can then regard the dashed box in Fig. 3.2 as a generalized "plant" with state matrices

$$\tilde{A}(1) = \begin{bmatrix} A & 0 \\ B_cC & A_c \end{bmatrix}, \quad \tilde{A}(2) = \begin{bmatrix} A & 0 \\ 0 & A_c + B_cC_c \end{bmatrix}$$

for the "ON" and "OFF" of the erasure channel respectively.

To simplify the subsequent analysis, we further assume that the compensator is chosen such that A_c and $A_c + B_cC_c$ are stable. Under these additional conditions and with account of Theorem 3.2.6 we have the corresponding Bode's integral theorem.

Theorem 3.3.1. Consider the NCS in Fig. 3.2, and assume that the signal $\mathbf{u}$ is Gaussian and stationary. The following relation holds for all causal controllers $\mathcal{K}$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(S_{\mathbf{d},\mathbf{e}}(\omega)\right) d\omega \geq \sum_j \Re\left(\log \lambda_j\left(A\right)\right)^+ . \tag{3.10}$$

Proof. The proof is a simple application of Theorem 3.2.12, and is therefore omitted here. □

Remark 3.3.2. This theorem characterizes the control design limitation for NCS with random packet dropout. Given the stable compensator, the right hand side in (3.10) shows that the lower bound of the closed loop performance is determined solely by the degree of instability of A. This observation suggests that, considering the relatively loose definition of stability in (3.2), packet dropout does not make the system "more" unstable. However, the dropout may add up to the performance limitation in other forms, for which a close scrutiny is required. This result agrees with the previous work on data-rate limited control [59].

3.4 Monetary Policy Limits Analysis

We now turn our attention to the field of macroeconomics, where extensive study has been conducted to investigate the use of feedback, in terms of monetary and fiscal policies, to achieve certain objectives, such as financial stability and high employment growth. For example, the celebrated Taylor Rule [60] suggests the short-term nominal interest rate as an appropriate weighted linear combination of deviations of inflation and GDP (Gross Domestic Product) from their *target values*. Following this seminal work, attention has been drawn to the area of optimal feedback policy design and analysis. Recently, Brock [61] proposed a frequency domain approach to asses the intrinsic tradeoffs between various control objectives, such as minimizing inflation, interests rate and economy output volatilities. More specifically, he employed Bode type integral to identify the impact of control rules on the shaping of fluctuations in frequency domain, subject to fundamental limits from the plant. In this section, we extend the method in [61] to the case when the economy is modeled as switching dynamics taking values randomly between regimes corresponding to an ergodic finite state Markov chain [62]. Besides, the information theoretic nature of our framework allows for a convenient incorporation of the information theoretic modeling of *rational inattention* [20] into our limitation analysis. Rational inattention is a well-observed phenomena that individual people have due limited information-processing capability, which is believed to contribute to some important aspects of macroeconomic fluctuations.

3.4.1 Bode's integral for Markov switching AR model

Consider the following typical AR(1) model, usually considered in monetary policy literature [63]:

$$\mathbf{x}(k) = a(\boldsymbol{\sigma}_k)\mathbf{x}(k-1) + \varsigma(\boldsymbol{\sigma}_k)\mathbf{u}(k) + \boldsymbol{\epsilon}(k), \tag{3.11}$$

where $\boldsymbol{\sigma}(k) \in \mathcal{N}$ is an ergodic Markov chain with transition matrix $P = [p_{ij}]\,, i, j \in \mathcal{N}$, $a(n), b(n) \in \mathbb{R}$; $\boldsymbol{\epsilon}(k)$ is a zero-mean Gaussian process; the the state of interest $\mathbf{x}(k)$ is simplified as a policy rule and is chosen as a general nonlinear function $\mathbf{u}(k) = f(\mathbf{x}_0^{k-1})$.

To unveil the role of the sensitivity function in this setup, we suppose that the policymaker wishes to minimize the variance of the state $\mathbf{E}\mathbf{x}^2(k)$ under the chosen control $\mathbf{u}(k)$. Notice that the closed-loop is assumed to be stationary. Then we have

$$\begin{aligned} E\mathbf{x}^2(k) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{\mathbf{x}}(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |\tilde{G}(j\omega)|^2 f_{\mathbf{e}}(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |\tilde{G}(j\omega)|^2 f_{\boldsymbol{\epsilon}}(\omega)|S_{\boldsymbol{\epsilon},\mathbf{e}}(\omega)|d\omega\,, \end{aligned} \tag{3.12}$$

where the first equality follows from the stationary of $\mathbf{x}$, and $\tilde{G}(\cdot)$ is the "transfer function" from $\mathbf{e}$ to $\mathbf{x}$, the detailed derivation of which can be found in [54] and [21]. The relation in the last equation in (3.12) helps in understanding the role of $S_{\boldsymbol{\epsilon},\mathbf{e}}(\omega)$, as it shapes the open loop response of $\boldsymbol{\epsilon}$ (i.e. $|\tilde{G}(j\omega)|^2 f_{\boldsymbol{\epsilon}}(\omega)$) into the controlled one in frequency domain. It is then natural to see that the limitation inherent to the control policy $f(\cdot)$ can be characterized by the spectrum of the sensitivity function. The constraint is now cast into the Bode integral formula, as it will be shown in the next theorem.

Theorem 3.4.1. Consider the model given in (3.11), and suppose that it is mean-square stable. Then the following Bode's integral inequality holds

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\log S_{\boldsymbol{\epsilon},\mathbf{e}}(\omega)d\omega \geq \left(\sum_{n}\pi_n \log|a(n)|\right)^{+}.$$

Proof. The proof is a simple application of Theorem 3.2.12. □

Remark 3.4.2. Comparing with the similar result obtained in [21], we present a lower bound on the performance limitation for a more general control policy rather than linear Taylor rules. Moreover, the calculation is simpler and can be easily extended to the higher order cases (AR(l), $l > 1$) with little modification.

Remark 3.4.3. This theorem provides a general formula only to motivate more theoretical development for various meaningful models as well as empirical validations.

3.4.2 Design limit under rational inattention

We can further exploit the design limitation problem by including *rational inattention*, which is elegantly modeled as channel capacity in Shannon's theory's framework following recent prominent work [20]. In our context, we assume that the policy takes effect after passing through a communication channel with finite capacity, depicted in Fig. 3.3. Here, $\mathbf{e}(k)$ is the actual effect of the feedback policy $\mathbf{u}'$. We now recall the usual definition of the channel capacity $\mathscr{C}$ [4]:

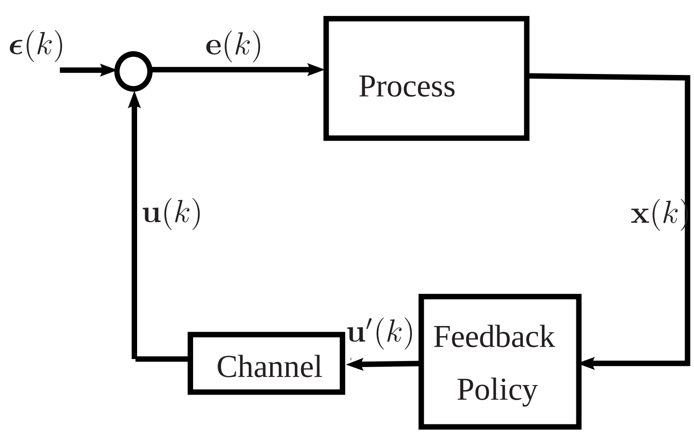

Figure 3.3: Policy design with limited information

$$\mathscr{C} := \sup_{\mathbf{u}'} \bar{I}(\mathbf{u}, \mathbf{u}')$$

The following theorem captures the design limits via an upper bound for the entropy rate $\bar{I}(\boldsymbol{\epsilon}; \mathbf{u})$.

Theorem 3.4.4. Consider the model given in (3.11), and suppose the it is mean-square stable. Then the following inequality holds

$$\bar{I}(\mathbf{u}; \boldsymbol{\epsilon}) \leq \mathscr{C} - \left(\sum_{n \in \mathcal{N}} \pi_n \log |a(n)| \right)^+ . \tag{3.13}$$

Proof. The information conservation law (3.5) and the stability of the closed loop imply that

$$\bar{h}(\boldsymbol{\epsilon}) \leq h(\mathbf{e}) - \left(\sum_{n \in \mathcal{N}} \pi_n \log |a(n)| \right)^+ .$$

The above inequality can be re-written as

$$\begin{aligned} \bar{I}(\boldsymbol{\epsilon}; \mathbf{u}) &= \bar{h}(\boldsymbol{\epsilon}) - \bar{h}(\boldsymbol{\epsilon} | \mathbf{u}) \\ &\leq \bar{h}(\mathbf{e}) - \bar{h}(\mathbf{e} | \mathbf{u}) - \left(\sum_{n \in \mathcal{N}} \pi_n \log |a(n)| \right)^+ \\ &= I(\mathbf{e}; \mathbf{u}) - \left(\sum_{n \in \mathcal{N}} \pi_n \log |a(n)| \right)^+ , \end{aligned} \tag{3.14}$$

where we have used the fact that $\bar{h}(\boldsymbol{\epsilon}|\mathbf{u}) = \bar{h}(\boldsymbol{\epsilon} + \mathbf{u}|\mathbf{u}) = \bar{h}(\mathbf{e}|\mathbf{u})$.

From (P3) and the definition of channel capacity we have

$$\bar{I}(\mathbf{e}; \mathbf{u}) \leq \bar{I}(\mathbf{u}; \mathbf{u}') \leq \mathscr{C} . \tag{3.15}$$

The proof is completed by combining (3.14) and (3.15). □

Remark 3.4.5. The nonnegativity of mutual information rate $\bar{I}(\mathbf{u}; \boldsymbol{\epsilon})$ further implies that one needs $\mathscr{C} \geq \left(\sum_{n \in \mathcal{N}} \pi_n \log |a(n)| \right)^+$ for closed loop stability. The same relation has been developed in various control chapters to provide a sufficient condition for stabilization of a closed loop with limited data-rate. In the language of macroeconomics, this relation can be alternatively interpreted that the level of information processing capability of the agents should be greater than degree of instability of the process, for which the control policy is to be designed.

If we assume that signals in the closed loop Fig. 3.3 are Gaussian (in which case, the channel may need to be a Gaussian additive one), one can also represent (3.16) in a log-integral fashion.

Corollary 3.4.6. Consider the model given in (3.11), and suppose the it is mean-square stable, and $\mathbf{u}, \boldsymbol{\epsilon}$ are Gaussian stationary. Then the following inequality holds

$$\begin{aligned} &-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(1-\frac{|f_{\boldsymbol{\epsilon},\mathbf{u}}(\omega)|^2}{f_{\boldsymbol{\epsilon}}(\omega)f_{\mathbf{u}}(\omega)}\right)d\omega \\ &\qquad \le \mathscr{C}-\left(\sum_{n\in\mathcal{N}}\pi_n\log|a(n)|\right)^+ . \end{aligned} \tag{3.16}$$

Proof. The proof is obviously implied by the fact that

$$\bar{I}(\boldsymbol{\epsilon};\mathbf{u}) = -\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(1-\frac{|f_{\boldsymbol{\epsilon},\mathbf{u}}(\omega)|^2}{f_{\boldsymbol{\epsilon}}(\omega)f_{\mathbf{u}}(\omega)}\right)d\omega$$

for Gaussian stationary processes $\mathbf{u}$ and $\boldsymbol{\epsilon}$ [17]. □

3.5 Conclusions

This chapter has developed a relatively complete Bode's integral formula for stochastic switched closed loops. Information theory has been employed as machinery to obtain a relationship among different system variables, which has in turn resulted in Bode's integral for stationary cases. Various algebraic conditions have been proposed to capture tight performance bounds. Application of this theoretic framework to the field of NCS as well as macroeconomics illustrates the usefulness of this fundamental result.

CHAPTER 4

LQR OVER ADDITIVE GAUSSIAN CHANNEL

4.1 Preliminaries & Problem Formulation

The problem formulation and related definitions are given in this section. We consider the system in Fig. 4.1, with the details of each component given below:

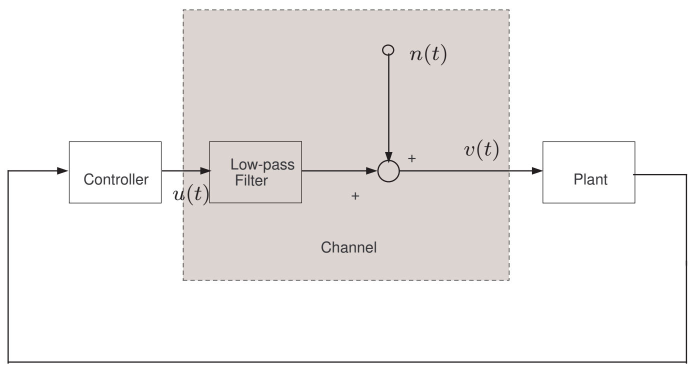

Figure 4.1: Closed-loop system

4.1.1 Plant

Consider the following single input LTI system as the plant to be controlled

$$dx(t) = Ax(t)dt + Bv(t)dt\,, \qquad t \geq 0, \tag{4.1}$$

where $x(t) \in \mathbb{R}^n$ is the state of the plant and $x(0) = x_0$, while $v(t) \in \mathbb{R}$ is the input signal.

4.1.2 Channel

As shown in Fig.4.1, a memoryless AWGN channel is located between the controller and the plant, where $n(t)$ is standard white noise.

The white noise process $n(t)$ could be viewed as a generalized derivative of a standard Brownian motion W_t. The Brownian motion is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_{t\geq 0})$, where Ω is the sample space, $\mathcal{F}$ is the σ algebra, the filtration $\mathcal{F}_t$ is an increasing sub σ-algebra to which W_t is adapted, and $\mathbb{P}$ is the probability measure.

We represent the channel dynamics as

$$\begin{aligned} dx_c(t) &= A_c x_c(t)dt + B_c u(t)dt \\ v(t)dt &= C_c x_c(t)dt + dW_t\,, \quad t \geq 0 \end{aligned} \tag{4.2}$$

where $x_c(t) \in \mathbb{R}^k$ is the state, $x_c(0) = x_{c0}$; $u(t) \in \mathbb{R}$ is the output of the controller and is the channel input and is further assumed to be $\mathcal{F}_t$ adapted; $v(t)$ is the channel output and is fed into the plant, while (A_c, B_c, C_c) is a realization of the low-pass filter, which characterizes the bandwidth of the channel.

The power constraint is an important characterization of an AWGN channel. It takes the form [48]

$$\mathbb{E}(|u(t)|^2) \leq \mathcal{P}\,, \quad \forall\, t \geq 0\,, \tag{4.3}$$

where $\mathbb{E}(\cdot)$ refers to the expectation operator on the aforementioned complete probability space, and $\mathcal{P} > 0$ is a pre-specified upper bound on the average power of the channel input $u(t)$.

Remark 4.1.1. An extra communication channel could also be located in between the output of the plant and the controller. Here, we assume that the connection between sensor(s) and controller is of unlimited communication ability (infinite bandwidth and noiseless).

4.1.3 Augmented System

To treat this communication/control interconnection as a whole, we introduce the following Itô-type linear (SDE)

$$d\xi(t) = \bar{A}\xi(t)dt + B_w dW_t + B_u u(t)dt\,, \qquad t \geq 0 \tag{4.4}$$

where $\xi(t) = [x^\top(t), x_c^\top(t)]^\top$, $\bar{A} = \begin{bmatrix} A & BC_c \\ 0 & A_c \end{bmatrix}$, $B_w = \begin{bmatrix} B \\ 0 \end{bmatrix}$, and $B_u = \begin{bmatrix} 0 \\ B_c \end{bmatrix}$. We make the standard assumption that $(\bar{A}, B_u)$ is controllable.

4.1.4 Control Objective

The class of admissible control signals $u(\cdot)$ will be defined similar to [64]:

$$\mathcal{U} = \Bigg\{ u(\cdot) : u(t) \text{ is } \mathcal{F}_t \text{ adapted;} \\ \limsup_{t\to\infty} \frac{\int_0^t |u(s)|^2 dt}{\int_0^t x(s) ds} < \infty \quad a.s. \Bigg\} \tag{4.5}$$

Consider the following cost-functional:

$$J = \limsup_{T\to\infty} \frac{1}{T} \int_0^T (\xi^\top(t) Q \xi(t) + \rho u^2(t)) dt \,, \tag{4.6}$$

where $Q = Q^\top > 0$, $\rho > 0$.

The control objective is to design an optimal controller such that the cost-functional J is minimized, subject to constraints of the communication channel. Specifically, for the system dynamics in (4.4) we address the following constrained stochastic linear quadratic control problem

$$\inf_{u(\cdot)\in\mathcal{U}} J \,, \tag{4.7}$$

subject to the power constraint in (4.3).

Remark 4.1.2. The linear quadratic regulator (LQR) problem with stochastic disturbance without any communication constraint has been thoroughly investigated in parallel with its deterministic counterpart in [65], [66], etc.

4.2 Controller Design via Linear Matrix Inequalities

In this section, the problem (4.7) is cast into an LMI optimization problem and is solved in the framework of an eigenvalue problem (EVP).

4.2.1 LMI Configuration of Stochastic LQR

The classic state feedback stochastic LQR problem for the system dynamics in (4.4) is defined as

$$\inf_{u(\cdot)=K\xi(\cdot)} J, \tag{4.8}$$

where K is the feedback gain matrix to be determined. It is well known that under the assumption that $(\bar{A}, B_u)$ is controllable, the optimal state feedback control could be expressed as

$$u^*(t) = -\frac{1}{\rho} B_u^\top P \xi(t)\,, \tag{4.9}$$

where $P > 0$ is the solution of the following algebraic Riccati equation

$$\bar{A}^\top P + P\bar{A} - \frac{1}{\rho} P B_u B_u^\top P + Q = 0\,. \tag{4.10}$$

The minimum of the cost-functional J is given by

$$J^*(u^*(\cdot)) = B_w^\top P B_w \quad a.s.. \tag{4.11}$$

The problem (4.8) can be alternatively solved using the following LMI EVP,

$$\min_{R,Y} \gamma \tag{4.12}$$

subject to

$$\begin{bmatrix} \bar{A}R + R\bar{A}^\top + B_u Y + Y^\top B_u^\top & RQ^{1/2} & \sqrt{\rho} Y^\top \\ (RQ^{1/2})^\top & -\mathbb{I} & \mathbf{0} \\ \sqrt{\rho} Y & \mathbf{0} & -1 \end{bmatrix} \le 0\,, \tag{4.13}$$

$$\begin{bmatrix} \gamma & B_w^\top \\ B_w & R \end{bmatrix} \ge 0\,, \tag{4.14}$$

and

$$R > 0\,, \tag{4.15}$$

where the matrices $R \in \mathbb{R}^{(n+k)\times(n+k)}$ and $Y^\top \in \mathbb{R}^{n+k}$ are decision variables, over which γ is optimized, and $\mathbb{I}$ stands for identity matrix of appropriate dimension.

Suppose γ^* is the minimum. Then $J^* = \gamma^*$, and the optimal state-feedback gain $K^* = Y^* R^{*-1}$, where $(K^*, R^*) = \arg\min \gamma^*$. The above EVP is derived following the same procedure for LMI representation of deterministic LQR, for which one can refer to [67] for details.

Remark 4.2.1. In the current scenario, the state used for the control law is $\xi(t)$, which is a stacked vector of $x(t)$ and $x_c(t)$. The availability of the channel state is a conventional assumption when a communication channel with feedback is considered

4.2.2 LMI Representation of Power Constraint

In what follows, LMI conditions for power constraint (4.3) are derived and summarized in the following lemma.

Theorem 4.2.2. Consider the system (4.4) with state feedback control $u(t) = -K\xi(t)$. The power constraint (4.3) is satisfied if for arbitrary $\epsilon > 0$ and $\lambda_{\max} \geq \lambda_{\min} > 0$ there exist $R \in \mathbb{R}^{(n+k)\times(n+k)}$ and $Y \in \mathbb{R}^{(n+k)\times 1}$ solving the following LMIs:

$$\lambda_{\min}\mathbb{I} \leq R \leq \lambda_{\max}\mathbb{I}\,, \tag{4.16}$$

$$\bar{A}R + R^\top \bar{A}^\top - B_u Y - Y^\top B_u^\top \leq -\epsilon\mathbb{I}, \tag{4.17}$$

and

$$\begin{bmatrix} \mathcal{P} & Y \\ Y^\top & \mu^{-1}R \end{bmatrix} \geq 0\,, \quad \mu = \frac{\lambda_{\max} B_w^\top B_w}{\epsilon \lambda_{\min}^3} + \frac{\lambda_{\max}\xi_0^\top \xi_0}{\lambda_{\min}^2}\,. \tag{4.18}$$

The corresponding control gain is obtained as $K = Y R^{-1}$.

Proof. Substituting the state feedback law into (4.4) we have

$$d\xi(t) = (\bar{A} - B_u K)\xi(t)dt + B_w dW_t\,. \tag{4.19}$$

This linear SDE has a unique strong solution [68] (Chapter 5.6)

$$\xi(t) = e^{\tilde{A}t}\xi_0 + \int_0^t e^{\tilde{A}(t-\tau)} B_w dW_t, \;\; t \geq 0\,, \tag{4.20}$$

where $\tilde{A} \triangleq \bar{A} - B_u K$. Then by applying Itô's isometry [68], it is straightforward to show that

$$\begin{aligned}
&\mathbb{E}(u^2(t)) \\
&= \mathbb{E}\left(\xi^\top(t)K^\top K\xi(t)\right) \\
&= \xi_0^\top e^{\tilde{A}^\top t}K^\top K e^{\tilde{A}t}\xi_0 \\
&+ B_w^\top \int_0^t e^{\tilde{A}^\top(t-\tau)}K^\top K e^{\tilde{A}(t-\tau)}B_w d\tau\,,
\end{aligned} \tag{4.21}$$

In order to find an upper bound for $E(u^2(t))$, we first note that (4.16) and (4.17) imply that

$$e^{\tilde{A}^\top t}e^{\tilde{A}t} \leq \frac{\lambda_{\max}}{\lambda_{\min}}e^{-\epsilon\lambda_{\min}t}\mathbb{I}\,, \qquad \forall t \geq 0\,, \tag{4.22}$$

where the inequality is obtained by the standard argument of upper-bounding a quadratic Lyapunov function (it equals $\xi^\top(t)R^{-1}\xi(t)$ in this case). Next, we derive an upper bound on the terms in the right-hand-side of (4.21). The first term could be bounded as

$$\begin{aligned}
&\xi_0^\top e^{\tilde{A}^\top t}K^\top K e^{\tilde{A}t}\xi_0 \\
&\leq \frac{\lambda_{\max}}{\lambda_{\min}}(\xi_0^\top\xi_0)(KK^\top)e^{-\epsilon\lambda_{\min}t} \\
&\leq \frac{\lambda_{\max}\xi_0^\top\xi_0}{\lambda_{\min}}KK^\top \\
&= \frac{\lambda_{\max}\xi_0^\top\xi_0}{\lambda_{\min}^2}K\lambda_{\min}K^\top \\
&\leq \frac{\lambda_{\max}\xi_0^\top\xi_0}{\lambda_{\min}^2}KRK^\top\,,
\end{aligned} \tag{4.23}$$

and repeating the same steps for the second term gives:

$$\begin{aligned}
&\int_0^t B_w^\top e^{\tilde{A}^\top(t-\tau)}K^\top K e^{\tilde{A}(t-\tau)}B_w d\tau \\
&\leq \frac{\lambda_{\max}}{\lambda_{\min}}KK^\top B_w^\top B_w \int_0^t e^{-\epsilon\lambda_{\min}(t-\tau)}d\tau \\
&\leq \frac{\lambda_{\max}B_w^\top B_w}{\epsilon\lambda_{\min}^2}KK^\top(1-e^{-\epsilon\lambda_{\min}t}) \\
&\leq \frac{\lambda_{\max}B_w^\top B_w}{\epsilon\lambda_{\min}^2}KK^\top \\
&\leq \frac{\lambda_{\max}B_w^\top B_w}{\epsilon\lambda_{\min}^3}KRK^\top\,.
\end{aligned} \tag{4.24}$$

Hence, it follows from (4.21) together with the bounds (4.23) and (4.24) that

$$\mathbb{E}\left(\xi^\top(t)K^\top K\xi(t)\right) \leq \mu KRK^\top\,, \tag{4.25}$$

Therefore, if we want the power constraint (4.3) to hold for state feedback $u(t)$, it is sufficient to have

$$\mu KRK^\top = \mu KRR^{-1}RK^\top = \mu YR^{-1}Y^\top \leq \mathcal{P}\,, \tag{4.26}$$

which is equivalent to (4.18) by using the well known Schur's complement. □

4.2.3 Communication Constrained LQR

The problem (4.7) is readily solved if the power-constraint LMI conditions (4.16), (4.17) and (4.18) are imposed on an EVP (4.12), corresponding to a stochastic LQR problem formulation. We have the following theorem.

Theorem 4.2.3. Consider the closed-loop system given by (4.1) and (4.2) and the LMIs given by (4.13), (4.14), (4.16), (4.17) and (4.18). The quadratic performance index J, subject to the dynamics of both the plant and the channel, is minimized by solving the following LMI EVP for R and Y

$$\min_{R,Y} \gamma \tag{4.27}$$

The optimal control gain is obtained as $K^* = Y^* R^{*-1}$.

Proof. The proof is completed by taking into account (4.16), (4.17) and (4.18) as additional LMI constraints for the EVP (4.12), where (4.15) is dropped because R is further bounded by (4.16). □

Remark 4.2.4. Notice that $\lambda_{\min}, \lambda_{\max}$ and ϵ are tuning parameters and can be adjusted to obtain a solution to the LMI EVP. More specifically, $\lambda_{\min}$ and $\lambda_{\max}$ specify the lower and upper bound of the matrix spectrum of R, and these two parameters can be chosen conservative (i.e. small $\lambda_{\min}$ and large $\lambda_{\max}$) as computational capability allows. The parameter ϵ reflects the negativeness of (4.17).

4.3 Numerical Example

In this section we consider a numerical example to evaluate the control design method of Theorem 4.2.3. We choose a 2nd order plant with the following matrices:

$$A = \begin{bmatrix} -0.5 & -1 \\ 0 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The parameters of the channel are chosen as:

$$A_c = -4, \quad B_c = 10, \quad C_c = 1.$$

The initial values are $x_0 = [-1, \ \ 2]^\top$ and $x_{c0} = 0$, which implies $\xi(0) = [-1, \ 2, \ 0]^\top$. The augmented system is written as

$$\begin{aligned} d\xi(t) = &\begin{bmatrix} -0.5 & -1 & 1 \\ 0 & -2 & 1 \\ 0 & 0 & -4 \end{bmatrix} \xi(t)dt + \begin{bmatrix} 0 \\ 0 \\ 10 \end{bmatrix} u(t)dt \\ &+ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} dW_t, \quad \xi_0 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}. \end{aligned} \tag{4.28}$$

We set the power constraint level $\mathcal{P} = 3$. Choose weight matrices

$$Q = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}, \quad \rho = 3,$$

and tuning parameters as $\lambda_{\min} = 0.3, \lambda_{\max} = 50$ and $\epsilon = 0.07$. Using Matlab toolbox YAMLP [69] as the LMI solver, and applying Theorem 4.2.3 we obtain the minimum $\gamma^* = 1.0442$, and corresponding matrices

$$R^* = \begin{bmatrix} 2.0034 & 1.8531 & -5.6960 \\ 1.8531 & 3.4384 & -3.6070 \\ -5.6960 & -3.6070 & 31.9690 \end{bmatrix},$$

$$Y^* = \begin{bmatrix} -0.003001 & -0.00220 & 0.0000313 \end{bmatrix}.$$

The control gain is then computed as

$$K^* = Y^* R^{*-1} = \begin{bmatrix} -0.0042 & 0.0010 & -0.0006 \end{bmatrix}.$$

4.4 Conclusion

In this chapter a new approach has been proposed to address the continuous-time linear quadratic control problem for LTI systems subject to AWGN channel constraints. The main result of the chapter is expressed as LMI EVP, the solution of which results in the optimal state feedback gain, minimizing a quadratic cost-functional. The key idea was to express both the control and the constraint as convex optimization problems. Further research will pursue dynamic feedback, plant uncertainties, and channel uncertainties.

CHAPTER 5

NOISE ATTENUATION OVER ADDITIVE GAUSSIAN CHANNELS

The chapter is organized as follows. In Section 5.1, the problem is formulated, and an LMI solution is provided. In Section 5.2 the method is extended to a MIMO channel, and in Section 5.3 a numerical example is given to illustrate the proposed algorithm. The chapter is concluded in Section 5.4.

Notation:

- The $\mathcal{H}_2$ norm of a transfer function matrix $G(s)$, denoted by $\|G\|_{\mathcal{H}_2}$, is obtained by $\|G\|_{\mathcal{H}_2} = \frac{1}{2\pi}\int_{-\infty}^{\infty}$ trace$(G(j\omega)G^*(j\omega))$, where $(\cdot)^*$ represents the conjugate transpose.

- The $\mathcal{H}_\infty$ norm of a transfer function matrix $G(s)$, denoted by $\|G\|_{\mathcal{H}_\infty}$, is obtained by $\|G\|_{\mathcal{H}_\infty} = \sup_\omega \bar{\sigma}\|G(j\omega)\|$, where $\bar{\sigma}(\cdot)$ gives the maximum singular value.

- The expectation operator is denoted by $\mathbf{E}(\cdot)$.

- The power spectral density (PSD) of a wide-sense stationary signal $x(t),\ t \geq 0$ is denoted by $f_e(\omega)$. If $e(t)$ is an n dimensional vector, then $f_e(\omega)$ is matrix.

5.1 Single Input Single Output Channel

We consider the problem of stabilizing an unstable plant over a noisy communication channel, while keeping a certain performance bound for the channel noise attenuation. We consider the following system:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \quad x(0) = x_0. \\ z(t) &= Cx(t) \end{aligned} \tag{5.1}$$

where state $x(t) \in \mathbb{R}^n$, control input $u(t) \in \mathbb{R}$, performance output $z(t)$. We also assume the system is initialized with a zero-mean Gaussian random variable x_0. Here we assume that (A, B, C) is a minimal realization. The closed loop is shown in Fig. 5.1.

The communication channel is assumed to be an infinite bandwidth AWGN channel as follows

$$u(t) = e(t) + n(t), \quad t \geq 0, \tag{5.2}$$

where $e(t) = -Kx(t)$ is the channel input and $K \in \mathbb{R}^{1\times n}$ is the control gain matrix, $u(t)$ is the channel output, and $n(t)$ is a zero-mean white Gaussian process with PSD σ_n^2. The power of the channel input signal is given by $\mathbf{E}(e^2(t))$, which can be alternatively expressed as

$$Ee^2(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f_e(\omega)d\omega\,.$$

A power constraint is imposed on the input of the AWGN channel as $\mathbf{E}e^2(t) \leq \mathcal{P}$, $\forall t \geq 0$, where $\mathcal{P} > 0$ is a pre-specified value, reflecting the hardware limitations or some other design requirements. We define the following Signal-to-Noise Ratio, or SNR of the channel (5.2) as

$$\mathsf{SNR} \triangleq \frac{\mathcal{P}}{\sigma_n^2}\,.$$

It has been shown that the channel capacity is $\mathsf{SNR}/2$ nat/sec [48].

Three important aspects of the closed-loop system are considered and analyzed in detail.

Closed loop stability

The closed loop system is stabilized by choosing the control gain K from the admissible set $\mathcal{K} \triangleq \{K : A - BK \text{ is Hurwitz }\}$.

Power Constraint

Denote the transfer function from $n(t)$ to the channel input $e(t)$, also known as complementary sensitivity, as $T_{en}(s)$. The following relation holds [29]:

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} f_e(\omega)d\omega = \|T_{en}\|_{\mathcal{H}_2}^2 f_n(\omega) = \|T_{en}\|_{\mathcal{H}_2}^2 \sigma_n^2 \,.$$

Therefore, the power constraint can be equivalently expressed as

$$\|T_{en}\|_{\mathcal{H}_2}^2 \leq \frac{\mathcal{P}}{\sigma_n^2} = \mathsf{SNR}\,.$$

Channel Noise Attenuation

We are also interested in the impact of the channel noise $n(t)$ on the measurement variable $z(t)$. Consider the closed loop depicted in Fig.5.1. We say that the channel noise attenuation is achieved with level $\gamma > 0$, if

$$\|T_{zn}\|_{\mathcal{H}_\infty} \leq \gamma\,,$$

where $T_{zn}(s)$ is the transfer function from $n(t)$ to $z(t)$. Observing the following relation

$$\|T_{zn}\|_{\mathcal{H}_\infty} = \sup_{\omega\in\mathbb{R}} |T_{zn}(j\omega)| = \sup_{\omega\in\mathbb{R}} \sqrt{\frac{f_z(\omega)}{f_n(\omega)}},$$

the quantity $\|T_{nz}\|_{\mathcal{H}_\infty}$ reflects the the maximum magnitude of $f_z(\omega)/\sigma_n^2$ over all frequencies. We note that the $\mathcal{H}_\infty$ norm is not induced by $\mathcal{L}_2$ norms of $z(t)$ and $n(t)$ in the conventional sense [70].

We address the following control problem: find a static state feedback control gain $K \in \mathcal{K}$, such that the required SNR is minimized subject to a desired noise attenuation level $\gamma > 0$.

Remark 5.1.1. State feedback is used for the simplicity of the presentation of the main ideas. More complex cases can be considered in a similar manner.

5.1.1 Tradeoff Between Signal-to-Noise Ratio and Channel Noise Attenuation

First Order Case

Consider the following first order unstable dynamics

$$\begin{aligned} \dot{x}(t) &= ax(t) + u(t)\,, \\ z(t) &= x(t)\,, \end{aligned} \tag{5.3}$$

where $a > 0$ and $u(t) = -kx(t) + n(t)\, k \geq a$. Let the noise attenuation level be $\gamma > 0$. We have the following theorem.

Theorem 5.1.2. The minimal channel SNR for the system (5.3) to be stable and satisfying the noise attenuation level $\|T_{zn}\|_{\mathcal{H}_\infty} \leq \gamma$ is given by

$$\mathsf{SNR} \geq \begin{cases} 2a & \gamma \geq \frac{1}{a}\,, \\ \frac{\gamma}{2}(a + \frac{1}{\gamma})^2 & 0 < \gamma < \frac{1}{a}\,. \end{cases} \tag{5.4}$$

The corresponding control gain is given as

$$k^\star = \begin{cases} 2a & \gamma \geq \frac{1}{a}\,, \\ a + \frac{1}{\gamma} & 0 < \gamma < \frac{1}{a}\,. \end{cases} \tag{5.5}$$

Proof. Calculate the inequality $\|T_{zn}\|_{\mathcal{H}_\infty} \leq \gamma$ as follows

$$\gamma \geq \sup_{\omega \in \mathbb{R}} |T_{zn}(j\omega)| = \sup_{\omega \in \mathbb{R}} \frac{1}{\sqrt{\omega^2 + (a-k)^2}} = \frac{1}{k-a}\,.$$

Then we have $k \geq a + 1/\gamma$ as an additional constraint for the minimization of the power of the channel input signal kx. This optimization problem is formulated and explicitly solved as

$$\begin{aligned} \inf_{k \geq a + 1/\gamma} \|kx\|^2_{\mathcal{H}_2} &= \inf_{k \geq a+1/\gamma} k\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\sigma_n^2}{\omega^2 + (k-a)^2} d\omega \\ &= \inf_{k \geq a+1/\gamma} \sigma_n^2 \frac{k^2}{k-a} \\ &= \begin{cases} 2a\sigma_n^2 & \gamma \geq \frac{1}{a}\,, \\ \frac{\gamma}{2}(a+\frac{1}{\gamma})^2\sigma_n^2 & 0 < \gamma < \frac{1}{a}\,. \end{cases} \end{aligned}$$

Eqs. (5.4) and (5.5) follow straightforwardly. □

Remark 5.1.3. This simple example gives us a chance to understand how much extra SNR (channel capacity) is required to attain a given channel noise attenuation level. As Eqn. (5.3) suggests, an extra amount of SNR is needed if the attenuation level γ is larger than $1/a$. In view of the the fact that the required channel capacity for closed-loop stability is a [29], the quantity $\max\left\{\frac{\gamma}{4}(a+\frac{1}{\gamma})^2-a,0\right\}$ can be regarded as the cost of extra channel capacity to attain the attenuation level γ.

Ch5second Order System: A Case Study

We go one step further and consider the second order system:

$$\dot{x}(t) = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)\,, \tag{5.6}$$

$$z(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t)\,. \tag{5.7}$$

The control gain is given by the matrix $K = \begin{bmatrix} k_1 & k_2 \end{bmatrix} \in \mathbb{R}^2$. The AWGN is given as

$$u(t) = -Kx(t) + n(t)\,.$$

The following proposition gives the explicit expression of $\|T_{zn}\|_{\mathcal{H}_\infty}$ in terms of k_1 and k_2.

Proposition 5.1.4. The $\|T_{zn}\|_{\mathcal{H}_\infty}$ for the closed-loop system composed of (5.6), the controller K and the AWGN is given as

$$\|T_{zn}\|_{\mathcal{H}_\infty} = \begin{cases} \frac{1}{2+k_1-k_2} & (k_1,k_2) \in \mathcal{S} \\ \frac{2}{(k_2-3)\sqrt{4k_1-(k_2-1)^2}} & (k_1,k_2) \in \mathcal{T}\,, \end{cases} \tag{5.8}$$

where

$$\mathcal{S} = \{k_1 = 3, k_2 = 3\} \cup \{(3 < k_1 < 5) \cap (3 < k_2 < k_1)\}$$
$$\bigcup \left\{(k_1 \geq 5) \cap (3 < k_2 < 2 + \sqrt{-1+2k_1})\right\}$$
$$\mathcal{T} = \{k_1 > 5\} \bigcap \left\{2 + \sqrt{-1+2k_1} < k_2 \leq k_1\right\}$$

Proof. We calculate $\|e\|^2_{\mathcal{H}_2}$ by using the following complex contour integral and the residue theorem [71]:

$$\begin{aligned}
&\|e\|^2_{\mathcal{H}_2} \\
&= \frac{\sigma_n^2}{2\pi}\int_{-\infty}^{\infty} T_{en}(j\omega)T_{en}(-j\omega)d\omega \\
&= \frac{\sigma_n^2}{2\pi j}\oint_{\gamma} T_{en}(s)T_{en}(-s)ds \\
&= \sigma_n^2(Res(T_{en}(s)T_{en}(-s);p_1) + Res(T_{en}(s)T_{en}(-s);p_2)) \\
&= \sigma_n^2\frac{k_2^3-(k_1+3)k_2^2+2k_1k_2-k_1^2}{2(-3+k_2)(-2-k_1+k_2)}\,,
\end{aligned}$$

where γ represents the contour $[-j\omega R, j\omega R] \cup \{R\exp(j\theta) : -\pi/2 < \theta < \pi/2\}$ with large enough radius $R > 0$, and $Res(\cdot; p_i)$ denotes the residue evaluated at the poles $p_i, i = 1, 2$. During the course of calculation we have used the fact that

$$T_{en}(s) = K(s\mathbb{I} - A + BK)^{-1}B = \frac{k_2 s + k_1 - k_2}{s^2 + (k_2-3)s + k_1 - k_2 + 2}\,.$$

The conclusion is therefore reached by noticing that SNR $= \|e\|^2_{\mathcal{H}_2}/\sigma_n^2$. □

Now we proceed to calculate the power of channel input e as summarized by the following proposition.

Proposition 5.1.5. The power of the channel input, in terms of k_1 and k_2 can be written as

$$\int_{-\infty}^{\infty} f_e(\omega)d\omega = \frac{k_2^3-(k_1+3)k_2^2+2k_1k_2-k_1^2}{2(-3+k_2)(-2-k_1+k_2)}\,. \tag{5.9}$$

Proof. First note that

$$T_{zn}(s) = \frac{1}{s^2 + (k_2-3)s + k_1 - k_2 + 2}\,,$$

and by using Routh's criterion, the set of stabilizing control gains is obtained as

$$\mathcal{K} = \{[k_1 k_2] : \{k_1 \geq k_2\} \cap \{k_2 \geq 3\}\}\,. \tag{5.10}$$

The rest of the proof follows the procedure of solving the optimization problem

$$\sup_{k_1 \text{ and } k_2 \text{ satisfy } (5.10)} \sqrt{T_{zn}(j\omega)_{zn}(-j\omega)}\,.$$

The machinery used for this problem is reduced to calculus, and is dropped therefore.

Unlike the first order case, the increased degree of complexity in the second order case makes it very difficult to get an explicit solution for the problem, even though we have obtained the expressions of the corresponding $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms in (5.9) and (5.8) respectively. As an alternative, we illustrate the SNR (Channel capacity)/ performance tradeoff graphically in the following plots.

Fig. 5.2 shows the required SNR for the given control gain that satisfies the conditions given in (5.10). As we can see, without an additional constraint for noise attenuation, the minimal SNR takes the value $(k_1^\star = 6, k_2^\star = 6)$.

In Fig. 5.3, the effect of the enforced noise attenuation on the solution set of K is shown. The size of the feasibility set of K decreases along with γ, which is consistent with (5.5) for the the first order case.

5.1.2 Controller Design via Linear Matrix Inequality

In this section, we use LMI technique to solve the problem for the general case.

To start with, we introduce the following theorem for SNR minimization.

Lemma 5.1.6. Consider the closed loop shown in Fig. 5.1. The optimization problem

$$\inf_{K \in \mathcal{K}} \mathsf{SNR}\,,$$

is equivalent to the following LMI minimization problem

$$\begin{aligned} &\min_{X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}, \rho} \rho \\ &\text{subject to} \qquad X_{\mathsf{SNR}} > 0\,, \Phi(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}, \rho) \geq 0 \text{ and} \\ &\qquad\qquad\quad \Psi(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}) \leq 0\,, \end{aligned} \tag{5.11}$$

where $\rho \in \mathbb{R}$, $X_{\mathsf{SNR}} = X_{\mathsf{SNR}}^\top \in \mathbb{R}^{n\times n}$, $Y_{\mathsf{SNR}} \in \mathbb{R}^{1\times n}$,

$$\Phi(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}, \rho) \triangleq \begin{bmatrix} \rho & Y_{\mathsf{SNR}} \\ Y_{\mathsf{SNR}}^\top & X_{\mathsf{SNR}} \end{bmatrix},$$

and

$$\Psi(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}) \triangleq \begin{bmatrix} \begin{pmatrix} AX_{\mathsf{SNR}} + X_{\mathsf{SNR}}A^\top \\ -BY_{\mathsf{SNR}} - Y_{\mathsf{SNR}}^\top B^\top \end{pmatrix} & X_{\mathsf{SNR}}B \\ B^\top X_{\mathsf{SNR}} & -1 \end{bmatrix}.$$

The optimal control gain is obtained as

$$K^\star_{\mathsf{SNR}} = Y^\star_{\mathsf{SNR}}(X^\star_{\mathsf{SNR}})^{-1}\,,$$

where $X^\star_{\mathsf{SNR}}$ and $Y^\star_{\mathsf{SNR}}$ are the optimal solutions to the problem (5.11).

Proof. The proof is based on the classic LMI solution to the state feedback $\mathcal{H}_2$ optimization synthesis problem for the following auxiliary deterministic closed loop, composed of the plant $G(s)$ and the controller $K(s)$:

$$G(s) = \left[\begin{array}{c|c} A & B \\ \hline 0 & 1 \\ \mathbb{I} & 0 \end{array}\right]\,, \quad K(s) = K \in \mathbb{R}^{1\times n}\,,$$

with the objective function given by $\|K(s\mathbb{I} - A + BK)^{-1}B\|^2_{\mathcal{H}_2}$. The proof is completed by using the standard procedure given in [67]. $\square$

Similarly, the noise attenuation can also be cast into LMI conditions, given by the following lemma.

Lemma 5.1.7. Consider the closed loop in Fig. 5.1. The CNA level is less than γ, if and only if we can find matrices $0 < X_n = X_n^\top \in \mathbb{R}^{n\times n}$, $Y_n \in \mathbb{R}^{1\times n}$ that satisfy the following LMI feasibility condition

$$\Theta_\gamma(Y_n, X_n) \triangleq \begin{bmatrix} \begin{pmatrix} AX_n + X_nA^\top \\ -BY_n - Y_n^\top B^\top \end{pmatrix} & B & X_nC^\top \\ B^\top & -1 & 0 \\ CX_n & 0 & -\gamma^2 \end{bmatrix} \le 0\,. \tag{5.12}$$

The resulting control gain is obtained as

$$K_n = Y_nX_n^{-1}\,.$$

Proof. Consider the following auxiliary system G and the controller K respectively

$$G_n(s) = \left[\begin{array}{c|cc} A & B & B \\ \hline C & 0 & 0 \\ \mathbb{I} & 0 & 0 \end{array}\right]\,, \quad K(s) = K_n \in \mathbb{R}^{1\times n}\,.$$

The transfer function from the disturbance to the performance measurement is calculated as

$$S(s) = C(s\mathbb{I} - A + BK_n)^{-1}B\,,$$

which is identical to $T_{zn}(s)$. Subsequently we can use standard LMI arguments to obtain the feasibility sets of X_n and Y_n that satisfy $\|T_{zn}(s)\|_{\mathcal{H}_\infty} \leq \gamma$. The readers can refer to [67] for details. □

It is easy to see that the *problem* is equivalent to minimizing ρ over all matrices $X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}, X_n, Y_n, \rho$ that satisfy (5.11) and (5.12). While the optimization problems in (5.11) and (5.12) are convex themselves, the joint one is not convex. Therefore we enforce the condition

$$X = X_{\mathsf{SNR}} = X_n \text{ and } Y = Y_{\mathsf{SNR}} = Y_n$$

to obtain the convexity, admittedly with some degree of conservatism. Indeed, the same treatment is widely used in mixed $\mathcal{H}_2/\mathcal{H}_\infty$ problems, such as [72]. The above argument proves the following main theorem.

Theorem 5.1.8. Given a desired channel noise attenuation level γ, a lower bound for the required channel SNR of the closed loop system is obtained via the following LMI optimization problem:

$$\begin{array}{ll} \min\limits_{X,Y} & \rho \\ \text{Subject to} & X > 0\,, \Phi(X,Y,\rho) > 0 \\ & \Psi(X,Y) \leq 0 \text{ and } \Theta_\gamma(X,Y) \leq 0\,. \end{array} \tag{5.13}$$

The lower bound of the SNR is given as $\rho^\star$, which is the optimal value obtained in (5.13). The corresponding controller is given as

$$K^\star = Y^\star(X^\star)^{-1}\,,$$

where $X^\star$, $Y^\star$ are the resulting values of the decision variables X and Y respectively.

Remark 5.1.9. In this section, only the full state feedback is considered. However, the same approach can be easily extended to output feedback case.

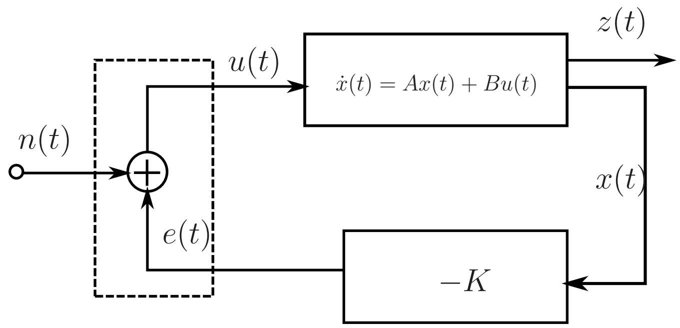

Figure 5.1: Closed-loop system

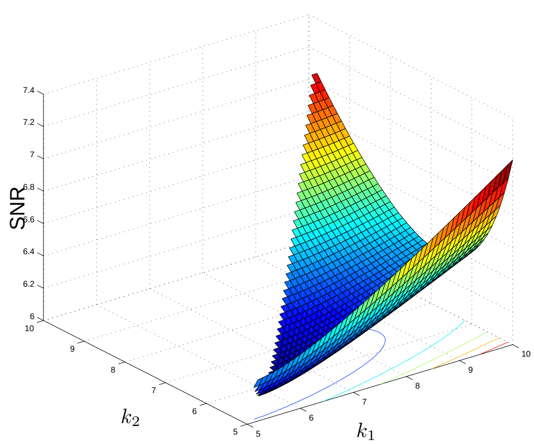

Figure 5.2: SNR v.s. Control gain

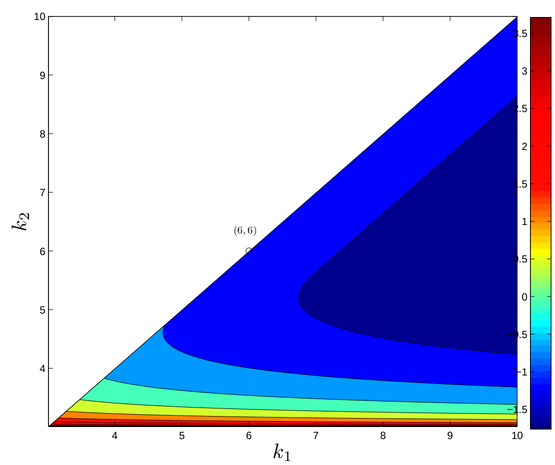

Figure 5.3: Feasibility sets for different $\log \gamma$.

5.2 Vector Gaussian Channel

Here we consider the case where the control signal is a vector, and it is transmitted through a vector Gaussian channel, which is also a simple case of a MIMO channel. In applications, this scenario represents the case, where actuators and controllers are geographically distributed and multiple transmitters and receivers are therefore employed to conduct the communication task, as shown in Fig. 5.4. From the perspective of wireless communication, a multiple access system with multiple antennas at the base-station allows several users, who are spatially separated, to communicate simultaneously. Moreover, the channel fading in the point-to-point communication can be overcome or even utilized by MIMO communication schemes [73].

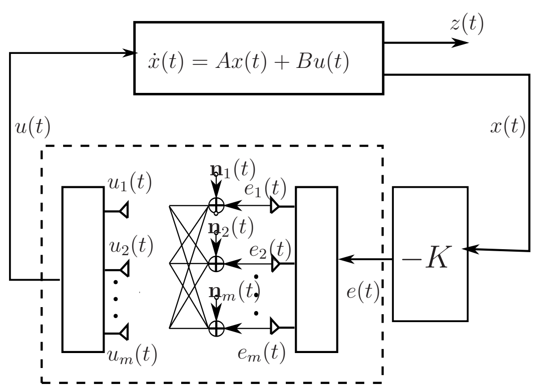

Figure 5.4: MIMO Channel

The channel is modeled as follows.

$$u(t) = He(t) + n(t) \quad t \geq 0, \tag{5.14}$$

where $e(t) \in \mathbb{R}^m$ is the channel input, and $u(t) \in \mathbb{R}^m$ is the channel output, $n(t)$ is a m dimensional Gaussian white noise process with $En(t) = 0$ and $En(t)n^\top = \sigma_n^2\mathbb{I}$, and $H \in \mathbb{R}^{m\times m}$ is a channel gain matrix, which is assumed to be deterministic here. The channel input is required to satisfy the following power constraint as

$$\mathbf{E}e^\top(t)e(t) = \text{trace}(\mathbf{E}(e(t)e^\top(t))) \leq \mathcal{P} \quad \forall t \geq 0$$

for some pre-specified input power level $\mathcal{P} > 0$. Similar to the scalar case, the power of the channel input $\mathbf{E}e^\top(t)e$ can be also represented as

$$\text{trace}(\mathbf{E}(e(t)e^\top(t))) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \text{trace} f_e(j\omega)d\omega\,.$$

Here the SNR is similarly defined as

$$\mathsf{SNR} := \frac{\mathcal{P}}{\sigma_n^2}\,.$$

5.2.1 State Feedback Stabilization

In this section, we design a controller/transmitter K, such that the closed loop system satisfies the power constraint $\mathcal{P}$. We can then formulate the following theorem for the solution of SNR constrained state feedback stabilization.

Lemma 5.2.1. Consider the feedback configuration in Fig. 5.4, where we have

$$\begin{aligned} \min \quad & \text{trace}(\Omega) \\ \text{subject to} \quad & X_{\mathsf{SNR}} > 0\,, \tilde{\Phi}(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}, \Omega) > 0 \text{ and} \\ & \tilde{\Psi}(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}) \le 0\,, \end{aligned} \tag{5.15}$$

in which $\Omega \in \mathbb{R}^{m\times m}$, $X_{\mathsf{SNR}} = X_{\mathsf{SNR}}^\top \in \mathbb{R}^{n\times n}$, $Y_{\mathsf{SNR}} \in \mathbb{R}^{m\times n}$,

$$\tilde{\Phi}(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}, \Omega) \triangleq \begin{bmatrix} \Omega & Y_{\mathsf{SNR}} \\ Y_{\mathsf{SNR}}^\top & X_{\mathsf{SNR}} \end{bmatrix},$$

and

$$\tilde{\Psi}(X_{\mathsf{SNR}}, Y_{\mathsf{SNR}}) \triangleq \begin{bmatrix} \begin{pmatrix} AX_{\mathsf{SNR}} + X_{\mathsf{SNR}}A^\top \\ -BY_{\mathsf{SNR}} - Y_{\mathsf{SNR}}^\top B^\top \end{pmatrix} & X_{\mathsf{SNR}}B \\ B^\top X_{\mathsf{SNR}} & -\mathbb{I} \end{bmatrix} \le 0.$$

The optimal control gain is obtained as

$$K_{\mathsf{SNR}}^\star = H^{-1}Y_{\mathsf{SNR}}^\star (X_{\mathsf{SNR}}^\star)^{-1}\,,$$

where $Y_{\mathsf{SNR}}^\star$ and $X_{\mathsf{SNR}}^\star$ are the solutions of the optimization problem.

Proof. Note that the power of the channel input can be represented as

$$\mathbf{E}e(t)e^\top(t) \quad = \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} \text{trace} f_e(j\omega)d\omega$$

$$
\begin{aligned}
&= \frac{\sigma^2}{2\pi}\int_{-\infty}^{\infty}\text{trace}(T_{ne}(j\omega)T_{ne}^*(j\omega)d\omega \\
&= \|T_{ne}\|_{\mathcal{H}_2}\sigma^2\,, \qquad (5.16)
\end{aligned}
$$

where the transfer function $T_{ne}(s)$ is written as

$$T_{ne}(s) = HK(s\mathbb{I} - A + BHK)^{-1}B\,. \qquad (5.17)$$

Therefore the problem is reduced to the following $\mathcal{H}_2$ optimization problem:

$$\inf_K \|T_{ne}\|_{\mathcal{H}_2}\,.$$

To minimize the channel SNR, we consider the following auxiliary deterministic closed loop, composed of the plant $G(s)$ and the controller $K(s)$:

$$G(s) = \left[\begin{array}{c|c} A & B \\ \hline 0 & 1 \\ \mathbb{I} & 0 \end{array}\right], \quad K(s) = HK, K \in \mathbb{R}^{m\times n}\,,$$

for which we minimize the $\mathcal{H}_2$ norm of $T_{ne}(s)$. The solution can be obtained by solving the standard $\mathcal{H}_2$ optimal control problem via LMIs [67]. □

Similar to the scalar case, the noise attenuation is also presented via relevant LMI conditions in the following lemma.

Lemma 5.2.2. Consider the closed loop in Fig. 5.4. The noise attenuation level is less than γ, if and only if we can find matrices $0 < X_n = X_n^\top \in \mathbb{R}^{n\times n}$, $Y_n \in \mathbb{R}^{m\times n}$ that satisfy the following LMI feasibility condition

$$
\begin{aligned}
&\tilde{\Theta}_\gamma(Y_n, X_n) \triangleq \\
&\begin{bmatrix} \begin{pmatrix} AX_n + X_nA^\top \\ -BY_n - Y_n^\top B^\top \end{pmatrix} & B & X_nC^\top \\ B^\top & -\mathbb{I} & 0 \\ CX_n & 0 & -\gamma^2\mathbb{I} \end{bmatrix} \le 0\,. \qquad (5.18)
\end{aligned}
$$

The resulting control gain is obtained as

$$K_n = H^{-1}Y_nX_n^{-1}\,.$$

Then we have the following theorem.

Theorem 5.2.3. Given the channel noise attenuation level γ, a lower bound for the required channel ρ of the closed loop system is obtained via the solution of the following LMI EVP problem:

$$\begin{array}{ll} \min\limits_{X,Y} & \text{trace}(\Omega) \\ \text{Subject to} & X > 0\,, \tilde{\Phi}(X,Y,\Omega) > 0 \\ & \tilde{\Psi}(X,Y) \leq 0 \;\text{ and } \tilde{\Theta}_\gamma(X,Y) \leq 0\,. \end{array} \tag{5.19}$$

The (sub)optimal value of SNR is given as trace$(\Omega^\star)$, and the corresponding controller is given as

$$K^\star = Y^\star (X^\star)^{-1}\,,$$

where $X^\star$, $Y^\star$ are the optimal values of the decision matrices X and Y respectively.

5.3 Numerical Example

In this section we will give a simple example to illustrate the proposed algorithm. We consider the following state space realization of a 3rd order LTI system:

$$A = \begin{bmatrix} 4 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0.5 \\ 0 & 0 \\ 1 & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}.$$

The vector Gaussian channel has two inputs and two outputs, where the Gaussian noise vector $n(t) \in \mathbb{R}^2$ and $\mathbb{E}n(t) = [0,0]^\top, \mathbb{E}n^\top(t)n(t) = \mathbb{I}$, and the channel matrix are given as

$$H = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}.$$

We first calculate the control gains(K) and the minimal channel input power ($\|T_{en}(j\omega)\|^2_{\mathcal{H}_2}$, for different values of γ. The result is summarized in the following table.

γ	$\min \|T_{en}\|^2_{\mathcal{H}_2}$	K
0.1	22.5095	$\begin{bmatrix} -2.9905 & 58.9790 & 19.0237 \\ 16.5871 & -8.4777 & -3.6414 \end{bmatrix}$
0.5	16.7534	$\begin{bmatrix} -2.5710 & 21.5600 & 9.6628 \\ 16.4738 & -0.2014 & -1.5698 \end{bmatrix}$
1	16.2293	$\begin{bmatrix} -2.4630 & 17.8123 & 8.7206 \\ 16.4472 & 0.7493 & -1.3309 \end{bmatrix}$
10	16.0064	$\begin{bmatrix} -2.3994 & 15.9983 & 8.2658 \\ 16.4279 & 1.2303 & -1.2101 \end{bmatrix}$

Upon obtaining the control gains for different γs we can compare the corresponding PSDs of the observation signal z, which are depicted in Fig. 5.5. As we

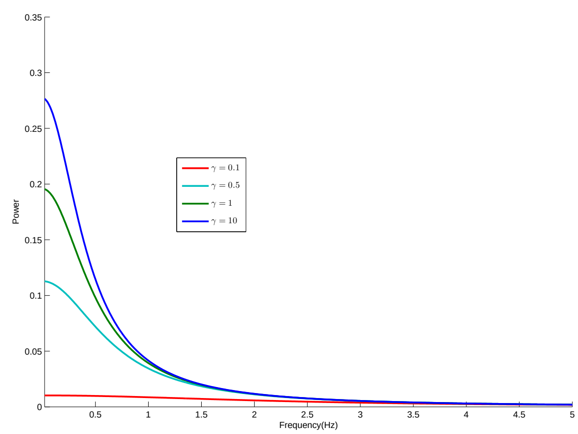

Figure 5.5: Power Spectral Density of z for different noise attenuation levels

can see from Fig. 5.5, setting γ lower implies that the impact of the channel noise on the observation signal z is smaller.

Fig. 5.6 shows the relation between the minimal SNR and the noise attenuation level γ.

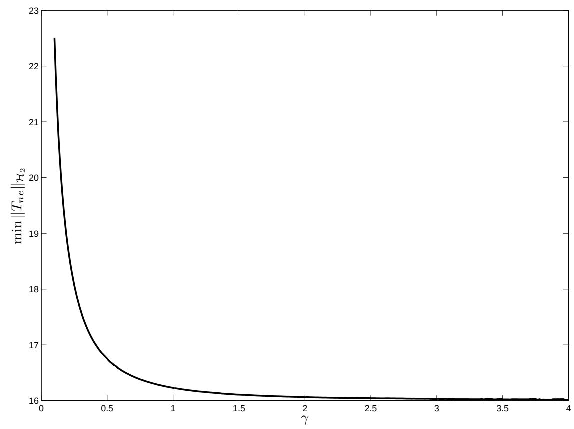

Figure 5.6: Minimal SNR (power of z) versus γ

5.4 Conclusion

In this chapter, we have considered the channel noise attenuation problem for feedback control over both scalar and vector Gaussian channels. An effective LMI approach is proposed and verified. Future development includes uncertain systems and output feedback.

CHAPTER 6

OPTIMAL ESTIMATION OVER GAUSSIAN CHANNELS WITH NOISELESS FEEDBACK

The chapter is organized as follows. In Section 6.1, we introduce the models for both the channel and the plant, and the design problem statement. Section 6.2 discusses a scalar version of the problem, which leads to the development of the solution in Section 6.3. A numerical example is analyzed in Section 6.4. We conclude the chapter with different problems for future research directions in Section 6.5.

6.1 Problem Formulation

In this section we state the problem formulation. The scheme is depicted in Fig. 6.1 where the transmitter has the access to the time-history of the channel output via a noiseless feedback.

- The plant of interest is given by the following n dimensional linear SDE

  $$\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t)\,, \mathbf{x}(0) = \mathbf{x}_0\,. \tag{6.1}$$

  where $A \in \mathbb{R}^{n\times n}$. To ensure the solution $\mathbf{x}(t)$ of (6.1) is Gaussian, the initial value $\mathbf{x}_0$ is also assumed to be Gaussian. Also, $\mathbf{E}\mathbf{x}_0\mathbf{x}_0^\top$ is not singular.

- The communication part of the closed loop is modeled as an additive white Gaussian channel

  $$d\mathbf{v}(t) = \mathbf{z}(t)dt + \sigma d\mathbf{W}(t)\,, \tag{6.2}$$

  where $\mathbf{z}(t)$ is the channel input generated by the signal $\mathbf{x}_0^t$, $\mathbf{W}(t)$ is a standard Wiener process and $\mathbf{v}(t)$ is the channel output. An average power constraint is imposed on the channel input:

  $$\limsup_{T\to\infty} \frac{1}{T}\int_0^T \mathbf{E}\mathbf{z}^2(t)dt \leq \mathscr{P}\,,$$

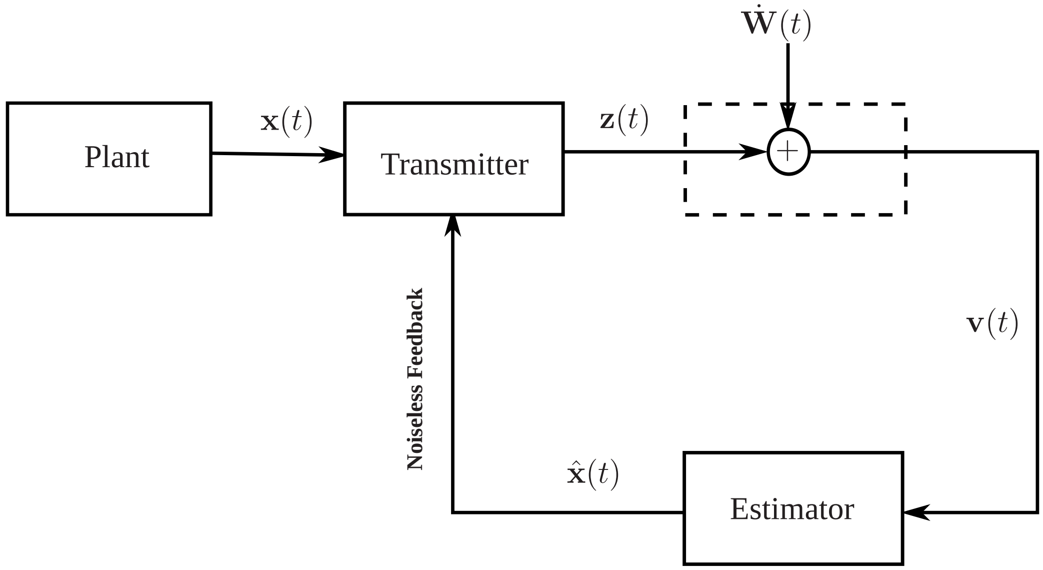

Figure 6.1: State Estimation via Noiseless Feedback

for some $\mathscr{P} > 0$. Slightly different from most of the communication theory literature, the power constraint here is defined over an infinite time horizon to get aligned with some notions in control theory such as asymptotic stability. We also define the noise-to-signal ratio of the channel as

$$\mathsf{SNR} \triangleq \frac{\mathscr{P}}{\sigma^2}\,.$$

It is well-known that the channel capacity is $\mathcal{C} = \mathsf{SNR}/2$ [48].

- The transmitter (encoder) is a causal map defined as $\mathbf{z}(t) \triangleq f(t, \mathbf{x}_0, \mathbf{v}_0^t)$. The receiver(decoder)/estimator is also a causal map $\hat{\mathbf{x}}(t) \triangleq g(t, \mathbf{v}_0^t)$, where $\hat{\mathbf{x}}(t)$ is the estimation of the state $\mathbf{x}(t)$. The error signal is defined as $\tilde{\mathbf{x}}(t) \triangleq \mathbf{x}(t) - \hat{\mathbf{x}}(t)$.

- As a standard assumption, all the random variables (processes) in this system are defined in a filtered probability space $(\Omega, \mathscr{F}, \mathscr{F}_t, \mathbb{P})$.

Definition 6.1.1. The unique solution $X(t)$ of a stochastic differential equation is said to be mean-square exponentially stable with convergence rate $\nu < 0$ if

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbf{E}\|X(t)\|^2 \leq \nu$$

The objective of joint estimation/communication design is to identify *a transmitter and receiver/estimator combination such that the error dynamics with state* $\tilde{\mathbf{x}}(t)$ *is mean-square exponentially stable with minimal decaying rate.*

6.2 Estimation, Communication and Control over Gaussian Channel: A Scalar case study

In this section we review a scalar estimation problem with communication constraint, which was originated by [48] and [36]. Some modifications and innovative observations are made to shed a light on the main result to be presented in the next section.

6.2.1 Transmitting a Gaussian Random Variable

We consider the simplest case, where an analog scalar Gaussian variable $\mathbf{e}$ is to be transmitted through a continuous-time AWNG channel. We further assume that the transmitter (encoder) takes the following affine structure for easy computation and Guassianity of f, given by

$$f(t, \mathbf{e}, \mathbf{v}_0^t) \triangleq \phi(t, \mathbf{v}_0^t) + \psi(t, \mathbf{v}_0^t)\mathbf{e}\,. \tag{6.3}$$

For this given structure of information transmission scheme, the minimal mean-square error for each time instance t is achieved by choosing the estimation $\hat{\mathbf{e}}(t) = \mathbf{E}[\mathbf{e}|\mathbf{v}_0^t]$, which is not readily calculable in general case. So one needs to show a way to construct the corresponding receiver/estimator, which yields $\hat{\mathbf{e}}(t)$. Upon that, constrained by the channel input power level $\mathscr{P}$, parameter optimization for f and g needs to be conducted to reach minimal mean square error. In other words, the problem of optimal estimation is solved in two steps:

1. For the given transmitter (6.3), obtain the estimation scheme g with output $\hat{\mathbf{e}}(t)$;
2. Solve the optimization problem $\min_{g,f} \mathbf{E}(\tilde{\mathbf{e}}^2(t))$ subject to power constraint $\mathscr{P}$.

The first step is straightforwardly obtained by the conditional Kalman-Bucy filter.

Lemma 6.2.1. Consider the linear transmission strategy in (6.3). Then

$$\begin{aligned} d\hat{\mathbf{e}}(t) &= \frac{1}{\sigma^2} P(t)\psi(t,\mathbf{v}_0^t)[d\mathbf{v}_t - \phi(t,\mathbf{v}_0^t)dt - \psi(t,\mathbf{v}_0^t)\hat{\mathbf{e}}(t)dt] \\ \frac{dP(t)}{dt} &= -\frac{1}{\sigma^2} P^2(t)\psi^2(t,\mathbf{v}_0^t)\,, \end{aligned} \tag{6.4}$$

where $P(t) \triangleq \mathbf{E}[(\tilde{\mathbf{e}}(t))^2|\mathbf{v}_0^t]$, $P(0) = \mathbf{E}(\tilde{m}(0))^2$ and $\hat{\mathbf{e}}(0) = \mathbf{E}\mathbf{e}$.

Proof. The proof is just an application of Kalman-Bucy filter for the dynamic system with $\mathbf{e}(t)$ as the system state and $\mathbf{v}(t)$ as the noise corrupted observation.

$$\begin{aligned} d\mathbf{e}(t) &= 0 \\ d\mathbf{v}(t) &= [\phi(t,\mathbf{v}_0^t) + \psi(t,\mathbf{v}_0^t)\mathbf{e}]dt + \sigma d\mathbf{W}(t)\,. \end{aligned}$$

□

The second step is solved by the following lemma.

Lemma 6.2.2. Within the class of linear transmission strategies, which satisfy the condition of (6.2.6) and the power constraint, optimal transmission strategy ϕ^* and ψ^* are given by

$$\phi^*(t,\mathbf{v}_0^t) = -\sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp\left(\frac{\mathsf{SNR}}{2}t\right)\hat{\mathbf{e}}(t)$$

$$\psi^*(t,\mathbf{v}_0^t) = \sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp\left(\frac{\mathsf{SNR}}{2}t\right).$$

The optimal mean square error for this strategy is

$$\mathbf{E}\tilde{\mathbf{e}}^2(t) = P(0)\exp\left(-\mathsf{SNR}t\right)$$

The proof of the lemma can be found in [36].

Remark 6.2.3. Not surprisingly, this feedback coding strategy design can be regarded as feedback stabilization problem, where the state to be stabilized, in the mean-square sense, is defined as $\tilde{\mathbf{e}}(t)$. The stabilization problems can be solved conveniently by using Lyapunov's indirect method. More specifically, one can employ the Lyapunov argument developed in stochastic setting by choosing the candidate Lyapunov function as $V(\tilde{\mathbf{e}}(t)) = \frac{1}{2}\tilde{\mathbf{e}}^2(t)$, and ensure its negative derivative by designing proper transmission schemes. The details of this approach are not discussed here.

Remark 6.2.4. It is also shown in [36] that the solution $\phi^*(t, \mathbf{v}_0^t) + \psi^*(t, \mathbf{v}_0^t)\mathbf{e}$ is optimal among nonlinear functionals of $\mathbf{e}$ (i.e. $f(t, \mathbf{e}, \mathbf{v}_0^t)$).

Remark 6.2.5. This feedback communication scheme can be regarded as an continuous-time extension of the S-K method.

6.2.2 Transmission of a signal

Next we go one step further by replacing the constant source $\mathbf{e}$ by a dynamic one $\mathbf{x}(t)$, evolving according to the linear scalar differential equation with parameter $\lambda \in \mathbb{R}$ and a Gaussian initial value $\mathbf{x}_0$

$$\frac{d\mathbf{x}(t)}{dt} = \lambda \mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0. \tag{6.5}$$

Following the same idea in (6.4), we can consider the Kalman-Bucy filter for the dynamics

$$\begin{aligned} d\mathbf{x}(t) &= \lambda \mathbf{x}(t) dt, \\ d\mathbf{v}(t) &= [\phi(t, \mathbf{v}_0^t) + \psi(t, \mathbf{v}_0^t)\mathbf{x}(t)]dt + \sigma d\mathbf{W}(t). \end{aligned}$$

Next, we proceed with the two-step strategy. The following lemma provides a structure of decoder/estimator, which yields the optimal estimation $\hat{\mathbf{x}}(t) = \mathbf{E}[\mathbf{x}(t)|\mathbf{v}_0^t]$.

Lemma 6.2.6. Consider the linear transmission strategy in (6.3) (where $\mathbf{e}$ is replaced by $\mathbf{x}$) and the source (6.5). Then the optimal estimation of $\mathbf{x}(t)$ is given as

$$\begin{aligned} d\hat{\mathbf{x}}(t) &= \lambda \hat{\mathbf{x}}(t) \\ &\quad + \frac{1}{\sigma^2} P(t)\psi(t, \mathbf{v}_0^t)[d\mathbf{v}_t - \phi(t, \mathbf{v}_0^t)dt - \psi(t, \mathbf{v}_0^t)\hat{\mathbf{x}}(t)dt] \\ \frac{dP(t)}{dt} &= 2\lambda P(t) - \frac{1}{\sigma^2} P^2(t)\psi^2(t, \mathbf{v}_0^t)\,, \end{aligned} \tag{6.6}$$

where $P(t) \triangleq \mathbf{E}[\tilde{\mathbf{x}}^2|\mathbf{v}_0^t]$, $P(0) = \mathbf{E}\mathbf{x}_0^2$ and $\hat{\mathbf{x}}(0) = \mathbf{E}\mathbf{x}_0$.

Next we proceed to the step two. Towards this end, the differential equation with equality of $P(t)$ in (6.6) is rewritten as

$$\dot{P}(t) = \left(\lambda - \frac{1}{\sigma^2} P(t)\psi^2(t, \mathbf{v}_0^t)\right) P(t)\,,$$

and solved by

$$P(t) = P(0) \exp\left(\int_0^t \left(2\lambda - \frac{1}{\sigma^2} P(\tau)\psi^2(\tau, \mathbf{v}_0^\tau)\right) d\tau\right)\,.$$

Taking the expectation and using Jensen's inequality, we have

$$E\tilde{\mathbf{x}}^2(t) = P(0) \exp\left(\int_0^t \left(2\lambda - \frac{1}{\sigma^2} \mathbf{E} P(\tau)\psi^2(\tau, \mathbf{v}_0^\tau)\right) d\tau\right)\,,$$

where Fubini's theorem is also used to interchange integration and expectation. The Lyapunov exponent can be calculated as

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T} \log \mathbf{E} P(T) \\
&\geq 2\lambda - \frac{1}{\sigma^2} \liminf_{T\to\infty} \frac{1}{T} \int_0^t \mathbf{E} P(t)\psi^2(t, \mathbf{v}_0^t, t) dt \\
&\geq 2\lambda - \frac{1}{\sigma^2} \limsup_{T\to\infty} \frac{1}{T} \int_0^t \mathbf{E} P(t)\psi^2(t, \mathbf{v}_0^t, t) dt\,.
\end{aligned} \tag{6.7}$$

It is clear that the minimization of $P(t)$ is reduced to the choice of ψ that minimizes $\frac{1}{\sigma^2} \limsup_{T\to\infty} \frac{1}{T} \int_0^t \mathbf{E} P(t)\psi^2(t, \mathbf{v}_0^t, t)dt$. Towards this end, we have

$$\begin{aligned}
\mathscr{P} &\geq \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{E}[\phi(t, \mathbf{v}_0^t) + \psi(t, \mathbf{v}_0^t)\mathbf{x}(t)]^2 \\
&= \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{E}[\phi(t, \mathbf{v}_0^t) + \psi(t, \mathbf{v}_0^t)\hat{\mathbf{x}}(t)]^2 \\
&+ \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{E}\psi^2(t, \mathbf{v}_0^t) P(t) dt \\
&\geq \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{E}\psi^2(t, \mathbf{v}_0^t) P(t) dt\,.
\end{aligned}$$

A lower bound of the Lyapunov exponent of $\mathbf{E}P(t)$ is given as

$$\limsup_{T\to\infty} \frac{1}{T} \log \mathbf{E} P(T) \geq 2\lambda - \frac{\mathscr{P}}{\sigma^2} = 2\lambda - \mathsf{SNR}\,. \tag{6.8}$$

The above lower bound can be achieved on

$$\psi^2(t, \mathbf{v}_0^t) P(t) = \mathscr{P}$$

and

$$\phi(t, \mathbf{v}_0^t) + b(t, \mathbf{v}_0^t) \hat{\mathbf{x}}(t) = 0 \,, \forall t \geq 0 \,,$$

which in turn gives the optimal solution of

$$\psi^*(t, \mathbf{v}_0^t) = \sigma \sqrt{\frac{\mathsf{SNR}}{P(0)}} \exp\left(\frac{\mathsf{SNR} - 2\lambda}{2} t\right)$$

and

$$\phi^*(t, \mathbf{v}_0^t) = -\sigma \sqrt{\frac{\mathsf{SNR}}{P(0)}} \exp\left(\frac{\mathsf{SNR} - 2\lambda}{2} t\right) \hat{\mathbf{x}}(t) \,.$$

Remark 6.2.7. Eqn. (6.8) shows that for the variance of $\tilde{\mathbf{x}}(t)$ to be exponentially decaying, one needs $\lambda < \frac{\mathsf{SNR}}{2} = \mathcal{C}$. In other words, converging estimation is achievable provided that the degree of instability of the source is less than the channel capacity. This observation can be roughly explained by Shannon's source-channel separation principle [4]. The unstable process produces extra information at the steady rate $\lambda (\geq 0)$, which needs to me transmitted in a timely manner for the vanishing of the mean square error (or rate distortion function). Therefore adequate channel capacity needs to be allocated. For an alternative in-depth treatment of unstable sources, by resorting to the concept of *any time capacity*, one is referred to [74].

6.2.3 Estimation Without Feedback

As a special case, the non-feedback communication scheme can be considered by proceeding to a similar argument as in the case when feedback is available. In fact, without the knowledge of $\mathbf{v}_0^t$, the optimal estimation of $\mathbf{x}(t)$, utilized on the transmitter's sid,e reduces to its expectation: $\mathbf{E}\mathbf{x}(t) = \exp(\lambda t) \mathbf{E}\mathbf{x}_0$ and $\phi(t, \mathbf{u}_0^t)$ becomes $\phi(t)$, which is a non-random function. Consequently the output of the estimator verifies the following dynamics:

$$\frac{dP(t)}{dt} = 2\lambda P(t) - \frac{1}{\sigma^2} P^2(t) b(t) \,,$$

which is solved by

$$P(t) = \frac{\exp(2\lambda t)}{P^{-1}(0) + \frac{1}{\sigma^2}\int_0^t \psi^2(\tau, \mathbf{v}_0^\tau)\exp(2\lambda\tau)d\tau}.$$

Similar to the previous case, we have the optimal solution

$$\phi^*(t) = -\sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp(-\lambda t)\mathbf{E}\mathbf{x}_0$$

and

$$\psi^*(t) = \sigma\sqrt{\frac{\mathsf{SNR}}{P(0)}}\exp(-\lambda t)$$

Remark 6.2.8. The following discussion further reveals the dependency of the optimal performance on the nature of the source dynamics:

- *Stable source (*$\lambda < 0$*)*: $P^*(t)$ is exponentially decaying at the rate $|\lambda|$, which is given by the inequality

  $$P^*(t) \le P(0)\exp(-|\lambda|t)$$

- *Neutrally stable source (*$\lambda = 0$*)*: $P(t)$ presents a much slower decay rate given by

  $$P^*(t) = \frac{P(0)}{1 + \mathsf{SNR}t}.$$

  The behavior of $P(t)$ in above equation is similar to the one that has been achieved by traditional sphere-packing coding strategy in discrete-time setting, with code word length n replaced by the time t.

- *Unstable source (*$\lambda > 0$*)*: $P(t)$ diverges with arbitrary instability rate, since

  $$P^*(t) = \frac{P(0)\exp(|\lambda|t)}{1 + \mathsf{SNR}t}.$$

  However, if only the finite horizon problem is considered, one can always find a global minimum.

6.3 Main Result: Optimal Estimation Over A Gaussian Channel

With the clear identification of the relation between communication and estimation in the previous section,we are now ready to tackle the main problem. The solution is given by using a water-filling type of argument.

6.3.1 Estimation Structure & a Dual Control Problem

Like in the scalar case, we first consider the optimal estimation problem for the vector dynamics

$$\begin{aligned} d\mathbf{x}(t) &= A\mathbf{x}(t)dt, \\ d\mathbf{v}(t) &= \phi(t,\mathbf{v}_0^t)dt + \psi^\top(t,\mathbf{v}_0^t)\mathbf{x}(t) + \sigma d\mathbf{W}(t)\,. \end{aligned}$$

The transmitter is expressed as $\phi(t,\mathbf{v}_0^t)dt+\psi^\top(t,\mathbf{v}_0^t)\mathbf{x}(t)$. The functions $\phi(t,\mathbf{v}_0^t) \in \mathbb{R}$ $\psi(t,\mathbf{v}_0^t) \in \mathbb{R}^n$ are nonlinear functions to be determined to minimize the Lyapunov index of the error variance, while ensuring the average power of channel input below the constrained level $\mathscr{P}$.

For the given transmitting scheme, the following Kalman-Bucy filter is adopted for the optimal estimation of $\mathbf{x}(t)$,

$$\begin{aligned} d\hat{\mathbf{x}}(t) &= A\hat{\mathbf{x}}(t)dt \\ &\quad + \frac{1}{\sigma^2}P(t)\psi(t,\mathbf{v}_0^t)[d\mathbf{v} - \phi(t,\mathbf{v}_0^t)dt - \psi^\top(t,\mathbf{v}_0^t)\hat{\mathbf{x}}(t)dt], \\ \dot{P}(t) &= AP(t) + P(t)A^\top \\ &\quad - \frac{1}{\sigma^2}P(t)\psi(t,\mathbf{v}_0^t)\psi^\top(t,\mathbf{v}_0^t)P(t)\,, \end{aligned} \tag{6.9}$$

where $P(t) := \mathbf{E}\left[\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^\top(t)|\mathbf{v}_0^t\right]$.

Remark 6.3.1. One can consider the dual control problem with plant dynamics given by

$$\begin{aligned} \frac{d\mathbf{x}(t)}{dt} &= A\mathbf{x}(t) + B\mathbf{u}(t), \\ d\mathbf{v}(t) &= \psi^\top(t,\mathbf{v}_0^t)\mathbf{x}(t)dt + \sigma d\mathbf{W}(t)\,, \end{aligned} \tag{6.10}$$

where the second equation models the AWGN channel identical to (6.2). If the control signal $\mathbf{u}(t)$ is designed via the typical LQG method [75], then the separation principle further shows that the variance of the error between the state and its estimated value is identical to $\mathbf{E}P(t)$ in (6.9). Therefore, to control the plant (6.10) over the AWGN channel, one can design a proper estimator to cope with the communication constraint, and the control part, which falls into the classical linear quadratic framework, is relatively independent, given the convergence of the estimation. Admittedly, the overall closed loop performance is fundamentally restricted by the communication-constrained estimation, no matter how well the controller is designed. On can further refer to [76] for the same property in general nonlinear systems. This estimation-control separation also explains why our focus is on the estimation part, whose relationship with communication constraint is unveiled in detail subsequently.

6.3.2 Solving The Estimation Problem: A water-filling approach

We first introduce a space $\mathcal{B}$, which is a real Hilbert space with internal product defined as

$$\langle \alpha, \gamma \rangle \triangleq \lim_{T\to\infty} \frac{1}{T} \int_0^T \alpha^\top(t)\gamma(t)dt \quad \alpha(\cdot), \gamma(\cdot) \in \mathcal{B}\,. \tag{6.11}$$

We say $\beta(\cdot) \in \mathcal{B}$, if $\langle \beta, \beta \rangle$ exits and is less than ∞. If $\beta(\cdot) \in \mathcal{B}$, then the $\lim_{T\to} \frac{1}{T}\int_0^T \beta(t)\beta^\top(t)$ exists.

Next, we define a new quantity $\beta(t) \triangleq \frac{1}{\sigma} P^{1/2}(t)\psi(t, \mathbf{v}_0^t)$, and assume that $\beta(\cdot) \in \mathcal{B}$.

Remark 6.3.2. Rigorously speaking, rather than a deterministic function of t as its notation suggests, $\beta(t)$ is a stochastic process on the σ-algebra generated by $\mathbf{v}_0^t$. However, we implicitly drop the randomness for three reasons: (1) We can always choose $\psi(t, \mathbf{v}_0^t) = \sigma P^{-1/2}(t)\beta(t)$ to make it non-stochastic; (2) The scalar cases in the previous section suggest that deterministic choices of $\beta(t)$ suffice for the optimality, which is also verified in the later discussion for this vector case; (3) This simplification reduces an otherwise accusive math discussion, while keeps the main point clear. For example, we see obviously that $\mathbf{E}P(t) = P(t)$, which will be useful in the later discussion.

The next lemma links Lyapunov exponent of the the variance of $\tilde{\mathbf{x}}$ with a matrix eigenvalue.

Lemma 6.3.3. If $P(0)$ is non-singular, and assume

$$\int_0^T \left(\lim_{T\to\infty} \frac{1}{T} \int_0^T \beta(t)\beta^\top(t)dt - \beta(t)\beta^\top(t) \right) dt \prec M \tag{6.12}$$

for some symmetric matrix M. then the following ine quality holds:

$$\begin{aligned} &\limsup_{t\to\infty} \frac{1}{t} \log \mathbf{E}\|\tilde{\mathbf{x}}(t)\|^2 \\ &\leq \lambda_{\max}\left(A^\top + A - \lim_{t\to\infty} \frac{1}{t} \int_0^t \beta(\tau)\beta^\top(\tau)d\tau \right) . \end{aligned} \tag{6.13}$$

The proof is follows the same line in [6].

Remark 6.3.4. Note that the assumption is not that strict. If one choose $\beta(t) = [\sqrt{2}\sin(t), \sqrt{2}\cos(t)]^\top$, it is easy to see

$$\int_0^T \left(\lim_{T\to\infty} \frac{1}{T} \int_0^T \beta(t)\beta^\top(t)dt - \beta(t)\beta^\top(t) \right) dt \prec \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Note that $\lambda_{\max}$ cannot made arbitrarily small due the power constraint, clearly shown by the following inequality

$$\begin{aligned} \mathscr{P} &\geq \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{E}[\phi(t, \mathbf{v}_0^t) + \psi^\top(t, \mathbf{v}_0^t)\mathbf{x}(t)]^2 dt \\ &\geq \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{E}[\phi(t, \mathbf{v}_0^t) + \psi^\top(t, \mathbf{v}_0^t)\hat{\mathbf{x}}(t)]^2 dt \\ &+ \mathbf{E}\psi^\top(t, \mathbf{v}_0^t)P(t)\psi(t, \mathbf{v}_0^t)dt \\ &\geq \limsup_{T\to\infty} \frac{1}{T} \int_0^T \mathbf{E}\psi^\top(t, \mathbf{v}_0^t)P(t)\psi(t, \mathbf{v}_0^t)dt \\ &= \sigma^2\langle \beta, \beta \rangle , \end{aligned} \tag{6.14}$$

where the second inequality follows from the orthogonality between $\tilde{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}(t)$.

Hence, an optimization problem could be formulated to achieve the lowest Lyapunov exponent upperbound by the choice of $\beta(\cdot)$.

$$
\begin{aligned}
& \inf_{\beta(\cdot)\in\mathcal{B}} \lambda_{\max}\left(A^\top + A - \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt\right) \\
s.t. \quad & \langle\beta,\beta\rangle \le \mathsf{SNR} \text{ and} \\
& A^\top + A - \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt \prec 0\,.
\end{aligned}
\tag{6.15}
$$

Another related optimization problem can be formulated in the same fashion, where the optimal $\beta(\cdot)$ must achieve a minimal channel SNR, subject to closed loop stability:

$$
\begin{aligned}
& \inf_{\beta(\cdot)\in\mathcal{B}} \langle\beta,\beta\rangle \\
s.t. \quad & A^\top + A - \lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt \prec 0\,.
\end{aligned}
$$

For both problems, once the optimal decision function $\beta^*(\cdot)$ is obtained, the optimal transmitter and estimator are straightforwardly obtained. Unfortunately, it is very hard, if not impossible to obtain $\beta^*(t)$ by using numerical routines, because these optimization problems are all inherently infinite-dimensional. Here we propose a solution inspired by the water-filling strategy.

Before jumping into the detailed development, an immediate observation can be made regarding the minimal SNR for mean square stability.

Proposition 6.3.5. If the error dynamics are mean-square exponentially stable, then channel SNR statistics for any causal transmission and decoding/control is given by

$$
\frac{\mathsf{SNR}}{2} > \frac{1}{2}\sum_i \lambda_i^+(A + A^\top) \ge \sum_j \Re^+\left(\lambda_j(A)\right) \tag{6.16}
$$

Proof of Proposition 6.3.5: Note that matrices $A+A^\top$, $\lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt$ and the difference of the two are Hermitian, so all their eigenvalues are real and can be ordered as $\lambda_1 \ge \lambda_2, ..., \ge \lambda_n$ for convenience. Then using Theorem III.4.1 of [77] and noting the fact that $\lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt - (A + A^\top) \succ 0$, we have

$$
\begin{aligned}
0 &< \sum_{i=1}^k \lambda_i\left(\lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt - (A + A^\top)\right) \\
&\le \sum_{i=1}^k \lambda_i\left(\lim_{T\to\infty}\frac{1}{T}\int_0^T \beta(t)\beta^\top(t)dt\right) - \sum_{i=1}^k \lambda_i\left(A + A^\top\right)\,,
\end{aligned}
\tag{6.17}
$$

for all $k \geq 1$. Particularly, the inequality (6.17) is also valid for $k = \kappa \triangleq \max_i \left\{i|\lambda_i(A + A^\top) \geq 0\right\}$, in which we have

$$\begin{aligned}
\sum_{i=1}^{\kappa} \lambda_i \left(A + A^\top\right) &< \sum_{i=1}^{\kappa} \lambda_i \left(\lim_{T\to\infty} \frac{1}{T} \int_0^T \beta(t)\beta^\top(t)dt\right) \\
&\leq \sum_{i=1}^{n} \lambda_i \left(\lim_{T\to\infty} \frac{1}{T} \int_0^T \beta(t)\beta^\top(t)dt\right) \qquad (6.18) \\
&= \lim_{T\to\infty} \frac{1}{T} \int_0^T \beta^\top(t)\beta(t)dt \leq \mathsf{SNR}\,.
\end{aligned}$$

The first inequality in (6.16) is straightforward to obtain. The second inequality is a direct application of Proposition III.5.3 of (3.22) in [77]. The detailed proof is omitted. □

Now we are ready to construct an optimal information transmission scheme. More specifically, given the channel SNR level, the smallest mean-square convergence rate ν of the state is obtained via the choice of $\beta(\cdot)$. The complete algorithm follows these steps.

Basis Construction

Choose a set of orthonormal basis functions $\beta_i(\cdot) \in B, i = 1, 2, ..., n$ such that

$$\langle \beta_i, \beta_j \rangle = \delta_{ij}, \;\; i, j = 1, 2, ..., n$$

where δ_{ij} is the Kronecker's delta. There are a number of ways to construct the basis functions, e.g. if $n = 2$, we can simply choose

$$\beta_1(t) = \sqrt{2}\sin(\omega t), \text{ and } \beta_2(t) = \sqrt{2}\cos(\omega t) \quad \omega > 0.$$

Weight Choice by Water-filling

Choose an orthonormal matrix $Q \in \mathbb{R}^{n\times n}$ such that

$$Q^\top(A + A^\top)Q = \text{ diag } \{\lambda_1, \lambda_2, ..., \lambda_n\} \,,$$

where λ_i is short for $\lambda_i(A + A^\top)$. Then $\beta(\cdot)$ can be parameterized by the basis constructed in 1) with a set of weighting factors $\eta_1, \eta_2, ..., \eta_n \geq 0$ as

$$Q^\top \beta(t) = [\eta_1 \beta_1(t), \eta_2 \beta_2(t), ..., \eta_n \beta_n(t)]^\top .$$

Based on this fact, the following identity is evident and will be useful later for

$$\langle \beta, \beta \rangle = \langle Q^\top \beta, Q^\top \beta \rangle = \sum_{i=1}^{n} \eta_i^2 .$$

Then the convergence rate minimization problem (6.15) can be reduced to the following finite dimensional case

$$\begin{aligned} &\min_{\eta_i, \nu} \nu \\ s.t. &\sum_{i=1}^{n} \eta_i^2 \leq \mathsf{SNR} \text{ and } (\lambda_i - \nu)^+ \leq \eta_i^2 , \end{aligned}$$

where the positivity of η_i^2 brings up $(\lambda_i - \nu)^+ \leq \eta_i^2$. This standard optimization problem can be solved by using the Lagrange multipliers $\xi_i \in \mathbb{R}, i = 1, 2, ..., n$ and $L \in \mathbb{R}$. The objective function is re-written as

$$J \triangleq \nu + \sum_{i=1}^{n} \xi_i((\lambda_i - \nu)^+ - \eta_i^2) + L \left(\sum_{i=1}^{n} \eta_i^2 - \mathsf{SNR} \right) .$$

Differentiating with respect to $\eta_1^2, ..., \eta_n^2$ and ν respectively, we have

$$\begin{aligned} 0 &= \frac{\partial J}{\partial \eta_i^2} = -\xi_i + L \\ 0 &= \frac{\partial J}{\partial \nu} = 1 - \sum_{i \in \mathbb{S}} \xi_i \, , \mathbb{S} \triangleq \{ i | (\lambda_i - \nu) \geq 0 \} \end{aligned}$$

Solving the set of equations and using Kuhn-Tucker conditions, we have the optimal assignment of the energy

$${\eta^*}_i^2 = (\lambda_i - \nu^*)^+ , \quad \sum_{i=1}^{n} {\eta^*}_i^2 = \mathsf{SNR}$$

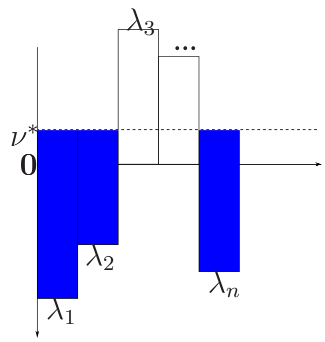

Figure 6.2: Water Filling For Optimal Energy Distribution

The optimal convergence rate ν^* solves

$$\sum_{i=1}^{n}(\lambda_i - \nu^*)^+ = \mathsf{SNR}$$

The solution is depicted graphically in Fig. 6.2. The vertical levels indicate the eigenvalues of the matrix $A + A^\top$, and the vertical axis is downward pointing. As the input power is increased from zero, we allocate the power to the eigenspace associated with the largest eigenvalue. When more power becomes available, it will be spilled over other eigenspaces to achieve an even ”water level”.

Optimal Transmitter and Estimator

Notice that (from last step)

$$\langle \beta^*, \beta^* \rangle = \sum_{i=1}^{n} {\eta^*}_i^2 = \mathsf{SNR}\,,$$

and the equality in (6.14) holds. Then we have the optimality achieved on

$$\phi^*(t, \mathbf{v}_0^t) + {\psi^*}^\top(t, \mathbf{v}_0^t)\hat{\mathbf{x}}(t) = 0\,.$$

Expressed in terms of $\beta^*(t)$, we have the optimal transmitter design:

$$\phi^*(t, \mathbf{v}_0^t) = -{\beta^*}^\top(t) P^{*-\frac{1}{2}}(t)\hat{\mathbf{x}}(t) \quad \psi^*(t, \mathbf{v}_0^t) = P^{*-\frac{1}{2}}(t)\beta^*(t)\,,$$

where $P^*(t)$ solves a variation of differential Lyapunov equation given by ($P^*(0) = P(0)$)

$$\dot{P}^*(t) = P^*(t)A + A^\top P^*(t) - P^{*\frac{1}{2}}(t)\beta^*(t)\beta^{*\top}(t)P^{*\frac{1}{2}}(t)\,. \tag{6.19}$$

and the estimator/receiver is given as

$$d\hat{\mathbf{x}}(t) = A\hat{\mathbf{x}}(t)dt + \frac{1}{\sigma^2}P^{*-\frac{1}{2}}(t)\beta^*(t)d\mathbf{v}(t)\,, \hat{\mathbf{x}}(0) = \hat{\mathbf{x}}_0$$

Remark 6.3.6. Note that the time profile of $P^*(t)$ (and hence $\psi^*(t, \mathbf{v}_0^t)$) can be determined off-line by integrating (6.19).

6.4 Simulation: Estimation via Amplitude Modulation

In this section we demonstrate our approach by using an analog amplitudes modulation (AM) method to transmit the estimation error. The schematic block diagram is shown in Fig. 6.3, where we do not assume any digitalization (A/D, D/A, quantization etc.) for simplicity. Here the plant is given as

$$\frac{d\mathbf{x}(t)}{dt} = \begin{bmatrix} 0 & 1 \\ -6 & 3.5 \end{bmatrix} \mathbf{x}(t)\,, \mathbf{x}(0) = [1 \quad 1]^\top\,.$$

The communication channel is corrupted by a standard white Gaussian noise ($\dot{\mathbf{W}}(t)$, $\sigma^2 = 1$) and is assumed to have the power constraint $\mathscr{P} = 13$ (SNR $= \mathscr{P}/\sigma^2 = 13$).

The design procedure follows the three steps proposed in the previous section, following an initialization stage:

1. The estimator is initialized with $\hat{\mathbf{x}}_0 = [0, 0]^\top$, and $P(0)$ is set to a 2×2 unit matrix;

2. We choose the basis functions as

$$\beta_1(t) = \sqrt{2}\sin(200\pi t) \text{ and } \beta_2(t) = \sqrt{2}\cos(200\pi t)$$

respectively .

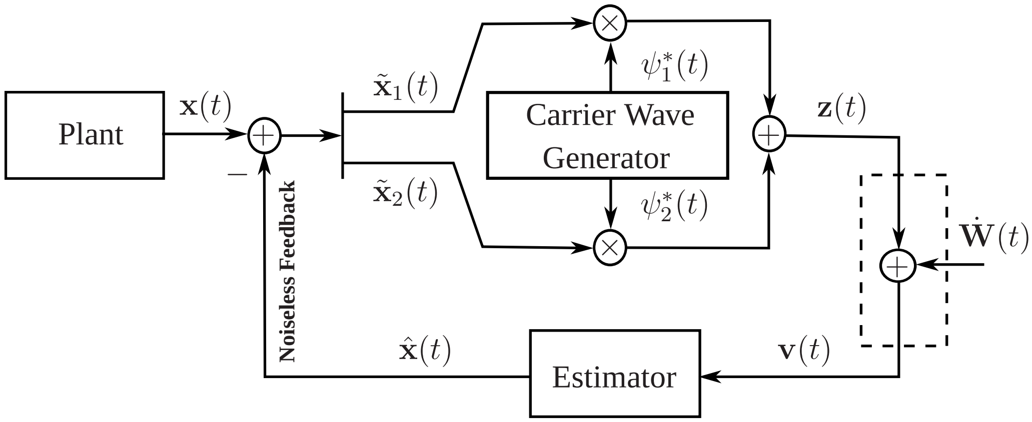

Figure 6.3: Feedback Amplitude Modulation Estimation

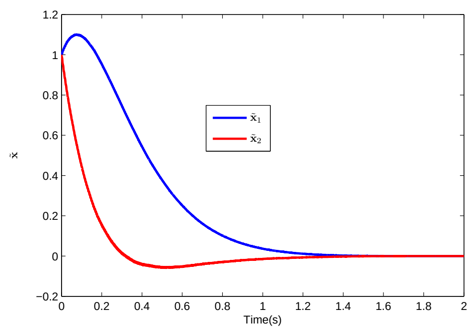

Figure 6.4: State Error

3. We conduct the water filling algorithm to determine the optimal convergence rate $\nu^* = -3$ and weights $\eta_1 = 0.6299\,, \eta_2 = 3.5501$. In turn we have

$$\beta^*(t) = \begin{bmatrix} -0.7901\sin(200\pi t) - 2.3186\cos(200\pi t) \\ 0.4114\sin(200\pi t) + 4.4532\cos(200\pi t) \end{bmatrix}$$

4. The carrier waves $\psi_1^*(t)$ and $\psi_2^*(t)$, as well as the estimator, can be generated by solving the matrix differential equation (Ricatti).

Figure 6.4 shows the time-history of the state error $\tilde{\mathbf{x}}(t)$; Fig. 6.5 shows the modulated channel input and Fig. 6.6 shows the noise-corrupted channel output.

The simulation result is consistent with the theory developed in this chapter and exhibits fast estimation error convergence in the presence of channel noise and power constraint. Compared with traditional amplitude modulation communications, where carrier waves are usually chosen as sinusoidal signals with constant

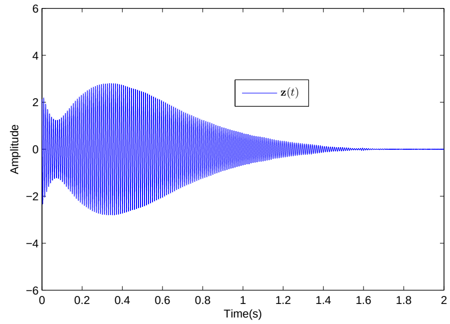

Figure 6.5: Channel Input

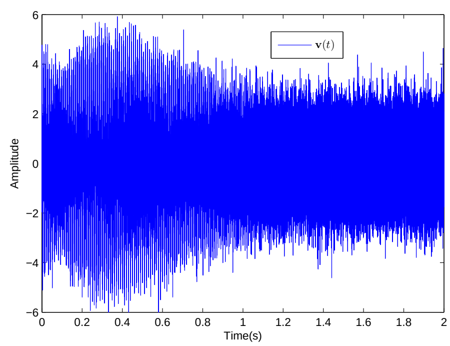

Figure 6.6: Channel Output

amplitudes, this method explicitly uses the knowledge of the signal dynamics (A) to generate a set of carrier waves to meet the needs of optimal estimation. This example also suggests that the method can be extended to more practical scenarios for the simplicity of amplitude modulation in communication systems.

6.5 Conclusion

In this chapter, we develop a design method to solve the optimal estimation problem with limited information. The objective is achieved by first fixing the structure of the transmitter and estimator by using conditional Kalman-Bucy filtering theory. Then the optimal parameters of the given structure are determined by a water-filling like technique by distributing the available channel input power to properly address the state-space of the dynamics to be estimated. The resulting communication/estimation scheme turns out to be surprisingly simple and fits into the conventional amplitude modulation framework with modified carrier waveforms, as shown in the example. The future research includes extension to digital communications and noisy feedbacks.

CHAPTER 7

FUTURE RESEARCH

In this dissertation, a framework has been laid out to facilitate the in-depth analysis of the closed loop trade-off in the presence of limited information. For the purpose of synthesis , several approaches have also been proposed to fit the existing control design methods into the systems with communication constraint. We list several directions as possible future research

- Bode-like formula for time-varying systems. A similar framework based on Chapter 2 can be readily utilized to derive a relevant information conservation law for the closed loop with a time-varying plant. The central issue relies on the "degree" of instability, which can be possibly characterized by a Lyapunov exponent. Not surprisingly, a certain dichotomy assumption should be enforced on the plant to obtain the Lyapunov exponent.

- Bode-like formula for continuous-time switched systems. This topic would combine the result of both Chapter 2 and Chapter 3. More specifically, when the regularity conditions similar in Chapter 2 are imposed on the continuous-time processes in the closed loop with Markov switching, the discrete-time result of Chapter 3 can be readily extended to continuous-time case.

- Control design in the presence of additive Gaussian channels. In Chapter 4 and 5, two approaches have been given for stationary and nonstationary cases respectively. Rather than designing a simple control gain, the future research along this line relies on the encoding and decoding schemes. Stochastic nonlinear control theory might be a suitable framework to work on.

REFERENCES

[1] N. Wiener, *Cybernetics*. J. Wiley, 1948.

[2] C. Shannon, "Coding theorems for a discrete source with fidelity criterion," 1960.

[3] R. Murray, K. Astrom, S. Boyd, R. Brockett, and G. Stein, "Future directions in control in an information-rich world," *IEEE Control Systems Magazine*, vol. 23, no. 2, pp. 20–33, 2003.

[4] T. Cover, J. Thomas, J. Wiley, and W. InterScience, *Elements of information theory*. Wiley Wiley-Interscience, 2006.

[5] H. Kitano, "Towards a theory of biological robustness," *Molecular systems biology*, vol. 3, no. 1, 2007.

[6] Y. Fang, K. Loparo, and X. Feng, "New estimates for solutions of lyapunov equations," *Automatic Control, IEEE Transactions on*, vol. 42, no. 3, pp. 408 –411, Mar. 1997.

[7] C. Shannon and W. Weaver, "A mathematical theory of communication," *Bell Syst. Tech. J*, vol. 27, pp. 379–423, 1948.

[8] R. Gallager, *Information theory and reliable communication*. John Wiley & Sons, Inc. New York, NY, USA, 1968.

[9] M. Seron, G. Goodwin, and J. Braslavsky, *Fundamental Limitations in Filtering Contro*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 1997.

[10] N. Martins and M. Dahleh, "Feedback control in the presence of noisy channels:Bode-like fundamental limitations of performance," *IEEE Transactions on Automatic Control*, vol. 52, no. 7, pp. 1604–1615, 2008.

[11] N. Martins, M. Dahleh, and J. Doyle, "Fundamental limitations of disturbance attenuation in the presence of side information," *IEEE Transactions on Automatic Control*, vol. 52, no. 1, pp. 56–66, 2007.

[12] N. Elia, "When Bode meets Shannon: control-oriented feedback communication schemes," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, p. 1477, 2004.

[13] N. Elia, “The information cost of loop shaping over Gaussian channels,” in *Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation*, 2005, pp. 1097–1102.

[14] K. Okano, S. Hara, and H. Ishii, “Characterization of a complementary sensitivity property in feedback control: An information theoretic approach,” *Automatica*, vol. 45, no. 2, pp. 504–509, 2009.

[15] A. Kolmogorov, “On the Shannon theory of information transmission in the case of continuous signals,” *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 102–108, 1956.

[16] A. Kolmogorov, “Theory of transmission of information,” *Amer. Math. Soc. Translation, Ser*, vol. 2, no. 33, pp. 291–321, 1963.

[17] M. Pinsker, *Information and information stability of random variables and processes*. Holden-Day, 1964.

[18] D. Liberzon, *Switching in systems and control*. Springer, 2003.

[19] G. Nair, S. Dey, and R. Evans, “Infimum data rates for stabilising Markov jump linear systems,” in *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, vol. 2. IEEE, 2003, pp. 1176–1181.

[20] C. Sims, “Implications of rational inattention,” *Journal of Monetary Economics*, vol. 50, no. 3, pp. 665–690, 2003.

[21] B. Pataracchia, “Design limits in regime-switching cases,” *Department of Economics University of Siena*, 2008.

[22] Y.-H. Kim, “Feedback capacity of stationary gaussian channels,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 57–85, Jan. 2010.

[23] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. Sastry, “Foundations of control and estimation over lossy networks,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[24] D. Li and N. Hovakimyan, “Bode’s integral for stochastic switched systems,” in *American Control Conference*, 2011, p. Accepted.

[25] C. Charalambous, A. Farhadi, and S. Denic, “Control of continuous-time linear Gaussian systems over additive Gaussian wireless fading channels: a separation principle,” *IEEE Transactions on Automatic Control*, vol. 53, no. 4, pp. 1013–1019, 2008.

[26] A. Rojas, J. Braslavsky, and R. Middleton, “Brief paper: Fundamental limitations in control over a communication channel,” *Automatica*, vol. 44, no. 12, pp. 3147–3151, 2008.

[27] S. Tatikonda and S. Mitter, “Control over noisy channels,” *IEEE Transactions on Automatic Control*, vol. 49, no. 7, pp. 1196–1201, 2004.

[28] M. Ruan and A. Choudhury, “Steady-state optimal controller with actuator noise variance linearly related to actuator signal variance,” *Automatic Control, IEEE Transactions on*, vol. 38, no. 1, pp. 133 –135, Jan. 1993.

[29] J. Braslavsky, R. Middleton, and J. Freudenberg, “Feedback stabilization over signal-to-noise ratio constrained channels,” *IEEE Transactions on Automatic Control*, vol. 52, no. 8, pp. 1391–1403, 2007.

[30] S. Tatikonda, A. Sahai, and S. Mitter, “Stochastic linear control over a communication channel,” *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1549–1561, 2004.

[31] E. Johannesson, A. Ghulchak, A. Rantzer, and B. Bernhardsson, “MIMO Encoder and Decoder Design for Signal Estimation,” in *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems–MTNS*, vol. 5, no. 9, 2010.

[32] E. Johannesson, A. Rantzer, B. Bernhardsson, and A. Ghulchak, “Encoder and decoder design for signal estimation,” in *American Control Conference (ACC), 2010*. IEEE, 2010, pp. 2132–2137.

[33] H. Robbins and S. Monro, “A stochastic approximation method,” *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[34] J. Schalkwijk and T. Kailath, “A coding scheme for additive noise channels with feedback–I: No bandwidth constraint,” *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 172–182, 1966.

[35] J. Liu and N. Elia, “Convergence of fundamental limitations in feedback communication, estimation, and feedback control over Gaussian channels,” *Arxiv preprint arXiv:0910.0320*, 2009.

[36] R. Lipster and A. Shiryaev, *Statistics of Random Processes, Vol. 2: Applications*. Springer Verlag, Berlin, 2001.

[37] S. Ihara, *Information theory for continuous systems*. World Scientific Pub Co Inc, 1993.

[38] M. Pinsker, V. Prelov, and S. Verdu, “Sensitivity of the ϵ-entropy of stationary continuous-time Gaussian processes,” *Problemy Peredachi Informatsii*, vol. 33, pp. 95–113, 1997.

[39] T. Cover, P. Gacs, and R. Gray, “Kolmogorov’s contributions to information theory and algorithmic complexity,” *The Annals of Probability*, pp. 840–865, 1989.

[40] K. Aström, *Introduction to stochastic control theory*. Academic Pr, 1970.

[41] A. Sayed and T. Kailath, "A survey of spectral factorization methods," *Numerical Linear Algebra with Applications*, vol. 8, no. 6-7, pp. 467–496, 2001.

[42] C. Baker and S. Ihara, "Information capacity of the stationary Gaussian channel," *IEEE Transactions on Information Theory*, vol. 37, no. 5, pp. 1314–1326, 1991.

[43] H. Bode, "Network analysis and feedback amplifier design," *Princeton, NJ*, 1945.

[44] J. Massey, "Causality, feedback and directed information," in *IEEE 1990 International Symposium on Information Theory and its Applications*, 1990, pp. 27–30.

[45] Y.-H. Kim, H. Permuter, and T. Weissman, "Directed information and causal estimation in continuous time," in *IEEE International Symposium on Information Theory, 2009.*, Jun. 2009, pp. 819 –823.

[46] B. Wu and E. Jonckheere, "A simplified approach to Bode's theorem for continuous-time and discrete-time systems," *IEEE Transactions on Automatic Control*, vol. 37, no. 11, pp. 1797–1802, 1992.

[47] D. Mustafa and K. Glover, *Minimum entropy H^∞ control*. Springer-Verlag New York, Inc., 1990.

[48] S. Ihara, *Information theory for continuous systems*. World Scientific Pub Co Inc, 1993.

[49] T. Kadota, M. Zakai, and J. Ziv, "Mutual information of the white gaussian channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 17, no. 4, pp. 368–371, Jul 1971.

[50] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[51] X. Feng, "Lyapunov exponents and stability of linear stochastic systems," Ph.D. dissertation, Case Western Reserve University, 1990.

[52] H. Radjavi and P. Rosenthal, *Simultaneous triangularization*. Springer Verlag, 2000.

[53] O. Costa and M. Fragoso, "Stability results for discrete-time linear systems with Markovian jumping parameters," *Journal of mathematical analysis and applications*, vol. 179, no. 1, pp. 154–178, 1993.

[54] C. Francq and J. Zakoļan, "Stationarity of multivariate Markov-switching ARMA models," *Journal of Econometrics*, vol. 102, no. 2, pp. 339–364, 2001.

[55] D. Li and N. Hovakimyan, "Bode-like integral for continuous-time closed-loop systems in the presence of limited information," *IEEE Transactions on Automatic Control (Submitted)*, 2010.

[56] V. Gupta, B. Hassibi, and R. Murray, "Optimal LQG control across packet-dropping links," *Systems & Control Letters*, vol. 56, no. 6, pp. 439–446, 2007.

[57] Q. Ling and M. Lemmon, "Power spectral analysis of networked control systems with data dropouts," *IEEE Transactions on Automatic control*, vol. 49, no. 6, pp. 955–959, 2004.

[58] N. Elia, "Remote stabilization over fading channels," *Systems & Control Letters*, vol. 54, no. 3, pp. 237–249, 2005.

[59] G. Nair, F. Fagnani, S. Zampieri, and R. Evans, "Feedback control under data rate constraints: An overview," *PROCEEDINGS-IEEE*, vol. 95, no. 1, p. 108, 2007.

[60] J. Taylor, *Monetary policy rules*. University of Chicago Press, 2001.

[61] W. Brock and S. Durlauf, "Elements of a theory of design limits to optimal policy," *The Manchester School*, vol. 72, no. s1, pp. 1–18, 2004.

[62] J. Hamilton, "Analysis of time series subject to changes in regime," *Journal of econometrics*, vol. 45, no. 1-2, pp. 39–70, 1990.

[63] J. Galí, *Monetary policy, inflation, and the business cycle: An introduction to the New Keynesian Framework*. Princeton Univ Pr, 2008.

[64] T. Duncan, L. Guo, and B. Pasik-Duncan, "Adaptive continuous-time linear quadratic Gaussian control," *IEEE Transactions on Automatic Control*, vol. 44, no. 9, pp. 1653–1662, 1999.

[65] W. Wonham, "On a matrix Riccati equation of stochastic control," *SIAM Journal on Control*, vol. 6, p. 681, 1968.

[66] W. Fleming and R. Rishel, "Deterministic and stochastic optimal control," *New York*, 1975.

[67] S. Boyd, L. Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM Studies in Applied Mathematics, 1994.

[68] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*. Springer, 1991.

[69] J. Löfberg, "Yalmip : A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. [Online]. Available: http://control.ee.ethz.ch/ joloef/yalmip.php

[70] K. Zhou and J. Doyle, *Essentials of robust control*. Prentice Hall New Jersey, 1998.

[71] J. Conway, *Functions of one complex variable II*. Springer, 1995.

[72] M. Chilali and P. Gahinet, "h_∞ design with pole placement constraints: an lmi approach," *IEEE Transactions on Automatic Control*, vol. 41, no. 3, pp. 358 –367, mar 1996.

[73] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge Univ Pr, 2005.

[74] A. SAHAI and S. MITTER, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link. part i: Scalar systems," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3369–3395, 2006.

[75] M. Davis, *Linear estimation and stochastic control*. Chapman and Hall Mathematics Series, 1977.

[76] S. Yu and P. Mehta, "Bode-like fundamental performance limitations in control of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1390–1405, 2010.

[77] R. Bhatia, *Matrix analysis*. Springer Verlag, 1997.