Networked Sensing, Estimation and Control Systems

Vijay Gupta
University of Notre Dame

Richard M. Murray
California Institute of Technology

Ling Shi
Hong Kong University of
Science and Technology

Bruno Sinopoli
Carnegie Mellon University

Contents

Chapter 7. Information Flow and Consensus   7-1

Chapter 8. Distributed Estimation   8-1

Chapter 9. Distributed Control   9-1

Bibliography   B-1

Preface

The area of "Networked Control Systems" has emerged over the past decade as a subdiscipline in control theory in which the flow of information in a system takes place across a communication network. Unlike traditional control systems, where computation and communications are usually ignored, the approaches that have been developed for networked control systems explicitly take into account various aspects of the communication channels that interconnect different parts of the overall system and the nature of the distributed computation that follows from this structure. This leads to a new set of tools and techniques for analysis and design of networked control systems that builds on the rich frameworks of communication theory, computer science and control theory.

This book is based on a series of courses that the authors have developed over the past several years, starting with a joint course taught at Caltech in Spring 2006. These courses were typically taken by students who have a good grounding in the basic techniques of control systems but may not have a strong background in computer science or some aspects of communication theory. While the level of mathematical detail in the book should allow it to be accessible to juniors or seniors in engineering, the treatment is tuned for first and second year graduate students in engineering or computer science. Some tutorial material on estimation theory is included, as well as a brief review of key concepts in graph theory that are needed primarily in the second half of the text.

The book is intended for researchers who are interested in the analysis and design of sensing, estimation and control systems in a networked setting. We focus primary on the effects of the network on the stability and performance of the system, including the effects of packet loss, time delay and distributed computation. We have attempted to provide a broad view of the field, in the hope that the text will be useful to a wide crossection of researchers. Most of the results are presented in the discrete time setting, with references to the literature for the continuous time analogs. We have also attempted to include a review of the current literature at the end of each chapter, with an emphasis on papers that are frequently referenced by others, along with some directions for future research, when appropriate. To keep the material focused, we have chosen to only touch on material on optimization-based control (e.g., receding horizon control) or protocols for distributed systems, although these are often an integral part of complex

networked control systems. References to the literature are given for readers interested in these important topics.

The book is organized into two main parts: a set of background chapters and the core material. Chapter 1 gives an introduction to the topic of networked control systems, including some driving application examples. Chapters 2–4 cover a collection of topics that are used throughout the remainder of the text. We assume familiarity with standard topics in estimation and control theory, including random processes, Kalman filtering and linear state space control theory, and provide only a quick review of this material in Chapter 2 to define the notation we will use and present some of the basic definitions and formulas. Chapters 3 and 4 complete the background chapters by giving concise overviews of the relevant results in information theory and Markovian jump linear systems, on which many of the later results of the book are built. These background chapters can be reviewed quickly for students and researchers already familiar with this material.

The core material on networked control systems is presented in Chapters 5 through 9. We begin by looking at the case of sensing, estimation and control of a single process across a communication channel, beging with the effects of rate limits in the channel in Chapter 5 and then the effects of packet loss in Chapter 6. Both of these chapters considers the cases where the communication channel affects on the measurements received from the sensor and where the channel affects both the measurements and the actuation commands. In Chapter 7 we begin to look at the problem of control over a graph, starting with an introduction to graph theory and the problem of consensus. Chapters 8 and 9 then go on to consider the distributed estimation and control problems, where one can have multiple processes, sensors, actuators, estimators and controllers distributed over a communications network. In each of these chapters on the core material we have attempted to present a unified view of many of the most recent and relevant results in network control, with the goal of establishing a foundation on which more specialized results of interest to specific groups can be covered.

The topics in the text have been taught by the authors and our colleagues in a variety of formats. In a semester-long, graduate course, it should be possible to cover most of the material in the book, assuming the students have good working knowledge of random processes, estimation theory and linear control systems. We have also used the material in the text for week-long short courses for masters and PhD students, where we cover the results in the background chapters in four 90 minute lectures, then spend 1–2 lectures on each of the remaining chapters. The material is fairly modular, so that the order of teaching the material can be varied according to the tastes of the instructor. The dependencies of the chapters are shown in Figure 1.
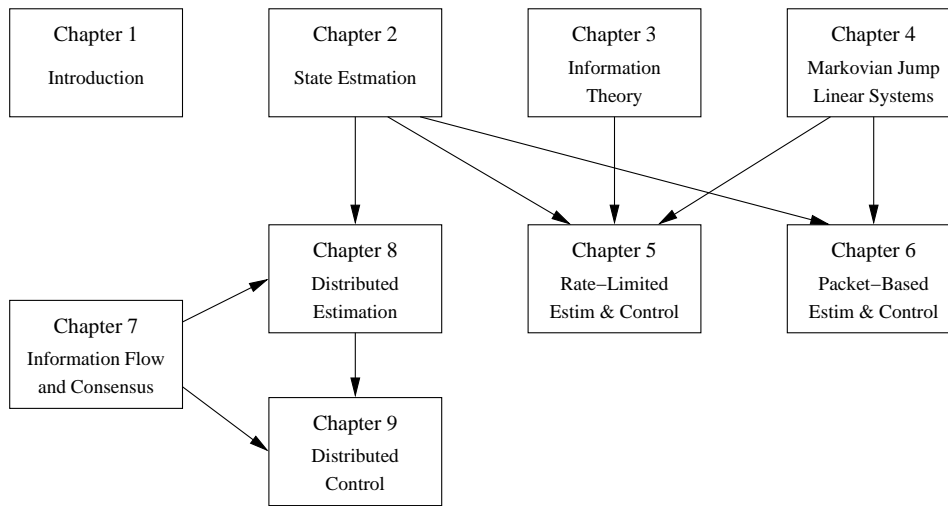
Figure 1: Dependencies of the chapters in the text.

Chapter 1
Introduction

Networked control is an emerging area of control theory driven by the increasing design, implementation and operation of control systems that makes use of communication networks to send information between the sensors, actuators and computational elements that make up a control system. In this chapter we provide an introduction to networked control systems (NCS), including a description of what is different about networked control versus traditional control design, some of the applications that are driving networked control systems research and engineering, and a list of some of the key subproblems in networked control that are the focus on the material in this text.

1.1 Overview of Networked Control Systems

Modern control theory is largely based on the abstraction that information ("signals") are transmitted along perfect communication channels and that computation is either instantaneous (continuous time) or periodic (discrete time). This abstraction has served the field well for 50 years and has led to many success stories in a wide variety of applications.

Future applications of control will be much more information-rich than those of the past and will involve networked communications, distributed computing, and higher levels of logic and decision-making (see [Mur03] for a recent analysis of future directions in this area). New theory, algorithms, and demonstrations must be developed in which the basic input/output signals are data packets that may arrive at variable times, not necessarily in order, and sometimes not at all. Networks between sensors, actuation, and computation must be taken into account, and algorithms must address the tradeoff between accuracy and computation time. Progress will require significantly more interaction between information theory, computer science, and control than ever before.

An emerging framework for networked control systems is shown in Figure 1.1. This architecture separates the traditional elements of sensing, estimation, control, and actuation for a given system across a network and also allows sharing of information between systems. As we will see in the examples below, careful decisions need to be made on how the individual components in this architecture are implemented and how the communica-
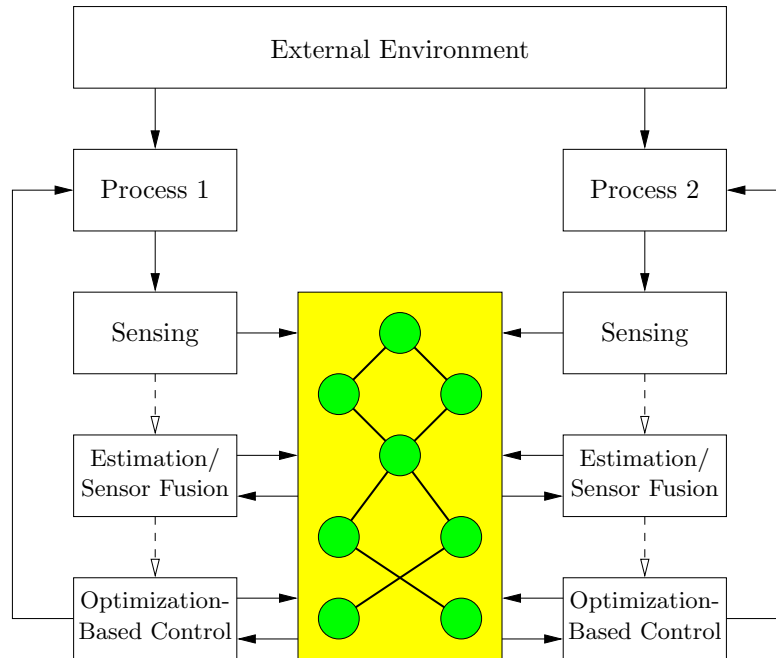
Figure 1.1: Emerging framework for networked control systems. Signals between control system modules for multiple processes are transmitted through a communication network.

tions across the networked elements is managed. This architecture can be used to model either a single system (using either half of the diagram) or multiple systems that interact through the network.

1.2 Application Examples

Embedded Systems

One example of the use of this architecture is autonomous operations for sensor-rich systems, such as unmanned, autonomous vehicles. As part of the 2004 and 2005 DARPA Grand Challenges, Caltech has developed two such vehicles ("Bob" and "Alice") that each make use of a networked control systems architecture. Alice, the 2005 vehicle, has six cameras, 4 LADAR units, an inertial meaurement unit (IMU), a GPS navigation system, and numerous internal temperature and vibration sensors. The raw data rate for Alice is approximately 1–3 Gb/s, which must be processed and acted upon at rates of up to 100 Hz in order to insure safe operation at high driving speeds.

The control system for Alice makes use of the architecture depicted in Figure 1.1, with distributed data fusion to determine elevation maps (for the
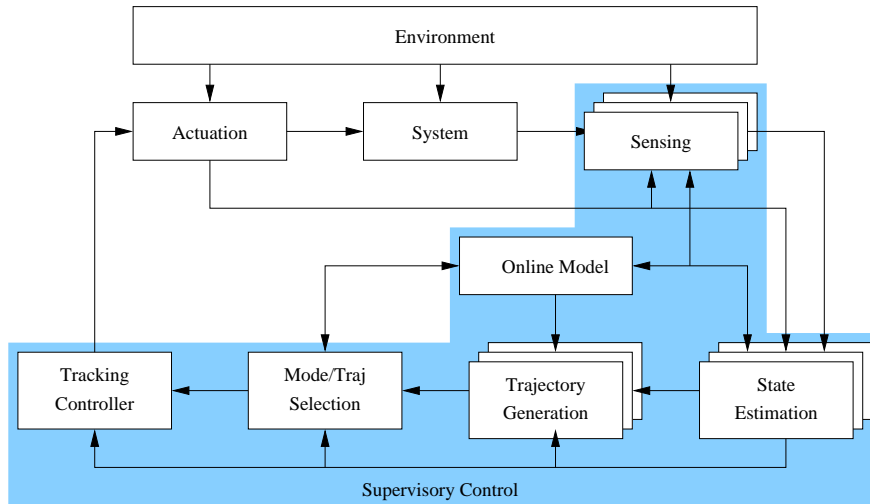
Figure 1.2: A detailed architecture for a networked control system, based on the control system for Alice [**?**].

height of the terrain in front of the vehicle), multiple optimization-based controllers to plan possible routes for the vehicle, and online modeling, fault management, and decision making to provide reliable and reconfigurable operation. Eight onboard computers distribute the computational load, sharing information at mixed rates across a 1 Gb/s switched network. System specifications call for reliable operation in the presence of up to 1 computer failure and 2 sensor failures, requiring careful coordination between computational elements.

A major challenge in Alice is determining how to send information between nodes. Because of the high data rates and computational loads on the CPUs, packets sent across the network are not always received and the system must be robust to various networking effects. The choice of protocols and design of the overall messaging system is currently informal and based on trial and error. As an example of the issues that must be resolved, certain packets of data are very important, such as packets containing raw sensor information from a portion of the terrain that is scanned only once. Other data can be dropped if needed, such as commanded trajectories (the old trajectory can be used for several sampling periods). Data from the inertial measurement unit must be received with minimum latency, while other data (a change in the temperature of the vehicle) is much less time critical. Substantial effort has been put into trying to make sure that the computations and network protocols complement each other and that loss of data and data latency does not degrade the performance of the system.

A more detailed architecture for a networked control system is shown in Figure 1.2. At the top of the figure, the standard elements for a control system are present: actuation, system dynamics, sensing and environmental

disturbances and noise. For many networked control systems, the amount of sensory information available is very large, requiring care in how this information is transmitted. Alice, for example, had between 1 and 3 gigabits/second (Gb/s) raw data rate, depending on the sensor suite taht was used. Another difference with traditional control systems is that the actuation subsystems are themselves embedded systems, capable of some amount of computation and local memory storage.

The primary control loop in a networked control system consists of state estimation, trajectory generation, and trajectory tracking. These elements can all represent relatively substantial computations (depending on the application) and are linked to each other through a number of network ports. In Alice, for example, the state estimation modules included a traditional inertial state estimator (combining GPS data with gryos and accelerometer measurements) as well as four computers that were estimating terrain information and computing a fused "speed map" that described the maximum allowable velocity that could be used in a given area of the terrain in front of it (more details on the software for Alice is given in Appendix **??**).

The information from the state estimators is used by trajectory generation algorithms that compute the desired state and inputs for the system to accomplish a task or minimize a cost function. The trajectory generation algorithms are responsible for taking into account actuator and state constraints on the sytem, as well as the nonlinear nature of the underlying process dynamics. A typical approach for these algorithms is to perform optimization-based control, in which one attempts to minimize a cost function subject to satisfying the constraints and dynamics. With the advances in computational power, it is often possible to run these optimization-based planners quickly enough that they can recompute the path from the current location in a "receding horizon" fashion, allowing feedback at the planner level. This is particularly useful to manage uncertainty in the cost function, for example when the cost is determined in real-time (as in the case of Alice, where the cost is based no the terrain that is being traversed).

As in the case of state estimation, networked control systems often use more than one trajectory generation algorithm running simultaneously. Since the physical system can only track one trajectory, some level of mode management and trajectory selection is required. This mode or trajectory selection logic is often under the control of higher levels of decision making (supervisory control).

The last element of the primary control loop is the trajectory tracking module, which is responsible for high frequency disturbance rejection and tracking. This module is itself a feedback system, using the state estimate and the desired trajectory to compute the actuation commands. In the context of a networked control system, the primary difference with traditional trajectory tracking algorithms is the need to run in a asynchronous execution environment, where reference trajectories and sensory measurements
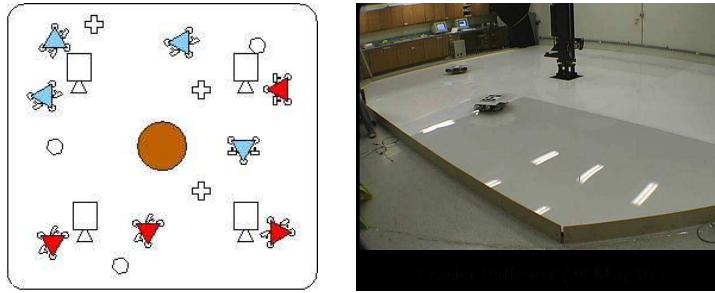
Figure 1.3: The Caltech Multi-Vehicle Wireless Testbed. The left figure shows the layout of the testbed area, including overhead cameras and fixed communication nodes (crosses and hexagons). The right picture is the current laboratory, with two vehicles shown.

may come in at varying rates, including short periods where no inputs may arrive (due to network delays, computational delays or fault handling in one of the other modules).

In addition to the elements of the primary control loop, networked control systems can also contain a number of modules responsible for higher levels of decision making. We loosely refer to these modules as "supervisory control": they are responsible for implementing various control system functions that involve choosing parameters used by the primary control loop (such as cost functions and communication rates), dealing with failures of hardware and software components, maintaining an online model of the system dynamics, and adapting the performance of the system based on observed behaviors and memory. While these elements are critical for the operation of a networked control systems, in this text we focus on the primary control loop, where the network effects are most directly relevant.

Sensor networks

Process Control

Cooperative Control

Another example of a networked control system is illustrated by the Caltech Multi-Vehicle Wireless Testbed (MVWT, shown in Figure 1.3), which consists of a collection of 8-12 vehicles performing cooperative tasks. The MVWT represents a slightly different instantiation of the architecture in Figure 1.1: each vehicle has a single processor with full access to local sensing and actuation, but information between vehicles must be sent across the network. The wireless commmuncation channels can exhibit significant degradation when multiple vehicles are attempting to communication and packet loss rates of 5-15% are not uncommon.

The issues in desiging a cooperative control policy for the MVWT vehicles faces many of the same challenges as those seen in Alice. Information

communicated between vehicles can be dropped, reordered or sent with variable delay. Sensor information required for overall situational awareness can be fused at multiple levels and/or in a distributed fashion. Again, the currently available protocols for network communications are not well tuned to operation in this type of environment. For example, bit errors in packets can result in losing the entire data packet, rather than passing the information to the applications layer where partial (lossy) information could still be used effectively.

1.3 NCS Subproblems

Chapter 2
State Estimation and Sensor Fusion

2.1 Review of Probability and Random Process

We assume the readers have some exposure to the theory of probability and random process. The material presented in this section only serves as a quick review of some basic concepts and tools from probability and random process that will be helpful to understand and derive some important results in subsequent sections and chapters. Good introductory books on probability and random process are:

Random Variables

Probability Space

Consider an experiment with many (possibly infinite) outcomes. All these outcomes form the *sample space* Ω. A subset $A \subset \Omega$ is called an *event*. Two events A_1, A_2 are called *mutually disjoint* if $A_1 \cap A_2 = \emptyset$. The *complement* of an event A is defined as $\bar{A} = \Omega \setminus A$.

A collection $\mathcal{F}$ of subsets of Ω is called a $\sigma-$**field** if it satisfies the following conditions:

1. $\emptyset \in \mathcal{F}$.

2. if $A_i \in \mathcal{F}, i = 1, 2, \ldots$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

3. if $A \in \mathcal{F}$, then $\bar{A} \in \mathcal{F}$.

A *probability measure* $P(\cdot)$ is a mapping from a $\sigma-$**field** $\mathcal{F}$ into the interval $[0, 1]$ such that the following axioms of probability are satisfied:

1. $P(A) \geq 0$ for all $A \subset \Omega$.

2. $P(\Omega) = 1$.

3. If $\{A_i, i = 1, 2, \ldots\}$ is a collection of disjoint members of $\mathcal{F}$, i.e., $A_i \cap A_j = \emptyset$ for all i, j, then $P(\cup A_i) = \sum_i P(A_i)$.

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*. From the axioms of probability, it follows that

$$P(A) \leq 1, \quad P(\emptyset) = 0, \quad P(\bar{A}) = 1 - P(A), \quad P(\cup A_i) \leq \sum_i P(A_i).$$

Conditional Probability, Independence, and Bayes' Rule

The joint probability of two events A and B is $P(A \cap B)$ which is often written as $P(AB)$ for simplicity. The conditional probability of A given B i.e., the probability that A occurs if B occurs in an experiment is

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad \text{assuming } P(B) \neq 0.$$

A and B are *mutually independent* if

$$P(AB) = P(A)P(B).$$

If $P(B) \neq 0$, the conditional probability $P(A|B)$ can be calculated from *Bayes' Rule* as

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

If $A_i, i = 1, 2, \ldots$ are mutually disjoint and $\cup A_i = \Omega$, then

$$P(B) = \sum_i P(B|A_i)P(A_i)$$

and

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)}.$$

Random Variable

A *random variable* is a function $X : \Omega \to \mathbb{R}$ with the property that

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \text{ for each } x \in \mathbb{R}.$$

The *cumulative distribution function* of a random variable X is a function $F_X : R \to [0, 1]$ given by

$$F_X(x) = P(X \leq x).$$

The cumulative distribution function F has the following properties

1. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.

2. If $x \leq y$, then $F_X(x) \leq F_X(y)$.

3. F_X is right-continuous.

When F_X is differentiable, we can define the associated *probability density function* $p_X(x)$ as

$$p_X(x) = \frac{dF_X(x)}{dx}.$$

The *joint cumulative distribution function* of two random variables X and Y, denoted as $F_{XY}(x, y)$, is given by

$$F_{XY}(x, y) = P(X \leq x) \cap P(Y \leq y).$$

If its derivative exists, the associated joint probability density function is given by

$$p_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y).$$

The definition and results extend trivially to three or more random variables.

Given $F_{XY}(x, y)$, the *marginal distribution functions* of X and Y can be calculated as

$$F_X(x) = P(X \leq x) = F_{XY}(x, \infty), \quad F_Y(y) = P(Y \leq y) = F_{XY}(\infty, y).$$

It follows that the *marginal density functions* of X and Y are

$$p_X(x) = \int_{-\infty}^{\infty} F_{XY}(x, y) dy, \quad p_Y(y) = \int_{-\infty}^{\infty} F_{XY}(x, y) dx.$$

The *conditional density function* of X given Y is given by

$$p_{X|Y}(x|y) = \frac{p_{XY}(x, y)}{p_Y(y)}.$$

The density function of X can also be calculated as

$$p_X(x) = \int_{-\infty}^{\infty} p_{X|Y}(x|y) p_Y(y) dy.$$

If X and Y are independent random variables, then the following statements holds and are equivalent to each other:

1. $F_{XY}(x, y) = F_X(x) F_Y(y)$.

2. $p_{XY}(x, y) = p_X(x) p_Y(y)$.

3. $p_{X|Y}(x|y) = p_X(x)$.

Statistical Properties of a Random Variable

A random variable X is completely specified by its distribution function $F_X(x)$ or density function $p_X(x)$. In many situations, $F_X(x)$ or $p_X(x)$ are difficult to obtain. It turns out the *mean* μ_X and *variance* σ_X^2 may provide us enough (useful) information about X. The mean and variance of a random variable X are defined as follows:

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} x p_X(x) dx,$$

$$\sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \int_{-\infty}^{\infty} (X - \mathbb{E}[X])^2 p_X(x) dx.$$

We denote $\mathbb{E}[\cdot]$ as the *expectation operator*. Since $\mathbb{E}[\cdot]$ is a linear operator, σ_X^2 can also be calculated as

$$\sigma_X^2 = \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2.$$

If X is a zero-mean random variable, i.e., $\mathbb{E}[X] = 0$, then $\sigma_X = \mathbb{E}[X^2]$. The kth *moment* of X is $m_k = \mathbb{E}[X^k]$ and the kth *central moment* is $\mu_k = \mathbb{E}\big[(X - \mathbb{E}[X])^k\big]$.

The *covariance* of two random variables X and Y is defined as $\mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]$. X and Y are *uncorrelated* if $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. If X and Y are uncorrelated, it is easy to verify that the covariance of X and Y is equal to zero. Clearly if X and Y are independent, then they are uncorrelated. However the converse does not hold in general.

The conditional expectation of X given $Y = y$ is

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|y)dx$$

which is a number that depends on the value of y. Similarly, the conditional expectation of X given Y is

$$\mathbb{E}[X|Y] = \int_{-\infty}^{\infty} x p_{X|Y}(x|Y)dx$$

which is also a random variable that depends on Y, i.e., it is a *function of the random variable* Y. The following property is very important and has great practical value in evaluating $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \mathbb{E}_Y\big[\mathbb{E}_X[X|Y]\big],$$

i.e., we first find the conditional expectation of X (conditioned on Y), and then remove the condition by taking the expectation with respect to Y. From this property, one can easily verify that if X and Y are independent, then

$$\mathbb{E}[X|Y] = \mathbb{E}[X].$$

Furthermore if X and Y are jointly independently of Z, then

$$\mathbb{E}[XY|Z] = \mathbb{E}[X|Z]\mathbb{E}[Y|Z].$$

Random Processes

A random process $X(t)$ is a generalization of a random variable. For a random variable, each experiment leads to a number (or a vector), while for a random process, each experiment leads to a function. For a fixed outcome $\omega \in \Omega$, one obtains the function $X(t, \omega)$, which is also called the *sample path* or *sample function* of the process. For a fixed t, $X(t, \omega)$ is a random variable with the underlying probability space Ω. The *mean process* of $X(t)$ is the time function $\mathbb{E}[X(t)]$. The autocorrelation of $X(t)$ is $\mathbb{E}[X(t_1)X(t_2)^T]$ and the autocovariance of $X(t)$ is $\mathbb{E}\big[\big(X(t_1) - m(t_1)\big)\big(X(t_2) - m(t_2)\big)^T\big]$.

Gaussian Random Variable and Random Process

A random process $X(t)$ is called a *Gaussian random process* if for any finite set $\{t_1, t_2, \ldots, t_N\}$, the random variables $\{X(t_1), X(t_2), \ldots, X(t_N)\}$ have a joint Gaussian distribution, i.e., their joint probability density function is given by

$$p_X(x) = \frac{1}{(2\pi)^{N/2}\sqrt{\det[\mathcal{C}_X]}}\exp\left[-\frac{1}{2}(x - m_X)^T \mathcal{C}_X^{-1}(x - m_X)\right] \qquad (2.1)$$

where $m_X = [m_X(t_1) \ m_X(t_2) \ \ldots \ m_X(t_N)]^T$ is the mean vector and $\mathcal{C}_X = \left[\text{cov}(X(t_i), X(t_j))\right]$ is the covariance matrix. Gaussian processes have the following properties.

Theorem 2.1. *Let $X(t)$ be a Gaussian process. Then*

1. *$X(t)$ is completely determined by m_X and $\mathcal{C}_X$.*

Theorem 2.2. *Let X and Y have a joint Gaussian distribution with mean and covariance given by*

$$\mu = \left[\begin{array}{c} \bar{x} \\ \bar{y} \end{array}\right] \text{ and } \Sigma = \left[\begin{array}{cc} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{array}\right].$$

Then X conditioned on $Y = y$ is Gaussian with mean and covariance given by

$$\mu_{X|Y=y} = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y - \bar{y}) \text{ and } \Sigma_{X|Y=y}\Sigma_x - \Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}.$$

In other words,

$$\mathbb{E}[X|Y = y] = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y - \bar{y}). \qquad (2.2)$$

The proof can be found in AndersonMoore1979.

Stability of stochastic systems

Sampling of a Continuous-time System

A wide variety of physical systems are modeled in the continuous-time domain. In this book, we focus on continuous-time systems with dynamics of the form

$$\frac{dx}{dt} = A_c x + B_c u + w, \quad y = C_c + v, \qquad (2.3)$$

where $x(t) \in \mathbb{R}^n$ is the state vector with unknown initial value $x(0)$, $u(t) \in \mathbb{R}^p$ is the input vector, $y(t) \in \mathbb{R}^m$ is the observation vector, and $w(t)$ and $v(t)$ are process disturbance and measurement noise. We assume $w(t)$ and $v(t)$ are mutually uncorrelated zero-mean Gaussian processes with autocovariances

$$\mathbb{E}[w(s)w(t)^T] = \delta_{st}\Sigma_{wc}, \quad \mathbb{E}[v(s)v(t)^T] = \delta_{st}\Sigma_{vc},$$

where $\delta_{st} = 1$ if $s = t$ and $\delta_{st} = 0$ otherwise.

As more and more controllers are implemented digitally, we need a procedure to convert the continuous-time system (2.3) into an equivalent discrete-time system. This procedure is called *sampling* or *discretization*. A frequently seen approach to implement the control law on a digital computer is to use a digital to analogue converter that holds the analog signal until the next time step, called *zero-order-hold* control.

Consider the following periodic sampling scheme: we sample the system (2.3) at time instances $t = k\tau$, $k = 0, 1, \ldots$, where $\tau > 0$ is the *sampling period*. It can be shown (see Astrom-Wittenmark) that the equivalent discrete-time system of (2.3) is given by

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad y_k = Cx_k + v_k, \tag{2.4}$$

where x_k and y_k correspond to $x(t)$ and $y(t)$ at time $t = k\tau$, and A, B and C are given by

$$A = e^{A_c \tau}, \quad B = \int_0^\tau e^{A_c t} dt B_c, \quad C = C_c. \tag{2.5}$$

In the discrete-time setting, the process and measurement noises are also uncorrelated zero-mean Gaussian random processes with covariance

$$\mathbb{E}[w_s w_k] = \delta_{sk} \Sigma_w, \quad \mathbb{E}[v_s v_k^T] = \delta_{sk} \Sigma_v,$$

where

$$\Sigma_w = \int_0^\tau e^{A_c t} \Sigma_{wc} e^{A_c^T t} dt, \quad \Sigma_v = \Sigma_{vc}.$$

Computing Σ_w directly from the above formula is sometimes difficult due to the integral of matrix exponentials. An easier approach to compute it is given as follows. Define M and N as

$$M = \begin{bmatrix} -A_c & \Sigma_{wc} \\ 0 & A_c^T \end{bmatrix} \tau, \quad N = e^M.$$

Then it is straightforward to show that

$$N = \begin{bmatrix} * & X^{-1}\Sigma_w \\ 0 & X^T \end{bmatrix}.$$

Therefore Σ_w can be computed from

$$\Sigma_w = (X^T)^T X^{-1} \Sigma_w,$$

i.e., Σ_w is obtained by multiplying the transpose of the lower-right submatrix of N with the upper-right submatrix of N.

Most of the results developed in this book also extend straightforward to cases where the sensor measurement y_k involves a direct input term, i.e.,

$$y_k = Cx_k + Du_k + v_k. \tag{2.6}$$

For simplicity, we shall use the system model as described by (2.4) for the remainder of the book unless otherwise explicitly stated.

Markov Chains

2.2 Optimal Estimation

Minimum Mean Square Error Estimator

Suppose we wish to know some quantity X, and we are not able to make a direct and accurate measurement of X. However we can make some indirect measurement Y that is related to X. Our task is to get an "optimal" estimate of X from Y.

One question that immediately arises before we attempt to solve the estimation problem is: what is a good estimate and when an estimate is "optimal"?

Intuitively a "good" estimate should make the estimation error $\hat{X} - X$ "small" since we wish to reconstruct X as perfectly as possible. An "optimal" estimate should make $\hat{X} - X$ the "smallest" among all other estimates. Many metrics can be used to define the size of the error $\hat{X} - X$ (hence we are able to say if it is "small" or not). Since $\hat{X} - X$ is typically a random variable, the metric that we shall use throughout the book is the following *mean squared error* (MSE)

$$\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)].$$

Therefore given $Y = y$ (i.e., the measurement that we take), our task is to construct the optimal estimate $\hat{X}$ that minimizes

$$\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y].$$

It turns out that the optimal $\hat{X}$ has a very simple form, given in the following theorem.

Theorem 2.3. *The optimal estimate $\hat{X}^*$ that minimizes*

$$\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y]$$

is given by the following conditional expectation of X

$$\hat{X}^* = \mathbb{E}[X|Y = y].$$

Proof. We can rewrite $\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y]$ as follows

$$\begin{aligned}
&\mathbb{E}[(\hat{X} - X)^T(\hat{X} - X)|Y = y] \\
={} &\mathbb{E}[X^TX|Y = y] - 2\hat{X}^T\mathbb{E}[X|Y = y] + \hat{X}^T\hat{X} \\
={} &\left(\hat{X} - \mathbb{E}[X|Y = y]\right)^T\left(\hat{X} - \mathbb{E}[X|Y = y]\right) + \mathbb{E}\left[X^TX - \mathbb{E}[X]^T\mathbb{E}[X^T]|Y = y\right].
\end{aligned}$$

Since $\mathbb{E}\left[X^TX - \mathbb{E}[X]^T\mathbb{E}[X^T]|Y = y\right]$ is independent of $\hat{X}$, we conclude that

$$\hat{X}^* = \mathbb{E}[X|Y = y].$$

$\square$

$\hat{X}^* = \mathbb{E}[X|Y = y]$ is also called the *minimum mean squared error* (MMSE) estimate of X.

Example 2.1 Estimate a Gaussian random variable

Consider the following equation

$$Y = X + N \tag{2.7}$$

where X and N are both scalar zero-mean Gaussian random variables with covariances σ_x and σ_n respectively. Further assume X and N are uncorrelated. Suppose we make a measurement of X and get y. The MMSE estimate of X is then given by

$$\hat{X} = \mathbb{E}[X|Y = y] = \frac{\sigma_x}{\sigma_x + \sigma_n}y.$$

$$\nabla$$

Kalman Filtering

Consider the following discrete-time linear time-invariant system

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad y_k = Cx_k + v_k, \tag{2.8}$$

where $x_k \in \mathbb{R}^n$ is the state vector with unknown initial value x_0, $u_k \in \mathbb{R}^p$ is the input vector, $y_k \in \mathbb{R}^m$ is the observation vector, and w_k and v_k are process and measurement noises (or disturbances).

Clearly nothing can be said on any estimator without defining a structure on w_k and v_k. In this book, we are particularly interested in w_k and v_k that have the following properties:

- w_k and v_k are zero-mean Gaussian random vectors;

- $\mathbb{E}[w_k w_j^T] = \delta_{kj}\Sigma_w$ with $\Sigma_w \geq 0$;

- $\mathbb{E}[v_k v_j^T] = \delta_{kj}\Sigma_v$ with $\Sigma_v > 0$;

- $\mathbb{E}[w_k v_j^T] = 0 \ \forall j, k$,

where $\delta_{kj} = 0$ if $k \neq j$ and $\delta_{kj} = 1$ otherwise. We also assume the initial value x_0 of system (2.8) is a zero-mean Gaussian random vector that is uncorrelated with w_k and v_k for all $k \geq 0$. The covariance of x_0 is given by $\Pi_0 \geq 0$. Furthermore we assume $(A, \sqrt{Q})$ is stabilizable.

Let $Y_k = \{y_0, y_1, \ldots, y_k\}$ be the measurements available at time k and $U_k = \{u_0, u_1, \ldots, u_k\}$ be the input applied to the system up to time k. We are interested in looking for the MMSE $\hat{x}_k$ of x_k at each time $k \geq 0$ given Y_k and U_{k-1}. From Theorem 2.3, we know that $\hat{x}_k$ is given by

$$\hat{x}_k = \mathbb{E}[x_k|Y_k, U_{k-1}], \tag{2.9}$$

and the corresponding error covariance P_k is given by

$$P_k = \mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T|Y_k, U_{k-1}]. \tag{2.10}$$

Calculating $\hat{x}_k$ and P_k according to equation (2.9) and (2.10) is not trivial and is computationally intensive as k increases. The celebrated Kalman filter provides a simple and elegant way to compute $\hat{x}_k$ and P_k recursively.

Assume that $\hat{x}_{k-1}$ and P_{k-1} defined as in equation (2.9) and (2.10) are available. Consider the one-step state prediction $\hat{x}_{k|k-1}$ (also called the *a priori state estimate*) given by

$$\hat{x}_{k|k-1} = \mathbb{E}[x_k|Y_{k-1}, U_{k-1}]$$

and the associated estimation error covariance (also called the *a priori error covariance*) $P_{k|k-1}$ given by

$$P_{k|k-1} = \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^T|Y_{k-1}, U_{k-1}].$$

From (2.8), we have

$$\begin{aligned}
\hat{x}_{k|k-1} &= \mathbb{E}[x_k|Y_{k-1}, U_{k-1}] \\
&= \mathbb{E}[Ax_{k-1} + Bu_{k-1} + w_{k-1}|Y_{k-1}, U_{k-1}] \\
&= A\hat{x}_{k-1} + Bu_{k-1},
\end{aligned} \tag{2.11}$$

where we use the fact that w_{k-1} is independent of any y_t ($t \leq k-1$) and the expectation operator is linear. Consequently,

$$P_{k|k-1} = AP_{k-1}A^T + \Sigma_w. \tag{2.12}$$

Now consider y_k conditioned on Y_{k-1} and U_{k-1} which has mean

$$\mathbb{E}[y_k|Y_{k-1}, U_{k-1}] = \mathbb{E}[Cx_k + v_k|Y_{k-1}, U_{k-1}] = C\hat{x}_{k|k-1}$$

and covariance

$$\mathbb{E}\left[\left(y_k - \mathbb{E}[y_k]\right)\left(y_k - \mathbb{E}[y_k]\right)^T|Y_{k-1}, U_{k-1}\right] = CP_{k|k-1}C^T + \Sigma_v,$$

where we have used the fact that v_k is independent of Y_{k-1}. The cross covariance of x_k and y_k conditioned on Y_{k-1} and U_{k-1} is given by

$$\mathbb{E}\left[\left(x_k - \mathbb{E}[x_k]\right)\left(y_k - \mathbb{E}[y_k]\right)^T|Y_{k-1}, U_{k-1}\right] = P_{k|k-1}C^T.$$

From the above analysis, we see that the random vector $[x_k' \quad y_k']'$ conditioned on Y_{k-1} and U_{k-1} is Gaussian with mean and covariance

$$\begin{bmatrix} \hat{x}_{k|k-1} \\ C\hat{x}_{k|k-1} \end{bmatrix} \text{ and } \begin{bmatrix} P_{k|k-1} & P_{k|k-1}C^T \\ CP_{k|k-1} & CP_{k|k-1}C^T + \Sigma_v \end{bmatrix}.$$

Therefore from Theorem 2.2, x_k conditioned on y_k (and on Y_{k-1} and U_{k-1}, i.e., conditioned on Y_k and U_{k-1}) has mean

$$\mathbb{E}[x_k|Y_k, U_{k-1}] = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1})$$

and covariance

$$(I - K_kC)P_{k|k-1}$$

where $K_k = P_{k|k-1}C^T[CP_{k|k-1}C^T + \Sigma_v]^{-1}$ is the so-called *Kalman gain*.

Let us summarize what we have said so far. Given the system (2.8), the MMSE estimate $\hat{x}_k$ of x_k is given by $\hat{x}_k = \mathbb{E}[x_k|Y_k, U_{k-1}]$, which can be computed recursively as follows

1. *time update:*

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1} + Bu_{k-1},$$
$$P_{k|k-1} = AP_{k-1}A^T + \Sigma_w.$$

2. *measurement update:*

$$K_k = P_{k|k-1}C^T[CP_{k|k-1}C^T + \Sigma_v]^{-1},$$
$$\hat{x}_k = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}),$$
$$P_k = (I - K_kC)P_{k|k-1}.$$

The initial values of the recursion are set as $\hat{x}_0 = 0$ and $P_0 = \Pi_0$. The Kalman filter essentially consists of the above two update steps.

Lemma 2.1. *The Kalman gain K_k and the error covariance P_k satisfy*

$$K_k = P_k C^T \Sigma_v^{-1}. \tag{2.13}$$

Proof. Since $P_k = (I - K_kC)P_{k|k-1}$, it suffices to show

$$(I - K_kC)P_{k|k-1}C^T\Sigma_v^{-1} = K_k$$

which is equivalent to

$$P_{k|k-1}C^T\Sigma_v^{-1} = K_k(I + CP_{k|k-1}C^T\Sigma_v^{-1})$$
$$\Longleftrightarrow P_{k|k-1}C^T\Sigma_v^{-1} = P_{k|k-1}C^T[CP_{k|k-1}C^T + \Sigma_v]^{-1}(I + CP_{k|k-1}C^T\Sigma_v^{-1})$$
$$\Longleftarrow \Sigma_v = (I + CP_{k|k-1}C^T\Sigma_v^{-1})^{-1}(CP_{k|k-1}C^T + \Sigma_v)$$
$$\Longleftrightarrow \Sigma_v = \Sigma_v(\Sigma_v + CP_{k|k-1}C^T)^{-1}(CP_{k|k-1}C^T + \Sigma_v)$$

where the last equation holds trivially. □

Alternate proof. K_k is defined as

$$K_k = P_{k|k-1}C^T(\Sigma_V + CP_{k|k-1}C^T)^{-1}.$$

Multiplying through by the inverse term on the right and expanding, we have

$$K_k(\Sigma_V + CP_{k|k-1}C^T) = P_{k|k-1}C^T,$$
$$K_k\Sigma_V + K_kCP_{k|k-1}C^T = P_{k|k-1}C^T,$$

and hence

$$K_k\Sigma_V = P_{k|k-1}C^T - K_kCP_{k|k-1}C^T,$$
$$= (I - K_kC)P_{k|k-1}C^T = P_{k|k}C^T.$$

The desired results follows by multiplying on the right by Σ_V^{-1}. □

To simplify the notations, let us define $h : \mathbb{S}_+^n \to \mathbb{S}_+^n$ as

$$h(X) \triangleq AXA^T + \Sigma_w, \tag{2.14}$$

and $\tilde{g} : \mathbb{S}_+^n \to \mathbb{S}_+^n$ as

$$\tilde{g}(X) \triangleq X - XC^T[CXC^T + \Sigma_v]^{-1}CX, \tag{2.15}$$

where $\mathbb{S}_+^n$ is the set of n by n positive semi-definite matrices. Further define $g : \mathbb{S}_+^n \to \mathbb{S}_+^n$ as

$$g(X) \triangleq h \circ \tilde{g} = AXA^T + \Sigma_w - AXC^T[CXC^T + \Sigma_v]^{-1}CXA. \tag{2.16}$$

For functions $f, f_1, f_2 : \mathbb{S}_+^n \to \mathbb{S}_+^n$, $f_1 \circ f_2$ is defined as

$$f_1 \circ f_2(X) \triangleq f_1\big(f_2(X)\big), \tag{2.17}$$

and f^t is defined as

$$f^t(X) \triangleq \underbrace{f \circ f \circ \cdots \circ f}_{t \text{ times}}(X). \tag{2.18}$$

With these definitions, it is straightforward to verify that $P_{k+1|k}$ and P_{k+1} satisfy

$$P_{k+1|k} = h(P_k),$$
$$P_{k+1|k} = g(P_{k|k-1}),$$
$$P_{k+1} = \tilde{g}(P_{k+1|k}),$$
$$P_{k+1} = \tilde{g} \circ h(P_k).$$

The equation $g(X) = X$, i.e.,

$$AXA^T + \Sigma_w - AXC^T[CXC^T + \Sigma_v]^{-1}CXA = X \tag{2.19}$$

is called the *Discrete-time Algebraic Riccati Equation* (DARE).

Properties of the Kalman filter

2.3 Optimal Control

2.4 Further Reading

Exercises

2.1 Show $\mathbb{E}[X|Y = y] = \frac{\sigma_x}{\sigma_x + \sigma_n} y$ in Example 2.1.

Chapter 3
Information Theory

Chapter 4
Markovian Jump Linear Systems

In this chapter, we present a short overview of Markovian jump linear systems. A more thorough and complete treatment is given in books such as [**?**]. As in other chapters, our focus will be on the Linear Quadratic Gaussian (LQG) control of such systems. As we shall see, even though such systems are non-linear, they can be analyzed using tools that are similar to those used in linear system analysis.

4.1 Introduction to Markovian Jump Linear Systems

A useful category of system models are those in which the system operates in multiple modes. Although each of the individual modes in linear, the switching between these modes introduces non-linearity into the overall system description. A general theory of such systems is developed in the hybrid systems community. However, much tighter results can be developed if a further assumptions holds, that the *mode switches are governed by a stochastic process that is statistically independent from the state values.* In the case when the stochastic process can be described by a Markov chain, the system is called a Markovian jump linear system. Although the individual modes of such systems may be continuous or discrete, we will concentrate on the latter case here.

More formally, consider a discrete time discrete state Markov process with state $r(k) \in \{1, 2, \cdots, m\}$ at time k. Denote the transition probability $\text{Prob}(r(k+1) = j | r(k) = i)$ by q_{ij}, and the resultant transition probability matrix by Q. We will assume that the Markov chain is irreducible and recurrent. Also denote

$$\text{Prob}(r(k) = j) = \pi_j(k),$$

with $\pi_j(0)$ as given. The evolution of a Markovian jump linear system (MJLS), denoted by $\mathcal{S}_1$ for future reference, can be described by the following equations

$$x(k+1) = A_{r(k)}x(k) + B_{r(k)}u(k) + F_{r(k)}w(k) \tag{4.1}$$
$$y(k) = C_{r(k)}x(k) + G_{r(k)}v(k),$$

where $w(k)$ is zero mean white Gaussian noise with covariance R_w, $v(k)$ is zero mean white Gaussian noise with covariance R_v and the notation

$X_{r(k)}$ implies that the matrix $X \in \{X_1, X_2, \cdots, X_m\}$ with the matrix X_i being chosen when $r(k) = i$. The initial state $x(0)$ is assumed to be a zero mean Gaussian random variable with variance $\Pi(0)$. For simplicity, we will consider $F_{r(k)} = G_{r(k)} \equiv I$ for all values of $r(k)$ in the sequel. We also assume that $x(0)$, $\{w(k)\}$, $\{v(k)\}$ and $\{r(k)\}$ are mutually independent. The particular case when $q_{ij} = q_j, \forall i, j$ (i.e., the random process governing the switching of the modes is a Bernoulli process) is sometimes referred to as a Bernoulli jump linear system.

Such systems have been studied for a long time in the fault isolation community, and have received new impetus with the advent of networked control systems. We now consider some examples of applicability of Markovian jump linear systems.

Example 4.1
Consider the following example of a failure prone production system, which is the discrete time equivalent of the model presented in [AK86]. Consider a manufacturing system producing a single commodity. There is a constant demand rate d for the commodity, and the goal of the manufacturing system is to try to meet this demand. The manufacturing system is, however, subject to occasional breakdowns and so at any time k, the system can be in one of two states: a functional ($r(k) = 1$) state and a breakdown ($r(k) = 2$) state. The transitions between these two states are usually modeled to occur as a Markov chain with given mean time between failures and mean repair time. When the manufacturing system is in the breakdown state it cannot produce the commodity, while if it is in the functional state it can produce at any rate u up to a maximum production rate $r > d > 0$. Let $x(k)$ be the inventory of the commodity at time k, i.e., $x(k) = $ (total production up to time k) - (total demand up to time k). Then the system is a Markovian jump linear system that evolves as

$$x(k+1) = \begin{cases} x(k) + u(k) - d & r(k) = 1 \\ x(k) - d & r(k) = 2, \end{cases}$$

where $u(k)$ is the controlled production rate. A negative $x(k)$ denotes backlog, and $u(k)$ satisfies a saturation constraint. ∇

Example 4.2
Consider a linear process evolving as

$$x(k+1) = Ax(k) + Bu(k) + w(k),$$

and being observed by a sensor of the form

$$y(k) = Cx(k) + v(k).$$

The measurements from the sensor are transmitted to an estimator across an analog erasure link. At any time k, the estimator receives measurement $y(k)$ with probability $1 - p$, and with a probability p no measurement is received.

As discussed in another chapter, this is a common model for a dynamic process being estimated across an analog erasure channel. This is a Bernoulli jump linear system with two modes $r(k) \in \{0, 1\}$. For both the modes, the system matrices $A_0 = A_1 = A$ and $B_0 = B_1 = B$. Mode 0 corresponds to no measurement being received and for this case $C_0 = 0$. Mode 1 corresponds to measurement being received, and for this case $C_1 = C$. $\qquad \nabla$

4.2 Stability of Markovian jump linear systems

In this section, we discuss the stability of autonomous Markovian jump linear systems. We will see that the necessary and sufficient condition for stability can be presented an algebraic condition in terms of the spectral radius of a suitable matrix. We will also present an equivalent condition in terms of a linear matrix inequality.

Since an Markovian jump linear systems is a stochastically varying system, numerous notions of stability may be defined. We will primarily be interested in mean square stability. Thus, define the state covariance $C(k) = E[x(k)x^T(k)]$, where the expectation is taken with respect to the initial state, process and measurement noise, and the discrete modes till time k. The system is stable if the steady state covariance is bounded, i.e., if $\lim_{k \to \infty} C(k) < C^\star$, where $C^\star$ is a constant matrix, and the inequality is understood in the positive definite sense.

The stability condition for Markovian jump linear systems is given by the following result.

Theorem 4.1. *Consider the system $\mathcal{S}_1$ with the control input $u(k) = 0$. The system is stable if and only if the condition*

$$\rho \left((Q^T \otimes I) \, diag(A_i \otimes A_i) \right) < 1$$

holds, where $\rho(M)$ is the spectral radius of matrix M, Q is the transition probability matrix of the Markov chain governing the mode switches of the system, $\otimes$ denotes the Kronecker product, I is the identity matrix of suitable dimensions, and $diag(A_i \otimes A_i)$ denotes a block diagonal matrix formed by using the matrices $A_i \otimes A_i$ for various mode values i.

Proof. Consider the term

$$C^i(k) = E[x(k)x^T(k)|r(k) = i]\pi_i(k),$$

so that the covariance is given by

$$C(k) = \sum_{i=1}^m C^i(k).$$

We will study the evolution of terms $C^i(k)$. Conditioning on the state value

at time $k - 1$ yields

$$C^i(k) = \sum_{j=1}^{m} \text{Prob}(r(k-1) = j | r(k) = i) \pi_i(k) E[x(k)x^T(k) | r(k) = i, r(k-1) = j]$$

$$= \sum_{j=1}^{m} \text{Prob}(r(k) = i | r(k-1) = j) \pi_j(k-1) E[x(k)x^T(k) | r(k) = i, r(k-1) = j]$$

$$= \sum_{j=1}^{m} q_{ji} \pi_j(k-1) E[x(k)x^T(k) | r(k-1) = j],$$

where in the second line we have used the Bayes law, and in the third line we have used the fact that given the Markov mode at time $k - 1$, $x(k)$ is conditionally independent of the Markov mode at time k. Now given the Markov mode at time $k - 1$, the covariance of the state at time k can be related to the covariance at time $k - 1$. Thus, we obtain

$$C^i(k) = \sum_{j=1}^{m} q_{ji} \pi_j(k-1) \Big(A_j E[x(k-1)x^T(k-1) | r(k-1) = j] A_j^T + R_w \Big)$$

$$= \sum_{j=1}^{m} q_{ji} A_j C^j(k-1) A_j^T + \sum_{j=1}^{m} q_{ji} \pi_j(k-1) R_w.$$

We can vectorize this equation and use the identity

$$\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B)$$

to obtain

$$\text{vec}(C^i(k)) = \sum_{j=1}^{m} q_{ji} (A_j \otimes A_j)\text{vec}(C^j(k-1)) + \pi_i(k)\text{vec}(R_w). \qquad (4.2)$$

For values of $i = 1, \cdots, m$, these coupled linear equations define the stability of $C(k)$. We can stack the vectors $\text{vec}(C^i(k))$ for various values of i, and obtain that the dynamical system recursion is governed by the matrix $\left((Q^T \otimes I)\text{diag}(A_i \otimes A_i) \right)$. Thus, we need to consider the spectral radius of this matrix. $\qquad\square$

For a Bernoulli jump linear system, the condition reduces to the following simple form.

Theorem 4.2. *Consider the system $\mathcal{S}_1$ with the control input $u(k) = 0$ and the additional assumption that the Markov transition probability matrix is such that for all states i and j, $q_{ij} = q_i$. The system is stable if and only if the condition*

$$\rho \left(E[A_i \otimes A_i] \right) < 1$$

holds, where the expectation is taken over the probabilities $\{q_i\}$.

Proof. In this case, we have $q_{ij} = q_j, \forall i$. Moreover, $r(k)$ and $x(k)$ are independent, so that $C^i(k) = C(k)\pi_i(k) = C(k)q_i(k)$. Thus, (4.2) yields

$$\text{vec}(C(k)) = \sum_{j=1}^{m}(A_j \otimes A_j)\text{vec}(C(k-1))q_j(k) + \text{vec}(R_w)$$

$$= E[A_i \otimes A_i]\text{vec}(C(k-1)) + \text{vec}(R_w),$$

which yields the desired stability condition.   □

Even though the above conditions are simple to write, the calculation of the spectral value may grow computationally expensive as the number of Markov states increases. We can present an alternate condition in terms of a linear matrix inequality as follows.

Theorem 4.3.

Proof.   □

4.3 LQG control

We will develop the LQG controller of Markovian jump linear systems in three steps. We will begin by considering the optimal linear quadratic regulator. We will then consider the optimal estimation problem for Markovian jump linear systems in the minimum mean squared error (MMSE). Finally, we will present a separation principle that will allow us to solve the LQG problem as a combination of the above filters.

Optimal Linear Quadratic Regulator

The Linear Quadratic Regulator (LQR) problem for the system $\mathcal{S}_1$ is posed by assuming that the noises $w(k)$ and $v(k)$ are not present. Moreover, the matrix $C_{r(k)} \equiv I$ for all choices of the mode $r(k)$. The problem aims at designing the control input $u(k)$ to minimize the finite horizon cost function

$$J_{LQR}(K) = \sum_{k=1}^{K}\left(E_{\{r(j)\}_{j=k+1}^{K}}\left[x^T(k)Qx(k) + u^T(k)Ru(k)\right]\right)$$

$$+ x^T(K+1)P(K+1)x(K+1),$$

where the expectation at time k is taken with respect to the future values of the Markov state realization, and $P(K+1)$, Q and R are all assumed to be positive definite. The controller at time k has access to control inputs $\{u(j)\}_{j=0}^{k-1}$, state values $\{x(j)\}_{j=0}^{k}$ and the Markov state values $\{r(j)\}_{j=0}^{k}$. Finally, the system is said to be stabilizable if the infinite horizon cost function $J_\infty \overset{def}{=} \lim_{K\to\infty}\frac{J_{LQR}}{K}$ is finite.

The solution to this problem can readily be obtained through dynamic programming arguments. The optimal control is given by the following result.

Theorem 4.4. *Consider the LQR problem posed above for the system $\mathcal{S}_1$.*

1. *At time k, if $r(k) = i$, then the optimal control input is given by*

$$u(k) = - \left(R + B_i^T P_i(k+1)B_i\right)^{-1} B_i^T P_i(k+1)A_i x(k),$$

where for $j = 1, 2, \cdots, m$,

$$
P_j(k) = \sum_{t=1}^{m} q_{tj} \Big(Q + A_t^T P_t(k+1)A_t \\
- A_t^T P_t(k+1)B_t \left(R + B_t^T P_t(k+1)B_t\right)^{-1} B_t^T P_t(k+1)A_t \Big), \quad (4.3)
$$

and $P_j(K+1) = P(K+1), \forall j = 1, 2, \cdots, m$.

2. *Assume that the Markov states reach a stationary probability distribution. A necessary and sufficient condition for stabilizability of the system is that there exist m positive definite matrices X_1, X_2, $\cdots$, X_m and m^2 matrices $K_{1,1}$, $K_{1,2}$, $\cdots$, $K_{1,m}$, $K_{2,1}$, $\cdots$, $K_{m,m}$ such that for all $j = 1, 2, \cdots, m$,*

$$X_j > \sum_{i=1}^{m} q_{ij} \left((A_i^T + K_{i,j}B_i^T)X_i(A_i^T + K_{i,j}B_i^T)^T + Q + K_{ij}RK_{ij}^T \right).$$

3. *A necessary condition for stabilizability is that*

$$q_{i,i}\rho(A_i)^2 < 1, \qquad \forall i = 1, 2, \cdots, m,$$

where $\rho(A_i)$ is the spectral radius of the matrix A_i that governs the dynamics of unstabilizable modes of the process in the i-th mode.

Proof. The proof follows by standard dynamic programming arguments. We begin by rewriting the cost function J_{LQR} to identify terms in the cost that depend on $x(K)$ and $u(K)$:

$$J_{LQR}(K) = \sum_{k=1}^{K-1} \left(E_{\{r(j)\}_{j=k+1}^{K}} \left[x^T(k)Qx(k) + u^T(k)Ru(k) \right] \right) + T(K)$$

$$T(K) = E_{r(K)} \left[x^T(K)Qx(K) + u^T(K)Ru(K) \right] + x^T(K+1)P(K+1)x(K+1).$$

We rewrite $T(K)$ by explicitly conditioning it on the value of $r(K)$.

$$T(K) = \sum_{i=1}^{m} \pi_i(K) \left(x^T(K)Qx(K) + u^T(K)Ru(K) + x^T(K+1)P_i(K+1)x(K+1)|r(K) = i \right),$$

where $P_i(K+1) = P(K+1), \forall i$. At the time of calculation of $u(K)$, the mode $r(K)$ is known. To choose the control input for any value of the mode,

we complete the square of each of the terms in the summation. For the i-th term we obtain

$$\begin{aligned}
\left(x^T(K)Qx(K)\right. &+ u^T(K)Ru(K) + x^T(K+1)P_i(K+1)x(K+1)|r(K) = i) \\
&= x^T(K)Qx(K) + u^T(K)Ru(K) + (A_ix(K) + B_iu(K))^T P_i(K+1)(A_ix(K) + B_iu(K)) \\
&= x^T(K)M_i(K)x(K) + (u(K) + S_i^{-1}(K)B_i^T P_i(K+1)A_ix(K))^T S_i(K)(u(K) + S_i^{-1}(K)B_i^T P_i(K+1)
\end{aligned}$$

where

$$S_i(K) = R + B_i^T P_i(K+1)B_i$$
$$M_i(K) = Q + A_i^T P_i(K+1)A_i - A_i^T P_i(K+1)B_i S_i^{-1}(K)B_i^T P_i(K+1)A_i.$$

Thus, the optimal choice of $u(K)$ for the case $r(K) = i$ is

$$u(K) = -S_i^{-1}(K)B_i^T P_i(K+1)x(K).$$

With the optimal choice of $u(K)$ for all values of $i = 1, \cdots, m$, the term $T(K)$ reduces to

$$\begin{aligned}
T(K) &= \sum_{i=1}^{m} \pi_i(K) \left(x^T(K)M_i(K)x(K)|r(K) = i\right) \\
&= \sum_{i=1}^{m} \pi_i(K) \sum_{j=1}^{m} q_{ji} \left(x^T(K)M_i(K)x(K)|r(K) = i, r(K-1) = j\right) \\
&= \sum_{j=1}^{m} \sum_{i=1}^{m} \pi_i(K)q_{ji} \left(x^T(K)M_i(K)x(K)|r(K-1) = j\right) \\
&= \sum_{j=1}^{m} \left(x^T(K)(\sum_{i=1}^{m} \pi_i(K)q_{ji}M_i(K))x(K)|r(K-1) = j\right) \\
&= \sum_{j=1}^{m} \left(x^T(K)\pi_j(K-1)P_j(K)x(K)|r(K-1) = j\right) \\
&= E_{r(K-1)} \left[x^T(K)P_j(K)x(K)\right],
\end{aligned}$$

where

$$\pi_j(K-1)P_j(K) = \sum_{i=1}^{m} \pi_i(K)q_{ji}M_i(K).$$

Thus, the cost function J_{LQR} can be rewritten as

$$J_{LQR}(K) = \sum_{k=1}^{K-2} \left(E_{\{r(j)\}_{j=k+1}^{K-1}} \left[x^T(k)Qx(k) + u^T(k)Ru(k)\right]\right) + T(K-1)$$
$$T(K-1) = E_{r(K-1)} \left[x^T(K-1)Qx(K-1) + u^T(K-1)Ru(K-1) + x^T(K)P_i(K)x(K)\right].$$

If we rewrite $T(K)$ by explicitly conditioning it on the value of $r(K-1)$,

$$T(K-1) = \sum_{i=1}^{m} \pi_i(K-1)\Big(x^T(K-1)Qx(K-1) + u^T(K-1)Ru(K-1)$$
$$+ x^T(K)P_i(K)x(K)|r(K-1) = i\Big),$$

we see that the problem of choosing $u(K-1)$ is formally identical to the problem that we solved above for choosing $u(K)$. Thus, the same argument can be repeated at any time step recursively. At a general time k, the control input $u(k)$ given $r(k) = i$ is given by

$$u(k) = -S_i^{-1}(k)B_i^T P_i(K+1)x(K),$$

where

$$S_i(k) = R + B_i^T P_i(k+1)B_i$$
$$\pi_j(K-1)P_j(K) = \sum_{i=1}^{m} \pi_i(K)q_{ji}M_i(K)$$
$$M_i(k) = Q + A_i^T P_i(k+1)A_i - A_i^T P_i(k+1)B_i S_i^{-1}(k)B_i^T P_i(k+1)A_i,$$

with boundary value $P_i(K+1) = P(K+1)\forall i$. This proves the first part of the theorem.

To prove the second and third parts, we need to study the stability of the terms $P_i(0)$ as the horizon $K \to \infty$. $\qquad\square$

The sufficient condition for stabilizability can also be cast in alternate forms as linear matrix inequalities, that can be efficiently solved, as follows.

Theorem 4.5.

Proof. $\qquad\square$

The above conditions reduce to simpler form for Bernoulli jump linear systems. For this case, the LQR and stabilizability problems can be solved to yield the following result.

Theorem 4.6. *Consider system $\mathcal{S}_1$ with the additional assumption that the Markov transition probability matrix is such that for all states i and j, $q_{ij} = q_i$ (in other words, the states are chosen independently and identically distributed from one time step to the next). Consider the LQR problem posed above for the system $\mathcal{S}_1$.*

1. *At time k, if $r(k) = i$, then the optimal control input is given by*

$$u(k) = -\left(R + B_i^T P(k+1)B_i\right)^{-1} B_i^T P(k+1)A_i x(k),$$

where

$$P(k) = \sum_{t=1}^{m} q_t \Big(Q + A_t^T P(k+1) A_t$$
$$- A_t^T P(k+1) B_t \left(R + B_t^T P(k+1) B_t \right)^{-1} B_t^T P(k+1) A_t \Big).$$

2. *Assume that the Markov states reach a stationary probability distribution. A sufficient condition for stabilizability of the system is that there exists a positive definite matrix X, and m matrices K_1, K_2, $\cdots$, K_m such that*

$$X > \sum_{i=1}^{m} q_i \left((A_i^T + K_i B_i^T) X (A_i^T + K_i B_i^T)^T + Q + K_i R K_i^T \right).$$

3. *A necessary condition for stabilizability is that*

$$q_i \rho(A_i)^2 < 1, \qquad \forall i = 1, 2, \cdots, m,$$

where $\rho(A_i)$ is the spectral radius of the matrix A_i that governs the dynamics of unstabilizable modes of the process in the i-th mode.

Proof. The result follows readily from the LQR solution of Markovian jump linear systems. Specifically, if we substitute $q_{tj} = q_j \forall t$ in (4.3), we see that all matrices $P_j(k)$ are identical for $j = 1, \cdots, m$. If we denote this value by $P(k)$, we obtain the desired form of the LQR control law. Similarly the stability conditions in the theorem also follow from those for Markovian jump linear systems in Theorem 4.4. $\square$

Optimal Minimum Mean Squared Error Estimator

The minimum mean squared error estimate problem for the system $\mathcal{S}_1$ is posed by assuming that the control $u_{r(k)}$ is identically zero. The objective is to identify at every time step k, an estimate $\hat{x}(k+1)$ of the state $x(k+1)$ that minimizes the mean squared error covariance

$$\Pi(k+1) = E_{\{w(j)\},\{v(j)\},x(0)} \left[(x(k+1) - \hat{x}(k+1))(x(k+1) - \hat{x}(k+1))^T \right],$$

where the expectation is taken with respect to the process and measurement noises, and the initial state value (but not the Markov state realization). The estimator at time k has access to observations $\{y(j)\}_{j=0}^{k}$ and the Markov state values $\{r(j)\}_{j=0}^{k}$. Moreover, the error covariance is said to be stable if the expected steady state error covariance $\lim_{k \to \infty} E_{\{r(j)\}_{j=0}^{k-1}}[\Pi(k)]$ is bounded, where the expectation is taken with respect to the Markov process.

Since the estimator has access to the Markov state values till time k, the optimal estimate can be calculated through a time-varying Kalman filter. Thus, if at time k, $r_k = i$, the estimate evolves as

$$\hat{x}(k+1) = A_i \hat{x}(k) + K(k) \left(y(k) - C_i \hat{x}(k) \right),$$

where

$$K(k) = A_i \Pi(k) C_i^T \left(C_i \Pi(k) C_i^T + R_v \right)^{-1}$$

$$\Pi(k+1) = A_i \Pi(k) A_i^T + R_w - A_i \Pi(k) C_i^T \left(C_i \Pi(k) C_i^T + R_v \right)^{-1} C_i \Pi(k) A_i^T.$$

The error covariance $\Pi(k)$ is available through the above calculations. However, calculating $E_{\{r(j)\}_{j=0}^{k-1}}[\Pi(k)]$ seems to be intractable. Instead, the normal approach is to consider an upper bound to this quantity[1] that will also help in obtaining sufficient conditions for the error covariance to be stable.

The intuition behind obtaining the upper bound is simple. The optimal estimator presented above optimally utilizes the information about the Markov states till time k. Consider an alternate estimator that at every time step k, averages over the values of the Markov states $r_0, \cdots, r_{k-1}$. Such an estimator is sub-optimal and the error covariance for this estimator forms an upper bound for $E_{\{r(j)\}_{j=0}^{k-1}}[\Pi(k)]$. A more formal derivation for the upper bound is presented below.

Theorem 4.7. *The term $E_{\{r(j)\}_{j=0}^{k-1}}[\Pi(k)]$ obtained from the optimal estimator is upper bounded by $M(k) = \sum_{j=1}^{m} M_j(k)$ where*

$$M_j(k) = \sum_{t=1}^{m} q_{tj} \Big(R_w + A_t M_t(k-1) A_t^T$$
$$- A_t M_t(k-1) C_t^T \left(R_v + C_t M_t(k-1) C_t^T \right)^{-1} C_t M_t(k-1) A_t^T \Big),$$

with $M_j(0) = \Pi(0) \ \forall j = 1, 2, \cdots, m$. Moreover, assume that the Markov states reach a stationary probability distribution. A sufficient condition for stabilizability of the system is that there exist m positive definite matrices $X_1, X_2, \cdots, X_m$ and m^2 matrices $K_{1,1}, K_{1,2}, \cdots, K_{1,m}, K_{2,1}, \cdots, K_{m,m}$ such that for all $j = 1, 2, \cdots, m$,

$$X_j > \sum_{i=1}^{m} q_{ij} \left((A_i + K_{i,j} C_i) X_i (A_i + K_{i,j} C_i)^T + R_w + K_{ij} R_v K_{ij}^T \right).$$

Finally, a necessary condition for stabilizability is that

$$q_{i,i} \rho(A_i)^2 < 1, \qquad \forall i = 1, 2, \cdots, m,$$

where $\rho(A_i)$ is the spectral radius of the matrix A_i that governs the dynamics of unobservable modes of the process in the i-th mode.

Proof. We begin by defining

$$M_j(k) = \pi_j(k-1) E \left[\Pi(k) | r(k-1) = j \right],$$

---

[1] We say that A is upperbounded by B if $B - A$ is positive semi-definite.

so that

$$E\left[\Pi(k)\right] = \sum_{i=1}^{m} M_j(k).$$

Now we can bound each term $M_j(k)$ as follows.

$$M_j(k+1) = \pi_j(k) \sum_{i=1}^{m} E\left[\Pi(k+1)|r(k)=j, r(k-1)=i\right] \mathrm{Prob}(r(k-1)=i|r(k)=j)$$

$$= \sum_{i=1}^{m} E\left[A_j\Pi(k)A_j^T + R_w - A_j\Pi(k)C_j^T(C_j\Pi(k)C_j^T + R_v)^{-1}C_j\Pi(k)A_j^T|r(k-1)=i\right] q_{ij}\pi_i(k -$$

since given $r(k-1)$, $\Pi(k)$ and $r(k)$ are independent. Further, note that the Riccati operator

$$f_j(M) = A_jMA_j^T + R_w - A_jMC_j^T(C_jMC_j^T + R_v)^{-1}C_jMA_j^T$$

is both concave and increasing. Since it is concave, Jensen's inequality yields

$$M_j(k+1) \leq \sum_{i=1}^{m} \left(A_jE[\Pi(k)|r(k-1)=i]A_j^T + R_w - A_jE[\Pi(k)|r(k-1)=i]C_j^T(C_jE[\Pi(k)|r(k-1)=i]C\right)$$

Now from the definition of $M_i(k-1)$ and the fact that $f_j(.)$ is an increasing operator, we obtain the required bound.

For the stability proof,   □

The special case of a Bernoulli jump linear systems can be obtained from the above result by substituting $q_{ij} = q_j \forall i$. We state the result below.

Theorem 4.8. *Consider the estimation problem posed above for the system $\mathcal{S}_1$ with the additional assumption that the Markov transition probability matrix is such that for all states i and j, $q_{ij} = q_i$ (in other words, the states are chosen independently and identically distributed from one time step to the next). The term $E_{\{r(j)\}_{j=0}^{k-1}}[\Pi(k)]$ obtained from the optimal estimator is upper bounded by $M(k)$ where*

$$M(k) = \sum_{t=1}^{m} q_t \left(R_w + A_t M(k-1)A_t^T \right.$$

$$\left. - A_t M(k-1)C_t^T \left(R_v + C_t M(k-1)C_t^T \right)^{-1} C_t M(k-1)A_t^T \right),$$

with $M(0) = \Pi(0)$. Further, a sufficient condition for stabilizability of the system is that there exists a positive definite matrix X, and m matrices K_1, K_2, $\cdots$, K_m such that

$$X > \sum_{i=1}^{m} q_i \left((A_i + K_iC_i)X(A_i + K_iC_i)^T + R_w + K_iR_vK_i^T \right).$$

Finally, a necessary condition for stabilizability is that

$$q_i \rho(A_i)^2 < 1, \qquad \forall i = 1, 2, \cdots, m,$$

where $\rho(A_i)$ is the spectral radius of the matrix A_i that governs the dynamics of unobservable modes of the process in the i-th mode.

Linear Quadratic Gaussian Control

Given the optimal linear quadratic regulator and minimum mean squared error estimator, the solution of the linear quadratic Gaussian control problem can be solved by utilizing a separation principle. The Linear Quadratic Gaussian (LQG) problem for the system $\mathcal{S}_1$ aims at designing the control input $u(k)$ to minimize the finite horizon cost function

$$J_{LQG} = E \left[\sum_{k=1}^{K} \left(x^T(k)Qx(k) + u^T(k)Ru(k) \right) + x^T(K+1)P(K+1)x(K+1) \right],$$

where the expectation at time k is taken with respect to the future values of the Markov state realization, the measurement and process noises, and the initial state. Further, the matrices $P(K+1)$, Q and R are all assumed to be positive definite. The controller at time k has access to control inputs $\{u(j)\}_{j=0}^{k-1}$, measurements $\{y(j)\}_{j=0}^{k}$ and the Markov state values $\{r(j)\}_{j=0}^{k}$. The system is said to be stabilizable if the infinite horizon cost function $J_\infty \overset{def}{=} \lim_{K \to \infty} \frac{J_{LQG}}{K}$ is finite.

The solution to this problem is provided by Theorems 4.4 and 4.7 because of the following separation principle.

Theorem 4.9. *Consider the LQG problem for the system $\mathcal{S}_1$. At time k, if $r(k) = i$, then the optimal control input is given by*

$$u(k) = - \left(R + B_i^T P_i(k+1)B_i \right)^{-1} B_i^T P_i(k+1)A_i \hat{x}(k),$$

where for $P_i(k)$ is calculated as in Theorem 4.4 and $\hat{x}(k)$ is calculated using a time-varying Kalman filter.

Proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Given this separation principle, the stabilizability conditions provided in Theorems 4.4 and 4.7 can then be combined to yield the stabilizability conditions for the LQG case as well. Finally, we note that a similar separation principle also holds for Bernoulli jump linear systems. Thus, the LQG problem can be solved for this case as well.

4.4 H_∞ Control

Include?

4.5 Further Resources

Chapter 5
Rate-Limited Estimation and Control

In this chapter, we consider the class of networked control systems in which the communication channel can be described by a digital noiseless channel. Such a channel imposes a limit on the number of bits that can be transmitted across it as a function of time; however, the transmission is perfect. As we shall see, there is a minimum bit rate required for the existence of encoders and decoders so that the plant can be stabilized across such a channel. In that sense, this problem is an analog of the source coding problem in information theory. However, the results from information theory are not directly applicable to the control scenario because of their reliance on large delays for the block codes to work. Nevertheless, concepts and insights from information theory will be used in the following discussion.

The chapter is organized as follows. We begin by describing the channel model in the next section.

5.1 Channel Model

By a digital noiseless channel, we will mean the following model. Consider a finite alphabet S of cardinality $M \geq 1$. At every time k, the channel accepts as input one symbol $s(k) \in S$. With a delay of d time steps, the channel outputs the symbol $r(k + d) = s(k)$. We will nominally consider the delay to be 0; however, we mention how the results can be extended to any finite value of the delay. Since the encoder for such a channel maps a continuous variable (e.g., the state value or the measurement) to a discrete variable (the input of the channel), it is often referred to as a quantizer.

An alternate viewpoint is to consider a channel that operates with a binary alphabet; however, at every time step, it can support a data rate $R = \log_2 M$ bits per sample. From this perspective, the channel model is that of a bit rate limited channel. We can also distinguish between channels that support a rate R at every time step, and those that support an average rate $R = \lim_{N \to \infty} \sum_{k=0}^{N} \frac{R(k)}{N}$, where $R(k)$ refers to the instantaneous rate (or number of bits supported by the channel) at time k.

5.2 Single Block Design

Consider a process with state $x(k) \in \mathbf{R}^n$ that evolves as

$$x(k+1) = Ax(k) + Bu(k) + w(k),$$

where $w(k)$ is process noise modeled as white and in a bounded region $\mathcal{W}$. The initial state $x(0)$ is also assumed to lie in a bounded region $\mathcal{X}$. For simplicity, we assume that the sensor can observe the state $x(k)$. The sensor transmits data to a controller across a digital noiseless channel with rate 2^M bits at every time step. The single block design problem refers to a situation in which the sensor quantizes the state space using M bits and transmits them to the controller. The controller aims to calculate a control input $u(k)$ to minimize the quadratic cost

$$J_T = \sum_{k=0}^{T} E[x^T(k)Qx(k) + u^T(k)Ru(k)] + x^T(T+1)P(T+1)x(T+1).$$

If the infinite horizon cost $\lim_{T \to \infty} \frac{J_T}{T}$ is bounded, we say that the process has been stabilized. Notice that in the single block design paradigm for this channel implies that the quantizer is given and the system designer specifies the decoder/controller. However, the quantizer can be of many different types as long as it satisfies the rate constraint. Some popular choices for quantizers are uniform or logarithmic with given range and step sizes.

The presence of a digital noiseless channel significantly complicates the analysis and design of control loops even for the LQG problem. For one, quantization is inherently a non-linear process and thus converts the problem to a non-linear control problem. Thus, there are only a limited number of results about optimal controller design. Another reason is that the quantization error introduced at any time step impacts the state value, and hence the quantization error, at all future time steps. This relation can become very complicated for arbitrary quantizers, possibly even leading to the control having a dual effect. For the cases when process noise is present, the possibility of state value becoming large enough to fall outside the quantizer range (termed quantizer overflow) is an additional complication.

The chief approach in single block design is to make a white noise approximation for the quantization error. Under this approximation, the possibility of quantizer overflow is ignored and each of the n elements in the state vector are assumed to be quantized independently using a uniform quantizer with step size δ, where δ is such that the total number of bits transmitted by the quantizer is M. Moreover, the quantization error $q(k)$ is assumed to be white and independent of $x(k)$. Since the quantization error for a uniform quantizer with step size δ has mean 0 and variance $\delta^2/12$, the effect of the above assumptions is to replace the quantizer with a sensor of the form

$$y(k) = x(k) + v(k),$$

where $v(k)$ is sensor noise modeled as bounded and white with mean zero and variance $\delta^2/12$. The controller design problem thus reduces to the design of a stabilizing controller for a linear system, which can be readily solved. Since the noises are not Gaussian, the performance optimal controller is harder to design.

Some initial results when the assumption of quantization error being either white or independent of the state value is not made are provided for uniform and logarithmic quantizers in [?]. The chief technical tool is the following *high rate approximation* result from source coding theory [?].

Theorem 5.1. *Given a scalar quantizer with mean squared error based distortion measure $d(x, y) = \|x - y\|^2$, the expected distortion of the random variable X that is being quantized can be bounded as*

$$\bar{d} \geq d_L = \frac{1}{12N^2} E[\lambda(X)^{-2}],$$

where $\lambda(X)$ is the asymptotic quantizer density normalized to unit integral, while N refers to the number of quantization levels. Further, the lower bound becomes tighter as the rate of the quantizer becomes high.

Thus, consider the quadratic cost J_T for a scalar plant

$$x(k + 1) = ax(k) + u(k) + w(k),$$

where the noise $w(k)$ and the initial state $x(0)$ are both bounded. Assume that there is no quantizer overflow, and that the control input is given by $u(k) = f\hat{x}(k)$, where $\hat{x}(k)$ is the estimate of the state at the decoder. Then, for a midpoint based uniform quantizer, $d_L = \frac{\delta^2}{12}$ where δ is the quantizer step size. Moreover, as Marco and Neuhoff [?] proved, for a high rate uniform quantizer,[1] $E[x(k)\delta(k)] \ll E[\delta^2(k)]$ and can thus be approximated by zero. Thus, at high rates the cost J_T evaluates to

$$J_T = (Q + Rf^2)E[x^2(0)]\sum_{k=0}^{T}(a+f)^{2k} + Rf^2(T+1)\frac{\delta^2}{12} + \frac{Q + Rf^2}{1 - (a + f)^2}\left(\frac{\delta^2 f^2}{12} + \Sigma_w^2\right)\sum_{k=0}^{T}\left(1 - (a + f)^{2k}\right).$$

The optimal controller can now be evaluated numerically. On the other hand, if a logarithmic quantizer with ratio g operating over the union of regions $[-a, \epsilon]$ and $[\epsilon, a]$ is used, the distortion can be evaluated to be

$$d_L = \frac{(\ln g)^2}{12} E[x^2(k)].$$

Using the Cauchy-Schwarz inequality

$$-\sqrt{E[\delta^2(k)]E[x^2(k)]} \leq E[\delta(k)x(k)] \leq \sqrt{E[\delta^2(k)]E[x^2(k)]},$$

---

[1] There are some additional technical conditions required, which hold in this case.

we can then obtain that

$$h_1 E[x^2(0)]\frac{1 - g_1^{T+1}}{1 - g_1} + \frac{h_1\sigma^2(T - 1 + g_1^{T+1})}{1 - g_1} \leq J_T \leq h_2 E[x^2(0)]\frac{1 - g_2^{T+1}}{1 - g_2} + \frac{h_2\sigma^2(T - 1 + g_2^{T+1})}{1 - g_2},$$

where

$$g_1 = (a + f)^2 + cf^2 - 2 \mid f(a + f) \mid \sqrt{c}$$
$$g_2 = (a + f)^2 + cf^2 + 2 \mid f(a + f) \mid \sqrt{c}$$
$$h_1 = Q + Rf^2 + Rcf^2 - 2Rf^2\sqrt{c}$$
$$h_2 = Q + Rf^2 + Rcf^2 + 2Rf^2\sqrt{c}$$

and $c = \frac{(\ln g)^2}{12}$.

For a quantizer with large enough rate, either of the above approaches yield reasonably accurate results. However, analytically, the problem is largely unsolved since the framework with the above approximations fail to capture some crucial features of the solution. For one, the white noise approximation implies that the system can be stabilized by a suitable control law with any non-zero rate supported by the digital noiseless channel (provided that the pair (A, B) is stabilizable). However, as we shall see in the next section, the data rate theorem implies that there is a minimum data rate that needs to be supported by the channel, otherwise the system cannot be stabilized even in the two block design paradigm. Moreover, the assumptions mentioned in this section fail to predict the chaotic nature of the state space trajectory that was identified by Delchamps.

5.3 Two Block Design

The two block design paradigm involves designing both an encoder at the input of the channel and a decoder at the output of the channel. As we shall see, for the digital noiseless channel, encoders and decoders that achieve stability with the minimum possible bit rate have been identified for a variety of stability notions and conditions on the encoder structure. However, designs that minimize a performance cost are largely unknown.

We begin by considering the plant structure with state $x(k) \in \mathbf{R}^n$ that evolves as

$$x(k + 1) = Ax(k) + Bu(k) + w(k), \tag{5.1}$$

where $u(k) \in \mathbf{R}^m$ is the control input. The state is observed by a sensor that generates measurements $y(k) \in \mathbf{R}^p$ of the form

$$y(k) = Cx(k) + v(k).$$

For different notions of stability, we will make different assumptions on the noises $w(k)$ and $v(k)$, and the initial state $x(0)$. We assume that the pair (A, B) is controllable and the pair (A, C) is observable.

The encoder at the input of the channel transmits a symbol $s(k)$ from the alphabet S (equivalently, M bits) at every time step. The message that is transmitted is a function of past transmissions and all measurements till time k, i.e.,

$$s(k) = \gamma(k, y(0), y(1), \cdots, y(k), s(0), s(1), \cdots, s(k-1)).$$

The channel transmits the symbol $s(k)$ without distortion, but with a constant delay of d time steps. The decoder generates a control input of the form

$$u(k) = \delta(k, s(0), s(1), \cdots, s(k-d)).$$

We begin by considering stability in the sense of constraining the state value to lie within a bounded set. To this end, assume that the noises $w(k)$ and $v(k)$ are deterministic but unknown sequences constrained to lie in bounded sets $\mathcal{W}$ and $\mathcal{V}$ respectively. Moreover assume that the initial condition $x(0)$ lies in the bounded set $\mathcal{X}_0$. Then, we consider the system to be stable if the worst case cost

$$J = \limsup_{k \to \infty} \{\|x(k)\| : x(0) \in \mathcal{X}_0, w(j) \in \mathcal{W}, v(j) \in \mathcal{V}, j = 0, 1, \cdots\}$$

is bounded.

Date Rate Theorem

The basic result in stability across digital noiseless channels is called the data rate theorem and is stated in terms of the intrinsic entropy of a system. The intrinsic entropy of a system is a measure of instability of a system and for the linear process in equation (5.1) is defined by the relation $H = \sum_i \log_2(\max(|\lambda_i(A)|, 1))$ where $\lambda_i(A)$ is the i-th eigenvalue of the matrix A. Since any mode of the process whose evolution is governed by an eigenvalue with magnitude less than 1 is stable and decays to zero even without any control input, for stabilization purpose, without loss of generality we can consider A to have all eigenvalues with magnitude strictly larger than 1.

Theorem 5.2 (Date Rate Theorem:). *Consider the two block design formulation with a causal encoder and decoder structure as defined above with the channel supporting a rate R.*

1. *If $R \leq H$ and the process noise has non-zero support, then $J \to \infty$ with any encoder and decoder design.*

2. *If $R > H$ then*

$$J > \frac{\beta^{-1/n} \lambda(\mathcal{W})^{1/n}}{1 - 2^{-(R-H)/n}},$$

where β is the volume of an n-dimensional sphere with unit radius, and $\lambda(\mathcal{W})$ is the measure of $\mathcal{W}$.

The proof of this theorem relies on considering the rate of increase in the volume of the set that the norm of the state value $x(k)$ can be in. The volume increases at every step because of the unstable eigenvalues, and decreases because of the information passed by the encoder. Note that the control value simply shifts this set, and cannot alter the volume since all previous control values are known to the controller. By balancing the rate of increase and decrease, the two conditions in the data rate theorem are obtained. To focus on the basic idea of the proof, we prove the data rate theorem for the special case when the process state is a scalar $(n = 1)$. For this special case, the result implies:

1. If $R \leq \log_2(A)$ and the process noise has non-zero support, then $J \to \infty$ with any encoder and decoder design.

2. If $R > \log_2(A)$ then

$$J > \frac{0.5\lambda(\mathcal{W})}{1 - 2^{-(R-H)}},$$

for any encoder and decoder design, where $\lambda(\mathcal{W})$ is the measure of $\mathcal{W}$.

Proof. Define by $\lambda(x(t))$ the length of the possibly disconnected region defined by the set of values that the state value can achieve at time t for various values of control inputs, $x(0)$ and the noise till time t. Also for given values of the signals transmitted by the encoder, define the region $\lambda(x(t) : \{s(j)\}_{j=0}^{t-1} = \{c(j)\}_{j=0}^{t-1})$ similarly. Finally define

$$m(t) = \max_{\{c(j)\}_{j=0}^{t-1} \in S} \lambda(x(t) : \{s(j)\}_{j=0}^{t-1} = \{c(j)\}_{j=0}^{t-1}).$$

We wish to study the evolution of $m(t)$. By definition,

$$m(t+1) = \max_{\{c(j)\}_{j=0}^{t} \in S} \lambda(x(t+1) : \{s(j)\}_{j=0}^{t} = \{c(j)\}_{j=0}^{t})$$

$$= \max_{\{c(j)\}_{j=0}^{t} \in S} \lambda(Ax(t) + B\delta(t, \{c(j)\}_{j=0}^{t-d}) + w(t) : \{s(j)\}_{j=0}^{t} = \{c(j)\}_{j=0}^{t}).$$

Now for given symbols $\{c(j)\}_{j=0}^{t-d}$, the control value is a constant and hence cannot affect the measure of the set. Moreover, the Brunn-Minkowski inequality implies that $\lambda^{1/n}(a+b) \geq \lambda^{1/n}(a) + \lambda^{1/n}(b)$, for any sets a and b in n-dimensions. Utilizing these two facts, we obtain

$$m(t+1) \geq A \max_{\{c(j)\}_{j=0}^{t} \in S} \lambda(x(t) : \{s(j)\}_{j=0}^{t} = \{c(j)\}_{j=0}^{t}) + \lambda(w(t)). \quad (5.2)$$

Now, notice that

$$\{x(t) : \{s(j)\}_{j=0}^{t-1} = \{c(j)\}_{j=0}^{t-1}\} = \bigcup_{\text{all possible values of } c(t)} \{x(t) : \{s(j)\}_{j=0}^{t} = \{c(j)\}_{j=0}^{t}\}.$$

Thus, Brunn-Minkowski inequality yields

$$\lambda(x(t) : \{s(j)\}_{j=0}^{t-1} = \{c(j)\}_{j=0}^{t-1}) \leq \sum_{\text{all possible values of } c(t)} \lambda(x(t) : \{s(j)\}_{j=0}^{t} = \{c(j)\}_{j=0}^{t})$$

$$M \max_{c(t) \in S} \lambda(x(t) : \{s(j)\}_{j=0}^{t} = \{c(j)\}_{j=0}^{t}),$$

where M is the number of symbols in the alphabet S. Using this in (5.2) yields

$$m(t+1) \geq A \max_{\{c(j)\}_{j=0}^{t-1} \in S} \frac{1}{M} \lambda(x(t) : \{s(j)\}_{j=0}^{t-1} = \{c(j)\}_{j=0}^{t-1}) + \lambda(w(t))$$

$$= \frac{A}{M} m(t) + \lambda(w(t))$$

$$= 2^{-(R-H)} m(t) + \lambda(w(t),$$

since $\frac{A}{M} = \frac{2^H}{2^R}$. If $R \leq H$, then as $t \to \infty$, $m(t)$ grows without bound and $J \to \infty$. If $R > H$, then we can solve for $m(t)$ explicitly and achieve the bound stated in the theorem as $t \to \infty$. This, the data rate theorem is proven for the scalar case. □

The proof for the n-dimensional state space is along similar lines by considering the evolution of the volume of $|x(t)|$. The technical changes required are:

- to relate the volume of the set to $\sup |x(t)|$ by using the expression

$$\lambda(\mathcal{T}) \leq \beta (\sup |\tau| : \tau \in \mathcal{T})^{1/n},$$

  for any set $\mathcal{T}$, where β is the volume of the n-dimensional sphere with unit radius.

- to define $m(t)$ as the $1/n$-th root of volume so that Brunn-Minkowski inequality is applicable.

- to use the relation $\lambda(Ax) = \det(A)\lambda(x)$.

The proof of the theorem can be easily modified to consider the case when the rate R is time-varying. By considering the evolution of $m(t)$ in terms of $m(0)$, we see that the data rate theorem holds if we define R to be the long term average data rate.

Tightness of Bounds

There are two questions that one can consider regarding the tightness of the bounds:

1. Is it possible to achieve stability with a data rate $R = H + \epsilon$ for any non-zero ϵ?

2. Is it possible to achieve the lower bound on the state value when $R > H$?

The first question can be answered in affirmative. We construct an encoder and decoder for the scalar case that achieves stability with rate $H + \epsilon$ and indicate how it can be extended to more general cases. Consider at time k, the state $x(k)$ to be in a region with support $l(k)$. The encoder uniformly quantizes the region using M bits and transmits the symbol related to the midpoint of the quantization cell containing the state value. The decoder knows the region being quantized since there is no stochasticity in the system. Thus, it knows exactly the midpoint of the quantization cell as transmitted by the encoder. It applies the control that translates the midpoint to the origin. Using this encoder decoder pair, the length $l(k)$ evolves as follows. It increases by a factor of A due to the state dynamics, and decreases by a factor of 2^M due to the quantizer. Thus,

$$l(k+1) = \frac{1}{M}(Al(k) + l_w(k)),$$

where $l_w(k)$ is the support of the noise $w(k)$. Thus, the length is bounded as $k \to \infty$ (in other words stability is achieved) if $\frac{A}{M} < 1$ or $R = H + \epsilon$ for any non-zero ϵ. For a vector plant, a similar encoder-decoder pair can be used for each individual mode as identified by a Jordan decomposition. By allotting bit rates suitably for all unstable modes, stability can be achieved for any rate $R > H$.

Regarding the second question, we notice that the lower bound on the norm of the state in case of $R > H$ is independent of the delay d. Thus, it can be expected that the bound is quite loose in general. While the presence of a finite delay cannot affect the stability condition, it does affect the performance in terms of the achievable norm of the state. One can modify the above proof by considering the evolution of $m(t)$ in terms of $m(t-d)$ to more accurately capture the effects of the delay. However, the effect of the rate R and the delay d do not separate out in a simple manner.

Even if $d = 0$, in general the bound is not tight for vector plants. For scalar plants, the encoder-decoder proposed above will achieve the bound with equality.

Other Notions of Stability

Now we briefly discuss how to consider alternate notions of stability. If the noises $w(k)$ and $v(k)$ are random variables, then the state $x(k)$ evolves stochastically. In that case, we might be interested in boundedness of a particular moment of the state. The most popular notion is that of mean square stability, i.e., we define the system to be stable if $E[x(k)x^T(k)]$ is bounded as $k \to \infty$. This stability notion can be analyzed using very similar tools as employed in the data rate theorem. Instead of volume of the set

in which the state can lie, we consider the evolution of the entropy power of the state. By using the entropy power inequality instead of the Brunn-Minkowski inequality, we can follow the same proof as that of the data rate theorem. Rather surprisingly, the condition on the minimum bit rate required for stability turns out to be identical to the deterministic case considered earlier.

However, unlike the deterministic case, a finite memory encoder-decoder pair is no longer sufficient to provide moment stability if the noise has infinite support. This result is related to the area over which the quantizer needs to operate. If the noise has a compact support, then given the region in which $x(k)$ can lie, the region that needs to be quantized for $x(k+1)$ is bounded. The encoder can thus vary its range at every time step and achieve stability. However, if the noise has unbounded support, then there is always a finite chance that the state at time $k+1$ falls outside the range of the quantizer. Such quantizer overflow leads to controller saturation. If the plant is unstable, the difference between where the state is, and the maximum value that can be handled by the controller exacerbates and quantizer overflow happens with increasing probability, ultimately leading to instability.

Stability with noises that have infinite support requires an encoder that adapts its range to allow the control signal to catch up. Moreover, the adaptation parameter can depend on the entire history of actions and have infinite set of values. A typical example of such a quantizer is the zoom in / zoom out quantizer. .

If there is no noise in the system, one can also aim for asymptotic stability. The first result in this direction is the insufficiency of a finite memory encoder / decoder pair to achieve asymptotic stability. This is because given any finite memory encoder / decoder, and a finite data rate, at any time t, the controller can only distinguish between finitely many state values. In other words, at any time t, there are a countable number of values for the initial state $x(0)$, that can be mapped to the origin. For any other value of the initial state, let, if possible, the system achieve asymptotic stability. This implies that if there exists a time T, such that for all $t > T$, the state value starting from this initial state satisfies $|x(t)| < \epsilon$, then there must exist some time $\tau > T$ such that $u(\tau)$ is nonzero. This is simply because the state value $x(t)$ is non-zero and hence needs to be driven to the origin by a suitable control input. Now, of the (countably many) distinct possible values possible for the control input at time τ, let $m = \min |u(\tau)|$. Moreover, choose $\epsilon = \frac{m}{1+|a|}$, where for simplicity we have choosen a scalar system with system parameter a. Then we can obtain a contradiction by noting that

$$m \leq |u(\tau)| = |x(\tau+1) - ax(\tau)| \leq |x(\tau+1)| + |ax(\tau)| \leq (1+|a|)\epsilon < m.$$

Thus, for any finite memory of the encoder and decoder, there are uncountably many initial values of the state such that the system trajectories starting from them do not achieve asymptotic stability. There are two chief

research directions that have emerged in light of this negative result:

1. Loosening the constraint of asymptotic stability to practical stability: In practical stability, the system satisfies the constraint $|x(t)| < \epsilon$ for any given ϵ for times in the range $T_1 < t < T_2$.

2. Considering encoders / decoders with infinite memory: The zoom in and out quantizers discussed above can achieve asymptotic stability by varying the region of state space that is quantized. As the state value moves closer to the origin, the range of the quantizer decreases. The achievement of increasingly finer quantization levels leads to asymptotic stability.

.

5.4 Extensions and Open Questions

The above discussion provides a sketch of the type of problems that have been analyzed and the results that are available for estimation and control across a digital noiseless channel. Since this is a research area that continues to see intense activity, there are a number of aligned problems that have also been looked at. We provide a discussion on some of these problems and outline a few open research questions.

Performance. Most of the material presented this far focussed on various notions of stability. Results on the design of control inputs to minimize a cost metric are more limited. As an example, consider the case when the noises are stochastic and Gaussian. One can consider the LQG problem for this case. The problem is difficult because of the non-linearity introduced by the quantizer. However, it can be proven that for the class of encoders that subtract the effect of previous control inputs (thus transmitting an innovation-like quantity), the certainty equivalence principle holds. Thus, for this class of encoders, the optimal control law is linear and its form is obtained by assuming that the controller has access to state information. However, instead of the state value, an estimate of the state value is used in the law to obtain the control input value. Moreover, there is no loss of optimality by restricting attention to this class of encoders.

However, a complete separation principle does not exist because of the non-linearity introduced by the quantizer. In other words, the estimate at the controller needs to be calculated for a distortion metric that depends on the input matrix. Thus, the estimator depends on the control value being calculated and is hard to characterize analytically. In fact, the LQG problem, and identifying the optimal controller for general cost functions, is still largely open. In fact, even the problems of identifying the optimal encoder for a given controller and identifying the optimal controller for a given encoder from a performance perspective do not yet have general solutions.

Noisy Digital Channels. If the channel is not noiseless, the problem becomes much harder. The easiest extension to consider is when the bits are erased by the channel with a certain probability at every time step. Stability conditions for such channels are obtained by extending the results from this chapter and that of analog erasure channels. Consider the case when the sensor data for a scalar process with process matrix A is transmitted to the controller over a channel such that the channel supports a rate of R bits at every time step, and the data packet is erased with a probability p in an independent and identically distributed manner at every time step. Then, a necessary and sufficient condition for existence of causal encoders and decoders that achieve second moment stability for the plant state is that $pA^2 + (1-p)\frac{A^2}{2^{2R}} < 1$. Note that in the limit $R \to \infty$, the condition reduces to $pA^2 < 1$ that is the stability condition for two block design with analog erasure channels. Similarly in the limit $p = 0$, we regain the condition for digital noiseless channel. The extension of the condition for vector plants is obtained using Jordan decomposition of the system matrix and considering each unstable mode separately.

If the noise in the channel can yield bit errors (rather than erasures), then a binary symmetric channel is more accurate. However, only very limited results are available even for stability over such channels.

Finite L_p Gain / Nonlinear systems.

Distributed Control. Since performance optimal distributed controllers are not available for arbitrary connection topologies even for the case of no limitations in terms of communication channels, it is not surprising that the problem is open if various components transmit data over digital noiseless channels. However, the stability problem has been looked at by many researchers and conditions are available in many different but equivalent forms.

5.5 Conclusions

In this chapter, we looked at control across a digital noiseless link. Stability conditions were identified in the two block framework. Some extensions and open problems were also looked at.

Chapter 6
Packet-Based Estimation and Control

Outline:

- Problem setup and useful lemmas

- Expected value of covariance

- Probabilistic bounds

- Markov models for packet loss (JLMS)

- Multi-channel and/or multi-sensor?

6.1 Introduction

This goal of this chapter is to analyze the problem of state estimation in the case where observations have to travel through a network to reach the estimator and may get lost in the process.

Consider the problem of navigating a vehicle based on the sensor web's estimate of its current position and velocity. The measurements underlying this estimate can be lost or delayed due to the unreliability of the wireless links. The question is, then, what is the amount of data loss that the control loop can tolerate to reliably perform the navigation task? And, can communication protocols be designed to satisfy this constraint? Answering these questions requires a generalization of classical control techniques that explicitly take into account the stochastic nature of the communication channel.

In this setting, the sensor network provides observed data that is used to estimate the state of a controlled system, and this estimate is then in turn used for control purposes. This chapter and the next one study the effect of data loss due to the unreliability of the network links.

The current chapter generalizes the most ubiquitous recursive estimation technique in control—the discrete Kalman filter [**?**]—modeling the arrival of an observation as a random process whose parameters are related to the characteristics of the communication channel, see Figure 6.8. In this setting the statistical convergence of the expected estimation error covariance is characterized and analyzed.

The classical theory relies on several assumptions that guarantee convergence of the Kalman filter. Consider the following discrete time linear
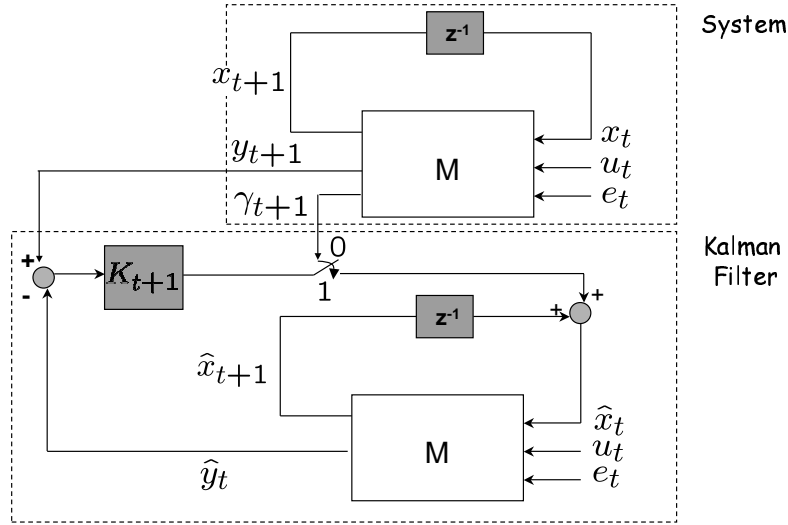
Figure 6.1: Overview of the system. The goal is to study the statistical convergence of the expected estimation error covariance of the discrete-time Kalman filter, where the observation, travelling over an unreliable communication channel, can be lost at each time step with probability $1 - \bar{\gamma}$.

dynamical system:

$$
\begin{aligned}
x_{t+1} &= Ax_t + w_t \\
y_t &= Cx_t + v_t,
\end{aligned}
\tag{6.1}
$$

where $x_t \in \mathbb{R}^n$ is the state vector, $y_t \in \mathbb{R}^m$ the output vector, $w_t \in \mathbb{R}^p$ and $v_t \in \mathbb{R}^m$ are Gaussian random vectors with zero mean and covariance matrices $Q \geq 0$ and $R > 0$, respectively. w_t is independent of w_s for $s < t$. Assume that the initial state, x_0, is also a Gaussian vector of zero mean and covariance Σ_0. Under the hypothesis of stabilizability of the pair (A, Q) and detectability of the pair (A, C), the estimation error covariance of the Kalman filter converges to a unique value from any initial condition [**?**].

The assumptions of the Kalman Filter have been relaxed in various ways. Extended Kalman filtering [**?**] attempts to cope with nonlinearities in the model; particle filtering [**?**] is also appropriate for nonlinear models and additionally does not require the noise model to be Gaussian. Recently, more general observation processes have been studied. In particular, in [**?**, **?**] the case in which observations are randomly spaced in time according

to a Poisson process has been studied, where the underlying dynamics evolve in continuous time. These authors showed the existence of a lower bound on the arrival rate of the observations below which it is possible to maintain the estimation error covariance below a fixed value, with high probability. However, the results were restricted to scalar SISO systems.

A similar approach is taken in this chapter. While the analysis falls within the framework of discrete time, it provides results for general n-dimensional MIMO systems. In particular, it considers a discrete-time system in which the arrival of an observation is a Bernoulli process with parameter $0 < \gamma < 1$, and, rather than asking for the estimation error covariance to be bounded with high probability, the study focuses on the asymptotic behavior (in time) of its average. The main contribution is to show that, depending on the eigenvalues of the matrix A, and on the structure of the matrix C, there exists a critical value γ_c, such that if the probability of arrival of an observation at time t is $\bar{\gamma} > \gamma_c$, then the expectation of the estimation error covariance is always finite (provided that the usual stabilizability and detectability hypotheses are satisfied). If $\bar{\gamma} \leq \gamma_c$, then the expectation of the estimation error covariance is unbounded. The following analysis provides explicit upper and lower bounds on γ_c, and shows that they are tight in some special cases.

Philosophically this result can be seen as another manifestation of the well known *uncertainty threshold principle* [**?**, **?**]. This principle states that optimum long-range control of a dynamical system with uncertainty parameters is possible if and only if the uncertainty does not exceed a given threshold. The uncertainty is modeled as white noise scalar sequences acting on the system and control matrices. In our case, the result is for optimal estimation, rather than optimal control, and the uncertainty is due to the random arrival of the observation, with the randomness arising from losses in the network.

6.2 Related Work

Studies on filtering with intermittent observations can be tracked back to Nahi [**?**] and Hadidi [**?**]. More recently, this problem has been studied using Jump Linear Systems (JLS) [**?**]. JLS are stochastic hybrid systems characterized by linear dynamics and discrete regime transitions modeled as Markov chains. In the work of Costa et al. [**?**] and Nilsson et al. [**?**, **?**] the Kalman filter with missing observations is modeled as a JLS switching between two discrete regimes: an open loop configuration and a closed loop configuration. Following this approach, these authors obtain convergence criteria for the expected estimation error covariance. However, they restrict their formulation to the steady state case, where the Kalman gain is constant, and they do not assume to know the switching sequence. The resulting

process is wide sense stationary [**?**], and this makes the exact computation of the transition probability and state error covariance possible. Other work on optimal, constant gain filtering can be found in the work of Wang et al. [**?**], who included the presence of system parameters uncertainty besides missing observations, and Smith et al. [**?**], who considered the fusion of multiple filters. Instead, we consider the general case of *time varying* Kalman gain. In the presence of missing observations, this filter has a smaller linear minimum mean square error (LMMSE) than its static counterpart, as it is detailed in Section 6.3.

The general case of time-varying Kalman filter with intermittent observations was also studied by Fortmann et al. [**?**], who derived stochastic equations for the state covariance error. However, they do not statistically characterize its convergence and provide only numerical evidence of the transition to instability, leaving a formal characterization of this as an open problem, which is addressed in this chapter. A somewhat different formulation was considered in [**?**], where the observations arrival have a bounded delay.

Finally, we point out that our analysis can also be viewed as an instance of Expectation-Maximization (EM) theory. EM is a general framework for doing Maximum Likelihood estimation in missing-data models [**?**]. Lauritzen [**?**] shows how EM can be used for general graphical models. In our case, however, the graph structure is a function of the missing data, as there is one graph for each pattern of missing data.

The chapter is organized as follows. In section 6.3 the problem of Kalman filtering with intermittent observations is formally defined. In section 6.4 upper and lower bounds on the expected estimation error covariance of the Kalman filter are provided, along with conditions on the observation arrival probability $\bar{\gamma}$ for which the upper bound converges to a fixed point, and for which the lower bound diverges. Section 6.5 describes some special cases and gives an intuitive understanding of the results. Section 6.6 compares the current approach to previous ones [**?**] based on jump linear systems.

6.3 Problem Formulation

Consider the canonical state estimation problem. The arrival of the observation at time t is modeled as a binary random variable γ_t, with probability distribution $p_{\gamma_t}(1) = \bar{\gamma}$, and with γ_t independent of γ_s if $t \neq s$. The output noise v_t is defined in the following way:

$$p(v_t|\gamma_t) = \left\{ \begin{array}{ccc} \mathcal{N}(0, R) & : & \gamma_t = 1 \\ \mathcal{N}(0, \sigma^2 I) & : & \gamma_t = 0, \end{array} \right.$$

for some σ^2. Therefore, the variance of the observation at time t is R if γ_t is 1, and $\sigma^2 I$ otherwise. In reality the absence of observation corresponds

to the limiting case of $\sigma \to \infty$. Following this approach the Kalman filter equations are re-derived using a "dummy" observation with a given variance when the real observation does not arrive, and then take the limit as $\sigma \to \infty$.

First define:

$$\hat{x}_{t|t} \triangleq \mathbb{E}[x_t | \mathbf{y_t}, \gamma_\mathbf{t}] \tag{6.2}$$

$$P_{t|t} \triangleq \mathbb{E}[(x_t - \hat{x}_t)(x_t - \hat{x}_t)' | \mathbf{y_t}, \gamma_\mathbf{t}] \tag{6.3}$$

$$\hat{x}_{t+1|t} \triangleq \mathbb{E}[x_{t+1} | \mathbf{y_t}, \gamma_\mathbf{t}] \tag{6.4}$$

$$P_{t+1|t} \triangleq \mathbb{E}[(x_{t+1} - \hat{x}_{t+1})(x_{t+1} - \hat{x}_{t+1})' | \mathbf{y_t}, \gamma_\mathbf{t}] \tag{6.5}$$

$$\hat{y}_{t+1|t} \triangleq \mathbb{E}[y_{t+1} | \mathbf{y_t}, \gamma_\mathbf{t}], \tag{6.6}$$

where the vectors $\mathbf{y_t}$ and $\gamma_\mathbf{t}$ are defined as: $\mathbf{y_t} \triangleq [y_0, \ldots, y_t]'$ and $\gamma_\mathbf{t} \triangleq [\gamma_0, \ldots, \gamma_t]'$. It is easy to see that:

$$\mathbb{E}[(y_{t+1} - \hat{y}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})' | \mathbf{y_t}, \gamma_\mathbf{t+1}] = CP_{t+1|t} \tag{6.7}$$

$$\mathbb{E}[(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})' | \mathbf{y_t}, \gamma_\mathbf{t+1}] = CP_{t+1|t}C' + \gamma_{t+1}R + (1 - \gamma_{t+1})\sigma^2 I \tag{6.8}$$

and it follows that the random variables x_{t+1} and y_{t+1}, conditioned on the output $\mathbf{y_t}$ and on the arrivals $\gamma_\mathbf{t+1}$, are jointly gaussian with mean

$$\mathbb{E}[x_{t+1}, y_{t+1} | \mathbf{y_t}, \gamma_\mathbf{t+1}] = \begin{pmatrix} \hat{x}_{t+1|t} \\ C\hat{x}_{t+1|t} \end{pmatrix},$$

and covariance

$$COV(x_{t+1}, y_{t+1} | \mathbf{y_t}, \gamma_\mathbf{t+1}) =$$
$$= \begin{pmatrix} P_{t+1|t} & P_{t+1|t}C' \\ CP_{t+1|t} & CP_{t+1|t}C' + \gamma_{t+1}R + (1 - \gamma_{t+1})\sigma^2 I \end{pmatrix}.$$

Hence, the Kalman filter equations are modified as follows:

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} \tag{6.9}$$

$$P_{t+1|t} = AP_{t|t}A' + Q \tag{6.10}$$

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + P_{t+1|t}C'(CP_{t+1|t}C' + \gamma_{t+1}R + (1 - \gamma_{t+1})\sigma^2 I)^{-1}(y_{t+1} - C\hat{x}_{t+1|t}) \tag{6.11}$$

$$P_{t+1|t+1} = P_{t+1|t} - P_{t+1|t}C'(CP_{t+1|t}C' + \gamma_{t+1}R + (1 - \gamma_{t+1})\sigma^2 I)^{-1}CP_{t+1|t}. \tag{6.12}$$

Taking the limit as $\sigma \to \infty$, the update equations (6.11) and (6.12) can be rewritten as follows:

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + \gamma_{t+1}K_{t+1}(y_{t+1} - C\hat{x}_{t+1|t}) \tag{6.13}$$

$$P_{t+1|t+1} = P_{t+1|t} - \gamma_{t+1}K_{t+1}CP_{t+1|t}, \tag{6.14}$$

where $K_{t+1} = P_{t+1|t}C'(CP_{t+1|t}C' + R)^{-1}$ is the Kalman gain matrix for the standard ARE. Note that performing this limit corresponds *exactly* to propagating the previous state when there is no observation update available

at time t. It is important to point out the main difference from the standard Kalman filter formulation: both $\hat{x}_{t+1|t+1}$ and $P_{t+1|t+1}$ are now random variables, being a function of γ_{t+1}, which is itself random.

Equations (6.13)-(6.66) give the minimum state error variance filter given the observations $\{y_t\}$ and their arrival sequence$\{\gamma_t\}$, i.e. $\hat{x}_t^{tm} = \mathbb{E}[x_t|y_t, \ldots, y_1, \gamma_t, \ldots, \gamma_1]$. As a consequence, the filter proposed in this paper is necessarily time-varying and stochastic since it depends on the arrival sequence. The filters that have been proposed so far using JLS theory [?, ?] give the minimum state error variance filters assuming that only the observations $\{y_t\}$ and the knowledge on the last arrival γ_t are available, i.e. $\hat{x}_t^{JLS} = \mathbb{E}[x_t|y_t, \ldots, y_1, \gamma_t]$. Therefore, the filter given by Equations (6.13)-(6.66) gives a better performance than its JLS counterparts, since it exploits additional information regarding the arrival sequence.

Given the new formulation, we now study the Riccati equation of the state error covariance matrix in the specific case when the arrival process of the observation is time-independent, i.e. $\bar{\gamma}_t = \bar{\gamma}$ for all time. This will allow us to provide deterministic upper and lower bounds on its expectation. We then characterize the convergence of these upper and lower bounds, as a function of the arrival probability $\bar{\gamma}$ of the observation.

6.4 Convergence conditions and transition to instability

It is easy to verify that the modified Kalman filter formulation in Equations (6.10) and (6.66) can be rewritten as follows:

$$P_{t+1} = AP_tA' + Q - \gamma_t \, AP_tC'(CP_tC' + R)^{-1}CP_tA', \qquad (6.15)$$

where we use the simplified notation $P_t = P_{t|t-1}$. Since the sequence $\{\gamma_t\}_0^\infty$ is random, the modified Kalman filter iteration is stochastic and cannot be determined off-line. Therefore, only statistical properties can be deduced.

In this section we show the existence of a critical value γ_c for the arrival probability of the observation update, such that for $\bar{\gamma} > \gamma_c$ the mean state covariance $\mathbb{E}[P_t]$ is bounded for all initial conditions, and for $\bar{\gamma} \leq \gamma_c$ the mean state covariance diverges for some initial condition. We also find a lower bound γ_{min}, and upper bound γ_{max}, for the critical probability γ_c, i.e., $\gamma_{min} \leq \gamma_c \leq \gamma_{max}$. The lower bound is expressed in closed form; the upper bound is the solution of a linear matrix inequality (LMI). In some special cases the two bounds coincide, giving a tight estimate. Finally, we present numerical algorithms to compute a lower bound $\bar{S}$, and upper bound $\bar{V}$, for $\lim_{t\to\infty} \mathbb{E}[P_t]$, when it is bounded.

First, we define the modified algebraic Riccati equation (MARE) for the Kalman filter with intermittent observations as follows,

$$g_{\bar{\gamma}}(X) = AXA' + Q - \bar{\gamma} \, AXC'(CXC' + R)^{-1}CXA'. \qquad (6.16)$$

Our results derive from two principal facts: the first is that concavity of the modified algebraic Riccati equation for our filter with intermittent observations allows use of Jensen's inequality to find an upper bound on the mean state covariance; the second is that all the operators we use to estimate upper and lower bounds are monotonically increasing, therefore if a fixed point exists, it is also stable.

We formally state all main results in form of theorems. Omitted proofs appear in the Appendix. The first theorem expresses convergence properties of the MARE.

Theorem 6.1. *Consider the operator*
$\phi(K, X) = (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}(FXF' + V)$, *where* $F = A + KC$, $V = Q + KRK'$. *Suppose there exists a matrix* $\tilde{K}$ *and a positive definite matrix* $\tilde{P}$ *such that*
$$\tilde{P} > 0 \quad and \quad \tilde{P} > \phi(\tilde{K}, \tilde{P})$$

Then,

- *for any initial condition* $P_0 \geq 0$, *the MARE converges, and the limit is independent of the initial condition:*
$$\lim_{t \to \infty} P_t = \lim_{t \to \infty} g_{\bar{\gamma}}^t(P_0) = \overline{P}$$

- $\overline{P}$ *is the unique positive semidefinite fixed point of the MARE.*

The next theorem states the existence of a sharp transition.

Theorem 6.2. *If* $(A, Q^{\frac{1}{2}})$ *is controllable,* (A, C) *is detectable, and* A *is unstable, then there exists a* $\gamma_c \in [0, 1)$ *such that*

$$\lim_{t \to \infty} \mathbb{E}[P_t] = +\infty \ for \ 0 \leq \bar{\gamma} \leq \gamma_c \quad and \quad \exists P_0 \geq 0 \qquad (6.17)$$
$$\mathbb{E}[P_t] \leq M_{P_0} \ \forall t \ for \ \gamma_c < \bar{\gamma} \leq 1 \quad and \quad \forall P_0 \geq 0 \qquad (6.18)$$

where $M_{P_0} > 0$ *depends on the initial condition* $P_0 \geq 0$[1].

The next theorem gives upper and lower bounds for the critical probability γ_c.

Theorem 6.3. *Let*

$$\gamma_{min} = \inf[\bar{\gamma} : \exists \hat{S} \mid \hat{S} = (1 - \bar{\gamma})A\hat{S}A' + Q] = 1 - \frac{1}{\alpha^2} \qquad (6.19)$$
$$\gamma_{max} = \inf[\bar{\gamma} : \exists(\hat{K}, \hat{X}) \mid \hat{X} > \phi(\hat{K}, \hat{X})] \qquad (6.20)$$

where $\alpha = \max_i |\sigma_i|$ *and* σ_i *are the eigenvalues of* A. *Then*

$$\gamma_{min} \leq \gamma_c \leq \gamma_{max}. \qquad (6.21)$$

---

[1]We use the notation $\lim_{t \to \infty} A_t = +\infty$ when the sequence $A_t \geq 0$ is not bounded; i.e., there is no matrix $M \geq 0$ such that $A_t \leq M, \forall t$.

Finally, the following theorem gives an estimate of the limit of the mean covariance matrix $\mathbb{E}[P_t]$, when this is bounded.

Theorem 6.4. *Assume that $(A, Q^{\frac{1}{2}})$ is controllable, (A, C) is detectable and $\bar{\gamma} > \gamma_{max}$, where γ_{max} is defined in Theorem 6.3. Then*

$$0 < S_t \leq \mathbb{E}[P_t] \leq V_t \quad \forall \, \mathbb{E}[P_0] \geq 0 \tag{6.22}$$

where $\lim_{t \to \infty} S_t = \bar{S}$ and $\lim_{t \to \infty} V_t = \bar{V}$, where $\bar{S}$ and $\bar{V}$ are solution of the respective algebraic equations
$\bar{S} = (1 - \bar{\gamma}) A \bar{S} A' + Q$ *and* $\bar{V} = g_{\bar{\gamma}}(\bar{V})$.

The previous theorems give lower and upper bounds for both the critical probability γ_c and for the mean error covariance $\mathbb{E}[P_t]$. The lower bound γ_{min} is expressed in closed form. We now resort to numerical algorithms for the computation of the remaining bounds $\gamma_{max}, \bar{S}$ and $\bar{V}$.

The computation of the upper bound γ_{max} can be reformulated as the iteration of an LMI feasibility problem. To establish this we need the following theorem:

Theorem 6.5. *If $(A, Q^{\frac{1}{2}})$ is controllable and (A, C) is detectable, then the following statements are equivalent:*

- $\exists \bar{X} \quad$ *such that* $\quad \bar{X} > g_{\bar{\gamma}}(\bar{X})$
- $\exists \bar{K}, \bar{X} > 0 \quad$ *such that* $\quad \bar{X} > \phi(\bar{K}, \bar{X})$
- $\exists \bar{Z}$ *and* $0 < \bar{Y} \leq I$ *such that*

$$\Psi_{\bar{\gamma}}(Y, Z) =$$
$$\begin{bmatrix} Y & \sqrt{\bar{\gamma}}(YA + ZC) & \sqrt{1 - \bar{\gamma}} Y A \\ \sqrt{\bar{\gamma}}(A'Y + C'Z') & Y & 0 \\ \sqrt{1 - \bar{\gamma}} A'Y & 0 & Y \end{bmatrix} > 0.$$

Proof. (a)$\Longrightarrow$(b) If $\bar{X} > g_{\bar{\gamma}}(\bar{X})$ exists, then $\bar{X} > 0$ by Lemma 6.1(g). Let $\bar{K} = K_{\bar{X}}$. Then $\bar{X} > g_{\bar{\gamma}}(\bar{X}) = \phi(\bar{K}, \bar{X})$ which proves the statement.
(b)$\Longrightarrow$(a) Clearly $\bar{X} > \phi(\bar{K}, \bar{X}) \geq g_{\bar{\gamma}}(\bar{X})$ which proves the statement.
(b)$\Longleftrightarrow$(c) Let $F = A + KC$, then:

$$X > (1 - \bar{\gamma}) A X A' + \bar{\gamma} F X F' + Q + \bar{\gamma} K R K'$$

is equivalent to

$$\begin{bmatrix} X - (1 - \bar{\gamma}) A X A' & \sqrt{\bar{\gamma}} F \\ \sqrt{\bar{\gamma}} F' & X^{-1} \end{bmatrix} > 0,$$

where we used the Schur complement decomposition and the fact that $X - (1 - \bar{\gamma}) A X A' \geq \bar{\gamma} F X F' + Q + \bar{\gamma} K R K' \geq Q > 0$. Using one more time the Schur complement decomposition on the first element of the matrix we obtain

$$\Theta = \begin{bmatrix} X & \sqrt{\bar{\gamma}} F & \sqrt{1 - \bar{\gamma}} A \\ \sqrt{\bar{\gamma}} F' & X^{-1} & 0 \\ \sqrt{1 - \bar{\gamma}} A' & 0 & X^{-1} \end{bmatrix} > 0.$$

This is equivalent to

$$\bar{\gamma} = \begin{bmatrix} X^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \Theta \begin{bmatrix} X^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} > 0$$

$$= \begin{bmatrix} X^{-1} & \sqrt{\bar{\gamma}}X^{-1}F & \sqrt{1-\bar{\gamma}}X^{-1}A \\ \sqrt{\bar{\gamma}}F'X^{-1} & X^{-1} & 0 \\ \sqrt{1-\bar{\gamma}}A'X^{-1} & 0 & X^{-1} \end{bmatrix} > 0.$$

Let us consider the change of variable $Y = X^{-1} > 0$ and $Z = X^{-1}K$, in which case the previous LMI is equivalent to:

$$\Psi(Y, Z) =$$

$$= \begin{bmatrix} Y & \sqrt{\bar{\gamma}}(YA + ZC) & \sqrt{1-\bar{\gamma}}YA \\ \sqrt{\bar{\gamma}}(A'Y + C'Z') & Y & 0 \\ \sqrt{1-\bar{\gamma}}A'Y & 0 & Y \end{bmatrix} > 0.$$

Since $\Psi(\alpha Y, \alpha K) = \alpha\Psi(Y, K)$, then Y can be restricted to $Y \leq I$, which completes the theorem. □

Combining theorems 6.3 and 6.5 we immediately have the following corollary

Corollary 6.5.1. *The upper bound γ_{max} is given by the solution of the following optimization problem,*

$$\gamma_{max} = \text{argmin}_{\bar{\gamma}} \Psi_{\bar{\gamma}}(Y, Z) > 0, \quad 0 \leq Y \leq I.$$

This is a quasi-convex optimization problem in the variables $(\bar{\gamma}, Y, Z)$ and the solution can be obtained by iterating LMI feasibility problems and using bisection for the variable $\bar{\gamma}$, as shown in [?].

The lower bound $\bar{S}$ for the mean covariance matrix can be easily obtained via standard Lyapunov Equation solvers. The upper bound $\bar{V}$ can be found by iterating the MARE or by solving a semidefinite programming (SDP) problem as shown in the following theorem.

Theorem 6.6. *If $\bar{\gamma} > \gamma_{max}$, then the matrix $\bar{V} = g_{\bar{\gamma}}(\bar{V})$ is given by:*

1. $\bar{V} = \lim_{t \to \infty} V_t$; $V_{t+1} = g_{\bar{\gamma}}(V_t)$ where $V_0 \geq 0$

2.

$$\begin{array}{ll} \text{argmax}_V & Trace(V) \\ \text{subject to} & \begin{bmatrix} AVA' - V + Q & \sqrt{\bar{\gamma}}AVC' \\ \sqrt{\bar{\gamma}}CVA' & CVC' + R \end{bmatrix} \geq 0, \quad V \geq 0 \end{array}$$

Proof. (a) It follows directly from Theorem 6.1.

(b) It can be obtained by using the Schur complement decomposition on the equation $V \leq g_{\bar{\gamma}}(V)$. Clearly the solution $\bar{V} = g_{\bar{\gamma}}(\bar{V})$ belongs to the

feasible set of the optimization problem. We now show that the solution of the optimization problem is the fixed point of the MARE. Suppose it is not, i.e., $\hat{V}$ solves the optimization problem but $\hat{V} \neq g_{\bar{\gamma}}(\hat{V})$. Since $\hat{V}$ is a feasible point of the optimization problem, then $\hat{V} < g_{\bar{\gamma}}(\hat{V}) = \hat{\hat{V}}$. However, this implies that $Trace(\hat{V}) < Trace(\hat{\hat{V}})$, which contradicts the hypothesis of optimality of matrix $\hat{V}$. Therefore $\hat{V} = g_{\bar{\gamma}}(\hat{V})$ and this concludes the theorem. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

6.5 Special Cases and Examples

In this section we present some special cases in which upper and lower bounds on the critical value γ_c coincide, and give some examples. From Theorem 6.1, it follows that if there exists a $\tilde{K}$ such that F is the zero matrix, then the convergence condition of the MARE is for $\bar{\gamma} > \gamma_c = 1 - 1/\alpha^2$, where $\alpha = \max_i |\sigma_i|$, and σ_i are the eigenvalues of A.

- **C is invertible**. In this case a choice of $K = -AC^{-1}$ makes $F = 0$. Note that the scalar case also falls under this category. Figure (2) shows a plot of the steady state of the upper and lower bounds versus $\bar{\gamma}$ in the scalar case. The discrete time LTI system used in this simulation has $A = -1.25$, $C = 1$, with v_t and w_t having zero mean and variance $R = 2.5$ and $Q = 1$, respectively. For this system we have $\gamma_c = 0.36$. The transition clearly appears in the figure, where we see that the steady state value of both upper and lower bound tends to infinity as $\bar{\gamma}$ approaches γ_c. The dashed line shows the lower bound, the solid line the upper bound, and the dash-dot line shows the asymptote.

- **A has a single unstable eigenvalue**. In this case, regardless of the dimension of C (and as long as the pair (A, C) is detectable), we can use Kalman decomposition to bring to zero the unstable part of F and thereby obtain tight bounds. Figure (3) shows a plot for the system
$$A = \begin{pmatrix} 1.25 & 1 & 0 \\ 0 & 0.9 & 7 \\ 0 & 0 & 0.6 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 2 \end{pmatrix}$$
with v_t and w_t having zero mean and variance $R = 2.5$ and $Q = 20 \cdot I_{3 \times 3}$, respectively. This time, the asymptotic value for trace of upper and lower bound is plotted versus $\bar{\gamma}$. Once again $\gamma_c = 0.36$.

In general F cannot always be made zero and we have shown that while a lower bound on γ_c can be written in closed form, an upper bound on γ_c is the result of a LMI. Figure (4) shows an example where upper and lower bounds have different convergence conditions. The system used for this simulation is $A = \begin{pmatrix} 1.25 & 0 \\ 1 & 1.1 \end{pmatrix}, C = \begin{pmatrix} 1 & 1 \end{pmatrix}$
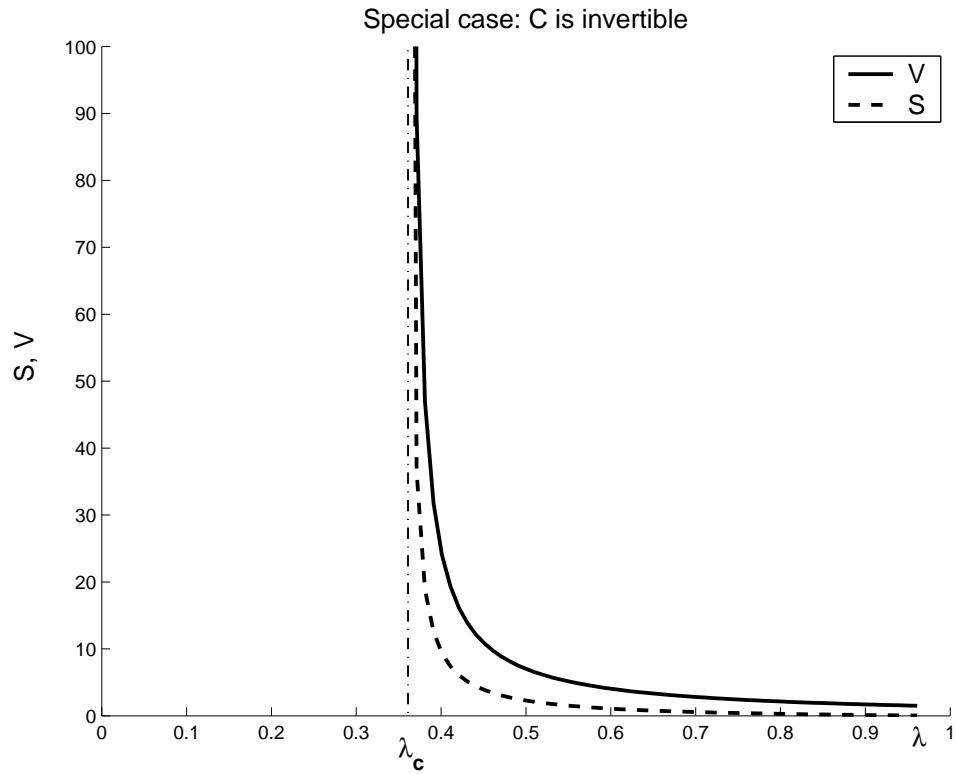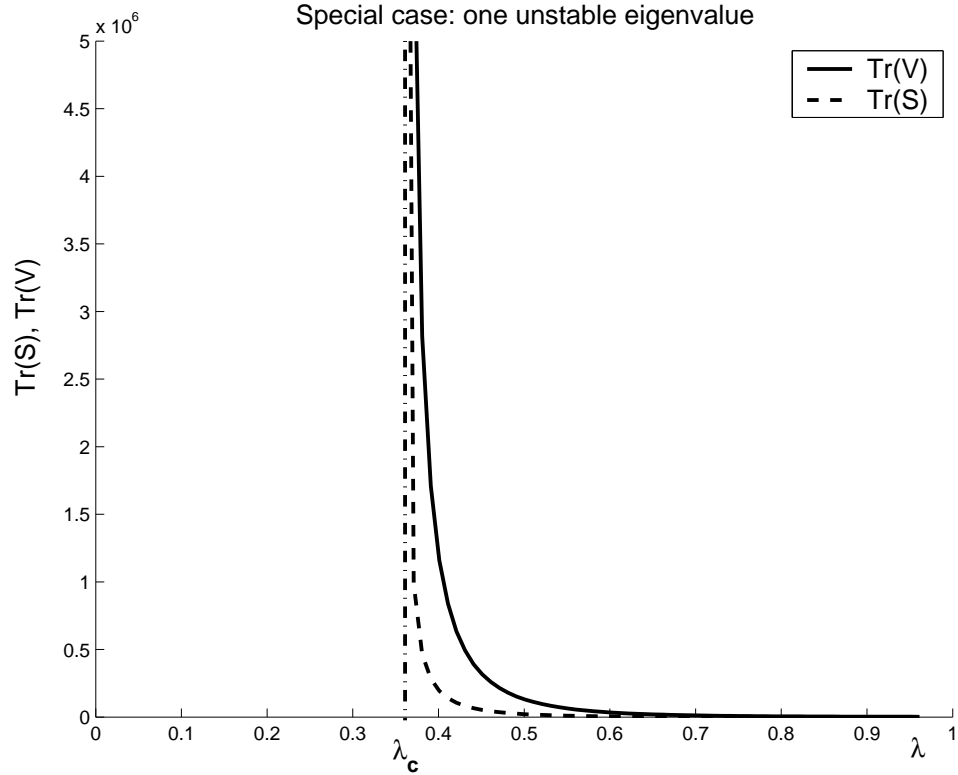
Figure 6.2: Example of transition to instability in the scalar case. The dashed line shows the asymptotic value of the lower bound $(\bar{S})$, the solid line the asymptotic value of the upper bound $(\bar{V})$, and the dash-dot line shows the asymptote (γ_c).

with v_t and w_t having zero mean and variance $R = 2.5$ and $Q = 20 \cdot I_{2\times2}$, respectively.

Finally, in Figure (5) we report results of another experiment, plotting the state estimation error for the scalar system used above at two similar values of $\bar{\gamma}$, one being below and one above the critical value. We note a dramatic change in the error at $\gamma_c \approx 0.36$. The figure on the left shows the estimation error with $\bar{\gamma} = 0.3$. The figure on the right shows the estimation error for the same system evolution with $\bar{\gamma} = 0.4$. In the first case the estimation error grows dramatically, making it practically useless for control purposes. In the second case, a small increase in $\bar{\gamma}$ reduces the estimation error by approximately three orders of magnitude.

Figure 6.3: Example of transition to instability with a single unstable eigenvalue in the MIMO case. The dashed line shows the asymptotic value of the trace of lower bound $(\bar{S})$, the solid line the asymptotic value of trace of the upper bound $(\bar{V})$, and the dash-dot line shows the asymptote (γ_c).

6.6 Static versus dynamic Kalman gain

In this section we compare the performance of filtering with static and dynamic gain for a scalar discrete system. For the static estimator we follow the jump linear system approach of [**?**]. The scalar static estimator case has been also worked out in [**?**].

Consider the dynamic state estimator

$$\hat{x}^d_{t+1} = A\hat{x}^d_t + \gamma_t K^d_t(y_t - \hat{y}_t)$$
$$K^d_t = AP_tC'(CP_tC' + R)^{-1}$$
$$P_{t+1} = AP_tA' + Q - \gamma_t K^d_t CP_tA' \tag{6.23}$$

where the Kalman gain K^d_t is time-varying. Also consider the static state estimator

$$\hat{x}^s_{t+1} = A\hat{x}^d_t + \gamma_t K_s(y_t - \hat{y}_t) \tag{6.24}$$
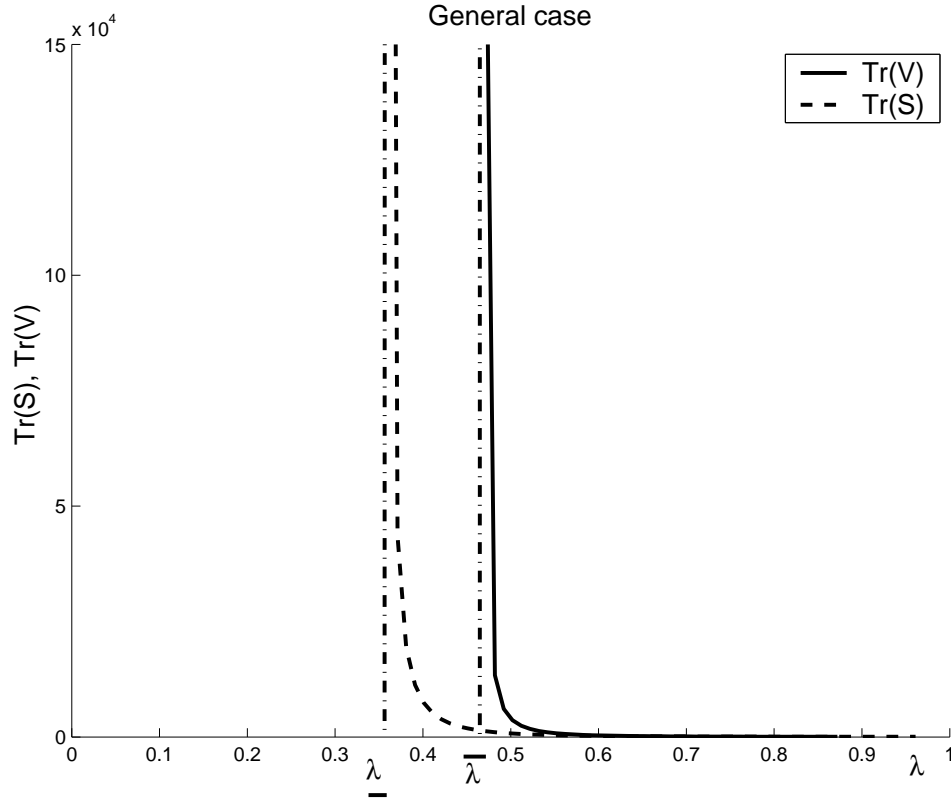
Figure 6.4: Transition to instability in the general case, with arbitrary A and C. In this case lower and upper bounds do not have the same asymptote.

where the estimator gain K_s is constant. If no data arrives, i.e. $\gamma_t = 0$, both estimators simply propagate the state estimate of the previous time-step.

The performance of the dynamic state estimator (6.23) has been analyzed in the previous sections. The performance of static state estimator (6.24), instead, can be readily obtained using jump linear system theory [**?**, **?**]. To do so, let us consider the estimator error $e_{t+1}^s \overset{\Delta}{=} x_{t+1} - \hat{x}_{t+1}^s$. Substituting Equations (6.1) for x_{t+1} and (6.24) for $\hat{x}_{t+1}^s$, we obtain the dynamics of the estimation error:

$$e_{t+1}^s = (A - \gamma_t K_s C)e_t^s + v_t + \gamma_t K_s w_t. \tag{6.25}$$

Using the same notation of Chapter 6 in Nilsson [**?**], where he considers the general system:

$$z_{k+1} = \Phi(r_k)z_k + \Gamma(r_k)e_k,$$

the system (6.25) can be seen as jump linear system switching between two
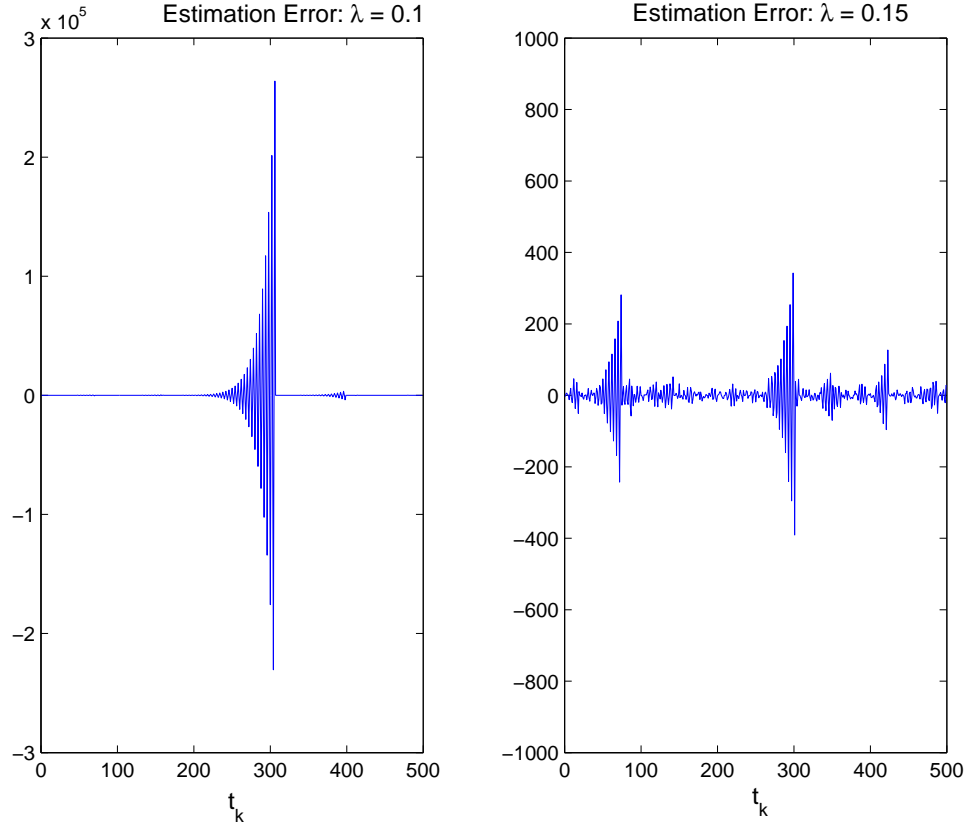
Figure 6.5: Estimation error for $\bar{\gamma}$ below (left) and above (right) the critical value

states $r_k \in \{1, 2\}$ given by:

$$\Phi(1) = A - K_s C \qquad \Gamma(1) = [1 \ K_s]$$
$$\Phi(2) = A \qquad\qquad \Gamma(2) = [1 \ 0],$$

where the noise covariance $\mathbb{E}[e_k e'_k] = R_e$, the transition probability matrix Q_π and the steady state probability distribution π^∞ are given by:

$$R_e = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}, \quad Q_\pi = \begin{bmatrix} \bar{\gamma} & 1 - \bar{\gamma} \\ \bar{\gamma} & 1 - \bar{\gamma} \end{bmatrix}, \quad \pi^\infty = \begin{bmatrix} \bar{\gamma} & 1 - \bar{\gamma} \end{bmatrix}.$$

Following the methodology proposed in Nilsson [**?**] is possible to show that the system above is mean square stable, i.e. $\lim_{t \to \infty} \mathbb{E}[(e_t^s)' e_t^s] = 0$ if and only if the transition probability is

$$\bar{\gamma} < \bar{\gamma}_s = \frac{1}{1 - \left(1 - \frac{K_s C}{A}\right)^2} \left(1 - \frac{1}{A^2}\right). \tag{6.26}$$

If the system is mean square stable, the steady state error covariance $P^s_\infty = \lim_{t\to\infty} \mathbb{E}[e^s_t(e^s_t)']$ is given by:

$$P^s_\infty = \frac{Q + K^2_s R}{1 - \bar{\gamma}(A - K_s C)^2 - (1 - \bar{\gamma})A^2}. \tag{6.27}$$

Calculations to obtain Equations (6.26) and (6.27) are tedious but straightforward, therefore they are omitted.

It is immediately evident that the critical transition probability $\bar{\gamma}_s$ of the estimator (6.24) using a static gain is always greater than the critical transition probability γ_c of the estimator (6.23) which adopts a dynamic gain, in fact

$$\bar{\gamma}_s = \gamma_c \frac{1}{1 - \left(1 - \frac{K_s C}{A}\right)^2}$$

and the two probabilities are equal only when $K_s = \frac{A}{C}$.

A natural choice for the static estimator gain K_s is the steady state Kalman gain K_{SS} of the closed loop system ($r = 1$), which is always different from $\frac{A}{C}$. For the scalar system considered in the previous section, where $A = -1.5$, $C = 1$, $Q = 1$, $R = 2.5$, this is given by $K_{SS} = -0.70$, while the gain for largest mean square stability range is $K_s = \frac{A}{C} = -1.25$. In the special case when the arrival probability is known, another natural choice for the estimator gain K is obtained by substituting the error covariance solution of $\bar{P} = g_{\bar{\gamma}}(\bar{P})$ into the equation for the Kalman filter gain $K_{\bar{\gamma}} = A\bar{P}C'(C\bar{P}C' + R)^{-1}$. For example, assuming $\bar{\gamma} = 0.6$, then $\bar{P} = 7.32$ and $K_{\bar{\gamma}} = -0.93$. Figure 6.6 shows all of these cases, comparing them with the upper bound on the state error covariance $\bar{V}$ of the dynamic estimator (6.23) that can be computed as indicated in Theorem 6. The steady state error covariance of the static predictor for the three different gains is always greater then our upper bound $\bar{V}$. This is not surprising, since the dynamic estimator is optimal over all possible estimators as shown in Section II. Note that the static predictor with static gain $K_{\bar{\gamma}}$ (designed for $\bar{\gamma} = 0.6$) achieves the same state error covariance predicted by our upper bound for the optimal dynamic filter when $\bar{\gamma} = 0.6$. However, the empirical error state covariance is on average better than the static filter, as shown in Figure 6.7. This is to be expected, since the solution of MARE gives only an upper bound of the true expected state covariance of the time-varying filter. Moreover, it is worth stressing that if the arrival probability is different from the one used to design the static gain, the performance of the static filter will degrade considerably, while the time-varying filter will still perform optimally since it does not require knowledge of $\bar{\gamma}$. From this example, it seems that the upper bound for the dynamic estimator $\bar{V}$ gives en estimate of the minimum steady state covariance that can be achieved with a static estimator for any given arrival probability if the static gain K_s is chosen optimally. Then the MARE could be used to find the minimum steady state covariance and
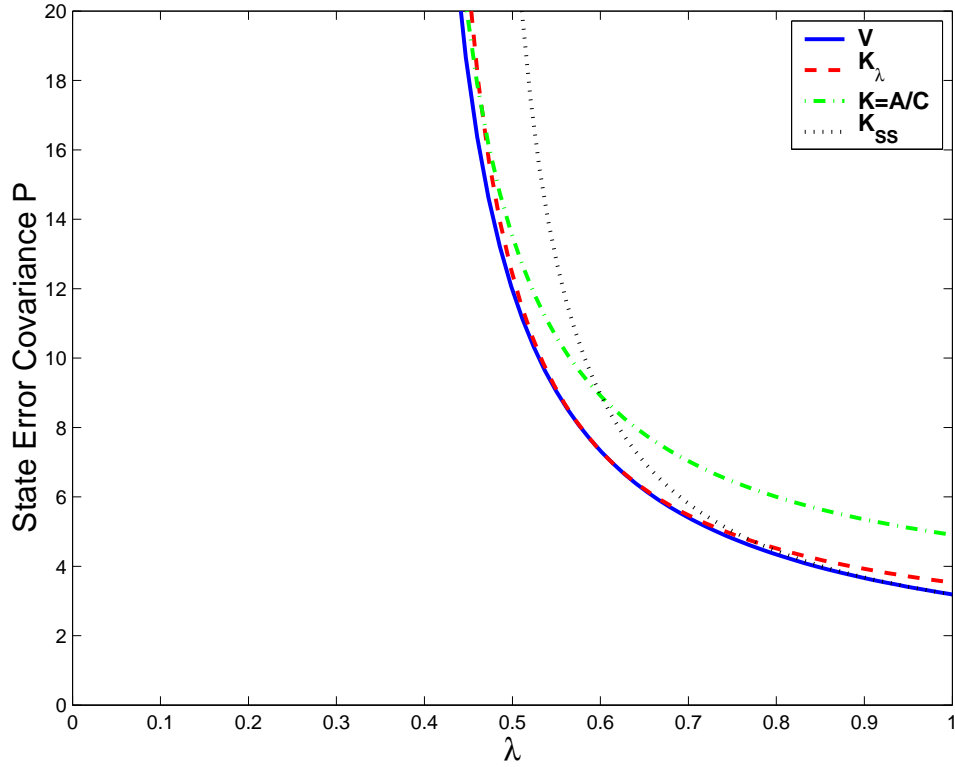
Figure 6.6: Error covariance bound $\bar{V}$ for dynamic predictor obtained from our theory and steady state error covariance for three natural static predictors obtained from JLS theory.

then the corresponding steady state modified Kalman gain, thus providing a useful tool for optimal static estimator design. Future work will explore this possibility.

6.7 Appendix A

In order to give complete proofs of our main theorems, we need to prove some preliminary lemmas. The first one shows some useful properties of the MARE.

Lemma 6.1. *Let the operator*

$$\phi(K, X) = (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}(FXF' + V) \qquad (6.28)$$

where $F = A + KC$, $V = Q + KRK'$. *Assume* $X \in \mathbb{S} = \{S \in \mathbb{R}^{n \times n} | S \geq 0\}$, $R > 0$, $Q \geq 0$, *and* $(A, Q^{\frac{1}{2}})$ *is controllable. Then the following facts are true:*

1. *With* $K_X = -AXC'(CXC' + R)^{-1}$, $g_{\bar{\gamma}}(X) = \phi(K_X, X)$
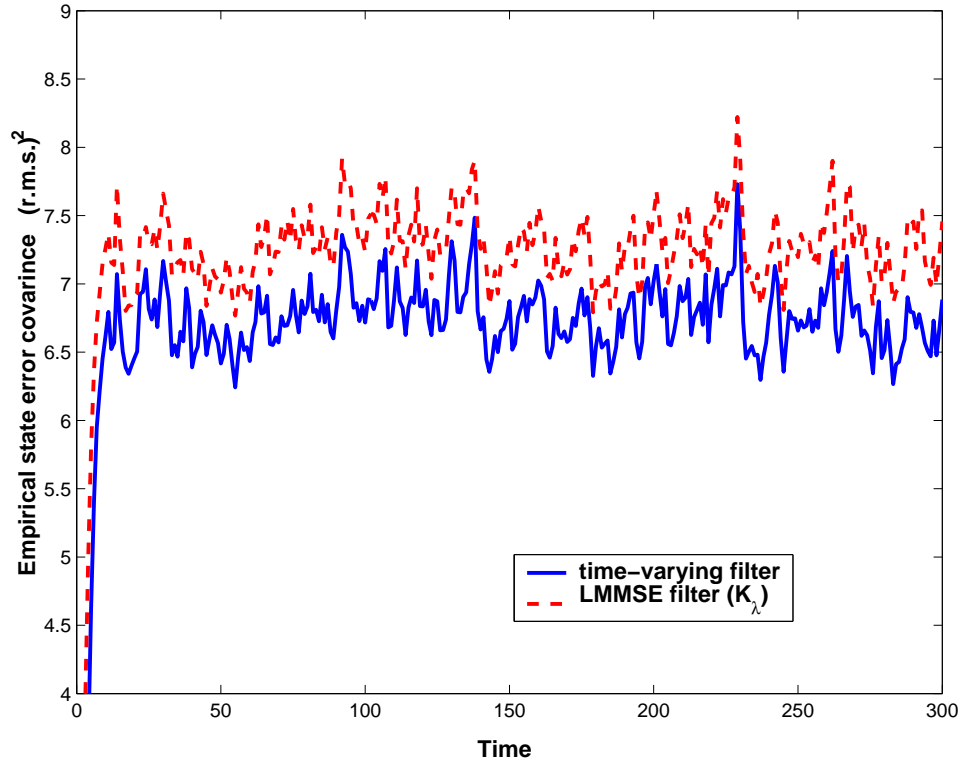
Figure 6.7: Empirical state error covariance of our time-varying filter and the linear minimum mean square error estimator (LMMSEE) [**?**] obtained by using the optimal static kalman gain $K_{\bar{\gamma}}$. The curves are obtained by averaging 10000 Monte Carlo simulations for $t = 1, \ldots, 300$, with the values of the input noise (v_t, w_t) and the arrival sequence γ_t generated randomly. Both filters were compared under the same conditions.

2. $g_{\bar{\gamma}}(X) = \min_K \phi(K, X) \leq \phi(K, X), \; \forall K$

3. If $X \leq Y$, then $g_{\bar{\gamma}}(X) \leq g_{\bar{\gamma}}(Y)$

4. If $\bar{\gamma}_1 \leq \bar{\gamma}_2$ then $g_{\bar{\gamma}_1}(X) \geq g_{\bar{\gamma}_2}(X)$

5. If $\alpha \in [0, 1]$, then
   $g_{\bar{\gamma}}(\alpha X + (1 - \alpha)Y) \geq \alpha g_{\bar{\gamma}}(X) + (1 - \alpha)g_{\bar{\gamma}}(Y)$

6. $g_{\bar{\gamma}}(X) \geq (1 - \bar{\gamma})AXA' + Q$

7. If $\bar{X} \geq g_{\bar{\gamma}}(\bar{X})$, then $\bar{X} > 0$

8. If X is a random variable, then
   $(1 - \bar{\gamma})A\mathbb{E}[X]A' + Q \leq \mathbb{E}[g_{\bar{\gamma}}(X)] \leq g_{\bar{\gamma}}(\mathbb{E}[X])$

Proof. (a) Define $F_X = A + K_X C$, and observe that

$$
\begin{aligned}
F_X X C' + K_X R &= (A + K_X C) X C' + K_X R \\
&= AXC' + K_X(CXC' + R) \\
&= 0.
\end{aligned}
$$

Next, we have

$$
\begin{aligned}
g_{\bar{\gamma}}(X) &= (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}(AXA' + Q - AXC'(CXC' + R)^{-1} CXA') \\
&= (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}(AXA' + Q + K_X CXA') \\
&= (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}(F_X XA' + Q) \\
&= (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}(F_X XA' + Q) + (F_X XC' + K_X R)K' \\
&= \phi(K_X, X)
\end{aligned}
$$

(b) Let $\psi(K, X) = (A + KC)X(A + KC)' + KRK' + Q$. Note that

$$
\text{argmin}_K \phi(K, X) = \text{argmin}_K FXF' + V = \text{argmin}_K \psi(X, K).
$$

Since $X, R \geq 0$, $\phi(K, X)$ is quadratic and convex in the variable K, therefore the minimizer can be found by solving $\frac{\partial \psi(K,X)}{\partial K} = 0$, which gives:

$$
2(A + KC)XC' + 2KR = 0 \Longrightarrow K = -AXC'(CXC' + R)^{-1}.
$$

Since the minimizer corresponds to K_X defined above, the fact follows from fact (1)

(c) Note that $\phi(K, X)$ is affine in X. Suppose $X \leq Y$. Then

$$
g_{\bar{\gamma}}(X) = \phi(K_X, X) \leq \phi(K_Y, X) \leq \phi(K_Y, Y) = g_{\bar{\gamma}}(Y).
$$

This completes the proof.

(d) Note that $AXC'(CXC' + R)^{-1}CXA \geq 0$. Then

$$
\begin{aligned}
g_{\bar{\gamma}_1}(X) &= AXA' + Q - \bar{\gamma}_1 AXC'(CXC' + R)^{-1}CXA \\
&\geq AXA' + Q - \bar{\gamma}_2 AXC'(CXC' + R)^{-1}CXA \\
&= g_{\bar{\gamma}_2}(X)
\end{aligned}
$$

(e) Let $Z = \alpha X + (1 - \alpha)Y$ where $\alpha \in [0, 1]$. Then we have

$$
\begin{aligned}
g_{\bar{\gamma}}(Z) &= \phi(K_Z, Z) \\
&= \alpha(A + K_Z C)X(A + K_Z C)' + (1 - \alpha)(A + K_Z C)Y(A + K_Z C)' + \\
&\quad + (\alpha + 1 - \alpha)(K_Z R K_Z' + Q) \\
&= \alpha\phi(K_Z, X) + (1 - \alpha)\phi(K_Z, Y) \\
&\geq \alpha\phi(K_X, X) + (1 - \alpha)\phi(K_Y, Y) \\
&= \alpha g_{\bar{\gamma}}(X) + (1 - \alpha)g_{\bar{\gamma}}(Y).
\end{aligned}
$$

(f) Note that $F_X X F_X' \geq 0$ and $KRK' \geq 0$ for all K and X. Then

$$g_{\bar{\gamma}}(X) = \phi(K_X, X) =$$
$$= (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}(F_X X F_X' + K_X RK_X' + Q)$$
$$\geq (1 - \bar{\gamma})(AXA' + Q) + \bar{\gamma}Q = (1 - \bar{\gamma})AXA' + Q.$$

(g) First observe that $\bar{X} \geq g_\lambda(\bar{X}) \geq 0$. Thus, to prove that $\bar{X} > 0$, we only need to establish that $\bar{X}$ is nonsingular. Suppose $0 \neq v \in \mathcal{N}(\bar{X})$, i.e. $\bar{X}v = 0$. Then

$$0 = v'\bar{X}v \geq v'g_\lambda(\bar{X})v$$
$$= (1 - \lambda)v'(A\bar{X}A' + Q)v + \lambda v'(F\bar{X}F' + Q)v$$

Positive semi-definiteness of $\bar{X}$ and Q implies that all the terms in the sum must be zero for the inequality to hold. Consequently we have

$$v'A\bar{X}A'v = 0 \implies \bar{X}A'v = 0 \implies A'v \in \mathcal{N}(\bar{X})$$

and

$$v'Qv = 0 \implies Qv = 0$$

As a result, the null space $\mathcal{N}(\bar{X})$ is A'-invariant. Therefore, $\mathcal{N}(\bar{X})$ contains an eigenvector of A', i.e. there exists $u \neq 0$ such that $\bar{X}u = 0$ and $A'u = \sigma u$. As before, we conclude that $Qu=0$. This implies (using the Popov-Belevich-Hautus(PBH) test) that the pair $(A, Q^{1/2})$ is not controllable, contradicting the hypothesis. Thus, $\mathcal{N}(\bar{X})$ is empty, proving that $\bar{X} > 0$.

(h) Using fact (f) and linearity of expectation we have

$$\mathbb{E}[g_{\bar{\gamma}}(X)] \geq \mathbb{E}[(1 - \bar{\gamma})AXA' + Q] = (1 - \bar{\gamma})A\mathbb{E}[X]A' + Q.$$

Fact (e) implies that the operator $g_{\bar{\gamma}}()$ is concave, therefore by Jensen's Inequality we have:

$$\mathbb{E}[g_{\bar{\gamma}}(X)] \leq g_{\bar{\gamma}}(\mathbb{E}[X]).$$

$\square$

Lemma 6.2. *Let $X_{t+1} = h(X_t)$ and $Y_{t+1} = h(Y_t)$. If $h(X)$ is a monotonically increasing function then:*

$$\begin{aligned}
X_1 \geq X_0 &\implies X_{t+1} \geq X_t, \quad \forall t \geq 0 \\
X_1 \leq X_0 &\implies X_{t+1} \leq X_t, \quad \forall t \geq 0 \\
X_0 \leq Y_0 &\implies X_t \leq Y_t, \quad \forall t \geq 0
\end{aligned}$$

Proof. This lemma can be readily proved by induction. It is true for $t = 0$, since $X_1 \geq X_0$ by definition. Now assume that $X_{t+1} \geq X_t$, then $X_{t+2} = h(X_{t+1}) \geq h(X_t) = X_{t+1}$ because of monotonicity of $h(\cdot)$. The proof for the other two cases is analogous. $\square$

It is important to note that while in the scalar case $X \in \mathbb{R}$ either $h(X) \leq X$ or $h(X) \geq X$; in the matrix case $X \in \mathbb{R}^{n \times n}$, it is not generally true that either $h(X) \geq X$ or $h(X) \leq X$. This is the source of the major technical

difficulty for the proof of convergence of sequences in higher dimensions. In this case convergence of a sequence $\{X_t\}_0^\infty$ is obtained by finding two other sequences, $\{Y_t\}_0^\infty, \{Z_t\}_0^\infty$ that bound X_t, i.e., $Y_t \leq X_t \leq Z_t, \forall t$, and then by showing that these two sequences converge to the same point.

The next two Lemmas show that when the MARE has a solution $\bar{P}$, this solution is also stable, i.e., every sequence based on the difference Riccati equation $P_{t+1} = g_{\bar{\gamma}}(P_t)$ converges to $\bar{P}$ for all initial positive semidefinite conditions $P_0 \geq 0$.

Lemma 6.3. *Define the linear operator*

$$\mathcal{L}(Y) = (1 - \bar{\gamma})(AYA') + \bar{\gamma}(FYF')$$

Suppose there exists $\overline{Y} > 0$ such that $\overline{Y} > \mathcal{L}(\overline{Y})$.

1. *For all $W \geq 0$,*
$$\lim_{k \to \infty} \mathcal{L}^k(W) = 0$$

2. *Let $U \geq 0$ and consider the linear system*
$$Y_{k+1} = \mathcal{L}(Y_k) + U \quad \text{initialized at} \quad Y_0.$$

Then, the sequence Y_k is bounded.

Proof. (a) First observe that $0 \leq \mathcal{L}(Y)$ for all $0 \leq Y$. Also, $X \leq Y$ implies $\mathcal{L}(X) \leq \mathcal{L}(Y)$. Choose $0 \leq r < 1$ such that $\mathcal{L}(\overline{Y}) < r\overline{Y}$. Choose $0 \leq m$ such that $W \leq m\overline{Y}$. Then,

$$0 \leq \mathcal{L}^k(W) \leq m\mathcal{L}^k(\overline{Y}) < mr^k\overline{Y}.$$

The assertion follows when we take the limit $r \to \infty$, on noticing that $0 \leq r < 1$.

(b) The solution of the linear iteration is

$$Y_k = \mathcal{L}^k(Y_0) + \sum_{t=0}^{k-1} \mathcal{L}^t(U)$$

$$\leq \left(m_{Y_0} r^k + \sum_{t=0}^{k-1} m_U r^t \right) \overline{Y}$$

$$\leq \left(m_{Y_0} r^k + \frac{m_U}{1 - r} \right) \overline{Y}$$

$$\leq \left(m_{Y_0} + \frac{m_U}{1 - r} \right) \overline{Y},$$

proving the claim. $\square$

Lemma 6.4. *Consider the operator $\phi(K, X)$ defined in Equation (6.28). Suppose there exists a matrix $\overline{K}$ and a positive definite matrix $\overline{P}$ such that*

$$\overline{P} > 0 \quad \text{and} \quad \overline{P} > \phi(\overline{K}, \overline{P}).$$

Then, for any P_0, *the sequence* $P_t = g_{\bar{\gamma}}^t(P_0)$ *is bounded, i.e. there exists* $M_{P_0} \geq 0$ *dependent of* P_0 *such that*

$$P_t \leq M \quad \text{for all} \quad t.$$

Proof. First define the matrices $\overline{F} = A + \overline{K}C$ and consider the linear operator

$$\mathcal{L}(Y) = (1 - \bar{\gamma})(AYA') + \bar{\gamma}(\overline{F}Y\overline{F}')$$

Observe that

$$\overline{P} > \phi(\overline{K}, \overline{P}) = \mathcal{L}(\overline{P}) + Q + \bar{\gamma}\overline{K}R\overline{K}' \geq \mathcal{L}(\overline{P}).$$

Thus, $\mathcal{L}$ meets the condition of Lemma 6.3. Finally, using fact (b) in Lemma 6.1 we have

$$P_{t+1} = g_{\bar{\gamma}}(P_t) \leq \phi(\overline{K}, P_t) = \mathcal{L}P_t + Q + \bar{\gamma}\overline{K}R\overline{K}' = \mathcal{L}(P_t) + U.$$

Since $U = \bar{\gamma}\overline{K}R\overline{K}' + Q \geq 0$, using Lemma 6.3, we conclude that the sequence P_t is bounded. $\square$

We are now ready to give proofs for Theorems 1-4.

Proof of Theorem 6.1

(a) We first show that the modified Riccati difference equation initialized at $Q_0 = 0$ converges. Let $Q_k = g_{\bar{\gamma}}^k(0)$. Note that $0 = Q_0 \leq Q_1$. It follows from Lemma 6.1(c) that

$$Q_1 = g_{\bar{\gamma}}(Q_0) \leq g_{\bar{\gamma}}(Q_1) = Q_2.$$

A simple inductive argument establishes that

$$0 = Q_0 \leq Q_1 \leq Q_2 \leq \cdots \leq M_{Q_0}.$$

Here, we have used Lemma 6.4 to bound the trajectory. We now have a monotone non-decreasing sequence of matrices bounded above. It is a simple matter to show that the sequence converges, i.e.

$$\lim_{k \to \infty} Q_k = \overline{P}.$$

Also, we see that $\overline{P}$ is a fixed point of the modified Riccati iteration:

$$\overline{P} = g_{\bar{\gamma}}(\overline{P}),$$

which establishes that it is *a* positive semi-definite solution of the MARE.

Next, we show that the Riccati iteration initialized at $R_0 \geq \overline{P}$ also converges, and to the same limit $\overline{P}$. First define the matrices

$$\overline{K} = -A\overline{P}C' \left(C\overline{P}C' + R\right)^{-1}, \quad \overline{F} = A + \overline{K}C$$

and consider the linear operator

$$\hat{\mathcal{L}}(Y) = (1 - \bar{\gamma})(AYA') + \bar{\gamma}(\overline{F}Y\overline{F}').$$

Observe that

$$\overline{P} = g_{\bar{\gamma}}(\overline{P}) = \mathcal{L}(\overline{P}) + Q + \overline{K}R\overline{K}' > \hat{\mathcal{L}}(\overline{P}).$$

Thus, $\hat{\mathcal{L}}$ meets the condition of Lemma 6.3. Consequently, for all $Y \geq 0$,

$$\lim_{k \to \infty} \hat{\mathcal{L}}^k(Y) = 0.$$

Now suppose $R_0 \geq \overline{P}$. Then,

$$R_1 = g_{\bar{\gamma}}(R_0) \geq g_{\bar{\gamma}}(\overline{P}) = \overline{P}.$$

A simple inductive argument establishes that

$$R_k \geq \overline{P} \quad \text{for all } k.$$

Observe that

$$\begin{aligned}
0 \leq (R_{k+1} - \overline{P}) &= g_{\bar{\gamma}}(R_k) - g_{\bar{\gamma}}(\overline{P}) \\
&= \phi(K_{R_k}, R_k) - \phi(K_{\overline{P}}, \overline{P}) \\
&\leq \phi(K_{\overline{P}}, R_k) - \phi(K_{\overline{P}}, \overline{P}) \\
&= (1 - \bar{\gamma})A(R_k - \overline{P})A' + \bar{\gamma}F_{\overline{P}}(R_k - \overline{P})F_{\overline{P}}' \\
&= \hat{\mathcal{L}}(R_k - \overline{P}).
\end{aligned}$$

Then, $0 \leq \lim_{k \to \infty}(R_{k+1} - \overline{P}) \leq 0$, proving the claim.

We now establish that the Riccati iteration converges to $\overline{P}$ for all initial conditions $P_0 \geq 0$. Define $Q_0 = 0$ and $R_0 = P_0 + \overline{P}$. Consider three Riccati iterations, initialized at Q_0, P_0, and R_0. Note that

$$Q_0 \leq P_0 \leq R_0.$$

It then follows from Lemma 6.2 that

$$Q_k \leq P_k \leq R_k \quad \text{for all } k.$$

We have already established that the Riccati equations P_k and R_k converge to $\overline{P}$. As a result, we have

$$\overline{P} = \lim_{k \to \infty} P_k \leq \lim_{k \to \infty} Q_k \leq \lim_{k \to \infty} R_k = \overline{P},$$

proving the claim.

(b) Finally, we establish that the MARE has a unique positive semi-definite solution. To this end, consider $\hat{P} = g_{\bar{\gamma}}(\hat{P})$ and the Riccati iteration initialized at $P_0 = \hat{P}$. This yields the constant sequence

$$\hat{P}, \hat{P}, \cdots$$

However, we have shown that every Riccati iteration converges to $\overline{P}$. Thus $\overline{P} = \hat{P}$.

Proof of Theorem 6.2

First we note that the two cases expressed by the theorem are indeed possible. If $\bar{\gamma} = 1$ the modified Riccati difference equation reduces to the standard Riccati difference equation, which is known to converge to a fixed point, under the theorem's hypotheses. Hence, the covariance matrix is always bounded in this case, for any initial condition $P_0 \geq 0$. If $\bar{\gamma} = 0$ then we reduce to open loop prediction, and if the matrix A is unstable, then the covariance matrix diverges for some initial condition $P_0 \geq 0$. Next, we show the existence of a single point of transition between the two cases. Fix a $0 < \bar{\gamma}_1 \leq 1$ such that $\mathbb{E}_{\bar{\gamma}_1}[P_t]$ is bounded for any initial condition $P_0 \geq 0$. Then, for any $\bar{\gamma}_2 \geq \bar{\gamma}_1$ $\mathbb{E}_{\bar{\gamma}_2}[P_t]$ is also bounded for all $P_0 \geq 0$. In fact we have

$$
\begin{aligned}
\mathbb{E}_{\bar{\gamma}_1}[P_{t+1}] &= \mathbb{E}_{\bar{\gamma}_1}[AP_t A' + Q - \gamma_{t+1} AP_t C'(CP_t C' + R)^{-1} CP_t A] \\
&= \mathbb{E}[AP_t A' + Q - \bar{\gamma}_1 AP_t C'(CP_t C' + R)^{-1} CP_t A] \\
&= \mathbb{E}[g_{\bar{\gamma}_1}(P_t)] \\
&\geq \mathbb{E}[g_{\bar{\gamma}_2}(P_t)] \\
&= \mathbb{E}_{\bar{\gamma}_2}[P_{t+1}],
\end{aligned}
$$

where we exploited fact (d) of Lemma 6.1 to write the above inequality . We can now choose

$$
\gamma_c = \{\inf \bar{\gamma}^* : \bar{\gamma} > \bar{\gamma}^* \Rightarrow \mathbb{E}_{\bar{\gamma}}[P_t] \text{is bounded, for all } P_0 \geq 0\},
$$

completing the proof.

Proof of Theorem 6.3

Define the Lyapunov operator $m(X) = \tilde{A} X \tilde{A}' + Q$ where $\tilde{A} = \sqrt{1 - \bar{\gamma}} A$. If $(A, Q^{\frac{1}{2}})$ is controllable, also $(\tilde{A}, Q^{\frac{1}{2}})$ is controllable. Therefore, it is well known that $\hat{S} = m(\hat{S})$ has a unique strictly positive definite solution $\hat{S} > 0$ if and only if $\max_i |\sigma_i(\tilde{A})| < 1$, i.e. $\sqrt{1 - \bar{\gamma}} \max_i |\sigma_i(A)| < 1$, from which follows $\gamma_m in = 1 - \frac{1}{\alpha^2}$. If $\max_i |\sigma_i(\tilde{A})| \geq 1$ it is also a well known fact that there is no positive semidefinite fixed point to the Lyapunov equation $\hat{S} = m(\hat{S})$, since $(\tilde{A}, Q^{\frac{1}{2}})$ is controllable.

Let us consider the difference equation $S_{t+1} = m(S_t)$, $S_0 = 0$. It is clear that $S_0 = 0 \leq Q = S_1$. Since the operator $m()$ is monotonic increasing, by Lemma 6.2 it follows that the sequence $\{S_t\}_0^\infty$ is monotonically increasing, i.e. $S_{t+1} \geq S_t$ for all t. If $\bar{\gamma} < \gamma_m in$ this sequence does not converge to a finite matrix $\bar{S}$, otherwise by continuity of the operator m we would have $\bar{S} = m(\bar{S})$, which is not possible. Since it is easy to show that a monotonically increasing sequence S_t that does not converge is also unbounded, then we have

$$
\lim_{t \to \infty} S_t = \infty.
$$

Let us consider now the mean covariance matrix $\mathbb{E}[P_t]$ initialized at

$\mathbb{E}[P_0] \geq 0$. Clearly $0 = S_0 \leq \mathbb{E}[P_0]$. Moreover it is also true that $S_t \leq \mathbb{E}[P_t]$ implies:

$$
\begin{aligned}
S_{t+1} &= (1 - \bar{\gamma}) A S_t A' + Q \\
&\leq (1 - \bar{\gamma}) A \mathbb{E}[P_t] A' + Q \\
&\leq \mathbb{E}[g_{\bar{\gamma}}(P_t)] \\
&= \mathbb{E}[P_{t+1}],
\end{aligned}
$$

where we used fact (h) from Lemma 6.1. By induction, it is easy to show that

$$
S_t \leq \mathbb{E}[P_t] \ \forall t, \ \forall \mathbb{E}[P_0] \geq 0 \Longrightarrow \lim_{t \to \infty} \mathbb{E}[P_t] \geq \lim_{t \to \infty} S_t = \infty.
$$

This implies that for any initial condition $\mathbb{E}[P_t]$ is unbounded for any $\bar{\gamma} < \gamma_m in$, therefore $\gamma_m in \leq \gamma_c$, which proves the first part of the Theorem.

Now consider the sequence $V_{t+1} = g_{\bar{\gamma}}(V_t)$, $V_0 = \mathbb{E}[P_0] \geq 0$. Clearly $\mathbb{E}[P_t] \leq V_t$ implies:

$$
\begin{aligned}
\mathbb{E}[P_{t+1}] &= \mathbb{E}[g_{\bar{\gamma}}(P_t)] \\
&\leq g_{\bar{\gamma}}(\mathbb{E}[P_t]) \\
&\leq [g_{\bar{\gamma}}(V_t)] \\
&= V_{t+1},
\end{aligned}
$$

where we used facts (c) and (h) from Lemma 6.1. Then a simple induction argument shows that $V_t \geq \mathbb{E}[P_t]$ for all t. Let us consider the case $\bar{\gamma} > \gamma_m ax$, therefore there exists $\hat{X}$ such that $\hat{X} \geq g_{\bar{\gamma}}(\hat{X})$. By Lemma 6.1(g) $\bar{X} > 0$, therefore all hypotheses of Lemma 6.3 are satisfied, which implies that

$$
\mathbb{E}[P_t] \leq V_t \leq M_{V_0} \ \forall t.
$$

This shows that $\gamma_c \leq \gamma_m ax$ and concludes the proof of the Theorem.

Proof of Theorem 6.4

Let us consider the sequences $S_{t+1} = (1 - \bar{\gamma}) A S_t A' + Q$, $S_0 = 0$ and $V_{t+1} = g_{\bar{\gamma}}(V_t)$, $V_0 = \mathbb{E}[P_0] \geq 0$. Using the same induction arguments in Theorem 6.3 it is easy to show that

$$
S_t \leq \mathbb{E}[P_t] \leq V_t \ \forall t.
$$

From Theorem 6.1 it also follows that $\lim_{t \to \infty} V_t = \bar{V}$, where $\bar{V} = g_{\bar{\gamma}}(V)$. As shown before the sequence S_t is monotonically increasing. Also it is bounded since $S_t \leq V_t \leq M$. Therefore $\lim_{t \to \infty} S_t = \bar{S}$, and by continuity $\bar{S} = (1 - \bar{\gamma}) A \bar{S} A' + Q$, which is a Lyapunov equation. Since $\sqrt{1 - \bar{\gamma}} A$ is stable and $(A, Q^{\frac{1}{2}})$ is controllable, then the solution of the Lyapunov equation is strictly positive definite, i.e. $\bar{S} > 0$. Adding all the results together we get

$$
0 < \bar{S} = \lim_{t \to \infty} S_t \leq \lim_{t \to \infty} \mathbb{E}[P_t] \leq \lim_{t \to \infty} V_t = \bar{V},
$$

which concludes the proof.

6.8 Packet-Based Control

Outline:

- Problem setup

- Communication protocols and information patterns

- TCP-Based Control

- UDP-Based Control

- Receding-Horizon networked control and actuation buffers

- Generalization to multi-channel

- Nonlinear estensions

6.9 Introduction

Today, an increasing number of applications demand remote control of plants over unreliable networks. The recent development of sensor web technology [?] enables the development of wireless sensor networks that can be immediately used for estimation and control of dynamical systems. In these systems, issues of communication delay, data loss, and time-synchronization play critical roles. Communication and control become very tightly coupled and these two issues cannot be addressed independent of one another during the design and analysis stages of development. Consider, for example, the problem of navigating a fleet of vehicles using observations from a sensor web. Wireless nodes collect their sensor measurements and send them to a computing unit. This unit, in turn, generates estimates of the state of each vehicle and computes inputs that are then delivered, using the same wireless channel, to the actuators onboard the vehicles. Due to the unreliability of the wireless channel, both observations underlying the estimate and control packets sent to each vehicle can be lost or delayed while travelling across the network. It needs to be determined exactly how much data loss a control loop can tolerate to reliably perform the navigation task. What's more, specific communication protocols need to be designed to satisfy this constraint. The goal of this chapter is to provide the first steps in meeting these requirements by examining the basic system-theoretic implications of using unreliable networks for control. This requires a generalization of classical control techniques that explicitly takes into account the stochastic nature of the communication channel.

Figure 6.8: Overview of the system. We study the statistical convergence properties of the expected state covariance of the discrete time LQG control system, where both the observation and the control signal, transmitted over an unreliable communication channel, can be lost at each time step with probability $1 - \bar{\gamma}$ and $1 - \bar{\nu}$ respectively.

Packet networks communication channels typically use one of two kinds of protocols: Transmission Control (TCP) or User Datagram (UDP). In the first case there is acknowledgement of received packets, while in the second case no confirmation feedback is provided on the communication link. In this chapter, we study the effect of data losses due to the unreliability of the network links under these two protocols. We generalize the Linear Quadratic Gaussian (LQG) optimal control problem to these problems by modeling the arrival of both observations and control packets as random processes whose parameters are related to the characteristics of the communication channel. Accordingly, two independent Bernoulli processes are considered, with parameters $\overline{\gamma}$ and $\overline{\nu}$, that govern packet losses between the sensors and the estimation-control unit, and between the latter and the actuation points (see Figure 6.8).

In our analysis, the distinction between the two classes of protocols resides exclusively in the availability of packet acknowledgements. Adopting the framework proposed by Imer *et al.* [**?**], we will refer therefore to TCP-like protocols if packet acknowledgements are available and to UDP-like protocols otherwise. We summarize our contributions as follows. For the

TCP-like case the classic separation principle holds, and consequently the controller and estimator can be designed independently. Moreover, the optimal controller is a linear function of the state. In sharp contrast, for the UDP-like case, a counter-example demonstrates that the optimal controller is in general non-linear. In the special case when the state is fully observable and the observation noise is zero, the optimal controller is indeed linear. We explicitly note that a similar, but slightly less general special case was previously analyzed in [?], where both observation and process noise are assumed to be zero and the input coefficient matrix to be invertible.

Our final set of results relate to convergence in the infinite horizon. Here, our previous results on estimation with missing observation packets [?] [?] are extended to the control case. We show the existence of a critical domain of values for the parameters of the Bernoulli arrival processes, $\overline{\nu}$ and $\overline{\gamma}$, outside which a transition to instability occurs and the optimal controller fails to stabilize the system. In particular, we show that under TCP-like protocols the critical arrival probabilities for the control and observation channel are independent of each other. This is another consequence of the fact that the separation principle holds for these protocols. In contrast, under UDP-like protocols the critical arrival probabilities for the control and observation channels are coupled. Here, the stability domain and performance of the optimal controller degrade considerably as compared with TCP-like protocols as shown in Figure 6.9.

Finally, we wish to mention some closely related research. The study of the stability of dynamical systems where components are connected asynchronously via communication channels has received considerable attention in the past few years and our contribution can be put in the context of the previous literature. In [?] and [?], the authors proposed to place an estimator, i.e. a Kalman filter, at the sensor side of the link without assuming any statistical model for the data loss process. In [?], Smith *et al.* considered a suboptimal but computationally efficient estimator that can be applied when the arrival process is modeled as a Markov chain, which is more general than a Bernoulli process. Other work includes Nilsson *et al.* [?][?] who present the LQG optimal regulator with bounded delays between sensors and controller, and between the controller and the actuator. In this work, bounds for the critical probability values are not provided. Additionally, there is no analytical solution for the optimal controller. The case where dropped measurements are replaced by zeros is considered by Hadjicostis and Touri [?], but only in the scalar case. Other approaches include using the last received sample for control [?], or designing a dropout compensator [?], which combines estimation and control in a single process. However, the former approach does not consider optimal control and the latter is limited to scalar systems. Yu *et al.* [?] studied the design of an optimal controller with a single control channel and deterministic dropout rates. Seiler *et al.* [?] considered Bernoulli packet losses only between the

Figure 6.9: Region of stability for UDP-like and TCP-like optimal control relative to measurement packet arrival probability γ, and the control packet arrival probability ν.

plant and the controller, and posed the controller design as an H_∞ optimization problem. Other authors [**?**] [**?**] [**?**] [**?**] model networked control systems with missing packets as Markovian jump linear systems (MJLSs), however this approach gives suboptimal controllers since the estimators are stationary. Finally, Elia [**?**][**?**] proposed to model the plant and the controller as deterministic time invariant discrete-time systems connected to zero-mean stochastic structured uncertainty. The variance of the stochastic perturbation is a function of the Bernoulli parameters, and the controller design is posed as an optimization problem to maximize mean-square stability of the closed loop system. This approach allows analysis of Multiple Input Multiple Output (MIMO) systems with many different controller and receiver compensation schemes [**?**], however, it does not include process and observation noise and the controller is restricted to be time-invariant, hence suboptimal. There is also extensive literature, inspired by Shannon's results on the maximum bit-rate that an imperfect channel can reliably carry. Here the goal is to determine the minimum bit-rate that is needed to stabilize a system

through feedback [?] [?] [?] [?] [?] [?] [?] [?] [?] [?]. This approach is some-what different from ours since in a packet-based communication network, such as ATMs, Ethernet and Bluetooth, bits are grouped into packets and are considered as a single entity. Nonetheless there are several similarities that are not yet fully explored.

This work considers the alternative approach where the external compensator feeding the controller is the optimal time varying Kalman gain. Moreover, this approach considers the general Multiple Input Multiple Output (MIMO) case, and gives some necessary and sufficient conditions for closed loop stability. The work of [?] is most closely related to ours. However, we consider the more general case when the matrix C is not the identity and there is noise in the observation and in the process. In addition, we also give stronger necessary and sufficient conditions for existence of solution for the infinite horizon LQG.

The remainder of this chapter is organized as follows. Section 2 provides a mathematical formulation of the problems we consider. Section 3 offers some preliminary results. Section 4 illustrates the TCP-like case, while the UDP-like case is studied in section 5. Finally, conclusions and directions for future work are offered in section 6.

6.10 Problem formulation

Consider the following linear stochastic system with intermittent observation and control packets:

$$x_{k+1} = Ax_k + Bu_k + w_k \qquad (6.29)$$

$$u_k^a = \nu_k u_k^c \qquad (6.30)$$

$$y_k = \gamma_k Cx_k + v_k, \qquad (6.31)$$

where u_k^a is the control input to the actuator, u_k^c is the desired control input computed by the controller, (x_0, w_k, v_k) are Gaussian, uncorrelated, white, with mean $(\bar{x}_0, 0, 0)$ and covariance (P_0, Q, R) respectively, and (γ_k, ν_k) are i.i.d. Bernoulli random variables with $P(\gamma_k = 1) = \bar{\gamma}$ and $P(\nu_k = 1) = \bar{\nu}$. The stochastic variable ν_k models the loss packets between the controller and the actuator: if the packet is correctly delivered then $u_k^a = u_k^c$, otherwise if it is lost then the actuator does nothing, i.e. $u_k^a = 0$. This compensation scheme is summarized by Equation (6.30). This modeling choice is not unique: for example if the control packet u_k^c is lost, then the actuator could use the previous control value, i.e. $u_k^a = u_{k-1}^a$. However, the latter control compensation is slightly more involved to analyze and it is left as future work. The stochastic variable γ_k models the packet loss between the sensor and the controller: if the packet is delivered then $y_k = Cx_k + v_k$, otherwise if it is lost then the controller reads pure noise, i.e. $y_k = v_k$. This observation model is summarized by Equation (6.31). A different observation formal-

ism was proposed in [**?**], where the missing observation was modeled as an observation for which the measurement noise had infinite covariance. It is possible to show that both models are equivalent, but the one considered here has the advantage to give rise to simpler analysis. This arises from the fact that when no packet is delivered, then the optimal estimator does not use the observation y_k at all, therefore its value is irrelevant.

Let us define the following information sets:

$$\mathcal{I}_k = \begin{cases} \mathcal{F}_k & \triangleq & \{\mathbf{y}^k, \boldsymbol{\gamma}^k, \boldsymbol{\nu}^{k-1}\}, & \text{TCP-like} \\ \mathcal{G}_k & \triangleq & \{\mathbf{y}^k, \boldsymbol{\gamma}^k\}, & \text{UDP-like} \end{cases} \tag{6.32}$$

where $\mathbf{y}^k = (y_k, y_{k-1}, \ldots, y_1)$, $\boldsymbol{\gamma}^k = (\gamma_k, \gamma_{k-1}, \ldots, \gamma_1)$, and $\boldsymbol{\nu}^k = (\nu_k, \nu_{k-1}, \ldots, \nu_1)$.

Consider also the following cost function:

$$J_N(\mathbf{u}^{N-1}, \bar{x}_0, P_0) = \mathbb{E}\left[x_N' W_N x_N + \sum_{k=0}^{N-1}(x_k' W_k x_k + \nu_k u_k' U_k u_k) \mid \mathbf{u}^{N-1}, \bar{x}_0, P_0\right] \tag{6.33}$$

where $\mathbf{u}^{N-1} = (u_{N-1}, u_{N-2}, \ldots, u_1)$. Note that we are weighting the input only if it is successfully received at the plant. In fact, if it is not received, the plant applies zero input and therefore there is no energy expenditure.

We now look for a control input sequence $\mathbf{u}^{*N-1}$ as a function of the admissible information set $\mathcal{I}_k$, i.e. $u_k = g_k(\mathcal{I}_k)$, that minimizes the functional defined in Equation (6.33), i.e.

$$J_N^*(\bar{x}_0, P_0) \triangleq \min_{\mathbf{u_k} = \mathbf{g}_k(\mathcal{I}_k)} J_N(\mathbf{u}^{N-1}, \bar{x}_0, P_0), \tag{6.34}$$

where $\mathcal{I}_k = \{\mathcal{F}_k, \mathcal{G}_k\}$ is one of the sets defined in Equation (6.32). The set $\mathcal{F}$ corresponds to the information provided under an acknowledgement-based communication protocols (TCP-like) in which successful or unsuccessful packet delivery at the receiver is acknowledged to the sender within the same sampling time period. The set $\mathcal{G}$ corresponds to the information available at the controller under communication protocols in which the sender receives no feedback about the delivery of the transmitted packet to the receiver (UDP-like). The UDP-like schemes are simpler to implement than the TCP-like schemes from a communication standpoint. Moreover UDP-like protocols includes broadcasting which you cannot do with TCP-like. However the price to pay is a less rich set of information. The goal of this chapter is to design optimal LQG controllers and to estimate their performance for each of these classes of protocols for a general discrete-time linear stochastic system.

6.11 Mathematical Preliminaries

Before proceeding, let us define the following variables:

$$
\begin{aligned}
\hat{x}_{k|k} &\triangleq \mathbb{E}[x_k \mid \mathcal{I}_k], \\
e_{k|k} &\triangleq x_k - \hat{x}_{k|k}, \\
P_{k|k} &\triangleq \mathbb{E}[e_{k|k}e'_{k|k} \mid \mathcal{I}_k].
\end{aligned}
\tag{6.35}
$$

Derivations below will make use of the following facts:

Lemma 6.5. *The following facts are true [?]:*

1. $\mathbb{E}\left[(x_k - \hat{x}_k)\hat{x}'_k \mid \mathcal{I}_k\right] = \mathbb{E}\left[e_{k|k}\hat{x}'_k \mid \mathcal{I}_k\right] = 0$

2. $\mathbb{E}\left[x'_k S x_k \mid \mathcal{I}_k\right] = \hat{x}'_k S \hat{x}_k + \text{trace}\left(SP_{k|k}\right) \quad \forall S \geq 0$

3. $\mathbb{E}\left[\mathbb{E}[\, g(x_{k+1}) \mid \mathcal{I}_{k+1}] \mid \mathcal{I}_k\right] = \mathbb{E}\left[g(x_{k+1}) \mid \mathcal{I}_k\right], \forall g(\cdot).$

Proof. (a) It follows directly from the definition. In fact:

$$
\begin{aligned}
\mathbb{E}\left[(x_k - \hat{x}_k)\hat{x}'_k \mid \mathcal{I}_k\right] &= \mathbb{E}\left[x_k\hat{x}'_k - \hat{x}_k\hat{x}'_k \mid \mathcal{I}_k\right] \\
&= \mathbb{E}\left[x_k \mid \mathcal{I}_k\right]\hat{x}'_k - \hat{x}_k\hat{x}'_k \\
&= 0
\end{aligned}
$$

(b) Using standard algebraic operations and the previous fact we have:

$$
\begin{aligned}
\mathbb{E}\left[x'_k S x_k \mid \mathcal{I}_k\right] &= \mathbb{E}\left[(x_k - \hat{x}_k + \hat{x}_k)'S(x_k - \hat{x}_k + \hat{x}_k) \mid \mathcal{I}_k\right] \\
&= \hat{x}'_k S \hat{x}_k + \mathbb{E}\left[(x_k - \hat{x}_k)'S(x_k - \hat{x}_k)\right] + 2\mathbb{E}\left[\hat{x}'_k S(x_k - \hat{x}_k) \mid \mathcal{I}_k\right] \\
&= \hat{x}'_k S \hat{x}_k + 2\text{trace}(S\mathbb{E}[(x_k - \hat{x}_k)\hat{x}'_k \mid \mathcal{I}_k]) + \text{trace}(S\mathbb{E}[(x_k - \hat{x}_k)(x_k - \hat{x}_k)' \mid \mathcal{I}_k]) \\
&= \hat{x}'_k S \hat{x}_k + \text{trace}\{SP_{k|k}\}
\end{aligned}
$$

(c) Let $g()$ any measurable function, (X, Y, Z) be any random vectors, and p their probability distribution, then

$$
\begin{aligned}
\mathbb{E}_{Y,Z}\left[g(X, Y, Z) \mid X\right] &= \int_Z \int_Y g(X, Y, Z)p(Y, Z|X)dY\,dZ \\
&= \int_Z \int_Y g(X, Y, Z)p(Y|Z, X)p(Z|X)dY\,dZ \\
&= \int_Z \left[\int_Y g(X, Y, Z)p(Y|Z, X)dY\right]p(Z|X)dZ \\
&= \mathbb{E}_Z\left[\, \mathbb{E}_Y\left[g(X, Y, Z) \mid Z, X\right] \mid X\right]
\end{aligned}
$$

where we used the Bayes' Rule. Since by hypothesis $\mathcal{I}_k \subseteq \mathcal{I}_{k+1}$, then fact (c) follows from the above equality by substituting $\mathcal{I}_k = X$ and $\mathcal{I}_{k+1} = (X, Z)$. $\square$

We now make the following computations that will be useful when de-

riving the equation for the optimal LQG controller.

$$
\begin{aligned}
\mathbb{E}[x'_{k+1}Sx_{k+1} \mid \mathcal{I}_k] &= \mathbb{E}[(Ax_k + \nu_k Bu_k + w_k)'S(Ax_k + \nu_k Bu_k + w_k) \mid \mathcal{I}_k] \\
&= \mathbb{E}[x'_k A'SAx_k + \nu_k^2 u'_k B'SBu_k + w'_k Sw_k + 2\nu_k u'_k B'SAx_k + 2(Ax_k + \nu_k Bu_k)w_k | \mathcal{I}_k] \\
&= \mathbb{E}[x'_k A'SAx_k | \mathcal{F}_k] + \bar{\nu} u'_k B'SBu_k + 2\bar{\nu} u'_k B'SA\,\mathbb{E}[x_k | \mathcal{I}_k] + \operatorname{trace}(S\mathbb{E}[w_k w'_k \mid \mathcal{F}_k]) \\
&= \mathbb{E}[x'_k A'SAx_k \mid \mathcal{I}_k] + \bar{\nu} u'_k B'SBu_k + 2\bar{\nu} u'_k B'SA\,\hat{x}_{k|k} + \operatorname{trace}(SQ) \qquad (6.36)
\end{aligned}
$$

where both the independence of ν_k, w_k, x_k, and the zero-mean property of w_k are exploited. The previous expectation holds true for both the information sets, i.e. $\mathcal{I}_k = \mathcal{F}_k$ or $\mathcal{I}_k = \mathcal{G}_k$. Also

$$
\begin{aligned}
\mathbb{E}[e'_{k|k}Te_{k|k} \mid \mathcal{I}_k] &= \operatorname{trace}(T\mathbb{E}[e_{k|k}e'_{k|k} \mid \mathcal{I}_k]) \\
&= \operatorname{trace}(TP_{k|k}), \quad \forall T \geq 0.
\end{aligned}
$$

6.12 LQG control for TCP-like protocols

First, equations for the optimal estimator are derived. They will be needed to solve the LQG controller design problem, as it will be shown later.

Estimator Design

Equations for optimal estimator are derived using similar arguments used for the standard Kalman filtering equations. The innovation step is given by:

$$
\begin{aligned}
\hat{x}_{k+1|k} &\triangleq \mathbb{E}[x_{k+1} | \nu_k, \mathcal{F}_k] = \mathbb{E}[Ax_k + \nu_k Bu_k + w_k | \nu_k, \mathcal{F}_k] \\
&= A\mathbb{E}[x_k | \mathcal{F}_k] + \nu_k Bu_k = A\hat{x}_{k|k} + \nu_k Bu_k \qquad (6.37)
\end{aligned}
$$

$$
\begin{aligned}
e_{k+1|k} &\triangleq x_{k+1} - \hat{x}_{k+1|k} \\
&= Ax_k + \nu_k Bu_k + w_k - (A\hat{x} + \nu_k Bu_k) \\
&= Ae_{k|k} + w_k \qquad (6.38)
\end{aligned}
$$

$$
\begin{aligned}
P_{k+1|k} &\triangleq \mathbb{E}[e_{k+1|k}e'_{k+1|k} \mid \nu_k, \mathcal{F}_k] \\
&= \mathbb{E}\left[\left(Ae_{k|k} + w_k\right)\left(Ae_{k|k} + w_k\right)' \mid \nu_k, \mathcal{F}_k \right] \\
&= A\mathbb{E}[e_{k|k}e'_{k|k} | \mathcal{F}_k]A' + \mathbb{E}[w_k w'_k] \\
&= AP_{k|k}A' + Q, \qquad (6.39)
\end{aligned}
$$

where the independence of w_k and $\mathcal{F}_k$, and the requirement that u_k is a deterministic function of $\mathcal{F}_k$, are used. Since $y_{k+1}, \gamma_{k+1}, w_k$ and $\mathcal{F}_k$ are independent, the correction step is given by:

$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + \gamma_{k+1} K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k}) \tag{6.40}$$

$$\begin{aligned}
e_{k+1|k+1} &\triangleq x_{k+1} - \hat{x}_{k+1|k+1} \\
&= x_{k+1} - \left(\hat{x}_{k+1|k} + \gamma_{k+1} K_{k+1}(Cx_{k+1} + v_{k+1} - C\hat{x}_{k+1|k}) \right) \\
&= (I - \gamma_{k+1} K_{k+1} C) e_{k+1|k} - \gamma_{k+1} K_{k+1} v_{k+1} \tag{6.41}
\end{aligned}$$

$$\begin{aligned}
P_{k+1|k+1} &= P_{k+1|k} - \gamma_{k+1} K_{k+1} C P_{k+1|k} \\
&= P_{k+1|k} - \gamma_{k+1} P_{k+1|k} C'(C P_{k+1|k} C' + R)^{-1} C P_{k+1|k} \tag{6.42}
\end{aligned}$$

$$K_{k+1} \triangleq P_{k+1|k} C'(C P_{k+1|k} C' + R)^{-1}, \tag{6.43}$$

where we simply applied the standard derivation for the time varying Kalman filter using the following time varying system matrices: $A_k = A$, $C_k = \gamma_k C$, and $\text{Cov}(v_k) = R$.

Controller design

Derivation of the optimal feedback control law and the corresponding value for the objective function will follow the dynamic programming approach based on the cost-to-go iterative procedure.

Define the optimal value function $V_k(x_k)$ as follows:

$$\begin{aligned}
V_N(x_N) &\triangleq \mathbb{E}[x_N' W_N x_N \mid \mathcal{F}_N] \\
V_k(x_k) &\triangleq \min_{u_k} \mathbb{E}[x_k' W_k x_k + \nu_k u_k' U_k u_k + V_{k+1}(x_{k+1}) | \mathcal{F}_k]. \tag{6.44}
\end{aligned}$$

where $k = N-1, \ldots, 1$. Using dynamic programming theory [?], one can show that $J_N^* = V_0(x_0)$. Under TCP-like protocols the following lemma holds true:

Lemma 6.6. *The value function $V_k(x_k)$ defined in Equations (6.44) for the system dynamics of Equations (6.29)-(6.29) under TCP-like protocols can be written as:*

$$V_k(x_k) = \mathbb{E}[\, x_k' S_k x_k \mid \mathcal{F}_k] + c_k, \quad k = N, \ldots, 0 \tag{6.45}$$

where the matrix S_k and the scalar c_k can be computed recursively as follows:

$$S_k = A' S_{k+1} A + W_k - \bar{\nu} A' S_{k+1} B (B' S_{k+1} B + U_k)^{-1} B' S_{k+1} A \tag{6.46}$$

$$c_k = \text{trace}\left((A' S_{k+1} A + W_k - S_k) P_{k|k} \right) + \text{trace}(S_{k+1} Q) + \mathbb{E}[c_{k+1} \mid \mathcal{F}_k] \tag{6.47}$$

with initial values $S_N = W_N$ and $c_N = 0$. Moreover the optimal control input is given by:

$$u_k = -(B' S_{k+1} B + U_k)^{-1} B' S_{k+1} A \, \hat{x}_{k|k} = L_k \, \hat{x}_{k|k}. \tag{6.48}$$

Proof. The proof follows an induction argument. The claim is certainly true for $k = N$ with the choice of parameters $S_N = W_N$ and $c_N = 0$. Suppose now

that the claim is true for $k+1$, i.e. $V_{k+1}(x_{k+1}) = \mathbb{E}[\,x'_{k+1}S_{k+1}x_{k+1} \mid \mathcal{F}_{k+1}] + c_{k+1}$. The value function at time step k is the following:

$$
\begin{aligned}
V_k(x_k) &= \min_{u_k} \mathbb{E}[x'_k W_k x_k + \nu_k u'_k U_k u_k + V_{k+1}(x_{k+1}) \mid \mathcal{F}_k] \\
&= \min_{u_k} \mathbb{E}[x'_k W_k x_k + \nu_k u'_k U_k u_k + \mathbb{E}[x'_{k+1}S_{k+1}x_{k+1} + c_{k+1} \mid \mathcal{F}_{k+1}] \mid \mathcal{F}_k] \\
&= \min_{u_k} \mathbb{E}[x'_k W_k x_k + \nu_k u'_k U_k u_k + x'_{k+1}S_{k+1}x_{k+1} + c_{k+1} \mid \mathcal{F}_k] \qquad (6.49) \\
&= \mathbb{E}[x'_k W_k x_k + x'_k A' S_{k+1} A x_k \mid \mathcal{F}_k] + \mathrm{trace}(S_{k+1}Q) + \mathbb{E}[c_{k+1} \mid \mathcal{F}_k] + \\
&\quad + \bar{\nu} \min_{u_k} \left(u'_k(U_k + B'S_{k+1}B)u_k + 2u'_k B'S_{k+1}A\,\hat{x}_{k|k} \right)
\end{aligned}
$$

where we used Lemma 1(c) to get the third equality, and Equation (6.36) to obtain the last equality. The value function is a quadratic function of the input, therefore the minimizer can be simply obtained by solving $\frac{\partial V_k}{\partial u_k} = 0$, which gives Equation (6.48). The optimal feedback is thus a simple linear function of the estimated state. If we substitute the minimizer back into Equation (6.49) we get:

$$
\begin{aligned}
V_k(x_k) &= \mathbb{E}[x'_k W_k x_k + x'_k A' S_{k+1} A x_k \mid \mathcal{I}_k] + \mathrm{trace}(S_{k+1}Q) + \mathbb{E}[c_{k+1} \mid \mathcal{I}_k] - \\
&\quad - \bar{\nu}\hat{x}'_{k|k}A'S_{k+1}B(U_k + B'S_{k+1}B)^{-1}B'S_{k+1}A\hat{x}_{k|k} \qquad\qquad (6.50) \\
&= \mathbb{E}[x'_k W_k x_k + x'_k A'S_{k+1}Ax_k - \bar{\nu}x'_k A'S_{k+1}B(U_k + B'S_{k+1}B)^{-1}B'S_{k+1}Ax_k \mid \mathcal{I}_k] + \\
&\quad + \mathrm{trace}(S_{k+1}Q) + \mathbb{E}[c_{k+1} \mid \mathcal{I}_k] + \bar{\nu}\,\mathrm{trace}(A'S_{k+1}B(U_k + B'S_{k+1}B)^{-1}B'S_{k+1}\,P_{k|k})
\end{aligned}
$$

where we used Lemma 1(b). Therefore, the claim given by Equation (6.45) is satisfied also for time step k for all x_k if and only if the Equations (6.46) and (6.47) are satisfied. $\qquad\square$

Since $J_N^*(\bar{x}_0, P_0) = V_0(x_0)$, from the lemma it follows that the cost function for the optimal LQG using TCP-like protocols is given by:

$$
J_N^* = \bar{x}'_0 S_0 \bar{x}_0 + \mathrm{trace}(S_0 P_0) + \sum_{k=0}^{N-1} \mathrm{trace}((A'S_{k+1}A + W_k - S_k)\mathbb{E}_\gamma[P_{k|k}] + S_{k+1}Q),
$$
$$(6.51)$$

where we used the fact $\mathbb{E}[x'_0 S_0 x_0] = \bar{x}'_0 S_0 \bar{x}_0 + \mathrm{trace}(S_0 P_0)$, and $\mathbb{E}_\gamma[\cdot]$ explicitly indicates that the expectation is calculated with respect to the arrival sequence $\{\gamma_k\}$.

It is important to remark that the error covariance matrices $\{P_{k|k}\}_{k=0}^{N}$ are stochastic since they depend on the sequence $\{\gamma_k\}$. Moreover, since the matrix $P_{k+1|k+1}$ is a nonlinear function of the previous time step matrix covariance $P_{k|k}$, as can be observed from Equations (6.39) and (6.43), the exact expected value of these matrices, $\mathbb{E}_\gamma[P_{k|k}]$, cannot be computed analytically, as shown in [?]. However, they can be bounded by computable deterministic quantities, as shown in [?] from which we can derive the following lemma:

Lemma 6.7 ([**?**]). *The expected error covariance matrix* $\mathbb{E}_\gamma[P_{k|k}]$ *satisfies the following bounds:*

$$\widetilde{P}_{k|k} \leq \mathbb{E}_\gamma[P_{k|k}] \leq \widehat{P}_{k|k} \qquad \forall k \geq 0, \tag{6.52}$$

where the matrices $\widehat{P}_{k|k}$ *and* $\widetilde{P}_{k|k}$ *can be computed as follows:*

$$\widehat{P}_{k+1|k} = A\widehat{P}_{k|k-1}A' + Q - \bar{\gamma}A\widehat{P}_{k|k-1}C'(C\widehat{P}_{k|k-1}C' + R)^{-1}C\widehat{P}_{k|k-1}A' \tag{6.53}$$

$$\widehat{P}_{k|k} = \widehat{P}_{k|k-1} - \bar{\gamma}\widehat{P}_{k|k-1}C'(C\widehat{P}_{k|k-1}C' + R)^{-1}C\widehat{P}_{k|k-1} \tag{6.54}$$

$$\widetilde{P}_{k+1|k} = (1 - \bar{\gamma})A\widetilde{P}_{k|k-1}A' + Q \tag{6.55}$$

$$\widetilde{P}_{k|k} = (1 - \bar{\gamma})\widetilde{P}_{k|k-1} \tag{6.56}$$

where the initial conditions are $\widehat{P}_{0|0} = \widetilde{P}_{0|0} = P_0$.

Proof. The proof is based on the observation that the matrices $P_{k+1|k}$ and $P_{k|k}$ are concave and monotonic functions of $P_{k|k-1}$. The proof is given in [**?**] and is thus omitted.   □

From this lemma it follows that also the minimum achievable cost J_N^*, given by Equation (6.51), cannot be computed analytically, but can bounded as follows:

$$J_N^{min} \leq J_N^* \leq J_N^{max} \tag{6.57}$$

$$J_N^{max} = \bar{x}_0'S_0\bar{x}_0 + \text{trace}(S_0P_0) + \sum_{k=0}^{N-1}\text{trace}(S_{k+1}Q)) + \sum_{k=0}^{N-1}\text{trace}\left((A'S_{k+1}A + W_k - S_k)\widehat{P}_{k|k}\right) \tag{6.58}$$

$$J_N^{min} = \bar{x}_0'S_0\bar{x}_0 + \text{trace}(S_0P_0) + \sum_{k=0}^{N-1}\text{trace}(S_{k+1}Q) + \sum_{k=0}^{N-1}\text{trace}\left((A'S_{k+1}A + W_k - S_k)\widetilde{P}_{k|k}\right) \tag{6.59}$$

Finite and Infinite Horizon LQG control

The results derived in the previous sections can be summarized in the following theorem:

Theorem 6.7. *Consider the system (6.29)-(6.31) and consider the problem of minimizing the cost function (6.33) within the class of admissible policies* $u_k = f(\mathcal{F}_k)$, *where* $\mathcal{F}_k$ *is the information available under TCP-like schemes, given in Equation (6.32). Then:*

1. *The separation principle still holds for TCP-like communication, since the optimal estimator, given by Equations (6.37),(6.39),(6.40),(6.42) and (6.43), is independent of the control input* u_k.

2. *The optimal estimator gain* K_k *is time-varying and stochastic since it depends on the past observation arrival sequence* $\{\gamma_j\}_{j=1}^k$.

3. *The optimal control input, given by Equations (6.48) and (6.46) with initial condition $S_N = W_N$, is a linear function of the estimated state $\hat{x}_{k|k}$, i.e. $u_k = L_k \hat{x}_{k|k}$, and is independent of the process sequences $\{\nu_k, \gamma_k\}$.*

Proof. The proof follows from the results given in the previous sections. □

The infinite horizon LQG can be obtained by taking the limit for $N \to +\infty$ of the previous equations. However, as explained above, the matrices $\{P_{k|k}\}$ depend nonlinearly on the specific realization of the observation sequence $\{\gamma_k\}$, therefore the expected error covariance matrices $\mathbb{E}_\gamma[P_{k|k}]$ and the minimal cost J_N^* cannot be computed analytically and do not seem to have limit [?]. Differently from standard LQG optimal regulator [?], the estimator gain does not converge to a steady state value, but is strongly time-varying due to its dependence on the arrival process $\{\gamma_k\}$. Moreover, while the standard LQG optimal regulator always stabilizes the original system, in the case of observation and control packet losses, the stability can be lost if the arrival probabilities $\bar{\nu}, \bar{\gamma}$ are below a certain threshold. This observation come from the study of existence of solution for a Modified Riccati Algebraic Equation (MARE), $S = \Pi(S, A, B, W, U, \nu)$, which was introduced by [?] and studied in [?], [?] and [?], where the nonlinear operator $\Pi(\cdot)$ is defined as follows:

$$\Pi(S, A, B, Q, R, \nu) \triangleq A'SA + W - \nu \, A'SB(B'SB + U)^{-1}B'SA \quad (6.60)$$

In particular, Equation (6.46), i.e. $S_{k+1} = \Pi(S_k, A, B, W, U, \nu)$, is the dual of the estimator equation presented in [?], i.e. $P_{k+1} = \Pi(P_k, A', C', Q, R, \gamma)$. The results about the MARE are summarized in the following lemma

Lemma 6.8. *Consider the modified Riccati equation defined in Equation (6.60). Let A be unstable, (A, B) be controllable, and $(A, W^{\frac{1}{2}})$ be observable. Then:*

1. *The MARE has a unique strictly positive definite solution S_∞ if and only if $\nu > \nu_c$, where ν_c is the critical arrival probability defined as:*

$$\nu_c \triangleq \inf_\nu \{0 \le \nu \le 1, S \ge 0) \,|\, S = \Pi(S, A, B, W, U, \nu)\}.$$

2. *The critical probability ν_c satisfy the following analytical bounds:*

$$p_{min} \le \nu_c \le p_{max}$$
$$p_{min} \triangleq 1 - \frac{1}{\max_i |\lambda_i^u(A)|^2}$$
$$p_{max} \triangleq 1 - \frac{1}{\prod_i |\lambda_i^u(A)|^2}$$

where $\lambda_i^u(A)$ are the unstable eigenvalues of A. Moreover, $\nu_c = p_{min}$ when B is square and invertible, and $\nu_c = p_{max}$ when B is rank one.

3. *The critical probability can be numerically computed via the solution of the following quasi-convex LMIs optimization problem:*

$$\nu_c = \mathrm{argmin}_{\bar{\nu}} \Psi_\nu(Y, Z) > 0, \quad 0 \leq Y \leq I.$$

$$\Psi_\nu(Y, Z) = \begin{bmatrix} Y & \sqrt{\nu}(YA' + ZB') & \sqrt{1 - \nu}YA' \\ \sqrt{\nu}(AY + BZ') & Y & 0 \\ \sqrt{1 - \nu}AY & 0 & Y \end{bmatrix}$$

4. If $\nu > \nu_c$, then $\lim_{k \to +\infty} S_k = S_\infty$ for all initial conditions $S_0 \geq 0$, where

$$S_{k+1} = \Pi(S_k, A, B, W, U, \nu)$$

.

Proof. The proof of facts (a),(c), and (d) can be found in [**?**]. The proof $\nu_c = p_{min}$ when B is square and invertible can be found in [**?**], and the proof $\nu_c = p_{max}$ when B is rank one in [**?**]. □

In [**?**] statistical analysis of the optimal estimator was given, which we report here for convenience:

Theorem 6.8 ([**?**])**.** *Consider the system (6.29)-(6.31) and the optimal estimator under TCP-like protocols, given by Equations (6.37),(6.39),(6.40),(6.42) and (6.43). Assume that $(A, Q^{\frac{1}{2}})$ is controllable, (A, C) is observable, and A is unstable. Then there exists a critical observation arrival probability γ_c, such that the expectation of estimator error covariance is bounded if and only if the observation arrival probability is greater than the critical arrival probability, i.e.*

$$\mathbb{E}_\gamma[P_{k|k}] \leq M \; \forall k \text{ iff } \bar{\gamma} > \gamma_c.$$

where M is a positive definite matrix possibly dependent on P_0. Moreover, it is possible to compute a lower and an upper bound for the critical observation arrival probability γ_c, i.e.:

$$p_{min} \leq \gamma_c \leq \gamma_{max} \leq p_{max}$$

, where:

$$\gamma_{max} \overset{\Delta}{=} \inf_{\gamma} \{0 \leq \gamma \leq 1, P \geq 0) \,|\, P = \Pi(P, A', C', Q, R, \gamma)\},$$

where p_{min} and p_{max} are defined in Lemma 6.8.

Proof. The proof can be found in [**?**] and is therefore omitted. □

Using the previous theorem and the results from the previous section, we can prove the following theorem for the infinite horizon optimal LQG under TCP-like protocols:

Theorem 6.9. *Consider the same system as defined in the previous theorem with the following additional hypothesis: $W_N = W_k = W$ and $U_k = U$. Moreover, let (A, B) and $(A, Q^{\frac{1}{2}})$ be controllable, and let (A, C) and $(A, W^{\frac{1}{2}})$ be observable. Moreover, suppose that $\bar{\nu} > \nu_c$ and $\bar{\gamma} > \gamma_{max}$, where ν_c and*

γ_{max} are defined in Lemma 6.8 and in Theorem 6.8, respectively. Then we have:

1. The infinite horizon optimal controller gain is constant:

$$\lim_{k \to \infty} L_k = L_\infty = -(B'S_\infty B + U)^{-1}B'S_\infty A \qquad (6.61)$$

2. The infinite horizon optimal estimator gain K_k, given by Equation (6.43), is stochastic and time-varying since it depends on the past observation arrival sequence $\{\gamma_j\}_{j=1}^{k}$.

3. The expected minimum cost can be bounded by two deterministic sequences:

$$\frac{1}{N}J_N^{min} \leq \frac{1}{N}J_N^* \leq \frac{1}{N}J_N^{max} \qquad (6.62)$$

where J_N^{min}, J_N^{max} converge to the following values:

$$J_\infty^{max} \triangleq \lim_{N \to +\infty} \frac{1}{N}J_N^{max}$$
$$= \operatorname{trace}((A'S_\infty A + W - S_\infty)(\widehat{P}_\infty - \bar{\gamma}\widehat{P}_\infty C'(C\widehat{P}_\infty C' + R)^{-1}C\widehat{P}_\infty)) + \operatorname{trace}(S_\infty Q)$$
$$J_\infty^{min} \triangleq \lim_{N \to +\infty} \frac{1}{N}J_N^{min}$$
$$= (1 - \bar{\gamma})\operatorname{trace}\left((A'S_\infty A + W - S_\infty)\widetilde{P}_\infty\right) + \operatorname{trace}(S_\infty Q),$$

and the matrices $S_\infty, \overline{P}_\infty, \underline{P}_\infty$ are the positive definite solutions of the following equations:

$$S_\infty = A'S_\infty A + W - \bar{\nu}\,A'S_\infty B(B'S_\infty B + U)^{-1}B'S_\infty A$$
$$\overline{P}_\infty = A\overline{P}_\infty A' + Q - \bar{\gamma}\,A\overline{P}_\infty C'(C\overline{P}_\infty C' + R)^{-1}C\overline{P}_\infty A'$$
$$\underline{P}_\infty = (1 - \bar{\gamma})A\underline{P}_\infty A' + Q$$

Proof. (a) Since by hypothesis $\bar{\nu} > \nu_c$, from Lemma 6.8(d) follows that $\lim_{k \to +\infty} S_k = S_\infty$. Therefore Equation (6.61) follows from Equation (6.48).

(b) This follows from the dependence on the arrival sequence $\{\gamma_k\}$ of the optimal state estimator given by Equations (6.37),(6.39),(6.40),(6.42) and (6.43). Since $\bar{\nu} > \nu_c$

(c) Equation (6.53) can be written in terms of the MARE as:

$$\widehat{P}_{k+1|k} = \Pi(\widehat{P}_{k|k-1}, A', C', Q, R, \gamma)$$

, therefore since $\bar{\gamma} > \gamma_{max}$ from Lemma 6.8(d) it follows that $\lim_{k \to +\infty} \widehat{P}_{k|k-1} = \overline{P}_\infty$, where $\overline{P}_\infty$ is the solution of the MARE $\overline{P}_\infty = \Pi(\overline{P}_\infty, A', C', Q, R, \gamma)$. Also $\lim_{k \to +\infty} \widetilde{P}_{k|k-1} = \underline{P}_\infty$, where $\widetilde{P}_{k|k-1}$ is defined in Equation (6.55) and $\underline{P}_\infty$ is the solution of the Lyapunov equation $\widehat{P}_\infty = \tilde{A}\widehat{P}_\infty\tilde{A}' + Q$, where $\tilde{A} = \sqrt{1 - \bar{\gamma}}A$. Such solution clearly exists since $\sqrt{1 - \bar{\gamma}} < \frac{1}{p_{min}} = \frac{1}{\max_i |\lambda_i^u(A)|}$ and thus the matrix $\tilde{A}$ is strictly stable. From Equations (6.54) and (6.56)

it follows that $\lim_{k \to +\infty} \widehat{P}_{k|k} = \overline{P}_\infty - \bar{\gamma} \overline{P}_\infty C'(C\overline{P}_\infty C' + R)^{-1} C\overline{P}_\infty$ and $\lim_{k \to +\infty} \widetilde{P}_{k|k} = (1 - \bar{\gamma})\underline{P}_\infty$. Also $\lim_{k \to +\infty} S_{k+1} = \lim_{k \to +\infty} S_k = S_\infty$. Finally from Equations (6.57) - (6.59) and the previous observations follow the claim.

$\square$

6.13 LQG control for UDP-like protocols

In this section equations for the optimal estimator and controller design for the case of communication protocols that do not provide any kind of acknowledgment of successful packet delivery (UDP-like). This case corresponds to the information set G_k, as defined in Equation (6.32). Some of the derivations are analogous to the previous section and are therefore skipped.

Estimator Design

We derive the equations for the optimal estimator using similar arguments to the standard Kalman filtering equations. The innovation step is given by:

$$
\begin{aligned}
\hat{x}_{k+1|k} &\triangleq \mathbb{E}[x_{k+1}|\mathcal{G}_k] = \mathbb{E}[Ax_k + \nu_k Bu_k + w_k|\mathcal{G}_k] \\
&= A\mathbb{E}[x_k|\mathcal{G}_k] + \mathbb{E}[\nu_k]Bu_k \\
&= A\hat{x}_{k|k} + \bar{\nu}Bu_k \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (6.63)
\end{aligned}
$$

$$
\begin{aligned}
e_{k+1|k} &\triangleq x_{k+1} - \hat{x}_{k+1|k} \\
&= Ax_k + \nu_k Bu_k + w_k - (A\hat{x}_{k|k} + \bar{\nu}Bu_k) \\
&= Ae_{k|k} + (\nu_k - \nu)Bu_k + w_k \quad\quad\quad\quad\quad (6.64)
\end{aligned}
$$

$$
\begin{aligned}
P_{k+1|k} &\triangleq \mathbb{E}[e_{k+1|k}e'_{k+1|k} \,|\mathcal{G}_k] \\
&= A\mathbb{E}[e_{k|k}e'_{k|k}|\mathcal{G}_k]A' + \mathbb{E}[(\nu_k - \nu)^2]Bu_k u'_k B' + \mathbb{E}[w_k w'_k] \\
&= AP_{k|k}A' + \bar{\nu}(1 - \bar{\nu})Bu_k u'_k B' + Q, \quad\quad\quad (6.65)
\end{aligned}
$$

where we used the independence and zero-mean of w_k, $(\nu_k - \bar{\nu})$, and $\mathcal{G}_k$, and the fact that u_k is a deterministic function of the information set $\mathcal{G}_k$. Note how under UDP-like communication, differently from TCP-like, the error covariance $P_{k+1|k}$ depends explicitly on the control input u_k. This is the main difference with control feedback systems under TCP-like protocols.

The correction step is the same as for the TCP case:

$$
\begin{aligned}
\hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + \gamma_{k+1}K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k}) \\
P_{k+1|k+1} &= P_{k+1|k} - \gamma_{k+1}K_{k+1}CP_{k+1|k}, \quad\quad\quad (6.66)
\end{aligned}
$$

$$
K_{k+1} \triangleq P_{k+1|k}C'(CP_{k+1|k}C' + R)^{-1}, \quad\quad\quad (6.67)
$$

where again we considered a time varying system with $A_k = A$ and $C_k = \gamma_k C$ as we did for the optimal estimator under TCP-like protocols.

Controller design: General case

In this section, we show that the optimal LQG controller, under UDP-like communication protocols, is in general not a linear function of the state estimate, and that the estimator and controller design cannot be separated anymore. To show this, we construct a counter-example considering a simple scalar system and we proceed using the dynamic programming approach. Let us consider the scalar system where $A = 1, B = 1, C = 1, W_N = W_k = 1, U_k = 0, R = 1, Q = 0$. Similarly to the TCP case, we define the value function, $V_k(x_k)$, as in Equations (6.44) where we just need to substitute the information set $\mathcal{F}_k$ with $\mathcal{G}_k$. For $k = N$, the value function is given by $V_N(x_N) = \mathbb{E}[x_N' W_N x_N \mid \mathcal{G}_N] = \mathbb{E}[x_N^2 \mid \mathcal{G}_N]$. For $k = N - 1$ we have:

$$
\begin{aligned}
V_{N-1}(x_{N-1}) &= \min_{u_{N-1}} \mathbb{E}[x_{N-1}^2 + V_N(x_N) \mid \mathcal{G}_{N-1}] \\
&= \min_{u_{N-1}} \mathbb{E}[x_{N-1}^2 + x_N^2 \mid \mathcal{G}_{N-1}] \\
&= \min_{u_{N-1}} \mathbb{E}[x_{N-1}^2 + (x_{N-1} + \nu_{N-1} u_{N-1})^2 \mid \mathcal{G}_{N-1}] \\
&= \min_{u_{N-1}} \left(\mathbb{E}[2x_{N-1}^2 \mid \mathcal{G}_{N-1}] + \mathbb{E}[\nu_{N-1}^2] u_{N-1}^2 + 2u_{N-1} \mathbb{E}[\nu_{N-1}] \mathbb{E}[x_{N-1} \mid \mathcal{G}_{N-1}] \right) \\
&= \min_{u_{N-1}} \left(\mathbb{E}[2x_{N-1}^2 \mid \mathcal{G}_{N-1}] + \bar{\nu} u_{N-1}^2 + 2\bar{\nu} u_{N-1} \hat{x}_{N-1|N-1} \right),
\end{aligned}
$$

where we used the independence of ν_{N-1} and $\mathcal{G}_{N-1}$, and the fact that u_{N-1} is a deterministic function of the information set $\mathcal{G}_{N-1}$. The cost is a quadratic function of the input u_{N-1}, therefore the minimizer can be simply obtained by finding $\frac{\partial V_{N-1}}{\partial u_{N-1}} = 0$, which is given by $u_{N-1}^* = -\hat{x}_{N-1|N-1}$. If we substitute back u_{N-1}^* into the value function we have:

$$
\begin{aligned}
V_{N-1}(x_{N-1}) &= \mathbb{E}[2x_{N-1}^2 \mid \mathcal{G}_{N-1}] - \bar{\nu} \hat{x}_{N-1|N-1}^2 \\
&= \mathbb{E}[(2 - \bar{\nu}) x_{N-1}^2 \mid \mathcal{G}_{N-1}] + \bar{\nu} P_{N-1|N-1}
\end{aligned}
$$

where we used Lemma 6.5(b). Before proceeding note that:

$$P_{N-1|N-1} = P_{N-1|N-2} - \gamma_{N-1} \frac{P_{N-1|N-2}^2}{P_{N-1|N-2} + 1}$$

$$= P_{N-1|N-2} - \gamma_{N-1} \left(P_{N-1|N-2} - 1 + \frac{1}{P_{N-1|N-2} + 1} \right)$$

$$= (1 - \gamma_{N-1}) \left(P_{N-2|N-2} + \bar{\nu}(1 - \bar{\nu})u_{N-2}^2 \right) + \gamma_{N-1} +$$

$$+ \gamma_{N-1} \frac{1}{P_{N-2|N-2} + \bar{\nu}(1 - \bar{\nu})u_{N-2}^2 + 1}$$

$$\mathbb{E}[P_{N-1|N-1}|\mathcal{G}_{N-2}] = (1 - \bar{\gamma}) \left(P_{N-2|N-2} + \bar{\nu}(1 - \bar{\nu})u_{N-2}^2 \right) + \bar{\gamma} + \bar{\gamma}\frac{1}{P_{N-2|N-2} + \bar{\nu}(1 - \bar{\nu})u_{N-2}^2 + 1}$$

$$\mathbb{E}[x_{N-1}^2|\mathcal{G}_{N-2}] = \mathbb{E}[(x_{N-2} + \nu_{N-2}u_{N-2})^2|\mathcal{G}_{N-2}]$$

$$= \mathbb{E}[x_{N-2}^2|\mathcal{G}_{N-2}] + 2\mathbb{E}[\nu_{N-2}]\mathbb{E}[x_{N-2}|\mathcal{G}_{N-2}]u_{N-2} + \mathbb{E}[\nu_{N-2}]u_{N-2}^2$$

$$= \mathbb{E}[x_{N-2}^2|\mathcal{G}_{N-2}] + 2\bar{\nu}\hat{x}_{N-2|N-2}u_{N-2} + \bar{\nu}u_{N-2}^2,$$

where we used Equations (6.65)-(6.67), and the fact that u_{N-2} and $P_{N-2|N-2}$ are a deterministic function of the information set $\mathcal{G}_{N-2}$. Using the previous equations we proceed to compute the value function for $k = N - 2$:

$$V_{N-2}(x_{N-2}) = \min_{u_{N-2}} \mathbb{E}[x_{N-2}^2 + V_{N-1}(x_{N-1}) \mid \mathcal{G}_{N-2}]$$

$$= \min_{u_{N-2}} \mathbb{E}[x_{N-2}^2 + (2 - \bar{\nu})x_{N-1}^2 + \bar{\nu}P_{N-1|N-1} \mid \mathcal{G}_{N-2}]$$

$$= (3 - \bar{\nu})\mathbb{E}[x_{N-2}^2|\mathcal{G}_{N-2}] + \bar{\nu}(1 - \bar{\gamma})P_{N-2|N-2} + \bar{\nu}\bar{\gamma} +$$

$$+ \min_{u_{N-1}} \left(2\bar{\nu}(2 - \bar{\nu})\hat{x}_{N-2|N-2}u_{N-2} + \bar{\nu}(2 - \bar{\nu})u_{N-2}^2 +$$

$$+ \bar{\nu}^2(1 - \bar{\nu})(1 - \bar{\gamma})u_{N-2}^2 + \bar{\nu}\bar{\gamma}\frac{1}{P_{N-2|N-2} + \bar{\nu}(1 - \bar{\nu})u_{N-2}^2 + 1} \right)$$

The first three terms inside the round parenthesis are convex quadratic functions of the control input u_{N-2}, however the last term is not. Therefore, the minimizer u_{N-2}^* is, in general, a non-linear function of the information set $\mathcal{G}_k$. The nonlinearity of the optimal controller arises from the fact that the correction error covariance matrix $P_{k+1|k+1}$ is a non-linear function of the innovation error covariance $P_{k+1|k}$, as it can be seen in Equations (6.66) and (6.67). The only case when $P_{k+1|k+1}$ is linear in $P_{k+1|k}$ is when measurement noise covariance $R = 0$ and the observation matrix C is square and invertible, from which follows that the optimal control is linear in the estimated states. However it is important to remark that the separation principle still does not hold even for this special case, since the control input affects the estimator error covariance.

We can summarize these results in the following theorem:

Theorem 6.10. *Let us consider the stochastic system defined in Equa-*

tions (6.29) with horizon $N \geq 2$. Then:

1. *The separation principle does not hold since the estimator error covariance depends on the control input, as shown in Equation (6.65).*

2. *The optimal control feedback $u_k = g_k^*(\mathcal{G}_k)$ that minimizes the cost functional defined in Equation (6.33) under UDP-like protocols is, in general, a non-linear function of information set $\mathcal{G}_k$.*

3. *The optimal control feedback $u_k = g_k^*(\mathcal{G}_k)$ is a linear function of the estimated state $\hat{x}_{k|k}$ if and only if the matrix C is invertible and there is no measurement noise.*

The next section will compute explicitly the optimal control for the special case and will give necessary and sufficient conditions for stability and performance of the infinite horizon scenario.

Special Case: R=0 and C invertible

Without loss of generality we can assume $C = I$, since the linear transformation $z = Cx$ would give an equivalent system where the matrix C is the identity. Let us now consider the case when there is no measurement noise, i.e. $R = 0$. These assumption mean that it is possible to measure the state x_k when a packet is delivered. In this case the estimator Equations (6.65)-(6.67) simplify as follows:

$$K_{k+1} = I \tag{6.68}$$

$$P_{k+1|k+1} = (1 - \gamma_{k+1})P_{k+1|k}$$
$$= (1 - \gamma_{k+1})(A'P_{k|k}A + Q + \bar{\nu}(1 - \bar{\nu})Bu_k u_k' B') \tag{6.69}$$

$$\mathbb{E}[P_{k+1|k+1}|\mathcal{G}_k] = (1 - \bar{\gamma})(A'P_{k|k}A + Q + \bar{\nu}(1 - \bar{\nu})Bu_k u_k' B') \tag{6.70}$$

where in the last equation we used independence of γ_{k+1} and $\mathcal{G}_k$, and we used the fact that $P_{k|k}$ is a deterministic function of $\mathcal{G}_k$.

Similarly to what done in the analysis of TCP-like optimal control, we claim that the value function $V_k^*(x_k)$ can be written as follows:

$$V_k(x_k) = \hat{x}_{k|k}' S_k \hat{x}_{k|k} + \text{trace}(T_k P_{k|k}) + \text{trace}(D_k Q) \tag{6.71}$$

for $k = N, \ldots, 0$. This is clearly true for $k = N$, in fact we have:

$$V_N(x_N) = \mathbb{E}[x_N' W_N x_N | \mathcal{G}_N] = \hat{x}_{N|N}' W_N \hat{x}_{N|N} + \text{trace}(W_N P_{N|N})$$

where we used Lemma 6.5(b), therefore the statement is satisfied by $S_N = W_N, T_N = W_N, D_N = 0$. Note that Equation (6.71) can be rewritten as follows:

$$V_k(x_k) = \mathbb{E}[x_k' S_k x_k | \mathcal{G}_k] + \text{trace}((T_k - S_k)P_{k|k}) + \text{trace}(D_k Q)$$

where we used once again Lemma 6.5(b). Moreover, to simplify notation we define $H_k \triangleq (T_k - S_k)$. Let us suppose that Equation (6.71) is true for $k+1$

and let us show by induction it holds true for k:

$$
\begin{aligned}
V_k(x_k) &= \min_{u_k} \mathbb{E}[x_k' W_k x_k + \nu_k u_k' U_k u_k + V_{k+1}(x_{k+1}) \mid \mathcal{G}_k] \\
&= \min_{u_k} \left(\mathbb{E}[x_k' W_k x_k + \nu_k u_k' U_k u_k + x_{k+1}' S_{k+1} x_{k+1} + \mathrm{trace}(H_{k+1} P_{k+1|k+1}) + \mathrm{trace}(D_{k+1} Q) \mid \mathcal{G}_k] \right) \\
&= \mathbb{E}[x_k'(W_k + A'S_{k+1}A)x_k | \mathcal{G}_k] + \mathrm{trace}(S_{k+1}Q) + (1-\bar{\gamma})\mathrm{trace}(H_{k+1}(A'P_{k|k}A+Q)) + \mathrm{trace}(D_{k+1}Q) + \\
&\quad + \min_{u_k} \left(\bar{\nu} u_k' U_k u_k + \bar{\nu} u_k' B' S_{k+1} B u_k + 2\bar{\nu} u_k' B' S_{k+1} A \hat{x}_{k|k} + \bar{\nu}(1-\bar{\nu})(1-\bar{\gamma})\mathrm{trace}(H_{k+1} B u_k u_k' B') \right) \\
&= \mathbb{E}[x_k'(W_k + A'S_{k+1}A)x_k | \mathcal{G}_k] + \mathrm{trace}((D_{k+1} + (1-\bar{\gamma})H_{k+1})Q) + (1-\bar{\gamma})\mathrm{trace}(AH_{k+1}A'P_{k|k}) + \\
&\quad + \mathrm{trace}(S_{k+1}Q) + \bar{\nu} \min_{u_k} \left(u_k'(U_k + B'(S_{k+1}+(1-\bar{\nu})(1-\bar{\gamma})H_{k+1})B)u_k + 2u_k' B' S_{k+1} A \hat{x}_{k|k} \right) \\
&= \hat{x}_{k|k}'(W_k + A'S_{k+1}A)\hat{x}_{k|k} + \mathrm{trace}((D_{k+1} + (1-\bar{\gamma})T_{k+1} + \bar{\gamma}S_{k+1})Q) + \\
&\quad + \mathrm{trace}((W_k + \bar{\gamma}A'S_{k+1}A + (1-\bar{\gamma})AT_{k+1}A')P_{k|k}) + \\
&\quad + \bar{\nu} \min_{u_k} \left(u_k'(U_k + B'((1-\bar{\alpha})S_{k+1} + \bar{\alpha}T_{k+1})B)u_k + 2u_k' B' S_{k+1} A \hat{x}_{k|k} \right),
\end{aligned}
$$

where we defined $\bar{\alpha} = (1-\bar{\nu})(1-\bar{\gamma})$, we used Lemma 6.5(c) to get the second equality, and Equations (6.36) and (6.70) to get the last equality. Since the quantity inside the big round parenthesis a convex quadratic function, the minimizer is the solution of $\frac{\partial V_k}{\partial u_k} = 0$ which is given by:

$$
u_k^* = \left(U_k + B'((1-\bar{\alpha})S_{k+1} + \bar{\alpha}T_{k+1})B \right)^{-1} B' S_{k+1} A \hat{x}_{k|k} \qquad (6.72)
$$

$$
= L_k \hat{x}_{k|k} \qquad (6.73)
$$

which is linear function of the estimated state $\hat{x}_{k|k}$. Substituting back into the value function we get:

$$
\begin{aligned}
V_k(x_k) &= \hat{x}_{k|k}'(W_k + A'S_{k+1}A)\hat{x}_{k|k} + \mathrm{trace}((D_{k+1} + (1-\bar{\gamma})T_{k+1} + \bar{\gamma}S_{k+1})Q) + \\
&\quad + \mathrm{trace}((W_k + A'S_{k+1}A + (1-\bar{\gamma})AT_{k+1}A')P_{k|k}) - \bar{\nu}\hat{x}_{k|k}'A'S_{k+1}BL_k\hat{x}_{k|k} \\
&= \hat{x}_{k|k}'(W_k + \bar{\gamma}A'S_{k+1}A - \bar{\nu}\hat{x}_{k|k}'A'S_{k+1}BL_k)\hat{x}_{k|k} + \mathrm{trace}((D_{k+1} + (1-\bar{\gamma})T_{k+1} + \bar{\gamma}S_{k+1})Q) + \\
&\quad + \mathrm{trace}((W_k + A'S_{k+1}A + (1-\bar{\gamma})AT_{k+1}A')P_{k|k}),
\end{aligned}
$$

where we used Lemma 6.5(b) in the last equality. From the last equation we see that the value function can be written as in Equation (6.71) if and only if the following equations are satisfied:

$$
\begin{aligned}
S_k &= A'S_{k+1}A + W_k - \bar{\nu}A'S_{k+1}B\left(U_k + B'((1-\bar{\alpha})S_{k+1} + \bar{\alpha}T_{k+1})B\right)^{-1}B'S_{k+1}A \\
&= \Phi_{\gamma,\nu}^S(S_{k+1}, T_{k+1}) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (6.74) \\
T_k &= (1-\bar{\gamma})A'T_{k+1}A + \bar{\gamma}A'S_{k+1}A + W_k \\
&= \Phi_{\gamma,\nu}^T(S_{k+1}, T_{k+1}) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (6.75) \\
D_k &= (1-\bar{\gamma})T_{k+1} + \bar{\gamma}S_{k+1} + D_{k+1} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (6.76)
\end{aligned}
$$

The optimal minimal cost for the finite horizon, $J_N^* = V_0(x_0)$ is then

given by:

$$J_N^* = \bar{x}_0' S_0 \bar{x}_0 + \text{trace}(S_0 P_0) + \sum_{k=1}^{N} \text{trace}\Big(\big((1-\bar{\gamma})T_k + \bar{\gamma}S_k\big)Q\Big) \qquad (6.77)$$

For the infinite horizon optimal controller, necessary and sufficient condition for the average minimal cost $J_\infty \triangleq \lim_{N\to+\infty} \frac{1}{N} J_N^*$ to be finite is that the coupled iterative Equations (6.74) and (6.75) should converge to a finite value S_∞ and T_∞ as $N \to +\infty$. In the work of Imer *et al.* [**?**] similar equations were derived for the optimal LQG control under UDP for the same framework with the additional conditions $Q = 0$ and B square and invertible. They find necessary and sufficient conditions for those equations to converge. Unfortunately, these conditions do not hold for the general case when B in not square. This is a very frequent situation in control systems, where in general we simply have (A, B) controllable.

Theorem 6.11. *Also, assume that the pair $(A, W^{1/2})$ is observable. Consider the following operator:*

$$\Upsilon(S, T, L) = A'SA + W + 2\bar{\nu}A'SBL + \bar{\nu}L'\Big(U + B'\big((1-\bar{\alpha})S + \bar{\alpha}T\big)B\Big)L \quad (6.78)$$

Then the following claims are equivalent:

1. *There exist a matrix $\tilde{L}$ and positive definite matrices $\tilde{S}$ and $\tilde{T}$ such that:*

$$\tilde{S} > 0, \ \tilde{T} > 0, \ \tilde{S} = \Upsilon(\tilde{S}, \tilde{T}, \tilde{L}), \ \tilde{T} = \Phi^T(\tilde{S}, \tilde{T})$$

2. *Consider the sequences:*

$$S_{k+1} = \Phi^S(S_k, T_k), \quad T_{k+1} = \Phi^T(S_k, T_k)$$

where the operators $\Phi^S(\cdot), \Phi^T(\cdot)$ are defined in Equations (6.74) and (6.75). For any initial condition $S_0, T_0 \geq 0$ we have

$$\lim_{k\to\infty} S_k = S_\infty, \quad \lim_{k\to\infty} T_k = T_\infty$$

and S_∞, T_∞ are the unique positive definite solution of the following equations

$$S_\infty > 0, \ T_\infty > 0, \ S_\infty = \Phi^S(S_\infty, T_\infty), \quad T_\infty = \Phi^T(S_\infty, T_\infty)$$

The convergence of Equations (6.74) and (6.75) depend on the control and observation arrival probabilities $\bar{\gamma}, \bar{\nu}$. General analytical conditions for convergence are not available, but some necessary and sufficient conditions can be found.

Lemma 6.9. *Let us consider the fixed points of Equations (6.74) and (6.75), i.e. $S = \Phi^S(S, T), T = \Phi^T(S, T)$ where $S, T \geq 0$. Let A be unstable. A necessary condition for existence of solution is*
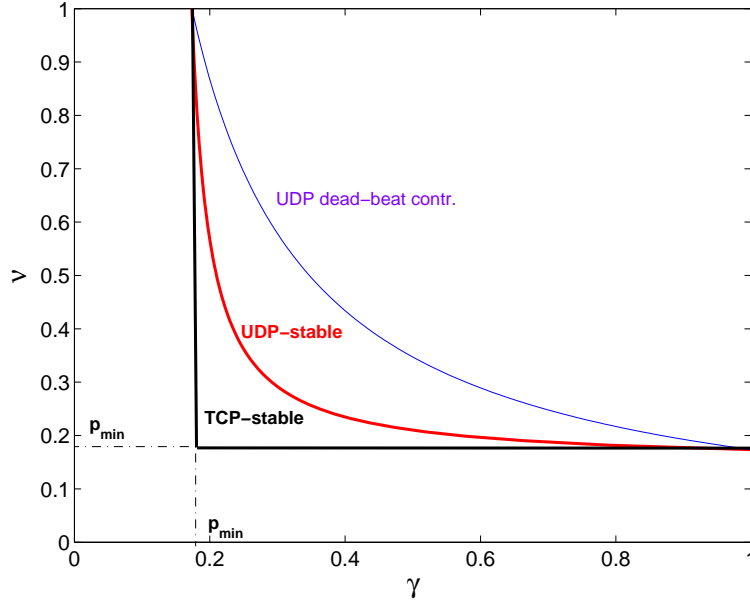
$$|A|^2(\bar{\gamma} + \bar{\nu} - 2\bar{\gamma}\bar{\nu}) < \bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} \qquad (6.79)$$

where $|A| \triangleq \max_i |\lambda_i(A)|$ is the largest eigenvalue of the matrix A.

Lemma 6.10. *Let us consider the fixed points of Equations (6.74) and (6.75), i.e. $S = \Phi^S(S,T), T = \Phi^T(S,T)$ where $S, T \geq 0$. Let A be unstable, $(A, W^{1/2})$ observable and B square and invertible. Then a sufficient condition for existence of solution is*

$$|A|^2(\bar{\gamma} + \bar{\nu} - 2\bar{\gamma}\bar{\nu}) < \bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} \qquad (6.80)$$

where $|A| \triangleq \max_i |\lambda_i(A)|$ is the largest eigenvalue of the matrix A.

Figure 6.10: Region of convergence for UDP-like and TCP-like optimal control in the scalar case. The bounds are tight in the scalar case. The thin solid line corresponds to the boundary of the stability region for a dead-beat controller under UDP-like protocols as given by [**?**], which is much more restrictive than what can be achieved with optimal UDP controllers.

A graphical representation of the stability bounds are shown in Figure 6.10, where we considered a scalar system with parameters $|A| = 1.1$, which gives the critical probability $p_{min} = 1 - 1/|A|^2 = 1.173$ as defined in Theorem 6.8. The critical arrival probabilities for TCP-like optimal control are $\gamma_c = \nu_c = p_{min}$. The boundary for the stability region of optimal control under UDP-like protocols given in Lemma 6.10 can be written also as $\bar{\nu} > \frac{\bar{\gamma}(A^2-1)}{\bar{\gamma}(2A^2-1)+1-A^2}$ for $\bar{\gamma} > p_{min}$. It is important to remark that the stability region of optimal control under UDP-like protocols is larger than the stability region obtained using a dead-beat controller proposed in [**?**], i.e. $u_k = -\gamma_k B^{-1} A y_k = -\gamma_k B^{-1} A x_k$, which is given by $\bar{\gamma}\bar{\nu} > 1 - 1/|A|^2$ and

graphically shown in Figure 6.10 . This is not surprising since the dead-beat controller is rather aggressive and requires a large gain L, which increases the estimator error covariance in Equation (6.70). Indeed, as shown in the constructive proof of Lemma 6.10, controllers with similar structure but smaller gains, i.e. $u_k = -\eta\gamma_k B^{-1} Ay_k = -\eta\gamma_k B^{-1} Ax_k$ where $\eta < 1$, have a larger region of stability.

We can summarize the results of this section in the following theorem

Theorem 6.12. *Consider the system (6.29)-(6.31) and consider the problem of minimizing the cost function (6.33) within the class of admissible policies $u_k = f(\mathcal{G}_k)$, where $\mathcal{G}_k$ is the information available under TCP-like schemes, given in Equation (6.32). Assume also that $R = 0$ and C is square and invertible. Then:*

1. *The optimal estimator gain is constant and in particular $K_k = I$ if $C = I$.*

2. *The infinite horizon optimal control exists if and only if there exists positive definite matrices $S_\infty, T_\infty > 0$ such that $S_\infty = \Phi^S(S_\infty, T_\infty)$ and $T_\infty = \Phi^T(S_\infty, T_\infty)$, where Φ^T and Φ^S are defined in Equations (6.74) and (6.75).*

3. *The infinite horizon optimal controller gain is constant:*

$$\lim_{k\to\infty} L_k = L_\infty = -(B'(\bar{\alpha}T_\infty + (1-\bar{\alpha})S_\infty)B + U)^{-1}B'S_\infty A \qquad (6.81)$$

4. *A necessary condition for existence of $S_\infty, T_\infty > 0$ is*

$$|A|^2(\bar{\gamma} + \bar{\nu} - 2\bar{\gamma}\bar{\nu}) < \bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} \qquad (6.82)$$

where $|A| \triangleq \max_i |\lambda_i(A)|$ is the largest eigenvalue of the matrix A. This condition is also sufficient if B is square and invertible.

5. *The expected minimum cost converges:*

$$J_\infty^* = \lim_{k\to\infty} \frac{1}{N}J_N^* = \mathrm{trace}\big((1-\bar{\gamma})T_\infty + \bar{\gamma}S_\infty)Q\big) \qquad (6.83)$$

In the scenario considered in this section when $R = 0$ and C is invertible, it is possible to directly compare the performance of optimal control under TCP-like and UDP-like protocols in terms of the infinite horizon cost J_∞^*. Let us consider for example the scalar system with the following parameters $A = 1.1, B = C = Q = W = U = 1, R = 0$. For simplicity also consider symmetric communication channels for sensor reading and control inputs, i.e. $\bar{\nu} = \bar{\gamma}$. Using results from Theorem 6.9 and Theorem 6.12 we can compute the infinite horizon cost using optimal controllers under UDP-like and an upper bound on the cost under TCP-like communication protocols, which are shown in Fig. 6.11. As expected optimal control performance under TCP-like is better than UDP-like, however the two curves are comparable for moderate packet loss. Although the TCP-like curve is only an upper bound of the true expected cost, it has been observed to be rather close to the empirical cost [**?**]. The observation that TCP-like and UDP-like
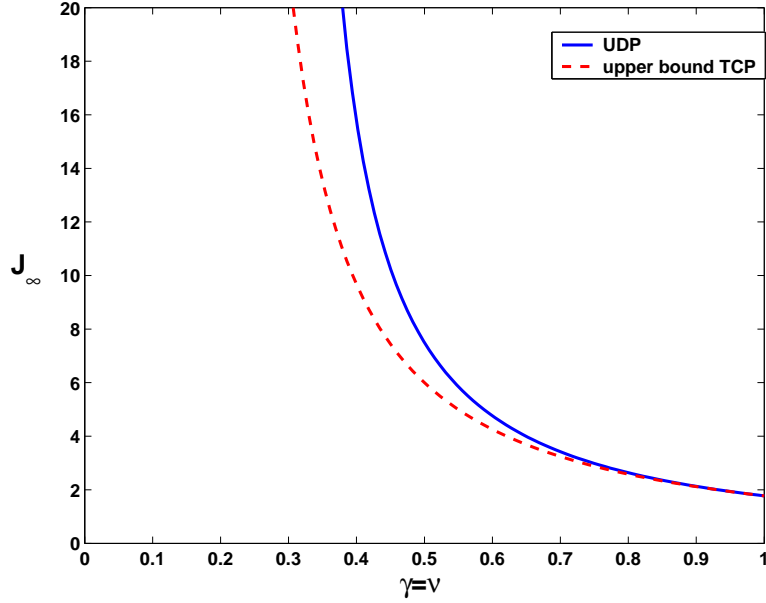
Figure 6.11: Exact infinite horizon cost using optimal LQG control under UDP-like and upper bound under TCP-like communication protocols in the scalar case.

optimal control performances seem remarkably close is extremely valuable since UDP-like protocols are much simpler to implement than TCP-like.

6.14 Appendix

Lemma 6.11. *Let* $S, T \in \mathbb{M} = \{M \in \mathbb{R}^{n \times n} | M \geq 0\}$. *Consider the operators* $\Phi^S(S, T)$, *and* $\Phi^T(S, T)$ *as defined in Equations (6.74) and (6.75), and consider the sequences* $S_{k+1} = \Phi^S(S_k, T_k)$ *and* $T_{k+1} = \Phi^T(S_k, T_k)$. *Consider* $L^*_{S,T} = -\big(U + B'\big((1 - \bar{\alpha})S + \bar{\alpha}T\big)B\big)^{-1} B'SA$. *operators*
Then the following facts are true:

1.

$$\Upsilon(S, T, L) = (1 - \frac{\bar{\nu}}{1 - \bar{\alpha}})A'SA + W + \frac{\bar{\nu}}{1 - \bar{\alpha}}\big(A + (1 - \bar{\alpha})BL\big)'S\big(A + (1 - \bar{\alpha})BL\big) + \bar{\nu}L'UL + \bar{\nu}\bar{\alpha}L'B'TBL$$

2. $\Phi^S(S, T) = \min_L \Upsilon(S, T, L)$

3. $0 \leq \Upsilon(S, T, L^*_{S,T}) = \Phi^S(S, T) \leq \Upsilon(S, T, L) \ \forall L$

4. *If* $S_{k+1} > S_k$ *and* $T_{k+1} > T_k$, *then* $S_{k+2} > S_{k+1}$ *and* $T_{k+2} > T_{k+1}$.

5. *If the pair* $(A, W^{1/2})$ *is observable and* $S = \Phi^S(S, T)$ *and* $T = \Phi^T(S, T)$, *then* $S > 0$ *and* $T > 0$.

Proof. Fact (a) can be easily checked by direct substitution.

(b) If U is invertible then it is easy to verify by substitution that

$$\Upsilon(S,T,L) = \Phi^S(S,T) + \bar{\nu}(L - L^*_{S,T})'\Big(U + B'\big((1-\bar{\alpha})S + \bar{\alpha}T\big)B\Big)(L - L^*_{S,T})$$
$$\geq \Phi^S(S,T)$$

(c) The non-negativeness follows from the observation that $\Upsilon(S,T,L)$ is a sum of positive semi-definite matrices. In fact $(1 - \frac{\bar{\nu}}{1-\bar{\alpha}}) = \frac{\bar{\gamma}(1-\bar{\nu})}{\bar{\nu}+\bar{\gamma}(1-\bar{\nu})} \geq 0$ and $0 \leq \bar{\alpha} \leq 1$.
The equality $\Upsilon(S,T,L^*_{S,T}) = \Phi^S(S,T)$ can be verified by direct substitution. The last inequality follows directly from Fact (b).

(d)

$$\begin{aligned}
S_{k+2} &= \Phi^S(S_{k+1}, T_{k+1}) = \Upsilon(S_{k+1}, T_{k+1}, L^*_{S_{k+1},T_{k+1}}) \\
&\geq \Upsilon(S_k, T_k, L^*_{S_{k+1},T_{k+1}}) \geq \Upsilon(S_k, T_k, L^*_{S_k,T_k}) \\
&= \Phi^S(S_k, T_k) = S_{k+1} \\
T_{k+2} &= \Phi^T(S_{k+1}, T_{k+1}) \geq \Phi^T(S_k, T_k) = T_{k+1}
\end{aligned}$$

(e) First observe that $S = \Phi^S(S,T) \geq 0$ and $T = \Phi^T(S,T) \geq 0$. Thus, to prove that $S,T > 0$, we only need to establish that S,T are nonsingular. Suppose they are singular, the there exist vectors $0 \neq v_s \in \mathcal{N}(S)$ and $0 \neq v_t \in \mathcal{N}(T)$, i.e. $Sv_s = 0$ and $Tv_t = 0$, where $\mathcal{N}(\cdot)$ indicates the null space. Then

$$\begin{aligned}
0 &= v'_s S v_s = v'_s \Phi^S(S,T) v_s = v'_s \Upsilon(S,T,L^*_{S,T}) v_s \\
&= (1 - \tfrac{\bar{\nu}}{1-\bar{\alpha}}) v'_s A' S A v_s + v'_s W v_s + \star
\end{aligned}$$

where $\star$ indicates other terms. Since all the terms are positive semi-definite matrices, this implies that all the term must be zero:

$$\begin{aligned}
v'_s A' S A v_s = 0 &\implies S A v_s = 0 \implies A v_s \in \mathcal{N}(S) \\
v'_s W v_s = 0 &\implies W^{1/2} v_s = 0
\end{aligned}$$

As a result, the null space $\mathcal{N}(S)$ is A-invariant. Therefore, $\mathcal{N}(S)$ contains an eigenvector of A, i.e. there exists $u \neq 0$ such that $Su = 0$ and $Au = \sigma u$. As before, we conclude that $Wu=0$. This implies (using the PBH test) that the pair $(A, W^{1/2})$ is not observable, contradicting the hypothesis. Thus, $\mathcal{N}(S)$ is empty, proving that $S > 0$. The same argument can be used to prove that also $T > 0$. $\qquad\square$

Proof of Theorem 6.11

(a)$\Rightarrow$(b) The main idea of the proof consists in the proving of the convergence of several monotonic sequences. Consider the sequences $V_{k+1} = \Upsilon(V_k, Z_k, \tilde{L})$ and $Z_{k+1} = \Phi^T(V_k, Z_k)$ with initial conditions $V_0 = Z_0 = 0$. It is easy to verify by substitution that $V_1 = W + \bar{\nu}\tilde{L}'U\tilde{L} \geq 0 = V_0$ and $Z_1 = W \geq 0 = Z_0$. Lemma 6.11(a) shows that the operator $\Upsilon(V, Z, \tilde{L})$ is linear and monotonically increasing in V and Z, i.e.

$(V_{k+1} \geq V_k, Z_{k+1} \geq Z_k) \Rightarrow (V_{k+2} \geq V_{k+1}, Z_{k+2} \geq Z_{k+1})$. Also the operator $\Phi^T(V, Z)$ is linear and monotonically increasing in V and Z. Since $V_1 \geq V_0$ and $Z_1 \geq Z_0$, using an induction argument we have that $V_{k+1} \geq V_k, Z_{k+1} \geq Z_k$ for all time k, i.e. the sequences are monotonically increasing. These sequences are also bounded, in fact $(V_0 \leq \tilde{S}), (Z_0 \leq \tilde{T}) \Rightarrow (V_1 = \Upsilon(0, 0, \tilde{L}) \leq \Upsilon(\tilde{S}, \tilde{T}, \tilde{L}) = \tilde{S}), (Z_1 = \Phi^T(0, 0) \leq \Phi^T(\tilde{S}, \tilde{T}) = \tilde{T})$ and the same argument can be inductively used to show that $V_k \leq \tilde{S}$ and $Z_k \leq \tilde{T}$ for all K. Consider now the sequences S_k, T_k as defined in the theorem initialized with $S_0 = T_0 = 0$. By direct substitution we find that $S_1 = W \geq 0 = S_0$ and $T_1 = W \geq 0 = T_0$. By Lemma 6.11(d) follows that the sequences S_k, T_k are monotonically increasing. Moreover, by Lemma 6.11(c) it follows that $(S_k \leq V_k, T_k \leq Z_k) \Rightarrow (S_{k+1} = \Phi^S(S_k, T_k) \leq \Upsilon(S_k, T_k, \tilde{L}) \leq \Upsilon(V_k, Z_k, \tilde{L}) = V_{k+1}), T_{k+1} = \Phi^T(S_k, T_k) \leq \Phi^T(V_k, Z_k) = Z_{k+1})$. Since this is verified for $k = 0$, it inductively follows that $(S_k \leq V_k, T_k \leq Z_k)$ for all k. Finally since V_k, Z_k are bounded, we have that $(S_k \leq \tilde{S}, T_k \leq \tilde{T}$. Since $S_k, T_k)$ are monotonically increasing and bounded, it follows that $\lim_{k \to \infty} S_k = S_\infty$ and $\lim_{k \to \infty} T_k = T_\infty$, where S_∞, T_∞ are semi-definite matrices. From this it easily follows that these matrices have the property $S_\infty = \Phi^S(S_\infty, T_\infty), T_\infty = \Phi^T(S_\infty, T_\infty)$. Definite positiveness of S_∞ follows from Lemma 6.11(e) using the hypothesis that $(A, W^{1/2})$ is observable. The same argument can be used to prove that $T_\infty > 0$. Finally proof of uniqueness of solution and convergence for all initial conditions S_0, T_0 can be obtained similarly to Theorem 1 in [?] and it is therefore omitted.

(b)$\Rightarrow$(a)

This part follows easily by choosing $\tilde{L} = L^*_{S_\infty, T_\infty}$, where L^* is defined in Lemma 6.11. Using Lemma 6.11(c) we have $S_\infty = \Phi^S(S_\infty, T_\infty) = \Upsilon(S_\infty, T_\infty, \tilde{L})$, therefore the statement is verified using $\tilde{S} = S_\infty$ and $\tilde{T} = T_\infty$.

Proof of Lemma 6.9

To prove the necessity condition, it is sufficient to show that there exist some initial conditions $S_0, T_0 \geq 0$ for which the sequences $S_{k+1} = \Phi^S(S_k, T_k), T_{k+1} = \Phi^T(S_k, T_k)$ are unbounded, i.e. $\lim_{k \to \infty} S_k = \lim_{k \to \infty} T_k = \infty$. To do so, suppose that at some time-step k we have $S_k \geq s_k vv'$ and $T_k \geq t_k vv'$, where $s_k, t_k > 0$, and v is the eigenvector corresponding to the largest eigenvalue of A', i.e. $A'v = \lambda_{max}v$ and $|\lambda_{max}| = |A'| = |A|$. Then we

have:

$$
\begin{aligned}
S_{k+1} = \Phi^S(S_k, T_k) &\geq \Phi^S(s_k vv', t_k vv') \\
&= \min_L \Upsilon(s_k vv', t_k vv', L) \\
&= \min_L \Big(s_k A' vv' A + W + 2 s_k \bar{\nu} A' vv' B L + \\
&\qquad\qquad + \bar{\nu} L' \big(U + B' \big((1 - \bar{\alpha}) s_k vv' + \bar{\alpha} t_k vv' \big) B \big) L \Big) \\
&\geq \min_L \Big(s_k |A|^2 vv' + 2 s_k \bar{\nu} \lambda_{max} vv' B L + \\
&\qquad\qquad + \bar{\nu} L' B' \big((1 - \bar{\alpha}) s_k vv' + \bar{\alpha} t_k vv' \big) B L \Big) \\
&= \min_L \Big(s_k |A|^2 vv' - \frac{|A|^2 \bar{\nu} s_k^2}{\xi_k} vv' + \\
&\qquad\qquad + \bar{\nu} \xi_k (\lambda_{max} s_k^2 I + \frac{1}{\xi_k} B L)' vv' (\lambda_{max} s_k^2 I + \frac{1}{\xi_k} B L) \Big) \\
&\geq s_k |A|^2 vv' - \frac{|A|^2 \bar{\nu} s_k^2}{(1 - \bar{\alpha}) s_k + \bar{\alpha} t_k} vv' \\
&= |A|^2 s_k \Big(1 - \frac{\bar{\nu} s_k}{(1 - \bar{\alpha}) s_k + \bar{\alpha} t_k} \Big) vv' \\
&= s_{k+1} vv'
\end{aligned}
$$

where I is the identity matrix and $\xi_k = (1 - \bar{\alpha}) s_k + \bar{\alpha} t_k$. Similarly we have:

$$
\begin{aligned}
T_{k+1} = \Phi^T(S_k, T_k) &\geq \Phi^T(s_k vv', t_k vv') \\
&= (1 - \bar{\gamma}) t_k A' vv' A + \bar{\gamma} s_k A' vv' A + W \\
&\geq (1 - \bar{\gamma}) t_k |A^2| vv' + \bar{\gamma} s_k |A|^2 vv' \\
&= |A|^2 \big((1 - \bar{\gamma}) t_k + \bar{\gamma} s_k \big) vv' \\
&= t_{k+1} vv'
\end{aligned}
$$

We can summarize the previous results as follows:

$$
\begin{aligned}
(S_k vv', T_k \geq t_k vv') &\Rightarrow (S_{k+1} \geq s_{k+1} vv', T_{k+1} \geq t_{k+1} vv') \\
s_{k+1} = \phi^s(s_k, t_k) &= |A|^2 s_k \Big(1 - \frac{\bar{\nu} s_k}{(1 - \bar{\alpha}) s_k + \bar{\alpha} t_k} \Big), \\
t_{k+1} = \phi^t(s_k, t_k) &= |A|^2 \big((1 - \bar{\gamma}) t_k + \bar{\gamma} s_k \big)
\end{aligned}
$$

Let us define the following sequences:

$$
\begin{aligned}
S_{k+1} = \Phi^S(S_k, T_k), \quad &T_{k+1} = \Phi^T(S_k, T_k), \quad S_0 = T_0 = vv' \\
s_{k+1} = \phi^s(s_k, t_k), \quad &t_{k+1} = \phi^t(s_k, t_k), \quad s_0 = t_0 = 1 \\
\tilde{S}_k = s_k vv', \quad &\tilde{T}_k = t_k vv'
\end{aligned}
$$

From the previous derivations, we have that $S_k \geq \tilde{S}_k, T_k \geq \tilde{T}_k$ for all time k. Therefore, it is sufficient to find when the scalar sequences s_k, t_k diverges to find the necessary conditions. It should be evident also that the operators $\phi^s(s,t), \phi^t(s,t)$ are monotonic in their arguments. Also, it should be clear that the only fixed points of $s = \phi^s(s,t), t = \phi^t(s,t)$ are $s = t = 0$. Therefore we should find when the origin is an unstable equilibrium point, since in this case $\lim_{k \to \infty} s_k, t_k = \infty$. Note that $t = \phi^t(s,t)$ can be written as:

$$t = \Phi^T(s,t) = (1 - \bar{\gamma})|A|^2 t + \bar{\gamma}|A|^2 s$$
$$= \psi(s) = \frac{\bar{\gamma}|A|^2 s}{1 - (1 - \bar{\gamma})|A|^2}$$

with the additional assumption $1 - (1 - \bar{\gamma})A^2 > 0$. A necessary condition for the stability of the origin is that the origin of restricted map $z_{k+1} = \phi(z_k, \psi(z_k))$ is stable. The restricted map is given by:

$$z_{k+1} = |A|^2 z_k \left(1 - \bar{\nu} \frac{z_k}{(1 - \bar{\alpha})z_k + \bar{\alpha}\frac{\bar{\gamma}|A|^2}{1-(1-\bar{\gamma})A^2}z_k} \right)$$
$$= |A|^2 \left(1 - \frac{\bar{\nu}}{(1 - \bar{\alpha}) + \bar{\alpha}\frac{\bar{\gamma}|A|^2}{1-(1-\bar{\gamma})A^2}} \right) z_k$$
$$= |A|^2 \left(1 - \frac{\bar{\nu}(1 - (1 - \bar{\gamma})|A|^2)}{\bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2} \right) z_k$$
$$= \left(\frac{\bar{\gamma}(1 - \bar{\nu})|A|^2}{\bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2} \right) z_k$$

This is a linear map and it is stable only if the term inside the parenthesis is smaller than unity, i.e.

$$\left(\frac{\bar{\gamma}(1 - \bar{\nu})|A|^2}{\bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2} \right) < 1$$
$$\bar{\gamma}(1 - \bar{\nu})|A|^2 < \bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2$$
$$|A|^2(\bar{\gamma} + \bar{\nu} - 2\bar{\gamma}\bar{\nu}) < \bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu}$$

which concludes the lemma.

Proof of Lemma 6.10

The proof is constructive. In fact, we find a control feedback gain $\tilde{L}$ that satisfies the conditions stated in Theorem 6.11(a). Let $\tilde{L} = -\eta B^{-1}A$ where $\eta > 0$ is a positive scalar that is to be determined. Also consider $S = sI, T = tI$, where I is the identity matrix and $s, t > 0$ are positive scalars. Then we

have

$$\Upsilon(sI, tI, \tilde{L}) = A'sA + W - 2\bar{\nu}\eta A'sA + \bar{\nu}A'B^{-'}UB^{-1}A +$$
$$+\bar{\nu}\eta^2 A'\big((1 - \bar{\alpha})s + \bar{\alpha}t\big)A$$
$$\leq |A|^2\Big(s - 2\bar{\nu}s\eta + \bar{\nu}\big((1 - \bar{\alpha})s + \bar{\alpha}t\big)\eta^2\Big)I + wI$$
$$= \varphi^s(s, t, \eta)I \qquad\qquad (6.84)$$
$$\Phi^T(sI, tI) = \bar{\gamma}A'sA + (1 - \bar{\gamma})A'tA + W$$
$$\leq \big(\bar{\gamma}|A|^2 s + (1 - \bar{\gamma})|A|^2 t\big)I + wI$$
$$\leq \varphi^t(s, t)I \qquad\qquad (6.85)$$

where $w = |W + \bar{\nu}A'B^{-'}UB^{-1}A| > 0$ and I is the identity matrix. Let us consider the following scalar operators and sequences:

$$\varphi^s(s, t, \eta) = |A|^2(1 - 2\bar{\nu}\eta + \bar{\nu}(1 - \bar{\alpha})\eta^2)s + \bar{\nu}\bar{\alpha}\eta^2 t + w$$
$$\varphi^t(s, t) = \bar{\gamma}|A|^2 s + (1 - \bar{\gamma})|A|^2 t + w$$
$$s_{k+1} = \varphi^s(s_k, t_k, \eta), \quad t_{k+1} = \varphi^t(s_k, t_k), \quad s_0 = t_0 = 0$$

The operators are clearly monotonically increasing in s, t, and since $s_1 = \varphi^s(s_0, t_0, \eta) = w \geq s_0$ and $t_1 = \varphi^t(s_0, t_0) = w \geq t_0$, it follows that the sequences s_k, t_k are monotonically increasing. If these sequences are bounded, then they must converge to $\tilde{s}, \tilde{t}$. Therefore s_k, t_k are bounded if and only if there exist $\tilde{s}, \tilde{t} > 0$ such that $\tilde{s} = \varphi^s(\tilde{s}, \tilde{t}, \eta)$ and $\tilde{t} = \varphi^t(\tilde{s}, \tilde{t})$. Let us find the fixed points:

$$\tilde{t} = \varphi^t(\tilde{s}, \tilde{t}) \Rightarrow$$
$$\tilde{t} = \frac{\bar{\gamma}|A|^2}{1 - (1 - \bar{\gamma})|A|^2}\tilde{s} + w_t$$

where $w_t \triangleq \frac{w}{1 - (1 - \bar{\gamma})|A|^2} > 0$, and we must have $1 - (1 - \bar{\gamma})|A|^2 > 0$ to guarantee that $\tilde{t} > 0$. Substituting back into the operator φ^s we have:

$$\tilde{s} = |A|^2(1 - 2\bar{\nu}\eta + \bar{\nu}(1 - \bar{\alpha})\eta^2)\tilde{s} + \bar{\nu}\bar{\alpha}\eta^2\frac{\bar{\gamma}|A|^2}{1 - (1 - \bar{\gamma})|A|^2}\tilde{s} +$$
$$+\bar{\nu}\bar{\alpha}\eta^2 w_t + w$$
$$= |A|^2\left(1 - 2\bar{\nu}\eta + \bar{\nu}\Big((1 - \bar{\alpha}) + \frac{\bar{\gamma}\bar{\alpha}|A|^2}{1 - (1 - \bar{\gamma})|A|^2}\Big)\eta^2\right)\tilde{s} + w(\eta)$$
$$= |A|^2\left(1 - 2\bar{\nu}\eta + \bar{\nu}\frac{\bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2}{1 - (1 - \bar{\gamma})|A|^2}\eta^2\right)\tilde{s} + w(\eta)$$
$$= a(\eta)\tilde{s} + w(\eta)$$

where $w(\eta) \triangleq \bar{\nu}\bar{\alpha}\eta^2 w_t + w > 0$. For a positive solution $\tilde{s}$ to exist, we must have $a(\eta) < 1$. Since $a(\eta)$ is a convex function of the free parameter η, we can try to increase the basin of existence of solutions by choosing

$\eta^* = \operatorname{argmin}_\eta a(\eta)$, which can be found by solving $\frac{da}{d\eta}(\eta^*) = 0$ and is given by:

$$\eta^* = \frac{1 - (1 - \bar{\gamma})|A|^2}{\bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2}$$

Therefore a sufficient condition for existence of solutions are given by:

$$a(\eta^*) < 1$$

$$|A|^2 \left(1 - \bar{\nu}\frac{1 - (1 - \bar{\gamma})|A|^2}{\bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2}\right) < 1$$

$$\left(\frac{\bar{\gamma}(1 - \bar{\nu})|A|^2}{\bar{\gamma} + \bar{\nu} - \bar{\gamma}\bar{\nu} - \bar{\nu}(1 - \bar{\gamma})|A|^2}\right) < 1$$

which is the same bound for the necessary condition of convergence in Lemma 6.9.

If this condition is satisfied then $\lim_{k\to\infty} s_k = \tilde{s}$ and $\lim_{k\to\infty} t_k = \tilde{t}$. Let us consider now the sequences $\bar{S}_k = s_k I$, $\bar{T}_k = t_k I$, $S_{k+1} = \Upsilon(S_k, T_k, \tilde{L})$ and $T_{k+1} = \Phi^T(S_k, T_k)$, where $\tilde{L} = -\eta^* B^{-1}A$, $S_0 = T_0 = 0$, and s_k, t_k where defined above. These sequences are all monotonically increasing. From Equations (6.84) and (6.85) it follows that $(S_k \leq s_k I, T_k \leq t_k I) \Rightarrow (S_{k+1} =\leq s_{k+1}I, T_{k+1} \leq t_k I)$. Since this is verified for $k = 0$ we can claim that $S_k < \tilde{s}I$ and $T_k < \tilde{t}I$ for all k. Since S_k, T_k are monotonically increasing and bounded, then they must converge to positive semi-definite matrices $\tilde{S}, \tilde{T} \geq 0$ which solve the equations $\tilde{S} = \Upsilon(\tilde{S}, \tilde{T}, \tilde{L})$ and $\tilde{T} = \Phi^T(\tilde{S}, \tilde{T})$. Since, by hypothesis, the pair $(A, W^{1/2})$ is observable, using similar arguments of Lemma 6.11(e), it is possible to show that $\tilde{S}, \tilde{T} > 0$. Therefore $\tilde{S}, \tilde{T}, \tilde{L}$ satisfy the conditions of statement (a) Theorem 6.11, from which if follows statement (b) of the same theorem. This implies that the sufficient conditions derived here guarantee the claim of the lemma.

6.15 Introduction

In the second part of the chapter, we will consider the two block design with an analog erasure channel inside the control loop. As discussed earlier, the analog erasure model (also referred to as the packet erasure or packet loss model) can be described as follows. The channel operates in discrete time steps. At every time step, the channel accepts as input a finite dimensional real vector $r(k)$. The value of the output of the channel $y(k)$ is chosen according to an *erasure process*. At every time step, the erasure process assumes either the value T or the value R. If the value at time k is T, $y(k + 1) = r(k)$ and a successful transmission is said to have occurred. Otherwise, $y(k + 1) = \phi$ and an erasure event, or a packet loss, is said to have occurred at time k. The symbol ϕ denotes that the receiver does not receive any data; however, the receiver is aware that an erasure event

has occurred at that time. Note that we have assumed that the channel introduces a constant delay of one time step.

While an analog erasure model has an infinite capacity in an information theoretic sense, it is often a useful representation for the cases when the communication protocols allow for large data packets to be transmitted at every time step. For instance, the minimum size of an ethernet data packet is 72 bytes. This is much more space for carrying information than usually required inside a control loop. If the data packets allow for transmission of control and sensing data to a high fidelity, the quantization effects are often ignored and an analog erasure model adopted.

To begin with, consider a set-up in which the linear time invariant process evolves as

$$x(k+1) = Ax(k) + Bu(k) + w(k),$$

where the state $x(k) \in \mathbf{R}^n$, the control variable $u(k) \in \mathbf{R}^m$ and $w(k)$ is process noise considered to be white and Gaussian with zero mean and covariance $R_w > 0$. The initial condition $x(0)$ is assumed to be Gaussian with zero mean and covariance $P(0)$. The process is observed using a sensor of the form

$$y(k) = Cx(k) + v(k),$$

where $v(k)$ is measurement noise that is again white Gaussian with zero mean and covariance $R_v > 0$. We suppose all the sources of randomness in the system (initial condition, process and measurement noise, and the erasure process for the channels) to be independent. The inputs are chosen to minimize the cost

$$J_{LQG} = E \left[\sum_{k=1}^{K} \left(x^T(k)Qx(k) + u^T(k)Ru(k) \right) + x^T(K+1)P(K+1)x(K+1) \right],$$

where the expectation at time k is taken with respect to the future values of the packet erasure events, the initial condition, and the measurement and process noises. Further, the matrices $P(K+1)$, Q and R are all assumed to be positive definite. The pair (A, B) is assumed to be controllable, and the pair (A, C) is assumed to be observable.

As with the one-block design, we can consider two cases:

1. When there is only one channel in the control loop, present between the sensor and the controller. Such a situation can arise, e.g., when the controller is co-located with the process and the sensor is remote, or the controller has access to large transmission power.

2. When there are two channels present. In addition to the sensor-controller channel, there is an additional channel between the controller and the actuator. In this case, it is also important to specify the action that the actuator takes when it does not receive a packet. The action depends on the amount of processing, memory and information

> about the process that is assumed to be available at the actuator. We consider the simplest choice, which is to apply zero control input if no packet was received. Other actions by the actuator can be treated in a similar fashion.

For most of the following discussion we assume that the erasures on the two channels occur in an i.i.d. fashion, independently of each other. However, erasures according to a Markov model, or correlated across the channels can be easily considered.

Two block design

As discussed earlier, two block design involves designing both an encoder at the input of a channel, and a decoder at the output to minimize the estimation / control cost. Note that the decoder for the sensor-controller channel and the encoder for the controller-actuator channel are merely the controller.

To begin with, we will consider only the sensor-controller channel to be present. To proceed, we must define the class of encoders that we will consider. The information theoretic capacity of an analog erasure channel is infinite. Thus, the only constraints we impose on the encoder are that the transmitted vector is some causal (possibly time-varying) function of the measurements available to the encoder until time k and that the dimension of the vector is finite. We will sometimes refer to the encoder as an encoding algorithm. For the controller-actuator channel, the choice of decoder will depend on the knowledge and processing available at the actuator. We will consider the case of this channel being present briefly towards the end.

At every time step k, the encoder at the sensor calculates a vector $s(k) = f\left(k, \{y(j)\}_{j=0}^{k}\right)$ and transmits it. Note that we have not assumed that the encoder has access to any acknowledgements from the decoder about which transmissions have been successful. However, we will show that the presence of such acknowledgements does not improve the optimal performance achievable by a suitable encoder.

Denote by $\mathcal{I}(k)$ the information set that the decoder can utilize to calculate the control at time k. As an example, if no erasures were happening, $\mathcal{I}(k) = \{y(0), y(1), \cdots, y(k-1)\}$. More generally, given any packet erasure pattern, we can define a time stamp $t_s(k)$ at every time step k such that the erasures did not allow any information transmitted by the encoder after time $t_s(k)$ to reach the decoder. Without loss of generality, we can restrict attention to information-set feedback controllers. For a given information set $\mathcal{I}(.)$, denote the minimal value of the cost J_{LQG} that can be achieved with the optimal controller design by $J_{LQG}^{\star}(\mathcal{I})$, and the smallest sigma algebra generated by the information set as $\mathbf{I}(.)$. If two information sets $I^1(.)$ and $I^2(.)$ are such that $\mathbf{I}^1(k) \subseteq \mathbf{I}^2(k)$, we have $J_{LQG}^{\star}(\mathcal{I}^2) \leq J_{LQG}^{\star}(\mathcal{I}^1)$.

Consider an algorithm $\mathcal{A}_1$ in which at every time step k, the sensor transmits all measurements $y(0), y(1), \cdots, y(k)$ to the decoder. Note that this algorithm is not a valid encoding algorithm since the dimension of the transmitted vector is not bounded, as k increases. However, with this algorithm, for any drop sequence, the decoder has access to an information set of the form $\mathcal{I}^{\max}(k) = \{y(0), y(1), \cdots, y(t_s(k))\}$, where $t_s(k) \leq k - 1$ is the time stamp defined above. This is the maximal information set that the decoder can have access to with any algorithm, in the sense that $\mathbf{I}(k) \subseteq \mathbf{I}^{\max}(k)$, for any other algorithm that yields the information set $\mathcal{I}(k)$. Thus, one way to achieve the optimal value of the cost function is to utilize an algorithm that makes $\mathcal{I}^{\max}(k)$ available to the sensor at every time k along with a controller that optimally utilizes this set. Further, one such encoder algorithm is $\mathcal{A}_1$. However, as discussed above, $\mathcal{A}_1$ is not a valid encoding algorithm. Surprisingly, as shown below, we can achieve the same performance with an algorithm that transmits a vector with finite dimension.

We begin with the following separation principle when the decoder has access to the maximal information set. Denote by $\hat{\alpha}(k|\beta(k))$ the minimum mean squared error (MMSE) estimate of the random variable $\alpha(k)$ based on the information $\beta(k)$.

Theorem 6.13 (Separation Principle with Maximal Information Set). *Consider the control problem as defined above, when the decoder has access to the maximal information set $\mathcal{I}^{\max}(k)$ at every time step. Then, the optimal control input is given by*

$$u(k) = \hat{u}_{LQ}\left(k|\mathcal{I}^{\max}(k), \{u(j)\}_{j=0}^{k-1}\right),$$

where $u_{LQ}(k)$ is the optimal LQ control law.

The proof of this result is similar to the standard separation principle (see, e.g., [**?**, Chapter 9]) and is omitted here. For our setting, the importance of this result lies in the fact that it recognizes that $\hat{u}_{LQ}\left(k|\mathcal{I}^{\max}(k), \{u(j)\}_{j=0}^{k-1}\right)$ (or, in turn, $\hat{x}_{LQ}\left(k|\mathcal{I}^{\max}(k), \{u(j)\}_{j=0}^{k-1}\right)$) is a sufficient statistic to calculate the control input that achieves the minimum possible cost for *any* encoding algorithm. Utilizing the fact that the optimal MMSE estimate of $x(k)$ is linear in the effects of the maximal information set and the previous control inputs, we can identify the quantity that the encoder should transmit that depends only on the measurements. We have the following result.

Theorem 6.14 (Separation of the Effect of the Control Inputs). *The quantity $\hat{x}_{LQ}\left(k|\mathcal{I}^{\max}(k), \{u(j)\}_{j=0}^{k-1}\right)$ can be calculated as*

$$\hat{x}_{LQ}\left(k|\mathcal{I}^{\max}(k), \{u(j)\}_{j=0}^{k-1}\right) = \bar{x}_{LQ}\left(k|\mathcal{I}^{\max}(k)\right) + \psi(k),$$

where $\bar{x}_{LQ}\left(k|\mathcal{I}^{\max}(k)\right)$ depends only on $\mathcal{I}^{\max}(k)$ but not on the control in-

puts and $\psi(k)$ depends only on the control inputs $\{u(j)\}_{j=0}^{k-1}$. Further both $\bar{x}_{LQ}(k|\mathcal{I}^{\max}(k))$ and $\psi(k)$ can be calculated recursively.

Proof. The proof follows readily from noting that $\hat{x}_{LQ}\left(k|\mathcal{I}^{\max}(k), \{u(j)\}_{j=0}^{k-1}\right)$ can be obtained from the Kalman filter which is affine in both measurements and control inputs. We can identify

$$\bar{x}_{LQ}(k|\mathcal{I}^{\max}(k)) = A^{k-t_s(k)-1}\breve{x}(t_s(k)+1|t_s(k))$$

$$\psi(k) = A^{k-t_s(k)-1}\breve{\psi}(t_s(k)+1) + \sum_{i=0}^{k-t_s(k)-2} A^i Bu(k-i-1),$$

where $\breve{x}(j+1|j)$ evolves as

$$M^{-1}(j|j) = M^{-1}(j|j-1) + C^T R_v^{-1} C$$
$$M^{-1}(j|j)\breve{x}(j|j) = M^{-1}(j|j-1)\breve{x}(j|j-1) + C^T R_v^{-1} y(j)$$
$$M(j|j-1) = AM(j-1|j-1)A^T + R_w$$
$$\breve{x}(j|j-1) = A\breve{x}(j-1|j-1),$$

with the initial conditions $\breve{x}(0|-1) = 0$ and $M(0|-1) = \Pi(0)$, and $\breve{\psi}(j)$ evolves as

$$\breve{\psi}(j) = Bu(j-1) + \Gamma(j-1)\breve{\psi}(j-1)$$
$$\Gamma(j) = AM^{-1}(j-1|j-1)M(j-1|j-2),$$

with the initial condition $\breve{\psi}(0) = 0$. $\qquad\square$

Now consider the following algorithm $\mathcal{A}_2$. At every time step k, the encoder calculates and transmits the quantity $\breve{x}(k|k)$ using the algorithm in the above proof. The decoder calculates the quantity $\psi(k)$. If the transmission is successful, the decoder calculates

$$\hat{x}_{LQ}\left(k+1|\mathcal{I}^{\max}(k+1), \{u(j)\}_{j=0}^k\right) = \bar{x}_{LQ}(k+1|\mathcal{I}^{\max}(k+1)) + \psi(k)$$

$$= A\breve{x}(k|k) + \psi(k).$$

If the transmission is unsuccessful, the decoder calculates

$$\hat{x}_{LQ}\left(k+1|\mathcal{I}^{\max}(k+1), \{u(j)\}_{j=0}^k\right) = A^{k-t_s(k)}\bar{x}_{LQ}(k+1|\mathcal{I}^{\max}(t_s(k)+1)) + \psi(k),$$

where the quantity $\bar{x}_{LQ}(k+1|\mathcal{I}^{\max}(t_s(k)+1))$ is stored in the memory from the last successful transmission (note that only the last successful transmission needs to be stored). Using the Theorems 6.13 and 6.14 clearly allows us to state the following result.

Theorem 6.15 (Optimality of the Algorithm $\mathcal{A}_2$). *Algorithm $\mathcal{A}_2$ is optimal in the sense that it allows the controller to calculate the control input $u(k)$ that minimizes J_{LQG}.*

Proof. At every time step, the algorithm $\mathcal{A}_2$ makes $\hat{x}_{LQ}\left(k+1|\mathcal{I}^{\max}(k+1),\{u(j)\}_{j=0}^{k}\right)$ available to the controller. Thus, the controller can calculate the same control input as with the algorithm $\mathcal{A}_1$ which together with an LQ controller yields the minimum value of J_{LQG}. □

Note that the optimal algorithm is non-linear (in particular, it is a switched linear system). This is not unexpected, in view of the non-classical information pattern in the problem.

Remarks

• *Boundedness of the Transmitted Quantity:* It should be emphasized that the quantity $\breve{x}(k|k)$ that the encoder transmits is <u>not</u> the estimate of $x(k)$ (or the state of some hypothetical open loop process) based only on the measurements $y(0), \cdots, y(k)$. In particular under the constraint on the erasure probability that we derive later, the state $x(k)$ is stable and hence the measurements $y(k)$ are bounded. *Thus, the quantity $\breve{x}(k|k)$ is bounded.* This can also be seen from the recursive filter used in the proof of Theorem 6.14. If the closed loop system $x(k)$ is unstable due to high erasure probabilities, $\breve{x}(k|k)$ would, of course, not be bounded. However, the optimality result implies that the system cannot be stabilized by transmitting any other bounded quantity (such as measurements).

• *Optimality for any Erasure Pattern and the 'Washing Away' Effect:* The optimality of the algorithm required no assumption about the erasure statistics. The optimality result holds for an arbitrary erasure sequence, and at every time step (not merely in an average sense). Moreover, any successful transmission 'washes away' the effect of the previous erasures in the sense that it ensures that the control input is identical to the case as if all previous transmissions were successful.

• *Presence of Delays:* We assumed that the communication channel introduces a constant delay of one time step. However, the same algorithm continues to remain optimal even if the channel introduces larger (or even time-varying) delays, as long as there is the provision of a time stamp from the encoder regarding the time it transmits any vector. The decoder uses the packet it receives at any time step only if it was transmitted later than the quantity it has stored from the previous time steps. If this is not true due to packet re-ordering, the decoder continues to use the quantity stored from previous time steps. Further, if the delays are finite, the stability conditions derived below remain unchanged. Infinite delays are equivalent to packet erasures, and can be handled by using the same framework.

Stability and Performance: Both the stability and performance of the system with this optimal coding algorithm in place can be analyzed by assuming specific models for the erasure process. For pedagogical ease, we adopt the i.i.d. erasure model, with an erasure occurring at any time step with probability p. Due to the separation principle, to obtain the stability conditions, we need to consider the conditions under which the LQ control

cost for the system, and the covariance of the estimation error between the state of the process $x(k)$ and the estimate at the controller $\hat{x}(k)$ remain bounded, as time k increases. Under the controllability and observability assumptions the LQ cost remains bounded, if the control value does. Define the estimation error and its covariance as

$$e(k) = x(k) - \hat{x}(k)$$
$$P(k) = E\left[e(k)e^T(k)\right],$$

where the expectation is taken with respect to the process and measurement noises, and the initial condition (but not the erasure process). Due to the 'washing away' effect of the algorithm, the error of the estimate at the decoder evolves as

$$e(k+1) = \begin{cases} \bar{e}(k+1) & \text{no erasure} \\ Ae(k) & \text{erasure event,} \end{cases}$$

where $\bar{e}(k)$ is the error between $x(k)$ and the estimate of $x(k)$ given all control inputs $\{u(j)\}_{j=0}^{k-1}$ and measurements $\{y(j)\}_{j=0}^{k-1}$. Thus, the error covariance evolves as

$$P(k+1) = \begin{cases} M(k+1) & \text{with probability } 1-p \\ AP(k)A^T + R_w & \text{with probability } p, \end{cases}$$

where $M(k)$ is the covariance of the error $\bar{e}(k)$. Thus, we obtain

$$E[P(k+1)] = (1-p)M(k+1) + pR_w + pAE[P(k)]A^T,$$

where the extra expectation for the error covariance is taken over the erasure process in the channel. Since the system is observable, $M(k)$ converges exponentially to a steady state value $M^\star$. Thus, the necessary and sufficient condition for the convergence of the above discrete algebraic Lyapunov recursion is

$$p\rho(A)^2 < 1,$$

where $\rho(A)$ is the spectral radius of A. Due to the optimality of the algorithm considered above, this condition is necessary for stability of the system with any causal encoding algorithm. In particular, for the strategy of simply transmitting the latest measurement from the sensor as considered in the one block design, this condition turns out to be necessary for stability (though not sufficient for a general process model). For achieving stability with this condition, we require an encoding strategy, such as the recursive algorithm provided above.

This analysis can be generalized to more general erasure models. For example, for a Gilbert-Eliot type channel model, the necessary and sufficient condition for stability is given by

$$q_{00}\rho(A)^2 < 1,$$

where q_{00} is the conditional probability of an erasure event at time $k + 1$, provided an erasure occurs at time k. In addition, by calculating the terms $E[P(k)]$ and the LQ control cost of the system with full state information, the performance J_{LQG} can also be calculated through the separation principle proved above. The value of the cost function thus achieved provides a lower bound to the value of the cost function achievable using any other encoding or control algorithm, for the same probability of erasure. An alternative viewpoint is to consider the encoding algorithm above as a means for transmitting data with lesser frequency to achieve the same level of performance, than, e.g., transmitting measurements to the controller.

Higher Order Moments: It can be seen that the treatment above can be extended to consider the stability of higher order moments of the estimation error, or the state value. In fact, the entire steady state probability distribution function of the estimation error can be calculated.

Extensions and Open Questions

The above framework was explained for a very simple set-up of an LQG problem. It is natural to consider its generalization to other models by removing various assumptions. We consider some of these assumptions below. We also point out some of the open questions.

• *Channel between Controller and Actuator:* The encoding algorithm presented above continues to remain optimal when a channel is present between the controller and the actuator (as considered in Figure **??**), as long as there is a provision for acknowledgement from actuator to controller for any successful transmission, and the protocol that the actuator follows in case of an erasure is known at the controller. This is because these two assumptions are enough for the separation principle to hold. If no such acknowledgement is available, the control input begins to have a dual effect and the optimal algorithm is still unknown. Moreover, the problem of designing the optimal encoder for the controller-actuator channel can also be considered. This design will intimately depend on the information that is assumed to be known at the actuator (e.g., the cost function, the system matrices and so on). Algorithms that optimize the cost function for such information sets are largely unknown. A simpler version of the problem would involve either

- analyzing the stability and performance gains for given encoding and decoding algorithms employed by the controller and the actuator respectively, or,

- considering algorithms that are stability optimal, in the sense of designing recursive algorithms that achieve the largest stability region for any possible causal encoding algorithm.

Both these directions have seen research activity. For the first direction,

algorithms typically involve transmitting some future control inputs at every time step, or the actuator using some linear combination of past control inputs if an erasure occurs. The second direction has identified the stability conditions that are necessary for any causal algorithm. Moreover, recursive designs that can achieve stability when these conditions are satisfied have also been identified. Surprisingly, the design is in the form of a *universal actuator* that does not require access to the model of the plant. Even if such knowledge were available, the stability conditions do not change. Thus, the design is stability optimal.

- *Presence of a Communication Network:* So far we have concentrated on the case when the sensor and the controller are connected using a single communication channel. A typical scenario, particularly in a wireless context, would instead involve a communication network with multiple such channels. If no encoding algorithm is implemented, and every node in the network (including the sensor) transmits simply the measurements, the network can be replaced by a giant erasure channel with the equivalent erasure probability being some measure of the reliability of the network. However, the performance degrades rapidly as the network size increases. If encoding is permitted, such an equivalence breaks down. The optimal algorithm is an extension of the single channel case, and is provided in [**?**]. The stability and performance calculations are considerably more involved. However, the stability condition has an interesting interpretation in terms of the capacity for fluid networks. The necessary and sufficient condition for stability can be expressed as the inequality

$$p_{\text{max-cut}}\rho(A)^2 < 1,$$

where $p_{\text{max-cut}}$ is the max-cut probability calculated in a manner similar to the min-cut capacity of fluid networks. We construct cut-sets by dividing the nodes in the network into two sets with one set containing the sensor, and the other the controller. For each cut-set, we calculate the cut-set erasure probability by multiplying the erasure probabilities of all the channels from the set containing the sensor to the set containing the controller. The maximum such cut-set erasure probability (over all possible cut-sets) denotes the max-cut probability of the network. The improvement in the performance and stability region of the system by using the encoding algorithm increases drastically with the size and the complexity of the network.

- *Multiple Sensors:* Another direction in which the above framework can be extended is to consider multiple sensors observing the same process. As with the case with one sensor, one can identify the necessary stability conditions and a lower bound for the achievable cost function with any causal coding algorithm. These stability conditions are also sufficient and recursive algorithms for achieving stability when these conditions are satisfied have been identified. These conditions are a natural extension of the stability conditions for the single sensor case. As an example, for the case of two

sensors described by sensing matrices C_1 and C_2 that transmit data to the controller across erasure channels for which erasure events are i.i.d. with probabilities p_1 and p_2 respectively, the stability conditions are given by the set

$$p_2\rho(A_1)^2 < 1$$
$$p_1\rho(A_2)^2 < 1$$
$$p_1 p_2 \rho(A)^2 < 1,$$

where $\rho(A_i)$ denotes the spectral radius of the unobservable part of the system matrix A, when the pair (A, C_i) is represented in the observability canonical form. However, the problem of identifying distributed encoding algorithms to be followed at each sensor for achieving the lower bounds on the achieved cost function remains largely open. This problem is related to the track fusion problem that considers identifying algorithms for optimal fusion of information from multiple sensors that interact intermittently (e.g., see [**?**]). That transmitting estimates based on local data from each sensor is not optimal is long known. While algorithms that achieve a performance close to the lower bound of the cost function have been identified, a complete solution is not available.

• *Inclusion of More Communication Effects:* Our discussion has focussed on modeling the loss of data transmitted over the channel. In our discussion of the optimal encoding algorithms, we also briefly considered the possibility of data being delayed or received out of order. An important direction for future work is to consider other effects due to communication channels. Both from a theoretical perspective, and for many applications such as underwater systems, an important effect is to impose a limit on the number of bits that can be communicated for every successful transmission. Some recent work [**?, ?**] has considered the analog digital channel in which the channel supports n bits per time step and transmits them with a certain probability p at every time step. Stability conditions for such a channel have been identified and are a natural combination of the stability conditions for the analog erasure channel above and the ones for a noiseless digital channel, as considered elsewhere in the book. The performance of optimal encoding algorithms and the optimal performance that is achievable remain unknown. Another channel effect that has largely been ignored is the addition of channel noise to the data received successfully.

• *More General Performance Criteria:* Our treatment focussed on a particular performance measure - a quadratic cost, and the stability notions emanating from that measure. Other cost functions may be relevant in applications. Thus the cost function may be related to target tracking, measures such as H_2 or H_∞ [**?**], or some combination of communication and control costs. The analysis and optimal encoding algorithms for such measures are expected to differ significantly. An an example, for target

tracking, the properties of the reference signal that needs to be tracked can be expected to play a significant role. Similarly, for H_∞ related costs, the sufficient statistic, and hence the encoding algorithms to transmit it, may be vastly different than the LQG case. Finally, a distributed control problem with multiple processes, sensors and actuators is a natural direction to consider.

• *More General Plant Dynamics:* The final direction is to consider plant dynamics that are more general than the linear model that we have considered. Moving to models such as jump linear systems, hybrid systems, and general non-linear systems will provide new challenges and results. As an example, for non-linear plants concepts such as spectral radius no longer hold. Thus, the analysis techniques are likely to be different and measures such as Lyapunov exponents and the Lipschitz constant for the dynamics will likely become important.

Chapter 7
Information Flow and Consensus

In this chapter we move from the problem of estimation and control of a single system across a communications channel to the challenge of sensing, estimation and control of a multi-agent system, with the information available to the agents represented by a graph of interconnections. We begin with a review of the relevant concepts in graph theory, focused on the use of algebraic techniques to characterize the properties of the interconnection structure. We then apply these concepts to study the problem of a group of agents reaching consensus on a shared property of the system.

7.1 Graph Theory

In this section we give a brief overview of the field of graph theory, focused on some of the algebraic methods that characterize the properties of the graph in terms of a set of matrices associated with it. These techniques will be very important for helping understand the interactions between dynamic agents across a graph, including the consensus problem in this chapter and the distributed estimation and control problems in the subsequent chapters. More detailed treatments are available in a number of textbooks, including Diestel [**?**], Godsil and Royle [**?**], and Horn and Johnson [**?**]. This section is based in part on a set of course notes originally developed by Reza Olfati-Saber [**?**].

Basic Definitions

We define a *directed graph* as a pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consisting of a set of *vertices* $\mathcal{V}$ and a set of *edges* $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. We represent a vertex (or *node*) as an element $v_i \in \mathcal{V}$ and an edge (or *link*) as a connection between two vertices, $e_{ij} = (v_i, v_j) \in \mathcal{E}$. We write $|\mathcal{V}|$ for the number of vertices in the graph, also known as the *order* of the graph. qAn edge has an orientation given by the ordering of the vertices, so the edge e_{ij} is distinct from the edge e_{ji}. We call v_i the head of the edge and v_j the tail. A directed graph is also referred to as a *digraph*.

We say that two vertices v_i and v_j are *adjacent* if there exists and edge $e = (v_i, v_j)$ and vertex v_j is called a *neighbor* of v_i. We write $\mathcal{N}_i = $ set of all neighbors of v_i and we say that a graph $\mathcal{G}$ is *complete* if all vertices are adjacent to each other. We define the *out-degree* of a vertex v_i, written
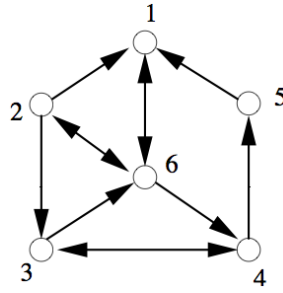
Figure 7.1: A graph with 6 vertices.

$\deg_{\text{out}}(V_i)$, as the number of edges whose head is v_i. Similarly, the *in-degree* of a vertex v_i, $\deg_{\text{in}}(v_i)$ is the number of edges with tail v_i.

Example 7.1 6 node graph

Consider a graph given by the vertices $\mathcal{V} = \{1, 2, 3, 4, 5, 6\}$ and a set of edges

$$\mathcal{E} = \{(1,6), (2,1), (2,3), (2,6), (6,2), (3,4),$$
$$(3,6), (4,3), (4,5), (5,1), (6,1), (6,2), (6,4)\},$$

as shown in Figure 7.1. Node 1 has an in-degree of 3 and an out-degree of 1. Its neighbor set is given by $\mathcal{N}_1 = \{v_6\}$. Node 2 has an in-degree of 1 and an out-degree of 3. Its neighbor set is given by $\mathcal{N}_2 = \{v_1, v_2, v_6\}$. ∇

In many instances the orientation of the edges in a graph will not matter and we can ignore the ordering of the verticies in an edge $e_{ij} = (v_i, v_j)$. Formally, we will consider a graph to be *undirected* if $e_{ij} \in \mathcal{E}$ implies that $e_{ji} \in \mathcal{E}$. In these cases it will often be easier to simply say that the graph is undirected and consider an edge e_{ij} to be equivalent to an edge e_{ji}. For an undirected graph the indegree and outdegree are the same, so we simply refer to the *degree of a vertex*. An undirected graph is *regular* (or *k-regular*) if all vertices of a graph have the same degree k. A directed graph is *balanced* if the out-degree is equal to the in-degree at each vertex.

Example 7.2

Figure 7.3 shows three examples of graphs. ∇

Connectedness of Graphs

A key set of properties of a graph have to do with whether there exists paths that connect its nodes. Formally, a *path* is a subgraph $\pi = (\mathcal{V}, \mathcal{E}_\pi) \subset \mathcal{G}$ with distinct vertices $\mathcal{V} = \{v_1, v_2, \ldots, v_m\}$ and

$$\mathcal{E}_\pi := \{(v_1, v_2), (v_2, v_3), \ldots, (v_{m-1}, v_m)\}.$$

The *length* of π is defined as $|\mathcal{E}_\pi| = m - 1$. A *cycle* (or *m-cycle*) $C = (\mathcal{V}, \mathcal{E}_C)$ is a path (of length m) with an extra edge $(v_m, v_1) \in \mathcal{E}$. We define the
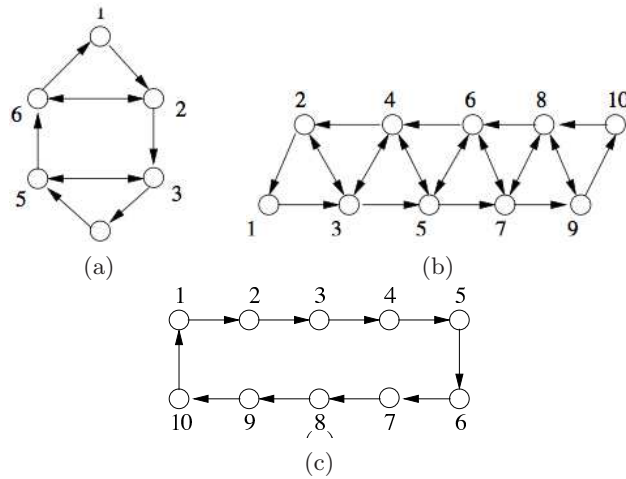
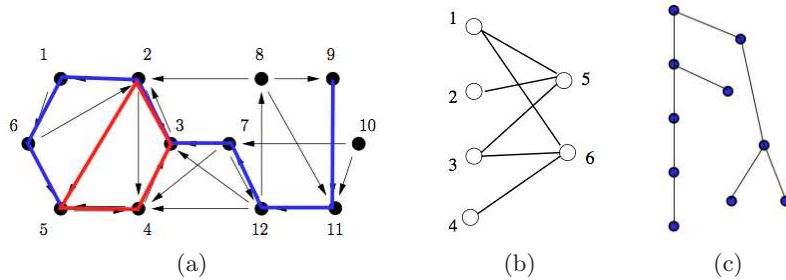**Figure 7.2:** Examples of graphs with different properties.

Figure 7.3: Examples of graphs with different properties.

distance between two vertices v and w as the length of the shortest path between them.

An undirected graph $\mathcal{G}$ is called *connected* if there exists a path π between any two distinct vertices of $\mathcal{G}$. For a connected graph $\mathcal{G}$, the length of the maximum distance between two vertices is called the *diameter* of $\mathcal{G}$. A graph with no cycles is called *acyclic*. A *tree* is a connected acyclic graph.

A digraph is called *strongly connected* if there exists a directed path π between any two distinct vertices of $\mathcal{G}$. A digraph is called *weakly connected* if there exists an undirected path between any two distinct vertices of $\mathcal{G}$.

Example 7.3
Figure **??** shows examples of graphs and their connectedness properties.
∇

Matrices Associated with a Graph

In order to characterize the properties of a graph, we will use matrices to represent the structure of the graph. The properties of these matrices can

then be related back to the properties of the graph.

The *adjacency matrix* $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ of a graph $\mathcal{G}$ of order n is given by:

$$a_{ij} := \begin{cases} 1 & \text{if } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

The *degree matrix* of a graph as a diagonal $n \times n$ $(n = |\mathcal{V}|)$ matrix

$$\Delta = \text{diag}\{\deg_{\text{out}}(v_i)\}$$

with diagonal elements equal to the out-degree of each vertex and zero everywhere else. The *Laplacian matrix* L of a graph is defined as

$$L = \Delta - A$$

. It follows from the definition that the row sums of the Laplacian are all 0.

Example 7.4 6 node graph

Consider the graph shown in Example **??**. The adjancy matrix and Laplacian are given by

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}, \qquad L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ -1 & 3 & -1 & 0 & 0 & -1 \\ 0 & 0 & 2 & -1 & 0 & -1 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & -1 & 0 & 3 \end{bmatrix}.$$

$$\nabla$$

Periodic Graphics and Weighted Graphs

A graph with the property that the set of all cycle lengths has a common divisor $k > 1$ is called *k-periodic*. A graph without cycles is said to be *acyclic*.

A *weighted graph* is graph $(\mathcal{V}, \mathcal{E})$ together with a map $\varphi : \mathcal{E} \to \mathbb{R}$ that assigns a real number $w_{ij} = \varphi(e_{ij})$ called a *weight* to an edge $e_{ij} = (v_i, v_j) \in \mathcal{E}$. The set of all weights associated with $\mathcal{E}$ is denoted by $\mathcal{W}$. A weighted graph can be represented as a triplet $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$.

In some applications it is natural to "normalize" the Laplacian by the outdegree. We define the *weighted Laplacian* as

$$\tilde{L} := \Delta^{-1} L = I - \tilde{A} = I - \Delta^{-1} A$$

, where $\tilde{A} = \Delta^{-1} A$ (weighted adjacency matrix).

Example 7.5 Weighted Laplacian for formation graph

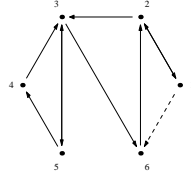Consider the graph in Figure 7.4. The weighted Laplacian is given by

Figure 7.4: Formation control graph.

$$L = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & 1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\nabla$$

Gershgorin Disk Theorem

Theorem 7.1 (Gershgorin Disk Theorem). *Let $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ and define the deleted absolute row sums of A as*

$$r_i := \sum_{j=1, j \neq i}^{n} |a_{ij}| \tag{7.1}$$

Then all the eigenvalues of A are located in the union of n disks

$$G(A) := \bigcup_{i=1}^{n} G_i(A), \quad \text{with} \quad G_i(A) := \{ z \in \mathbb{C} : |z - a_{ii}| \leq r_i \} \tag{7.2}$$

Furthermore, if a union of k of these n disks forms a connected region that is disjoint from all the remaining $n - k$ disks, then there are precisely k eigenvalues of A in this region.
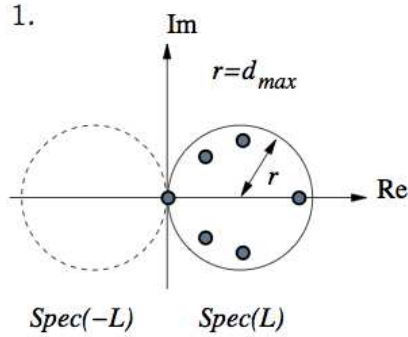
Sketch of proof. Let λ be an eigenvalue of A and let v be a corresponding eigenvector. Choose i such that $|v_i| = \max_j |v_j| > 0$. Since v is an eigenvector,

$$\lambda v_i = \sum_i A_{ij} v_j \quad \Longrightarrow \quad (\lambda - a_{ii}) v_i = \sum_{i \neq j} A_{ij} v_j$$

Now divide by $v_i \neq 0$ and take the absolute value to obtain

$$|\lambda - a_{ii}| = |\sum_{j \neq i} a_{ij} v_j| \leq \sum_{j \neq i} |a_{ij}| = r_i$$

$\square$

Figure 7.5: Graphical description of the Gershgorin disk theorem.

We can use the Gershgorin disk theorem to reason about the eigenvalues of the Laplacian and the weighted Laplacian.

Proposition 7.2. *Let L be the Laplacian matrix of a digraph $\mathcal{G}$ with maximum vertex out–degree of $d_{max} > 0$. Then all the eigenvalues of $A = -L$ are located in a disk*

$$B(\mathcal{G}) := \{s \in \mathbb{C} : |s + d_{max}| \leq d_{max}\} \qquad (7.3)$$

that is located in the closed LHP of s-plane and is tangent to the imaginary axis at $s = 0$.

Proposition 7.3. *Let $\tilde{L}$ be the weighted Laplacian matrix of a digraph $\mathcal{G}$. Then all the eigenvalues of $A = -L$ are located inside a disk of radius 1 that is located in the closed LHP of s-plane and is tangent to the imaginary axis at $s = 0$.*

Another property of the Laplacian is that its rank determines the connectivity of the graph.

Theorem 7.4 (Olfati-Saber)**.** *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ be a weighted digraph of order n with Laplacian L. If $\mathcal{G}$ is strongly connected, then $rank(L) = n - 1$.*

The proof for the directed case can be found in standard textbooks on graph theory, such as those listed at the beginning of this section.. The proof for the undirected case is available in [OSM04]. Note that for directed graphs, we require that $\mathcal{G}$ be strongly connected; the converse statement is not true.

Perron-Frobenius Theory

The spectrum of a matrix A is defined as $\text{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$, where λ_i, $i = 1, \ldots, n$ are the eigenvalues of A. The distance to the largest eigenvalue $\rho(A) = |\lambda_n| = \max_k |\lambda_k|$ is called the *spectral radius* of A.
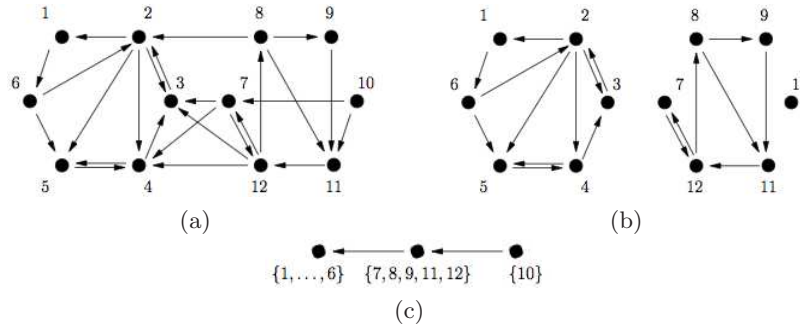
Figure 7.6: Irreducibility of a graph.

Theorem 7.5 (Perron's Theorem, 1907). *If $A \in \mathbb{R}^{n \times n}$ is a positive matrix $(A > 0)$, then*

1. $\rho(A) > 0$;
2. $r = \rho(A)$ *is an eigenvalue of A;*
3. *There exists a positive vector $x > 0$ such that $Ax = \rho(A)x$;*
4. $|\lambda| < \rho(A)$ *for every eigenvalue $\lambda \neq \rho(A)$ of A, i.e. $\rho(A)$ is the unique eigenvalue of maximum modulus; and*
5. $[\rho(A)^{-1}A]^m \to R$ *as $m \to +\infty$ where $R = xy^T$, $Ax = \rho(A)x$, $A^T y = \rho(A)y$, $x > 0$, $y > 0$, and $x^T y = 1$.*

Theorem 7.6 (Perron's Theorem for Non–Negative Matrices). *If $A \in \mathbb{R}^{n \times n}$ is a non-negative matrix $(A \geq 0)$, then $\rho(A)$ is an eigenvalue of A and there is a non–negative vector $x \geq 0$, $x \neq 0$, such that $Ax = \rho(A)x$.*

A directed graph is irreducible if, given any two vertices, there exists a path from the first vertex to the second. (Irreducible = strongly connected) A matrix is irreducible if it is not similar to a block upper triangular matrix via a permutation. A digraph is irreducible if and only if its adjacency matrix is irreducible.

Example 7.6
Consider the graph in Figure 7.6. ∇

Theorem 7.7 (Frobenius). *Let $A \in \mathbb{R}^{n \times n}$ and suppose that A is irreducible and non-negative. Then*

1. $\rho(A) > 0$;
2. $r = \rho(A)$ *is an eigenvalue of A;*
3. *There is a positive vector $x > 0$ such that $Ax = \rho(A)x$;*
4. $r = \rho(A)$ *is an algebraically simple eigenvalue of A; and*
5. *If A has h eigenvalues of modulus r, then these eigenvalues are all distinct roots of $\lambda^h - r^h = 0$.*

Using the Perron and Frobenius theorems, we can study additional properties of the Laplacian matrix of a graph. In particular, it can be shown
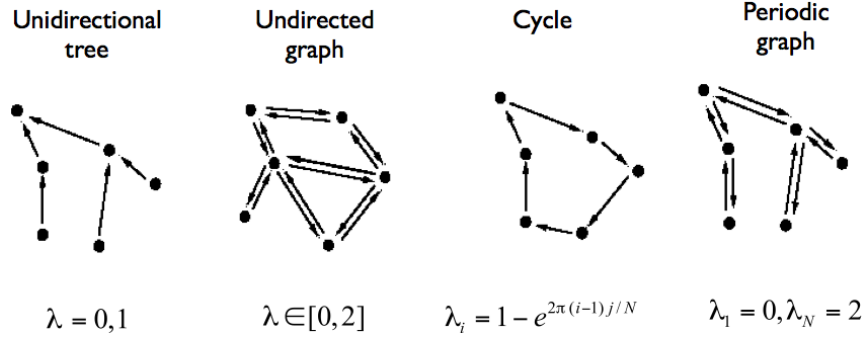
Unidirectional tree	Undirected graph	Cycle	Periodic graph
$\lambda = 0,1$	$\lambda \in [0,2]$	$\lambda_i = 1 - e^{2\pi(i-1)j/N}$	$\lambda_1 = 0, \lambda_N = 2$

Figure 7.7: Spectra of the Laplacian for classes of graphs.

that If $\mathcal{G}$ is strongly connected, the zero eigenvalue of L is simple. If $\mathcal{G}$ is aperiodic, all nonzero eigenvalues lie in the interior of the Gershgorin disk. If $\mathcal{G}$ is k-periodic, L has k evenly spaced eigenvalues on the boundary of the Gershgorin disk.

Theorem 7.8 (Variant of Courant-Fischer). *Let $A \in \mathbb{R}^{n \times n}$ be a Hermitian matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and let w_1 be the eigenvector of A associated with the eigenvalue λ_1. Then*

$$\lambda_2 = \min_{\substack{x \neq 0, x \in \mathbb{C}^n, \\ x \perp w_1}} \frac{x^* A x}{x^* x} = \min_{\substack{x^* x = 1, \\ x \perp w_1}} x^* A x \qquad (7.4)$$

Proof. Since A is Hermitian matrix, it is unitary diagonalizable (see Theorem **??**), i.e. $A = U\Lambda U^*$ where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Let $U = [w_1 | w_2 | \cdots | w_n]$ (w_k is the kth column of U). Then

$$x^* A x = x^* U \Lambda U^* x = (U^* x)^* \Lambda (U^* x)$$

$$= \sum_{i=1}^{n} \lambda_i |(U^* x)_i|^2 = \sum_{i=1}^{n} \lambda_i |w_i^* x|^2 = \sum_{i=2}^{n} \lambda_i |w_i^* x|^2 \quad (x \perp w_1)$$

$$\geq \lambda_2 \sum_{i=2}^{n} |w_i^* x|^2 = \lambda_2 \sum_{i=1}^{n} |w_i^* x|^2 \quad (x \perp w_1) = \lambda_2 \sum_{i=1}^{n} |(U^* x)_i|^2 \qquad (7.5)$$

$$= \lambda_2 (x^* U U^* x) = \lambda_2 x^* x$$

Thus, for $x \perp w_1$ and $x \neq 0$

$$x^* A x \geq \lambda_2 x^* x$$

where the equality is achieved with $x = w_2$. $\square$

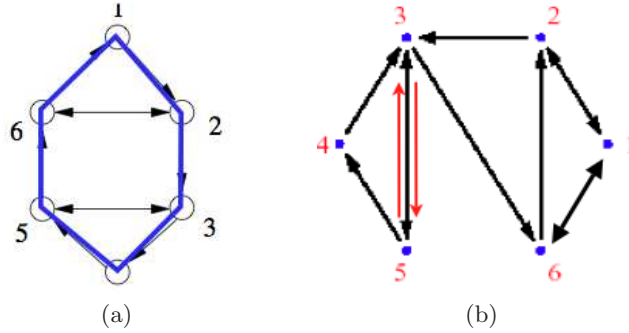The second eigenvalue of the Laplacian λ_2 is called the *algebraic connectivity* of L.

Figure 7.8: Cyclic separability.

Cyclically Separable Graphs

Definition 7.1 (Cyclic separability). A digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is *cyclically separable* if and only if there exists a partition of the set of edges $\mathcal{E} = \cup_{k=1}^{n_c} \mathcal{E}_k$ such that each partition $\mathcal{E}_k$ corresponds to either the edges of a cycle of the graph, or a pair of directed edges ij and ji that constitute an undirected edge. A graph that is not cyclically separable is called *cyclically inseparable*.

Lemma 7.1. *Let L be the Laplacian matrix of a cyclically separable digraph $\mathcal{G}$ and set $u = -Lx, x \in \mathbb{R}^n$. Then $\sum_{i=1}^{n} u_i = 0, \forall x \in \mathbb{R}^n$ and $\mathbf{1} = (1, \ldots, 1)^T$ is the left eigenvector of L.*

Proof. The proof follows from the fact that by definition of cyclic separability. We have

$$-\sum_{i=1}^{n} u_i = \sum_{ij \in \mathcal{E}} (x_j - x_i) = \sum_{k=1}^{n_c} \sum_{ij \in \mathcal{E}_k} (x_j - x_i) = 0$$

because the inner sum is zero over the edges of cycles and undirected edges of the graph. □

Example 7.7 Cyclic separability

∇

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a digraph. We say $\mathcal{G}$ is *balanced* if and only if the in–degree and out–degree of all vertices of $\mathcal{G}$ are equal, i.e.

$$\deg_{\text{out}}(v_i) = \deg_{\text{in}}(v_i), \quad \forall v_i \in \mathcal{V} \tag{7.6}$$

Let $\mathcal{G}$ be a digraph with a weighted adjacency matrix $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ that has the property $a_{ii} = 0$. Then, $\mathcal{G}$ is balanced if and only if $w_l = \mathbf{1}$.

Theorem 7.9. *A digraph is cyclically separable if and only if it is balanced.*

Proof. Assume the graph is cyclically separable. Then any arbitrary vertex v_i of the graph belongs to a finite number of cycles and/or undirected edges.

The main property of a cycle is that corresponding to any directed edge arriving at a vertex, there is one edge leaving that vertex and therefore the in–degree and out-degree of any vertex are equal, i.e. the graph is balanced.

Now, let us assume that the graph is balanced, we show that it is cyclically separable. Suppose the opposite holds, meaning that the graph is not cyclically separable. Then there exists a directed edge (v_k, v_l) of the graph which does not belong to any cycles and/or undirected edges. Set $x_i = 0, \forall i \neq l$ and let $x_l = 1$. Define $u = -Lx$, we have $u_i = 0, \forall i \neq k$ and $u_k = x_l - x_k = 1$ (notice that $u_l = 0$ since k is not an out–neighbor of l). Thus $\sum_{i=1}^{n} u_i = \mathbf{1}^T u = 1 \neq 0$. But we know that $\mathbf{1}$ eigenvector of L for any balanced graph, thus $\mathbf{1}^T u = -\mathbf{1}^T Lx = 0, \forall x$. This is a contradiction which means every directed edge of a balanced graph belongs to a cycle or an undirected edge, i.e. the graph is cyclically separable. □   □

7.2 Consensus algorithms

The *consensus problem* involves a group of agents reaching an agreement on a decision in a decentralized problem. In this sectoin we describe one approach to solving this problem, with the agents communicating on a graph.

Average Consensus

Consider a collection of N agents that communicate along a set of undirected links described by a graph $\mathcal{G}$. Each agent has a state x_i with initial value $x_i(0)$ and together they wish to determine the average of the initial states $\text{Ave}(x_0) = 1/N \sum x_0^i$.

The agents implement the following consensus algorithm:

$$x_{k+1}^i = \epsilon \sum_{j \in \mathcal{N}_i} (x_k^j - x_k^i) = -\epsilon |\mathcal{N}_i|(x_k^i - \text{Ave}(x_k^{\mathcal{N}_i}))$$

which is equivalent to the dynamical system

$$x_{k+1} == -\epsilon Lx_k.$$

Proposition 7.10. *If the graph is connected, there exists an ϵ such that the state of the agents converges to $x_i^* = Ave(x_0)$ exponentially fast.*

- Proposition 1 implies that the spectra of L controls the stability (and convergence) of the consensus protocol.

- To (partially) prove this theorem, we need to show that the eigenvalues of L are all positive.

$$\dot{x} = -Lx \qquad L = \Delta - A$$

Note first that the subspaced spanned by $\mathbf{1} = (1, 1, \ldots, 1)^T$ is an invariant subspace since $L \cdot \mathbf{1} = 0$ Assume that there are no other eigenvectors with eigenvalue 0. Hence it suffices to look at the convergence on the complementary subspace $\mathbf{1}^\perp$.

Let δ be the disagreement vector

$$\delta = x - \mathrm{Ave}(x(0)) \, \mathbf{1}$$

and take the square of the norm of δ as a Lyapunov function candidate, i.e. define

$$V(\delta) = \|\delta\|^2 = \delta^T \delta \tag{7.7}$$

Differentiating $V(\delta)$ along the solution of $\dot{\delta} = -L\delta$, we obtain

$$\dot{V}(\delta) = -2\delta^T L \delta < 0, \quad \forall \delta \neq 0, \tag{7.8}$$

where we have used the fact that $\mathcal{G}$ is connected and hence has only 1 zero eigenvalue (along $\mathbf{1}$). Thus, $\delta = 0$ is globally asymptotically stable and $\delta \to 0$ as $t \to +\infty$, i.e. $x^* = \lim_{t \to +\infty} x(t) = \alpha_0 \mathbf{1}$ because $\alpha(t) = \alpha_0 = \mathrm{Ave}(x(0)), \forall t > 0$. In other words, the average–consensus is globally asymptotically achieved.   □

For an undirected graph with Laplacian L, the rate of convergence for the consensus protocol is bounded by the second smallest eigenvalue λ_2

Corollary 7.10.1. *Consider a network of integrators with a directed information flow $\mathcal{G}$ and vertices that apply the consensus protocol. Then, $\alpha = Ave(x)$ is an invariant quantity if and only if $\mathcal{G}$ is balanced.*

Remarks.

- Balanced graphs generalized undirected graphs and retain many key properties

Consensus on Directed Graphs

Consensus over Communication Channels

Consensus for Idempotent Functions

7.3 Effects of Information Topology

7.4 Applications of Consensus

7.5 Further Reading

Exercises

Chapter 8
Distributed Estimation

In this chapter we consider the problem of state estimation in which we have a collection of sensors that are distributed across a network. We begin by exploring the problem of aggregating data from a decentralized network of sensors, either at a centralized node or across a fully connected network, where the goal is to minimize either communication or computation and the graph structure does not play a central role. We then consider more general distributed estimation problem where the graph is not completely connected, so that different agents on the network have different information at different times. We next investigate the case where the information can be lost as it is sent around the network, requiring the use of more advanced methods of design and analysis to accommodate the network dynamics. Finally, we provide some general remarks about when the estimation problem can be separated from the control problem, allow us to separately solve the (optimal) estimation problem. The next chapter looks in more detail at the distributed control problem, where we wish to design a feedback control system across a graph to solve a given task.

8.1 Decentralized Sensor Fusion

Note: The goal of this section is going to be to summary the "classical" results in distributed sensor fusion, focused mainly on the information form. [RMM, 19 Jun 09]

Consider a single process with multiple sensors connected together across a network, as shown in Figure 8.1. We assume that the system being observed has dynamics

$$x[k+1] = Ax[k] + Bu + w,$$

where $X \in \mathbb{R}^n$ represents the state, $u \in \mathbb{R}^p$ represents the (deterministic) input, $W \in \mathbb{R}^q$ represents process disturbances, $Y \in \mathbb{R}^q$ represents the system output and $W \in \mathbb{R}^q$ represents measurement noise. We would like to form an estimation of the state $\hat{x}$, either at each sensor or at the central hub.

The approach we take to solve this problem depends on the structure of the information pattern. If a centralized hub is available, all sensors can send data to the hub and a centralized Kalman filter can be used to compute
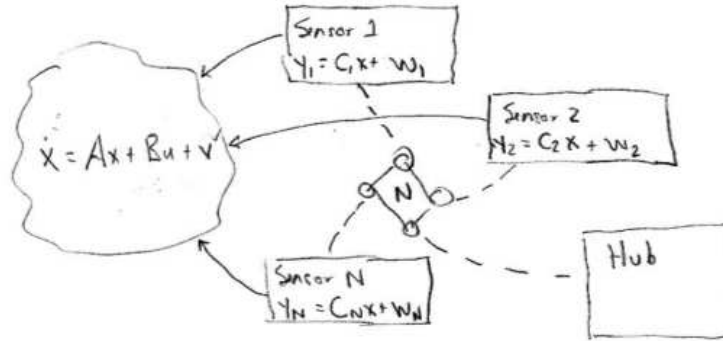
Figure 8.1: Schematic diagram of a distributed sensing system. The system on the left represents the system being observed. Multiple sensors take measurements and communication with each other across a communications network. A (optional) centralized hub collections information across the network.
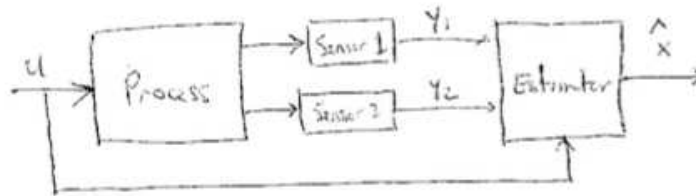
Figure 8.2: Sensor fusion

the estimate. An alternative, which is more efficient in some settings [**?**], is to have some of the information processing done at the sensor nodes. To see how this can be done, we reformulate the Kalman filter in an alternative form.

Consider the situation described in Figure 8.2, where we have an input/output dynamical system with multiple sensors capable of taking measurements. The problem of sensor fusion involves deciding how to best combine the measurements from the individual sensors in order to accurately estimate the process state X. Since different sensors may have different noise characteristics, evidently we should combine the sensors in a way that places more weight on sensors with lower noise. In addition, in some situations we may have different sensors available at different times, so that not all information is available on each measurement update.

To gain more insight into how the sensor data are combined, we investigate the functional form of $L[k|k]$. Suppose that each sensor takes a measurement of the form

$$Y^i = C^i X + V^i, \qquad i = 1, \ldots, q,$$

where the superscript i corresponds to the specific sensor. Let V^i be a zero

mean, white noise process with covariance $\sigma_i^2 = R_{V^i}(0)$. It follows from Lemma **??** that

$$L[k|k] = P[k|k]C^T R_W^{-1}.$$

First note that if $P[k|k]$ is small, indicating that our estimate of X is close to the actual value (in the MMSE sense), then $L[k|k]$ will be small due to the leading $P[k|k]$ term. Furthermore, the characteristics of the individual sensors are contained in the different σ_i^2 terms, which only appears in R_W. Expanding the gain matrix, we have

$$L[k|k] = P[k|k]C^T R_W^{-1}, \qquad R_W^{-1} = \begin{bmatrix} 1/\sigma_1^2 & & \\ & \ddots & \\ & & 1/\sigma_q^2 \end{bmatrix}.$$

We see from the form of R_W^{-1} that each sensor is inversely weighted by its covariance. Thus noisy sensors ($\sigma_i^2 \gg 1$) will have a small weight and require averaging over many iterations before their data can affect the state estimate. Conversely, if $\sigma_i^2 \ll 1$, the data is "trusted" and is used with higher weight in each iteration.

An alternative formulation of the Kalman filter is to make use of the inverse of the covariance matrix, called the *information matrix*, to represent the error of the estimate. It turns out that writing the state estimator in this form has several advantages both conceptually and when implementing distributed computations. This form of the Kalman filter is known as the *information filter*.

We begin by defining the information matrix I and the weighted state estimate $\hat{Z}$:

$$I[k|k] = P^{-1}[k|k], \qquad \hat{Z}[k|k] = P^{-1}[k|k]\hat{X}[k|k].$$

We also make use of the following quantities, which appear in the Kalman filter equations:

$$\Omega_i[k|k] = (C^i)^T R_{W^i}^{-1}[k|k]C^i, \qquad \Psi_i[k|k] = (C^i)^T R_{W^i}^{-1}[k|k]C^i \hat{X}[k|k].$$

Using these quantities, we can rewrite the Kalman filter equations as

Prediction   Correction

$$I[k|k-1] = \left(AI^{-1}[k-1|k-1]A^T + R_W\right)^{-1}, \qquad\qquad I[k|k] = I[k|k-1] + \sum_{i=1}^{q}\Omega_i[k|k],$$

$$\hat{Z}[k|k-1] = I[k|k-1]AI^{-1}[k-1|k-1]\hat{Z}[k-1|k-1] + Bu[k-1], \qquad \hat{Z}[k|k] = \hat{Z}[k|k-1] + \sum_{i=1}^{q}\Psi_i[k|k].$$

Note that these equations are in a particularly simple form, with the information matrix being updated by each sensor's Ω_i and similarly the state estimate being updated by each sensors Ψ_i.

Plots
(a)   (b)
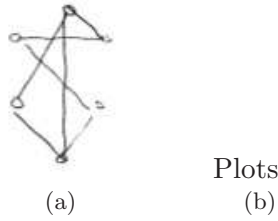
Figure 8.3: Distributed estimation on a graph.

Remarks:

1. Information form allows simple addition for correction step. Intuition: add information through additional data.

2. Sensor fusion: information content = inverse covariance (for each sensor)

3. Variable rate: incorporate new information whenever it arrives. No data $\implies$ no information $\implies$ prediction update only.

Another classical information pattern is that of a fully connected network. In this case, everyone can send either measurements to each other and we can reconstruct the state using local Kalman (or information) filters.

8.2 Distributed estimation on a graph

A more general case occurs when the information is distributed along a graph, as shown in Figure 8.3. Suppose that we have no central hub and we want each sensor to converge to a single global estimate. For simplicity we first consider the static system case, in which all nodes should converge to the optimal estimate

$$\hat{x}_\infty = \sum_{j=1}^{N} R_j^{-1} y^j.$$

As our starting point, we make use of the consensus algorithms described in Section **??**. The basic algorithm is as follows:

• Each node measures y^i

• Each node initializes its state to $x(0) = R_i^{-1} y^i$, where R_i is the covariance associated with sensor i

• Run the consensus protocol, which implies that each node converges to the optimal estimate.

From the results on consensus filters, it can be shown that the convergence rate is bounded by $e^{-\lambda_2 t}$, where λ_2 is the second smallest eigenvalue of the graph Laplacian.. Many extensions to this basic algorithm are possible, including the case of time-varying communication graphs, delays and intermittent communications.

The static algorithm can be extended to the case of a dynamic system in several different ways:

- Fixed graph, ala Durrant-Whyte et al

  - Communication measurements and run full KF

  - Communication local estimate + covariance and account for duplication [**?**]

  - Doesn't handle dropped packets, changing communications graph

- "Microfilter architecture" (Olfati-Saber)
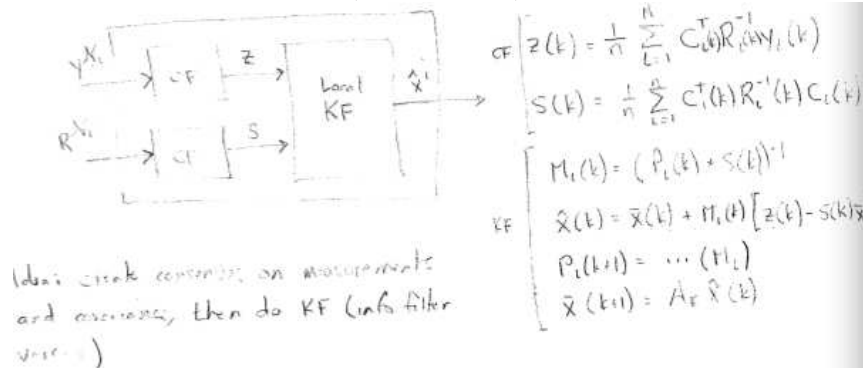
  - Need CF to convert quickly (compared to KF dynamics) and track measurements

  - Resulting filter is *approximate* (may not be optimal prior to convergence), but handles packet delays, etc (inherited from CF properties).

  - ??? sending measurements plus covariance matrices (if $R_i(k)$ not constant)

- Consensus on estimates

Proposition 8.1 (Olfati-Saber). *In the absence of noise, $\hat{x}^i \to x$.*

Proof.   □

Remarks:

1. Can write in discrete time

2. Only approximate KF; lose optimality during transient

3. Can handle varying graph, packet loss, time delay, etc

4. Only requires sensing estimates; P_i is *local* error covariance (doesn't account for neighbor covariance)

Final remarks (on distributed estimation):

1. Distributed KF on a fixed graph (star, completely connected, undirected) is well understood. Basically manipulate the information filter.

2. Alternative approach: use consensus filter on measurements or estimates. Lose optimality, but can handle network effects.

8.3 Distributed Estimation with Packet Loss

We now consider what happens if the graph describing the flow of information around the network is not along a fixed graph. We consider a number of cases, starting with the case in which we can only use a subset of the links on a network at a given time, and then moving to more complicated situations in which the network can drop packets in an unknown way.

8.4 Combining Estimation and Control

In many applications we wish to make use of our estimate of the state of a system for the purposes of controlling the behavior of the system. In this case, the system state depends on the action of the controller, which itself depends on the estimate of the state. In traditional control systems, it can be shown that a separation principle applies, in which we can design the controller assuming we measure the exact state and design the estimator without taking the specific form of the feedback controller into account. In this section we summarize the situations in which we can similar separation principles in distributed estimation principles. We defer the analysis and design of the distributed controllers to the next chapter.

8.5 Further Reading

The early literature on distributed estimation (and control) focused on extending optimal estimation techniques in which the information distribution network was either completely connected [**?**, **?**] or hierarchical [**?**]. A key question was how to incorporate data taken from a number of sensors into

either a centralized node or a (completely connected) set of agents. Much
of the early work was focused on the problem of target tracing across a dis-
tributed geographical area. A fairly general version of these decentralized
estimation results that made use of the information form of the Kalman
filter was presented by Rao et al. [**?**].

Chapter 9
Distributed Control

Outline:

- Problem discription: centralized, decentralized, distributed, coupled

- Stability and performance of networked systems (Fax stability criterion + sensitivity analysis by Z. Jin and S. Tonetti)

- (Sub) Optimal distributed control

- Spatially invariant systems (Dulerud, D'Andrea et al + Rotkowitz, Lall)

9.1 Introduction

9.2 Stability and performance of interconnected systems

9.3 Stability of interconnected sytems

Note: Plan to describe the stability conditions from Alex Fax that show how graph topology interacts with dynamics. Other possible things to include:

- Signal flow graphs (useful for computing sensitivity functions, ala Stefania)

- Formula for characteristic equation for the graph Laplacian (from Håkan)

[RMM, 19 Jun 09]

Suppose that each agent's dynamics are governed by

$$
\begin{aligned}
\dot{x}^i &= Ax^i + Bu^i \\
y^i &= Cx^i
\end{aligned}
\tag{9.1}
$$

Fax [FM04] considers a control law in which each system attempts to stabilize itself relative to its neighbors. This is accomplished by constructing an error for each system that is a weighted combination of the relative outputs of the neighbors:

$$
e^i = \sum_{j \in \mathcal{N}^i} \alpha_{ij}(y^j - y^i)
\tag{9.2}
$$

where α_{ij} is the relative weight. For simplicity, we consider uniform weighting here, so that $\alpha_{ij} = 1/|\mathcal{N}^i|$ where $|\mathcal{N}^i|$ is the number of neighbors of node i. The results are easily extended to the more general case.

Given the error (9.2), we apply a compensator that attempts to stabilize the overall system. For simplicity, we assume here that the stabilizer is given by a constant gain

$$u^i = Ke^i, \tag{9.3}$$

with $K \in \mathbb{R}^{m \times m}$ representing the compensation (gain) matrix. In practice, one can use a dynamic compensator to improve performance, but for analysis purposes we can just assume these dynamics are included in the system dynamics (9.1).

The interconnectedness of the system, represented by the neighbor sets $\mathcal{N}_i$ can be studied using tools from graph theory. In particular, for the the case of uniform weighting of the errors, it turns out that the combined error vector $e \in \mathbb{R}^{N \cdot m}$ can be written as

$$e = (\bar{L} \otimes I)x \tag{9.4}$$

where $\otimes$ represents the Kronecker product and $\bar{L}$ is the *weighted Laplacian* associated with the (directed) graph that models the neighbors of each node. The weighted Laplacian is a standard object in graph theory and can be defined as

$$\bar{L} = D^{-1}(D - A)$$

where D is a diagonal matrix whose entries are the out-degree of each node and A is the adjacency matrix for the graph (see [FM04] for more detail). Using this framework, Fax showed the following:

Theorem 9.1. *A local controller K stabilizes the formation dynamics in equation (9.1) with error (9.4) and gain K if and only if it stabilizes the set of N systems given by*

$$\begin{aligned}
\dot{x} &= Ax + B \cdot \lambda_i \cdot (Ky) \\
y &= Cx
\end{aligned} \tag{9.5}$$

where $\{\lambda_i\}$ are the eigenvalues of the weighted graph Laplacian $\bar{L}$.

Proof. We make use of the following notational conventions:

- $\widehat{A} = I_N \otimes A$: block diagonal matrix with A as elements
- $A_{(n)} = A \otimes I_n$: replace elemnts of A with $a_{ij}I_n$
- For $X \in \mathbb{R}^{r \times s}$ and $Y \in \mathbb{R}^{N \times N}$, $\widehat{X}Y_{(s)} = \widehat{Y}X_{(r)}$

Let T be a Schur transformation for L, so that $U = T^{-1}LT$ is upper triangular. Transform the (stacked) process states as $\tilde{x} = T_{(n)}x$ and the
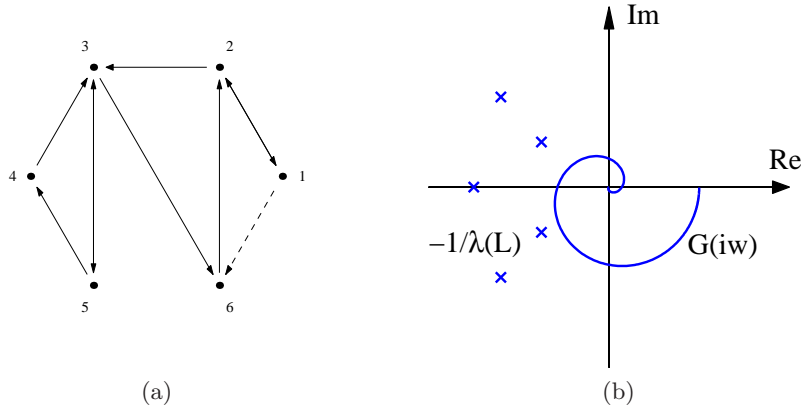
Figure 9.1: Interpretation of Theorem 1. The left figure shows the graph representation of the interconnected system and the right figure shows the corresponding Nyquist test. The addition of the dashed line to the graph moves the negative, inverse eigenvalues of $\bar{L}$ from the positions marked by circles to those marked by crosses.

(stacked) controller states as $\tilde{\xi} = T_{(n)}\xi$. The resulting dynamics become

$$\frac{d}{dt}\begin{bmatrix}\tilde{x}\\\tilde{\xi}\end{bmatrix} = \begin{bmatrix}\widehat{A} + \widehat{B}\widehat{K}\widehat{C}U_{(n)} & \widehat{B}\widehat{H}\\ \widehat{G}\widehat{C}U_{(n)} & F\end{bmatrix}\begin{bmatrix}\tilde{x}\\\tilde{\xi}\end{bmatrix}.$$

This system is upper triangular, and so stability is determined by the elements on the (block) diagonal:

$$\frac{d}{dt}\begin{bmatrix}\tilde{x}_j\\\tilde{\xi}_j\end{bmatrix} = \begin{bmatrix}A + BKC\lambda_j & BH\\ GC\lambda_j & F\end{bmatrix}\begin{bmatrix}\tilde{x}\\\tilde{\xi}\end{bmatrix}.$$

This is equivalent to coupling the process and controller with a gain λ_i.   □

This theorem has a very natural interpretation in terms of the Nyquist plot of dynamical system. In the standard Nyquist criterion, one checks for stability of a feedback system by plotting the open loop frequency response of the system in the complex plane and checking for net encirclements of the -1 point. The conditions in Theorem 1 correspond to replacing the -1 point with $-1/\lambda_i$ for each eigenvalue λ_i of $\bar{L}$. This interpretation is illustrated in Figure 9.1. The results can easily be extended to consider weightings that are nonuniform.

Theorem 9.1 illustrates how the *dynamics* of the system, as represented by equation (9.1), interacts with the *information flow* of the system, as represented by the graph Laplacian. In particular, we see that it is the eigenvalues of the Laplacian that are critical for determining stability of the overall system. Additional results in this framework allow tuning of the information flow (considered as both sensed and communicated signals) to improve the transient response of the system [FM04]. Extensions in a

stochastic setting [**?**, OSM04] allow analysis of interconnected systems whose dynamics are not identical and where the graph topology changes over time.

9.4 (Sub-) Optimal Control

9.5 Spatially Invariant Systems

Bibliography

[AK86]   R. Akella and P. R. Kumar. Optimal control of production rate in a failure
          prone manufacturing system. *IEEE Transactions on Automatic Control*, AC-
          31(2):116–126, February 1986.

[FM04]   J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle
          formations. *IEEE Transactions on Automatic Control*, 49(5):1465–1476, 2004.

[Mur03]   R. M. Murray, editor. *Control in an Information Rich World: Report of the
          Panel on Future Direcitons in Control, Dynamics and Systems*. SIAM, 2003.
          Available at `http://www.cds.caltech.edu/~murray/cdspanel`.

[OSM04]   R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents
          with switching topology and time-delays. *IEEE Transactions on Automatic
          Control*, 49(9):1520–1533, 2004.

Index