UNIVERSITY of CALIFORNIA
Santa Barbara

Communication scheduling methods for estimation over networks

A Dissertation submitted in partial satisfaction of the
requirements for the degree

Doctor of Philosophy
in
Electrical and Computer Engineering

by

Yonggang Xu

Committee in charge:

Professor João P. Hespanha, Chair
Professor Francesco Bullo
Professor Roy S. Smith
Professor Andrew R. Teel

March 2006

The dissertation of Yonggang Xu is approved.

Professor Francesco Bullo

Professor Roy S. Smith

Professor Andrew R. Teel

Professor João P. Hespanha, Committee Chair

November 2005

Communication scheduling methods for estimation over networks

by

Yonggang Xu

To the memory of my father

Shuiming Xu (1945–2004)

Acknowledgements

For the completion of this work I am grateful to my adviser Professor João Pedro Hespanha for his support and guidance. João has much for me to learn from, professionalism and dedication, among many others. The knowledge and skills I learned from him will be my great assets for all the years to come. Never would I forget of his warmth the first time I went to his office at the end of 2001.

It has been a great experience to be a member of Center for Control Engineering and Computation in UC Santa Barbara. From the professors, I learned not only knowledge, but also the willingness to share and collaborate. I am grateful to Professor Peter Kokotović, who showed me trails in Santa Barbara mountains and helped me making important decisions. I also want to thank Professor Brad Paden for having me here in UCSB.

Professors on my committee, Francesco Bullo, Roy Smith, and Andrew Teel, spent much of their Thanksgiving Holiday to review the dissertation draft. Their suggestions and critiques are essential and greatly appreciated.

All my colleagues and friends made my study and life so memorable in UCSB. On whiteboards, we studied real analysis and jump systems. At home we enjoyed international potluck dinners and birthday parties. On the local mountains and Grand Canyon trails were our footprints.

I owe the most to my family, my grandma, my late father, my mother, and my sister Yan. My parents worked hard and made every attempt for their children; they earned little yet managed to bring the family a vision, for which my dad made the ultimate sacrifice... Xiaoling supported me with love during various phases of the dissertation.

Curriculum Vitæ

Yonggang Xu

Yonggang Xu was born in Xiaoshan County, Zhejiang Province, China, in 1975.

EDUCATION

1998	B.S. Vehicle Engineering. Tsinghua University, Beijing, China.
2000	M.S. Vehicle Engineering. Tsinghua University, Beijing, China.
2002	M.S. Electrical Engineering. University of California, Santa Barbara

PROFESSIONAL EMPLOYMENT

1998–2000	Graduate Researcher, Tsinghua University, Beijing.
2000–2001	Teaching Assistant, University of California, Santa Barbara.
2001–2005	Research Assistant, University of California, Santa Barbara.

SELECTED PUBLICATIONS

Y. Xu and J.P. Hespanha. "Optimal communication logics for networked control systems." In *Proceedings of the 43rd IEEE Conference on Decision and Control*. pp 3527-3532, Dec 2004.

Y. Xu and J.P. Hespanha. "Design and analysis of communication logics in networked control systems." Book chapter in *Current Trends in Nonlinear Systems and Control*. Birkhauser, 2005.

Y. Xu and J.P. Hespanha. "Estimation under uncontrolled and controlled communications in networked control systems." In *Proceedings of the 44rd IEEE Conference on Decision and Control*. pp 842-847, Dec 2005.

J.P. Hespanha, P. Naghshtabrizi, and Y. Xu. "A Survey of recent results in networked control systems." To appear in the *Proceedings of IEEE*.

Abstract

Communication scheduling methods for estimation over networks

by

Yonggang Xu

This work focuses on data communication scheduling methods for estimation over networks. A communication scheduler is a mechanism that determines when a data packet should be sent to the network, given the information pattern available. For stochastic communication schedulers, decisions to send data are based on outcomes of random processes. Two issues are addressed in the dissertation: communication scheduler analysis and communication scheduler optimization.

We propose various types of stochastic communication schedulers based on counting processes. We model the dynamics as jump-diffusion processes and use stochastic analysis methods based on infinitesimal generators. The networks we consider include erasure networks and delayed networks.

We construct optimal stochastic communication schedulers that minimize network usage and maximize estimation performance. We formulate a long-term average cost problem with an unbounded per-stage cost function on a Borel state space, which is solved by dynamic programming techniques. The results shed light on how to schedule data communication for networked control systems (NCSs).

Another contribution is on data rate limitations. We derive conditions for the existence of stable estimators under both the *bit rate* and the *packet rate*

limitations. The conditions show a trade-off relation between the bit rate and the packet rate. We observe that if the packet size is large, then increasing it (and therefore increasing the bit rate) helps little, and the packet rate becomes the dominant factor. This observation encourages us to think in terms of *packets* instead of *bits*, as we do in most parts of the dissertation.

Contents

List of Figures

Chapter 1

Introduction

Networked Control Systems (NCSs) are spatially distributed systems in which communication between sensors, actuators, and controllers is supported by a shared communication network (see Figure 1.1).

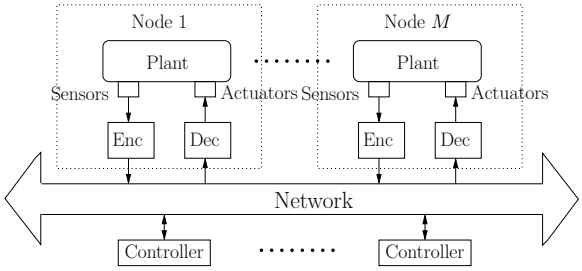


Figure 1.1. General NCS architecture.

In NCSs feedback is implemented by information flow over data networks. The research involves both the classic information theory and control theory. Murray *et al.* [32] identify *control over networks* as one of the key future directions for control. Readers are referred to [2, 17] for a general review.

NCSs have found numerous applications in a broad range of areas such as

mobile sensor networks [37], remote surgery [31], haptics collaboration over the Internet [16, 20, 48], and automated highway systems and unmanned aerial vehicles [45, 46].

This dissertation focuses on data communication scheduling methods for NCSs. A communication scheduler[1] is a mechanism that determines whether or not a node sends data to the network given information available. By the term *node*, we mean locally interconnected components, *e.g.*, a plant, its sensors/actuators, and its encoder/decoder (see Figure 1.1).

The mechanism can be either deterministic or stochastic. For a deterministic communication scheduler, the decision to send data is a deterministic function of the given information set. For a stochastic scheduler, the decision is based on outcomes of a certain random process. We will propose, analyze, and optimize communication schedulers.

Section 1.1 is on the motivation of the dissertation, from the viewpoint of network applications. In Section 1.2, we survey the literature on NCSs with emphasis on *estimation over networks* and *minimum packet rate*. In Section 1.3, we will give an overview of the main dissertation work. Finally, Section 1.4 outlines the chapters to come.

[1] In the previous papers, we used the terminology *communication logic*. The "logic" *schedules* when or whether to send data packets. In this perspective, the term *scheduler* may be more appropriate, which is adopted throughout the dissertation.

1.1 Motivation

The dissertation is on communication schedulers for estimation over networks. We give motivation behind this work.

1.1.1 Estimation over network problems

The applications of estimation over networks include remote sensing, space exploration, and sensor networks, among others. It is also a crucial component of certainty equivalence NCS controllers, which construct control signals based on state estimates of a remote systems. In several NCS scenarios, certainty equivalence controllers are not optimal, but they are still of a great practical interest due to the difficulties in designing optimal controllers. In the dissertation, we only consider estimation over networks. However, we can extend the idea to control problems.

In [60], Xu and Hespanha consider stabilizing certainty equivalence controllers for spatially distributed processes whose dynamics is decoupled but whose control objectives are not, *e.g.*, a group of autonomous aircraft flying in a geometric formation far enough from each other to have decoupled dynamics. In [51], Smith and Hadaegh consider the control of spacecraft, in which *parallel estimators* are used to construct local controls and estimators are updated via communication among the formation entities. They note that each craft having an estimate of the entire formation makes it possible to find an optimal control for the whole formation and for each craft to implement higher level functions such as collision avoidance.

1.1.2 Communication schedulers

The main reason to study communication schedulers is that data is transmitted packet by packet in most communication systems. We emphasize that a node makes communication decisions actively. We aim to reduce the communication load by properly choosing communication schedulers. In general, reducing packet rate (network load) improves the real-time network performance by reducing the packet dropout probability, shortening communication delays, and saving radio power, among others.

Communication schedulers can either be applied to existing communication protocols to better serve NCSs or be implemented in new protocols for NCSs applications. For the input-output stability analysis of NCSs for a broad class of medium access control (MACs) methods, the readers are referred to [36]. However, our emphasis is to explore the application potential of communication schedulers.

In carrier sense multiple access (CSMA) protocols, upon detecting that a collision happens, a sending node may choose to send data in a persistent or non-persistent fashion [54]. One algorithm is called the p-persistent CSMA, in which the node sends a packet with probability p, $0 \leq p \leq 1$. An important issue is to choose the value of u, *i.e.*, the probability of data re-sending immediately after a collision. Some research efforts are on how to choose the parameter p to improve network performance. It is shown via simulation in [8] that the *predictive* p-CSMA achieves a smaller collision rate, a lower mean packet delay, and a higher throughput under most traffic conditions.

We extend this idea in our framework, which provides a mechanism to choose

data sending probability u according to how *stringent* data communications are. Network usage is saved to nodes with highest priorities, while collision probability is reduced in the network overall. As we shall see in the dissertation, the parameter u pertains to the intensity function of a stochastic communication scheduler (Chapters 4) or the optimal communication policy (Chapter 5).

1.2 Literature review

There is an extensive list of NCSs research topics: sampling and quantization effects, encoder and decoder, minimum bit rate, minimum packet rate, network dynamics analysis, network protocol design, and controller design, among others. We will examine a few.

1.2.1 Sampling and quantization

Sampling schemes can either be time-driven or event-driven. Åström and Bernhardsson [53] compare the merits of the Riemann sampling (time-driven) and the Lebesgue sampling (event-driven) for one-dimensional systems. Yook *et al.* [63] propose that a node should broadcast the true value of the local plant state when it differs from the estimate known to the remote nodes by more than a given threshold. They show that this scheme results in a system that is bounded-input bounded-output stable. The relation between the threshold level and the message exchange rate is investigated through simulations.

A signal has to be quantized before being encoded and sent to a digital channel. A quantizer is a device that converts real numbers (an analog signal) into

a finite set of integers (a digital signal). Mathematically, a quantizer is a piecewise constant function, mapping a *quantization region* to a *quantization point*. Usually, a quantization region is a pre-specified rectangular shape. More efficient quantization schemes (for controls) have been developed over the years.

Both [13] and [21] advocate logarithmic-based quantization methods. Roughly speaking, quantization region becomes larger logarithmically as the distance from the origin grows. Quantization regions are allowed to evolve with time to capture the system dynamics [18].

Bullo and Liberzon [7] focus on designing the least destabilizing quantizer subject to a given information constraint. They show that quantizer design can be reduced to a version of the so-called *multicenter problem* from locational optimization, for which they develop iterative algorithms.

We use event-driven sampling methods for large part of the dissertation. And therefore it is related to [53] and [63]. Quantization is not the emphasis of the dissertation, as we assume that a data packet holds a real number with a precision sufficient for control or estimation purposes. However, we do allow bounded quantization errors in our analysis.

1.2.2 Minimum bit rate

Any communication network can only carry a finite amount of information per unit time. Inspired by Shannon's results on the *maximum bit rate* that a finite-bandwidth channel can reliably carry, a significant research effort has been devoted to the problem of determining the *minimum bit rate* to stabilize a system through feedback. The problem of determining the minimum information flow

needed for stabilization has been solved exactly for linear plants [59, 18, 34, 35, 57, 43] and preliminary results have been obtained for nonlinear plants [33, 28]. Other issues, such as robustness, have also been addressed [40, 27, 29, 55].

We will come back to this topic in Chapter 2. But we intend to depart from *bit rate* to the so-called *packet rate*.

1.2.3 Minimum packet rate

In digital communication networks, data is transmitted in atomic units called *packets*[2]. In parallel to bit rate, *packet rate* refers to the average number of data packets per unit time. We review *estimation over lossy networks* problems to demonstrate the effects of the packet rate.

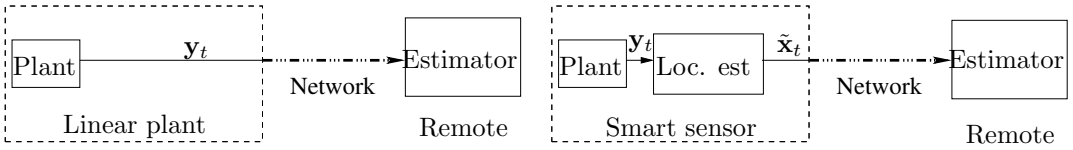
We consider erasure networks, in which a data packet can carry a real number without distortion, but that some packets may be lost. Two scenarios for state estimation over networks are investigated. In the one depicted in Figure 1.2(a), every raw sensor measurement $\mathbf{y}_t$ is sent to the remote estimator, but may not arrive there if there is a packet dropout [30, 50, 49]. Alternatively, in the scenario shown in Figure 1.2(b), the raw sensor measurements are processed locally and then sent to the remote estimator via the same network [60, 63, 62].

We restrict our attention to linear time-invariant (LTI) plants with Gaussian measurement noise and disturbance:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{y}_t = C\mathbf{x}_t + \mathbf{v}_t, \quad \forall\, t \in \mathbb{N},\ \mathbf{x}_t, \mathbf{w}_t \in \mathbb{R}^n,\ \mathbf{y}_t, \mathbf{v}_t \in \mathbb{R}^\ell, \quad (1.1)$$

where the initial state $\mathbf{x}_0$ is zero-mean Gaussian with covariance matrix Σ, and

[2]Here the term packet is generic. Depending on different networks, it may take the form of (variable-size) *packet* or (fixed-size) *cell*, among others [58].



(a) (b)

Figure 1.2. Scenarios for the state estimation over a network. In (a) the raw sensor measurements are sent to a remote estimator, whereas in (b) the measurements are processed locally before transmission.

the zero-mean Gaussian white noises $\mathbf{w}_t$ and $\mathbf{v}_t$ are mutually independent with covariance matrices $\Sigma_w \geq 0$ and $\Sigma_v > 0$, respectively. We assume that (C, A) is detectable and (A, Σ_w) is stabilizable.

We discuss the minimum packet rates required to construct estimators for configurations in Figure 1.2(a) as well as Figure 1.2(b).

Minimum packet rate without local calculation

First we consider the architecture in Figure 1.2(a), in which the measurements, $\mathbf{y}_t$, are sent. The optimal estimate of $\mathbf{x}_t$ ($t \in \mathbb{N}$) can be computed recursively using the following time-varying Kalman filter (TVKF) [23]:

$$\begin{aligned} \hat{\mathbf{x}}_{0|-1} &= 0, \\ \hat{\mathbf{x}}_{t|t} &= \hat{\mathbf{x}}_{t|t-1} + \boldsymbol{\nu}_t \mathbf{L}_t (\mathbf{y}_t - C\hat{\mathbf{x}}_{t|t-1}), \qquad \forall\, t \in \mathbb{N}, \\ \hat{\mathbf{x}}_{t+1|t} &= A\hat{\mathbf{x}}_{t|t}, \end{aligned} \tag{1.2}$$

where the matrix gain $\mathbf{L}_t$ is calculated recursively as follows

$$\begin{aligned}
\mathbf{P}_0 &= \Sigma, \\
\mathbf{L}_t &= \boldsymbol{\nu}_t \mathbf{P}_t C'(C\mathbf{P}_t C' + \Sigma_v)^{-1}, && \forall\ t \in \mathbb{N}, \\
\mathbf{P}_{t+1} &= A\mathbf{P}_t A' + \Sigma_w - A\mathbf{L}_t(C\mathbf{P}_t C' + \Sigma_v)\mathbf{L}_t' A',
\end{aligned}$$

where $\mathbf{P}_t$ ($\forall\ t \in \mathbb{N}$) is an estimation error covariance matrix, *i.e.*,

$$\mathbf{P}_t = \mathrm{E}[(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})(\mathbf{x}_t - \hat{\mathbf{x}}_{t|t-1})'].$$

Sinopoli *et al.* [49] study the performance of this Kalman filter when $\boldsymbol{\nu}_t$ is a Bernoulli process with probability of dropout ($\boldsymbol{\nu}_t = 0$) equal to $p \in [0, 1)$. They show the existence of a critical value p_c for the dropout rate p, above which the estimation error covariance matrix becomes unbounded. The critical value p_c satisfies $\underline{p} \leq p_c \leq \bar{p}$, where the upper bound, $\bar{p}$, is given by

$$\bar{p} = \frac{1}{(\max\{|\operatorname{eig}(A)|\})^2}, \tag{1.3}$$

and the lower bound, $\underline{p}$, is given by the solution to a linear matrix inequality. In special cases (*e.g.*, the matrix C is invertible) the upper bound in (1.3) is tight in the sense that $p_c = \bar{p}$.

Matveev and Savkin [30] consider multiple sensors, each independently sending its measurements to the estimator with some delay. This corresponds to the following plant model,

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{w}_t, \qquad \mathbf{y}_{s,t} = C_s\mathbf{x}_t + \mathbf{v}_{s,t}, \quad s \in \{1, \cdots, N\}, \qquad \forall\ t \in \mathbb{N}, \tag{1.4}$$

where $\mathbf{y}_{s,t}$ denotes the measurement collected by sensor s at time t. Assuming that the measurement $\mathbf{y}_{s,t}$ suffers a random delay of $\boldsymbol{\tau}_s(t)$, the optimal state

estimate at time t is given by

$$\hat{\mathbf{x}}_t = \mathrm{E}\left[\mathbf{x}_t \;\middle|\; \mathbf{y}_{s,\ell},\ \forall\ \ell \leq t - \boldsymbol{\tau}_s(\ell) \text{ s.t. } \boldsymbol{\theta}_{s,\ell} = 1\right],$$

where $\boldsymbol{\theta}_{s,\ell} = 1$ if $\mathbf{y}_{s,\ell}$ reaches its destination, and $\boldsymbol{\theta}_{s,\ell} = 0$ otherwise. They derive a recursive Kalman filter and provide conditions under which the estimation error process is almost surely stable [30]. These conditions are given in terms of the observability of $\mathbf{x}_t$ for specific realizations of the process $\boldsymbol{\theta}_{s,t}$.

Minimum packet rate for estimation with local computation

The encoder-estimator scheme in Figure 1.2(b) is motivated by the growing number of *smart sensors* with embedded processing units that are capable of local computation. In this context, [62] investigates the benefits of pre-processing measurements before transmission to the network. For the LTI plant (1.1), the smart sensor computes locally an optimal state estimate, $\tilde{\mathbf{x}}_{t|t} = \mathrm{E}[\mathbf{x}_{t|t}|\mathbf{y}_s, s \leq t]$, using a standard Kalman filter. The local node then transmits the local estimate $\tilde{\mathbf{x}}_{t|t}$ (instead of the raw measurement $\mathbf{y}_t$), which is used by the remote estimator to compute the optimal estimate $\hat{\mathbf{x}}_{t|s}$ of $\mathbf{x}_t$ ($t \in \mathbb{N}$) given all the data $\{\tilde{\mathbf{x}}_{r|r} : \boldsymbol{\nu}_r = 1, \forall\ \ r \leq s\}$ successfully received up to time $s \leq t$,

$$\hat{\mathbf{x}}_{t|s} = \mathrm{E}\left[\mathbf{x}_t \mid \tilde{\mathbf{x}}_{r|r}, \forall\ r \leq s \text{ s.t. } \boldsymbol{\nu}_r = 1\right].$$

The main advantage of using local computation is that each message $\tilde{\mathbf{x}}_{t|t}$ that successfully reaches the remote estimator encodes all the relevant information that can be extracted from *all raw measurements collected up to time t.* In general, this permits stability of the error process $\mathbf{e}_t$ for larger dropout probabilities than those permitted by the architecture in Figure 1.2(a), in which raw measurements are sent over the network. We will discuss further in Chapter 4.

1.3 Dissertation research

We characterize the main dissertation work as *communication scheduler analysis* and *communication scheduler optimization*. We also extend results on the minimum data rate problem, in which we consider both the *bit rate* and the *packet rate*. Now we take snapshots of the main results with informal explanations in the first part of this section. In the second part, we state the motivation from the viewpoint of control networks.

We mainly discuss estimation (of LTI systems) over networks problems. The estimation scheme consists of a local node, a remote estimator, and a network, as shown in Figure 1.3. The local node has a local estimator with the state $\tilde{\mathbf{x}}(t)$ and a communication scheduler. The communication scheduler decides whether or not a data packet should be sent to the network given information available. We focus on stochastic communication schedulers, in which decisions to send data are based on outcomes of a certain random process.

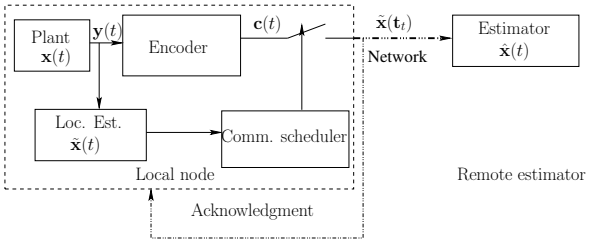


Figure 1.3. A remote estimation scheme with a communication scheduler

Communication scheduler analysis

In the continuous-time domain, a communication scheduler is modeled as a counting process (Chapter 3) with a certain intensity function. The intensity means

the average incremental rate of the counting process. For the special case of a Poisson process, the intensity function is the Poisson rate λ. The jump intensity function can be either independent of or dependent on the system dynamics. We call the former an *uncontrolled communication scheduler* and the latter a *controlled communication scheduler*.

We use a jump-diffusion process to model the estimation error $\mathbf{e}(t)$, which evolves as a diffusion process most of the time and resets to the neighborhood of the origin when there are data arrivals from the network. The continuous-time LTI system can be estimated over the network if the error $\mathbf{e}(t)$ is bounded in a statistical sense. Our basic conclusions are roughly stated as follows:

1. When the intensity λ is constant, the continuous-time LTI system can be estimated in the mean square sense if

$$\lambda > 2 \max\{\Re[\text{eig}(A)]\}.$$

   The equation above means that the data communication has to be fast compared to real part of the least stable LTI system pole. This is in contrast to the minimum bit rate problem, in which all the unstable poles are relevant [34, 18].

2. We give feedback mechanisms to communication schedulers by relating the estimation error to the intensity function $\lambda(\cdot)$, *e.g.*,

$$\lambda(e) = (e'Pe)^k,$$

   for any $P > 0$ and $k > 0$. We prove that this *controlled communication scheduler* always gives enough data packets for LTI system estimation.

We consider a much richer class of communication schedulers in Chapter 4. Also we address issues such as data packet losses and communication delays in the same framework.

Optimal communication schedulers

We summarize the main results on "optimal" communication schedulers. The objective is to minimize both the average estimation error variance and the communication cost, *i.e.*, the cost function takes the form of:

$$\text{cost} = \text{estimation error variance} + \text{communication load}.$$

In the discrete-time domain, the communication scheduler assigns a data sending probability u_t at each time t. We optimize the estimation scheme with respect to local estimators, communication schedulers, and remote estimators. However, the emphasis is on optimal communication schedulers. We take a dynamic programming (DP) approach. In essence, we address a long-term average cost problem with an unbounded per-stage cost. Our basic conclusions are roughly stated as follows:

1. If the network does not drop data packets, a DP equation exactly solves the optimal communication scheduling problem. We develop an iterative numerical method to solve the optimal communication scheduler and calculate the optimal cost. The optimal communication scheduler is stationary and deterministic, and it takes the form of:

$$u_t = \begin{cases} 0, & \text{if the estimation error is in } \Omega \\ 1, & \text{otherwise,} \end{cases}$$

where Ω is a set that can be pre-computed.

2. If the network drops data packet with probability p, we give both upper and lower bounds. The upper bound is related to the solution to a Lyapunov equation, while the lower bound is by recourse to results from the lossless network case.

Bit and packet: a unified framework

Consider the remote estimation scheme in Figure 1.3, in which the local node sends data packets to the remote estimator. The cardinality (the number of all possible *codeword*) of each packet is μ, and therefore the packet size is $\mathcal{R}_b := \log_2 \mu$ bits. Packets are sent with probability $1-p$, $0 \leq p \leq 1$, and therefore the average packet rate is $\mathcal{R} = 1-p$. For the remote estimation of a discrete-time LTI system (A, C), where $A \in \mathbb{R}^{n\times n}$, we have the following:

1. If integer $\mu > |\operatorname{eig}_i(A)|$ and $p < \frac{1}{|\operatorname{eig}_i(A)|^2}$, $1 \leq i \leq n$, the system can be estimated in the mean square sense over the channel, if and only if

$$\mu^2 > \prod_{|\operatorname{eig}_i(A)|\geq 1} \frac{(1-p)|\operatorname{eig}_i(A)|^2}{1-p|\operatorname{eig}_i(A)|^2}.$$

2. We observe that when the bit rate $\mathcal{R}_b$ is small, increasing it effectively reduces the packet rate requirement. However, it becomes less effective when the packet size $\mathcal{R}_b$ is large.

We will unify some key results on data rate limitations in the literature (see Chapter 2).

1.4 Summary

The dissertation embraces two issues on data communication scheduling methods:

1. We propose stochastic communication schedulers. Stochastic analysis methods are used to analyze the dynamics as jump-diffusion processes. The analysis methods are applicable to practical networks models.

2. We derive optimal stochastic communication schedulers that save network usage and improve estimation performance. Dynamic optimization techniques are applied to approach a long-term average cost problem with an unbounded per-stage cost function on a Borel state space. The results shed light on how to schedule data communication for NCSs.

Another contribution is on the extension of data rate limitations. In the context of an estimation over networks problem, we derive conditions for the existence of estimators in terms of the *bit rate* and the *packet rate*. The conditions also shed light on how to allocate the data packet size, which is instrumental in NCS communication protocol design.

In Chapter 2, we give a unified answer to both the bit rate and the packet rate requirements in the context of the estimation of discrete-time LTI systems over networks.

In Chapter 3, we briefly summarize results on jump-diffusion processes, which provides analysis tools for Chapter 4.

In Chapter 4, we propose stochastic communication schedulers for estimation over networks problems and analyze the estimator's stability properties. We consider plants with full and partial state measurements and pursue two objectives:

observability and finite network resource usage. We use *counting processes* such as the Poisson process to model the schedulers and use *jump-diffusion processes* to model estimator dynamics. The method based on *infinitesimal generator* allows us to analyze a large class of stochastic communication schedulers and various types of networks (*e.g.*, networks with delays and networks with packet dropouts).

In Chapter 5, we pursue optimal communication schedulers. For networks with no data dropouts, we solve an optimal communication problem via an average cost optimality equality (ACOE). To do that, we take advantage of some recurrent properties of the process and pose it as an essentially bounded per-stage cost problem. For general networks with a packet dropout probability p, we obtain both an upper bound and a lower bound on the optimal solution. The upper bound is obtained by solving an average cost optimality inequality (ACOI), and the lower bound is from an ACOE.

In Chapter 6, we give some concluding remarks and directions for future work.

Chapter 2

Rate limitations

This chapter is on data *rate limitations*, which include both the *bit rate* limitation and the *packet rate* limitation. The limitations are illustrated through the estimation of a discrete-time linear time-invariant (LTI) plant over networks. The local node sends data to the remote estimator, one *packet* per time step. Each packet contains the same fixed number of data *bits*. Due to network uncertainties, a packet may get lost with probability p ($0 \leq p \leq 1$), which results in an average packet rate of $1 - p$. We investigate whether it is feasible to construct estimators for the LTI systems over such communication channels.

In Section 2.1, we state the main results, which reveal trade-off relations between packet size and required data transmission reliability. In Section 2.2, we discuss several special cases and then make connections to results in the literature. In Section 2.3, we list a few observations, which convince us to go from *bits* to *packets*.

2.1 Rate limitations for LTI systems

In this section, we derive necessary and sufficient conditions for stability of a remote encoder-estimator scheme. For simplicity, we start with one-dimensional noiseless LTI systems, followed by systems with noise. Explicit encoder-estimator pairs are also constructed. The results are then extended to general LTI systems.

2.1.1 Problem setup

We consider the estimation over networks (see Figure 2.1.1). The network is viewed as a digital channel that transmits one symbol per second from an alphabet (cardinality $\mu \in \mathbb{N}$). One symbol is packed in one data packet, resulting a sending bit rate of [47, 10]:

$$\mathcal{R}_b := \log_2 \mu \quad \text{bps.}$$

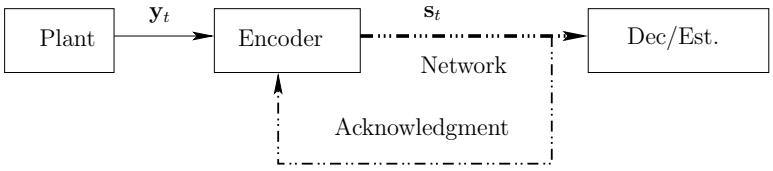


Figure 2.1. Estimation over networks

We consider a general family of *encoders*, which may have infinite memory. In particular, a symbol $\mathbf{s}_t$ is obtained by

$$\mathbf{s}_t = \mathcal{E}_t(\mathcal{Y}^t, \mathcal{S}^{t-1}), \qquad \forall\ t \geq 0, \tag{2.1}$$

where $\mathcal{E}_t(\cdot)$ is the encoding mapping, $\mathcal{Y}^t := \{\mathbf{y}_j \mid j \leq t\}$ denotes the sequence of all measurements up to time t, and $\mathcal{S}^{t-1} := \{\mathbf{s}_j \mid j \leq t-1\}$ denotes the sequence of all symbols sent by the encoder up to time $t-1$.

Similarly, we consider a general family of *decoders/estimators*. In particular, the estimate $\hat{\mathbf{x}}_t$ at time t is obtained by

$$\hat{\mathbf{x}}_t = \mathcal{D}_t(\bar{\mathcal{S}}^t), \qquad \forall\ t \geq 0, \tag{2.2}$$

where $\mathcal{D}_t(\cdot)$ is a decoding mapping and $\bar{\mathcal{S}}^t := \{\bar{\mathbf{s}}_j \mid j \leq t\}$ denotes the sequence of all symbols received by the decoder up to time t.

For "ideal channels", the symbols sent by the encoder always coincide with those received by the decoder, and therefore $\bar{\mathbf{s}}_t = \mathbf{s}_t$, for all $t \geq 0$. However, for *noisy channels* the symbols received may occasionally differ from the ones sent. We are interested in *erasure channels*, in which the symbols $\mathbf{s}_t$ sent to the network take values in the set $\{1, \cdots, \mu\}$, but the symbols $\bar{\mathbf{s}}_t$ received from the network take values in the set $\{1, \cdots, \mu, \emptyset\}$, where $\emptyset$ represents *erasure*. The symbol received differs from the symbol sent with probability p, independently and identically distributed (*i.i.d.*), and we have

$$\bar{\mathbf{s}}_t = \begin{cases} \mathbf{s}_t, & \text{with prob. } 1-p \\ \emptyset, & \text{with prob. } p \end{cases} \qquad \forall\ t \geq 0.$$

The encoder is informed when an erasure has occurred and can use this piece of information to encode subsequent symbols [10]. Channels may also introduce delays. For example, in *lossless τ-delay channels* the symbols are received τ time steps after they are sent, which corresponds to the following channel model,

$$\bar{\mathbf{s}}_t = \mathbf{s}_{t-\tau}, \qquad \forall t \geq 0.$$

The following channel is the basic network model in this chapter:

Definition 2.1 (1-delay p-erasure channel with feedback acknowledgment) *A channel is called a* 1-delay p-erasure channel with feedback acknowledgment *if*

it transfers one data packet (with cardinality μ) each time step, the output $\{\bar{\mathbf{s}}_t\}_{t\geq 1}$ and the input $\{\mathbf{s}_t\}_{t\geq 0}$ satisfy,

$$\bar{\mathbf{s}}_t = \begin{cases} \mathbf{s}_{t-1}, & \textit{with prob. } 1-p \\ \emptyset, & \textit{with prob. } p \end{cases} \qquad \forall\ t \geq 1, \tag{2.3}$$

and furthermore at time t the channel acknowledges the encoder whether the symbol $\mathbf{s}_{t-1}$ is received or not. For brevity, the channel is referred to as the erasure channel (2.3) *in the rest of this chapter.*

The expected number of packets the channel successfully transfers per second is $1-p$, which is defined to be the channel *packet rate*,

$$\mathcal{R} = 1 - p. \tag{2.4}$$

2.1.2 One-dimensional noiseless LTI systems

Consider one-dimensional discrete-time unstable noiseless systems

$$\begin{aligned} \mathbf{x}_{t+1} &= A\mathbf{x}_t, \quad \mathbf{x}_0 \in X_0, \\ \mathbf{y}_t &= \mathbf{x}_t, \end{aligned} \tag{2.5}$$

where $|A| > 1$, $\mathbf{x}_0$ is the only source of uncertainty, and X_0 is a closed interval[1]. The initial set X_0 is known to both the encoder and the estimator. We consider estimation of (2.5) over the erasure channel (2.3).

Assumption 2.1 *The initial condition $\mathbf{x}_0$ is uniformly distributed on X_0.*

[1]For unbounded X_0, one can use techniques based on entropy, as in [34].

From Assumption 2.1, $\mathbf{x}_t$ is uniformly distributed on $A^t X_0$. Furthermore, for any subset $\mathcal{B}_t \subset A^t X_0$, $\mathbf{x}_t \in \mathcal{B}_t$ implies that $\mathbf{x}_t$ is uniformly distributed on $\mathcal{B}_t$. Under Assumption 2.1, it can be shown that, for any μ-quantizer $\mathcal{Q}_\mu : X \to \mathbb{Z}_\mu$, the infimum,

$$\inf_{\mathcal{Q}_\mu} \max_{i \in \mathbb{Z}_\mu} \min_{\hat{\mathbf{x}}} \mathrm{E}\left[|\mathbf{x} - \hat{\mathbf{x}}\left(\mathcal{Q}_\mu(\mathbf{x})\right)|^m \mid \mathcal{Q}_\mu(\mathbf{x}) = i \right]$$

is achieved by a uniform quantization, where

$$\hat{\mathbf{x}} : \mathbb{Z}_\mu \to X$$

is an estimate of $\mathbf{x}$. Without loss of generality, we further assume that encoders are uniform.

Assumption 2.2 *Encoders are uniform,* i.e., *the bounded set X is equally partitioned into μ subsets, each of which is assigned a unique codeword.*

Necessary and sufficient conditions

We derive necessary and sufficient *observability* conditions for the estimation over networks problem. Stability is defined as boundedness of a signal in stochastic moment sense. Refer to [25] for general definitions on stochastic stabilities.

Definition 2.2 (Stability in the m^{th} moment) *A discrete-time random process $\mathbf{x}_t$ has initial distribution $\mathbf{x}_t = \mathbf{x}_0$ such that $\mathrm{E}[|\mathbf{x}_0|^m] < \delta$, for some $\delta > 0$. The process $\mathbf{x}_t$ is stable in the m^{th} moment, if for any $T > 0$, there exists $\epsilon > 0$, such that for any $t > T$, we have*

$$\mathrm{E}[|\mathbf{x}_t|^m] < \epsilon,$$

where $|\cdot|$ stands for the Euclidean norm. If in addition,

$$\lim_{t\to\infty} \mathrm{E}[|\mathbf{x}_t|^m] = 0,$$

the process is asymptotically stable in the m^{th} moment.

Obervability is defined accordingly:

Definition 2.3 (Observability in the m^{th} moment) *A linear system with the state $\mathbf{x}_t$ is (resp. asymptotically) observable over the* erasure channel (2.3) *in the m^{th} moment if there exists an encoder-decoder pair such that $\mathbf{x}_t - \hat{\mathbf{x}}_t$ is (resp. asymptotically) stable in the m^{th} moment. When $m = 2$, it is mean-square (resp. asymptotically) observable.*

Theorem 2.1 *For integer $\mu > |A| > 1$, $p \in [0, 1)$, and any $m \in (0, +\infty)$, the system* (2.5) *is asymptotically observable in the m^{th} moment over the erasure channel* (2.3) *if and only if*

$$p < \frac{\mu^m - |A|^m}{\mu^m |A|^m - |A|^m}, \quad \textit{or equivalently,} \quad \mu^m > \frac{(1-p)|A|^m}{1 - p|A|^m}.$$

We interpret the theorem in Figure 2.2, for which we choose $|A| = 2$ and $m = 2$. The system is asymptotically observable in the second moment in the region above the curve. Note that when the drop rate is higher, larger packet sizes are required for observability. We save further comments to Section 2.3, where we will interpret from another viewpoint. Now we prove Theorem 2.1.

Proof.[Theorem 2.1] Suppose that, at time t, the state is $\mathbf{x}_t$, the confidence estimation set is $X_{t|t}$ (the minimum set in which $\mathbf{x}_t$ lies, given the codewords

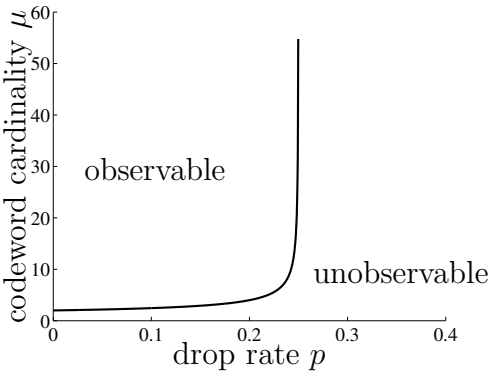


Figure 2.2. Graphical representation of Theorem 2.1.

up to time t), and the best estimate is $\hat{\mathbf{x}}_t$. Assume μ is an positive integer (see Remark 2.1 for general cases). At time $t+1$, if the codeword s_t is not received, the confidence set becomes $AX_{t|t}$. Otherwise, the confidence set becomes one of the μ equal partitions of $AX_{t|t}$. We have that

$$\begin{aligned}\mathrm{E}\left[|\mathbf{x}_{t+1}-\hat{\mathbf{x}}_{t+1}|^m\right] &= \mathrm{E}\left[|A|^m|\mathbf{x}_t-\hat{\mathbf{x}}_t|^m \mid \text{lost}\right] + \mathrm{E}\left[\left(\frac{|A|}{\mu}\right)^m |\mathbf{x}_t-\hat{\mathbf{x}}_t|^m \mid \text{received}\right] \\ &= \left(p|A|^m + (1-p)\left(\frac{|A|}{\mu}\right)^m\right)\mathrm{E}\left[|\mathbf{x}_t-\hat{\mathbf{x}}_t|^m\right],\end{aligned}$$

from which we conclude that for asymptotic observability in the m^{th} moment, it is necessary and sufficient that

$$p|A|^m + (1-p)\left(\frac{|A|}{\mu}\right)^m < 1, \tag{2.6}$$

which is then converted to the inequalities in the theorem. ■

Remark 2.1 *It is conjectured that Theorem 2.1 is valid for non-integer μ. If $\mu = s/t$ is a rational number, where integers s and t are coprime, we can extend the proof above by considering t time steps instead of one time step. This result may be further extend to real numbers as rational numbers are dense on the real axis.*

An encoder-estimator pair

The encoder-estimator pair constitutes a synthesis proof for Theorem 2.1. Denote the centroid of the confidence interval X_t as $\mathbf{z}_t$, the encoding precision as $\mathbf{\Delta}_t$, and the codeword as $\mathbf{c}_t$. Suppose the cardinality of the alphabet μ is an odd integer[2], and the alphabet is denoted by $\mathbb{Z}_\mu := \{\frac{-\mu+1}{2}, \cdots, 0, \cdots, \frac{\mu-1}{2}\}$. We construct the encoder-estimator pair recursively as follows:

- Encoder:

  1. At time 0, initialize $\mathbf{\Delta}_0$, $\mathbf{z}_0$, and $\mathbf{c}_0$:

$$\mathbf{\Delta}_0 = \frac{\max(X_0) - \min(X_0)}{\mu},$$

$$\mathbf{z}_0 = \text{centroid of } X_0,$$

$$\mathbf{c}_0 = 0.$$

  2. At time $t+1$, the sender gets the acknowledgment of whether $\mathbf{c}_t$ is received, from which the encoder calculates the centroid $\mathbf{z}_{t+1}$ and the precision $\mathbf{\Delta}_{t+1}$:

$$\mathbf{z}_{t+1} = \begin{cases} A(\mathbf{z}_t + \mathbf{c}_t\mathbf{\Delta}_t), & \text{if } \mathbf{c}_t \text{ is received,} \\ A\mathbf{z}_t, & \text{otherwise,} \end{cases} \tag{2.7}$$

$$\mathbf{\Delta}_{t+1} = \begin{cases} \frac{|A|}{\mu}\mathbf{\Delta}_t, & \text{if } \mathbf{c}_t \text{ is received,} \\ |A|\mathbf{\Delta}_t, & \text{otherwise.} \end{cases}$$

  3. Also at time $t+1$, the encoder finds the codeword $\mathbf{c}_{t+1}$ such that,

$$\mathbf{x}_{t+1} \in [\mathbf{z}_{t+1} + (\mathbf{c}_{t+1} - \frac{1}{2})\mathbf{\Delta}_{t+1}, \mathbf{z}_{t+1} + (\mathbf{c}_{t+1} + \frac{1}{2})\mathbf{\Delta}_{t+1}).$$

[2]Otherwise, the representation needs to be slightly changed.

4. Finally, at time $t+1$, the encoder sends $\mathbf{c}_{t+1}$ to the network.

- Estimator:

1. At time 0, fetch $\mathbf{\Delta}_0$ and $\mathbf{z}_0$ from the encoder and initialize $\hat{\mathbf{x}}_0$ to be 0.
2. At time $t+1$, the estimator may have received the codeword $\mathbf{c}_t$. The estimate is then calculated:

$$\hat{\mathbf{x}}_t = \begin{cases} \mathbf{z}_t + \mathbf{c}_t\mathbf{\Delta}_t, & \text{if } \mathbf{c}_t \text{ is received,} \\ \mathbf{z}_t, & \text{otherwise,} \end{cases}$$

$$\hat{\mathbf{x}}_{t+1} = A\hat{\mathbf{x}}_t.$$

3. Also at time $t+1$, the decoder updates the centroid $\mathbf{z}_{t+1}$ and the estimation precision $\mathbf{\Delta}_t$:

$$\mathbf{z}_{t+1} = \begin{cases} A(\mathbf{z}_t + \mathbf{c}_t\mathbf{\Delta}_t), & \text{if } \mathbf{c}_t \text{ is received,} \\ A\mathbf{z}_t, & \text{otherwise,} \end{cases}$$

$$\mathbf{\Delta}_{t+1} = \begin{cases} \frac{|A|}{\mu}\mathbf{\Delta}_t, & \text{if } \mathbf{c}_t \text{ is received,} \\ |A|\mathbf{\Delta}_t, & \text{otherwise.} \end{cases}$$

The encoder-estimator pair above, $\mathbf{z}_t$ and $\mathbf{\Delta}_t$ can be synchronized on the encoder and the estimator. The estimation error, $\mathbf{e}_t := \mathbf{x}_t - \hat{\mathbf{x}}_t$, satisfies

$$\mathbf{e}_t \in \begin{cases} (-\frac{1}{2}\mathbf{\Delta}_t, \frac{1}{2}\mathbf{\Delta}_t], & \text{if } \mathbf{c}_t \text{ is received,} \\ (\frac{-\mu+1}{2}\mathbf{\Delta}_t, \frac{\mu-1}{2}\mathbf{\Delta}_t], & \text{otherwise.} \end{cases}$$

Because $\mathbf{e}_t$ is uniformly distributed, it is asymptotically stable in the m^{th} moment if and only if $\mathbf{\Delta}_t$ is. Note that the estimation precision $\mathbf{\Delta}_t$ satisfies

$$\mathbf{\Delta}_{t+1} = \begin{cases} |A|\mathbf{\Delta}_t, & \text{with prob p,} \\ \frac{|A|}{\mu}\mathbf{\Delta}_t, & \text{with prob 1-p.} \end{cases}$$

For the m^{th} moment stability,

$$\begin{aligned}\mathrm{E}[\boldsymbol{\Delta}_{t+1}^m] &= \mathrm{E}[(|A|\boldsymbol{\Delta}_t)^m \mid \mathbf{c}_t \text{ is lost}] + \mathrm{E}\left[\left(\frac{|A|}{\mu}\right)^m \boldsymbol{\Delta}_t^m \mid \mathbf{c}_t \text{ is received}\right] \\ &= \left(p|A|^m + (1-p)\left(\frac{|A|}{\mu}\right)^m\right)\mathrm{E}[\boldsymbol{\Delta}_t^m],\end{aligned}$$

from which the condition (2.6) is sufficient.

2.1.3 One-dimensional LTI systems with noise

Consider a one-dimensional LTI system with bounded random noise $\mathbf{w}_t \in [-L_w, L_w]$ and $\mathbf{v}_t \in [-L_v, L_v]$,

$$\begin{aligned}\mathbf{x}_{t+1} &= A\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{x}_0 \in X_0, \\ \mathbf{y}_t &= \mathbf{x}_t + \mathbf{v}_t,\end{aligned} \tag{2.8}$$

where $|A| > 1$ and X_0 is a closed interval. The following theorem gives a result similar to Theorem 2.1:

Theorem 2.2 *For integer $\mu > |A| > 1$, $p \in [0, 1)$, and any $m \in (0, +\infty)$, the system* (2.8) *is observable in the m^{th} moment over the erasure channel* (2.3) *if and only if*

$$p < \frac{\mu^m - |A|^m}{\mu^m |A|^m - |A|^m}, \quad \textit{or equivalently,} \quad \mu^m > \frac{(1-p)|A|^m}{1 - p|A|^m}.$$

The necessary part of the theorem can be derived as in the proof of Theorem 2.1. For the sufficient part, we construct the same encoder-estimator pair as in the noiseless case. Note now the centroid $\mathbf{z}_t$ and precision $\boldsymbol{\Delta}_t$ are updated

according to,

$$\mathbf{z}_{t+1} = \begin{cases} A(\mathbf{z}_t + \mathbf{c}_t \boldsymbol{\Delta}_t), & \text{if } \mathbf{c}_t \text{ is received,} \\ A\mathbf{z}_t, & \text{otherwise,} \end{cases}$$

$$\boldsymbol{\Delta}_{t+1} = \begin{cases} \frac{1}{\mu}(|A|\boldsymbol{\Delta}_t + 2|A|L_w + 2L_v), & \text{if } \mathbf{c}_t \text{ is received,} \\ |A|\boldsymbol{\Delta}_t + 2|A|L_w + 2L_v, & \text{otherwise,} \end{cases}$$

in which $\boldsymbol{\Delta}_t$ is stable in the m^{th} moment if (2.6) is satisfied.

Remark 2.2 *Bounded noise does not impose additional data rate requirements. This is because the noise terms affect the estimate uncertainty additively, while the data infusions from the new packet arrivals result in uncertainty reduction multiplicatively. However, the asymptotic convergence is lost.*

2.1.4 Multi-dimensional LTI systems

Consider an n-dimensional system with bounded random noise $\mathbf{w}_t \in [-L_w^1, L_w^1] \times \cdots \times [-L_w^n, L_w^n]$ and $\mathbf{v}_t \in [-L_v^1, L_v^1] \times \cdots \times [-L_v^m, L_v^m]$,

$$\begin{aligned} \mathbf{x}_{t+1} &= A\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{x}_0 \in X_0, \\ \mathbf{y}_t &= C\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{2.9}$$

in which X_0 is a bounded set in $\mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m\times n}$, and (C, A) is an observable pair.

The proof of the following theorem runs parallel to that of Theorems 2.1, but requires a coordinate transformation [57] and the construction of a local Kalman estimator (as in Chapter 5). The proof is omitted here.

Theorem 2.3 *If integer* $\mu > |\operatorname{eig}_i(A)|$ *and* $p < \frac{1}{|\operatorname{eig}_i(A)|^m}$, $1 \leq i \leq n$, *for any* $m \in (0, +\infty)$, *the system* (2.9) *is observable in the* m^{th} *moment over the erasure channel* (2.3) *if and only if*

$$\mu^m > \prod_{|\operatorname{eig}_i(A)|\geq 1} \frac{(1-p)|\operatorname{eig}_i(A)|^m}{1-p|\operatorname{eig}_i(A)|^m}.$$

2.2 Special cases

In this section we apply the theorems to derive results that previously appeared in the literature.

2.2.1 Large packet sizes

If the packet size is very large, *i.e.*, $\mu \to \infty$, the condition in Theorem 2.3 becomes

$$p < \frac{1}{\max_i |\operatorname{eig}_i(A)|^m}, \tag{2.10}$$

which is consistent with [62].

2.2.2 Lossless networks

If the channel does not drop data packets, *i.e.*, $p = 0$, the condition in Theorem 2.3 leads to

$$\mu > \prod_{|\operatorname{eig}_i(A)|\geq 1} |\operatorname{eig}_i(A)|, \tag{2.11}$$

which is in parallel to the results from [34]. In Nair and Evans [34], with the same channel and information pattern, they consider the control of the LTI system:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t, \quad \mathbf{y}_t = C\mathbf{x}_t, \qquad t \geq 0,\ \mathbf{x}_t \in \mathbb{R}^n,\ \mathbf{u}_t \in \mathbb{R}^m,\ \mathbf{y}_t \in \mathbb{R}^\ell,$$

with the pair (A, B) controllable and the pair (C, A) observable. A random initial condition $\mathbf{x}_0$ is the only source of uncertainty. They prove necessary and sufficient conditions for stabilization over the network:

$$\mathcal{R}_b > \sum_{\mathrm{eig}_i(A)} \max\{0, \log_2 |\,\mathrm{eig}_i(A)|\}, \tag{2.12}$$

which is obtained by taking logarithm of (2.11).

2.2.3 Erasure networks

When m approaches 0 from above, i.e., $m \downarrow 0$, we obtain from Theorem 2.3 an almost sure[3] stability criterion,

$$\begin{aligned}
\mathcal{R}_b &> \lim_{m\downarrow 0} \frac{1}{m} \log_2 \prod_{|\,\mathrm{eig}_i(A)|\geq 1} \frac{(1-p)|\,\mathrm{eig}_i(A)|^m}{1-p|\,\mathrm{eig}_i(A)|^m} \\
&= \sum_{|\,\mathrm{eig}_i(A)|\geq 1} \lim_{m\downarrow 0} \frac{1}{m} \log_2 \frac{(1-p)|\,\mathrm{eig}_i(A)|^m}{1-p|\,\mathrm{eig}_i(A)|^m} \\
&= \sum_{|\,\mathrm{eig}_i(A)|\geq 1} \left[\lim_{m\downarrow 0} \frac{1}{m} \left[\log_2(1-p) - \log_2(1-p|\,\mathrm{eig}_i(A)|^m)\right] + \log_2 |\,\mathrm{eig}_i(A)| \right] \\
&= \frac{1}{1-p} \sum_{|\,\mathrm{eig}_i(A)|\geq 1} \log_2 |\,\mathrm{eig}_i(A)|,
\end{aligned}$$

which is consistent with [56]. Tatikonda and Mitter [56] consider the erasure channel in (2.3). They provide a sufficient condition for the almost sure observ-

[3]Stability with probability one. Refer to [25] for a definition.

ability of the system (2.9) if

$$\mathcal{R}_b > \frac{1}{1-p} \sum_{\text{eig}_i(A)} \max\{0, \log_2 |\, \text{eig}_i(A)|\}. \tag{2.13}$$

Although (2.13) is equivalent to (2.12) if $p = 0$, for $p > 0$, the inequality (2.13) is not sufficient for the existence of encoder-estimator that guarantees *mean-square* observability.

2.2.4 An example for rate limitations

We consider an LTI system to illustrate data rate limitations, including the bit rate ($\mathcal{R}_b$) and the packet rate ($\mathcal{R}$).

Example 2.1 *Consider the estimation of the following LTI plant,*

$$\mathbf{x}_{t+1} = \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \mathbf{x}_t + \mathbf{w}_t, \qquad \mathbf{y}_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x}_t + \mathbf{v}_t, \tag{2.14}$$

where $\mathbf{w}_t$ and $\mathbf{v}_t$ are random variables uniformly distributed on bounded sets. The network is as in Definition 2.1.

1. *When the channel does not drop packets ($p = 0$), for the system to have observability, the bit rate must satisfy* (2.12), i.e.,

$$\mathcal{R}_b > \log 2 + \log_2 \sqrt{2} = 1.5 \ \textit{bps}.$$

2. *For all $p \in [0, 1)$, to have almost sure observability, the data sending bit rate* (i.e. *the packet size*) *must satisfy* (2.13), i.e.,

$$\mathcal{R}_b > \frac{1}{1-p}(\log 2 + \log_2 \sqrt{2}) = \frac{1.5}{1-p} = \frac{1.5}{\mathcal{R}} \ \textit{bps},$$

where the last equality is from (2.4).

3. *If $p \in [0, \frac{1}{4})$ and the packet-size is fixed, to have observability in the second moment, we obtain from Theorem 2.3 that the data sending bit rate* (i.e., *the packet size*) *must satisfy,*

$$\begin{aligned}\mathcal{R}_b &> 1.5 + \frac{1}{2}\left(\log_2 \frac{1-p}{1-4p} + \log_2 \frac{1-p}{1-2p}\right) \\ &= 1.5 + \frac{1}{2}\left(\log_2 \frac{\mathcal{R}}{4\mathcal{R}-3} + \log_2 \frac{\mathcal{R}}{2\mathcal{R}-1}\right) \ \textit{bps},\end{aligned}$$

which can be proved to be larger than $\frac{1.5}{\mathcal{R}}$ bps (*as required for almost sure observability*).

2.3 From bits to packets

In this section, we show that under certain conditions, the packet rate is the determining factor for both observability and communication channel performance.

2.3.1 Bit rate vs. packet rate

We examine the roles of the bit rate and the packet rate in the observability of a linear plant over the channel (2.3). From Theorems 2.1– 2.3, we observe a trade-off relation between the bit rate and the packet rate: a larger bit rate $\mathcal{R}_b$ tolerates larger loss probability p, as long as $p < \frac{1}{\max_i |\operatorname{eig}_i(A)|^m}$; a larger packet rate $\mathcal{R}$, *i.e.*, smaller dropout rate p, allows smaller sending bit rate $\mathcal{R}_b$.

Consider the same one-dimensional case (Theorem 2.1) as in Figure 2.2 but with the appropriate coordinate transformation. Figure 2.3 depicts the observable and unobservable regions. The following observation can be made:

Remark 2.3 *For a fixed packet rate $\mathcal{R}$, once the bit rate $\mathcal{R}_b$ is reasonably large compared to the system dynamics, increasing $\mathcal{R}_b$ does little to expand the observable region, while increasing the packet rate is much more effective.*

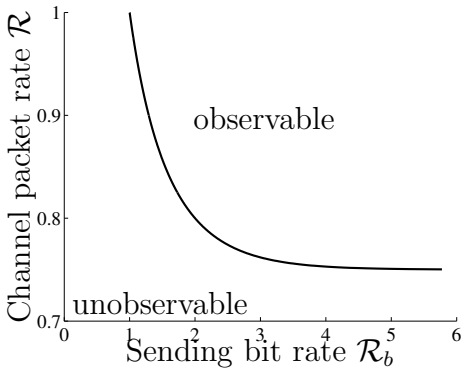


Figure 2.3. Packet rate vs. bit rate.

We may reach the conclusion above from sensitivity analysis. From the conditions in Theorem 2.1, the slope of the curve in Figure 2.3,

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\left(\frac{\mu^m - |A|^m}{\mu^m |A|^m - |A|^m}\right) = \frac{m(1-|A|^{-m})\mu^{m-1}}{(\mu^m - 1)^2},$$

approaches zero as μ grows.

2.3.2 Packet rate and communication performance

The number of packets is a major factor that determines the data traffic in a network. In most widely used communication protocols, the overhead of a packet is not reduced by decreasing the number of data-bits in the packet. The sending packet rate affects communication system performance in several ways, *e.g.*,

- Higher data traffic may induce longer communication queuing delay, which is undesirable in real-time systems.

- In shared networks, higher packet rate from each node results in higher collision probability among different nodes, and therefore more data dropouts.

- Radio is the primary source of energy consumption in non-mobile wireless sensor networks. An important criterion in assessing communication protocols is energy efficiency [9]. A smart communication scheduling method can extend the battery lifetime and therefore reduce the sensor network deployment costs.

2.4 Summary

For an estimation over the network problem, we derived conditions for observability in the m^{th} moment in terms of the packet size μ and the packet loss probability p. Roughly speaking, when the packet dropout rate is high, each packet has to be large enough to deliver sufficient information; when the packet size is small, the network must be reliable enough to transport large number of packets. But observability is impossible if either the loss probability is too large or the packet size is too small, no matter what the other factor is.

Those are important practical issues in communication protocol designs for NCSs, because data is sent in the form of *packets*, and one must allocate *bits* for each packet properly.

This chapter also serves as a transition to the chapters to come. From now on, we assume that *data is sent in the form of packets, and each packet is large enough to carry information available at the current time.*

Chapter 3

Jump-diffusion processes

We paraphrase some results in jump-diffusion processes, which will be applied in Chapter 4. Readers are referred to [39, 11, 22, 26] for general discussions.

We consider an n-dimensional jump-diffusion process $\mathbf{x}(t)$, satisfying the following jump-diffusion equation,

$$\mathrm{d}\mathbf{x}(t) = f(\mathbf{x}(t))\mathrm{d}t + \sigma(\mathbf{x}(t))\mathrm{d}\mathbf{w}(t) + \int_{\mathcal{M}} \rho(\mathbf{x}(t^-), \mathbf{z})\mathbf{N}(\mathrm{d}t, dz), \qquad (3.1)$$

which has three terms with clear physical interpretations. The first term is a *drift* term, in which $f : \mathbb{R}^n \to \mathbb{R}^n$ is a continuous function. The second is a *diffusion* term, in which $\mathbf{w}(t)$ is an standard Wiener process and $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times n}$. The Wiener process has covariance matrix $\Sigma_w := \sigma'\sigma > 0$. The third is a *jump* term, in which $\rho : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is the jump size, which depends on both the pre-jump state and an independent *mark* process $\mathbf{z}$, and $\mathbf{N}(\cdot, \mathcal{M})$ is a counting process to be defined in Section 3.1. The process $\mathbf{z}$ is distributed on a finite set $\mathcal{M}$.

Remark 3.1 *The independent marker process* $\mathbf{z}$ *gives the jump size distribution*

for $\mathbf{x}(t)$ *when the counting process* $\mathbf{N}$ *has an increment. We have*

$$\int_{\mathcal{M}} \mathbf{N}(\mathrm{d}t, \mathrm{d}z) = \mathbf{N}(\mathrm{d}t).$$

3.1 Counting process and its intensity function

We give a definition for a *counting process*, which is adapted from [1]. Refer to Appendix A for related definitions.

Definition 3.1 (Càdlàg) *A process* $\mathbf{x}(t)$ *is càdlàg if all its paths* (*or* realizations) *are right-continuous and admit left-hand limits.*

Definition 3.2 (Counting process) *Let a nonnegative function* $\lambda(t, y) : \mathbb{R}^+ \times \mathbb{R}^n \to \mathbb{R}^+$ *be continuous in* t *and finite for every* $t \in \mathbb{R}^+$ *and* $y \in \mathbb{R}^n$*. Let* $\mathbf{y}(t)$ *be a predictable process on the stochastic basis* $(\Omega, \mathcal{F}, \mathbf{F}, P)$*. We construct an* $\mathcal{F}_t$*-adapted counting process* $\mathbf{N}(t)$ *in two steps. First, construct a point process* $(\mathbf{T}_i(\omega))_{i\geq 0}$*,* $\omega \in \Omega$*, on the time axis by*

- $\mathbf{T}_0 = 0$*;*
- *Given* $(\mathbf{T}_i)_{0\leq i\leq j}$*, for* $t > \mathbf{T}_j$*,* $\mathbf{T}_{j+1}$ *is an* $\mathcal{F}_t$ *stopping time such that*

$$P\left(\mathbf{T}_{j+1} > t \mid \mathbf{y}(s) = y(s), s < t; \mathbf{T}_j = T_j\right) = \exp\left(-\int_{T_j}^{t} \lambda\left(s, y(s)\right) \mathrm{d}s\right).$$

The counting process $\mathbf{N}(t)$ *is an* $\mathcal{F}_t$*-adapted càdlàg process defined by*

$$\mathbf{N}(t) = \sum_{i=1}^{\infty} \mathbb{1}\left(\mathbf{T}_i \leq t\right), \tag{3.2}$$

where $\mathbb{1}(\cdot)$ *is the indicator function. The function* $\lambda(\cdot,\cdot)$ *is called the intensity function of* $\mathbf{N}(t)$.

Remark 3.2 *The process* $\mathbf{y}(t)$ *can be any* $\mathcal{F}_t$*-predictable process. Specifically,* $\mathbf{y}(t)$ *can be an observation of* hazardous events *that is related to the necessity of jumps. The process escapes hazardous situations by* jumping out.

A particularly interesting choice of the predictable process $\mathbf{y}(t)$ is

$$\mathbf{y}(t) = \mathbf{x}(t^-) := \lim_{s \uparrow t} \mathbf{x}(s),$$

which is well defined because $\mathbf{x}(t)$ is càdlàg. We will consider the homogeneous intensity function $\lambda(x^-)$, where x^- stands for $x(t^-)$ when t is implicit.

An important example of counting processes is a *Poisson process*, which is frequently used as a queue model.

Definition 3.3 (Poisson process) *If the function* $\lambda(\cdot,\cdot)$ *is* $\mathcal{F}_0$*-measurable,* i.e.*,* $\lambda = \lambda(t)$*, the counting process* $\mathbf{N}(t)$ *is called a general Poisson process. If* $\lambda(t)$ *is constant,* $\mathbf{N}(t)$ *is a Poisson process; otherwise it is an inhomogeneous Poisson process.*

3.2 Solutions to jump-diffusion equations

In this section, we study solutions to the jump-diffusion process (3.1), in which the jump intensity is a homogeneous function $\lambda(x(t^-))$. We impose the following assumptions:

Assumption 3.1 *For the jump-diffusion equation* (3.1)*:*

1. *There exists* $C_1 > 0$ *such that* $|f(x) - f(y)|^2 + |\sigma(x) - \sigma(y)|^2 \leq C_1|x - y|^2$, *for any* $x, y \in \mathbb{R}^n$.

2. *There exists* $C_2 > 0$ *such that* $|f(x)|^2 + |\sigma(x)|^2 \leq C_2(1 + |x|^2)$, *for any* $x \in \mathbb{R}^n$.

3. *The function* $\rho(x, z)$ *is in the form of*

$$\rho(x, z) = g(z) - x,$$

   where $g(z)$ *is a uniformly bounded continuous function for* $z \in \mathcal{M}$.

4. *The intensity function* $\lambda(\cdot)$ *is continuous, and* $\lambda(0) = 0$.

Remark 3.3 *Condition 1 is the global Lipschitz property, and Condition 2 guarantees almost linear growth. Condition 3 means that each jump resets the state* $\mathbf{x}(t)$ *to a bounded set.*

We define both *strong solutions* and *weak solutions* to (3.1). For that purpose, we write the equation in an integral form,

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t f\left(\mathbf{x}(t)\right) \mathrm{d}t + \int_0^t \sigma(\mathbf{x}(t)) \mathrm{d}\mathbf{w}(t) + \int_0^t \int_{\mathcal{M}} \rho(\mathbf{x}(t^-), z) \mathbf{N}(\mathrm{d}t, \mathrm{d}z). \tag{3.3}$$

Definition 3.4 (Strong solution) *Given a realization of both* $\mathbf{w}(t)$ *and* $\mathbf{N}(t)$, *any* $\mathbf{x}(t)$ *that satisfies* (3.3) *is called a strong solution to* (3.3).

Definition 3.5 (Weak solution) *Eq.* (3.1) *has a weak solution if for any initial probability distribution of* $\mathbf{x}(t)$*, there exists a stochastic base* $(\Omega, \mathcal{F}, \mathbf{F}, P)$ *with a filtration* $\{\mathcal{F}_t\}$*, an* $\mathcal{F}_t$*-adapted Wiener process, an* $\mathcal{F}_t$*-adapted counting process* $\mathbf{N}(t)$*, and an* $\mathcal{F}_t$*-adapted process* $\mathbf{x}(t)$ *satisfying* (3.3)*.*

Roughly speaking, a strong solution is defined path-wise, while a weak solution is in the sense of probability measures. To avoid more technicalities, the definition for weak solution, adapted from [26, pp 92], is not a strict one. The readers are referred to [22]. A strong solution is also a weak solution, but not the converse [38].

3.3 Infinitesimal generators

While it is usually difficult to solve (3.1), there are more tractable ways to analyze its stability properties. Our approach is based on *infinitesimal generators*.

Definition 3.6 (Infinitesimal generator) *[38] Given a twice continuously differentiable function* $V : \mathbb{R}^n \to \mathbb{R}^+$*, the generator* $\mathcal{L}$ *of a jump-diffusion process* $\mathbf{x}(t)$ *is defined by*

$$\mathcal{L}V(x) := \lim_{t \to s} \frac{\mathbf{E}\big[V(\mathbf{x}(t)) | \mathbf{x}(s) = x\big] - V(x)}{t - s}, \qquad \forall x \in \mathbb{R}^n,\ t > s \geq 0, \tag{3.4}$$

where $\mathbf{E}\big[V\left(\mathbf{x}(t)\right) | \mathbf{x}(s) = x\big]$ *denotes the expectation of* $V(\mathbf{x}(t))$ *given* $\mathbf{x}(s) = x$*. The set of functions* V *for which the limit exists at* x *is denoted by* $\mathcal{D}_{\mathcal{L}}(x)$*, while the set of functions for which the limit exists for all* $x \in \mathbb{R}^n$ *is denoted by* $\mathcal{D}_{\mathcal{L}}$*.*

Lemma 3.1 *If* $\mathbf{x}(t)$ *is a weak solution to* (3.1)*, then the following holds for any smooth function* $V(x) : \mathbb{R}^n \to \mathbb{R}^+$*,*

$$\mathcal{L}V(x) = \frac{\partial V(x)}{\partial x} \cdot f(x) + \frac{1}{2} \operatorname{trace}\left(\sigma(x)' \frac{\partial^2 V(x)}{\partial x^2} \sigma(x)\right) + \int_{\mathcal{M}} \left(V(x + \rho(x, z)) - V(x)\right) \lambda(x) \mathrm{d}z, \tag{3.5}$$

where $\frac{\partial V(x)}{\partial x}$ *and* $\frac{\partial^2 V(x)}{\partial x^2}$ *are the gradient vector and the Hessian matrix, respectively.*

Proof. See [19]. ■

In Chapter 4, we will have the following jump-diffusion process:

$$\mathrm{d}\mathbf{e}(t) = A\mathbf{e}(t)\mathrm{d}t + \sigma \mathrm{d}\mathbf{w}(t) + \big(\mathbf{v} - \mathbf{e}(t^-)\big)\mathrm{d}\mathbf{N}(t), \tag{3.6}$$

where $A \in \mathbb{R}^{n \times n}$, $\sigma \in \mathbb{R}^{n \times n}$, $\mathbf{v} \in \mathbb{R}^n$ is a random variable on a finite set $\mathcal{M}$ with probability density function $\mu(\cdot)$ and independent of $\mathbf{e}(t)$, and the counting process $\mathbf{N}(t)$ is defined in (3.2) with intensity function $\lambda(\mathbf{e}(t^-))$.

Theorem 3.1 *The generator for the jump-diffusion process described by* (3.6) *is given by*

$$\mathcal{L}V(e) = \frac{\partial V(e)}{\partial e} \cdot Ae + \frac{1}{2} \operatorname{trace}\left(\sigma' \frac{\partial^2 V(e)}{\partial e^2} \sigma\right) + \lambda(e)\Big(\int_{\mathcal{M}} V(v)\, \mathrm{d}\mu(v) - V(e)\Big). \tag{3.7}$$

In order to give readers a feeling of the physics of a jump-diffusion process and how to handle its different components, we provide a direct proof (without recourse to Lemma 3.1).

Proof.[Theorem 3.1] For $t > 0$, $\tau \ll 1$, and $\mathbf{e}(t) = e$,

$$\Pr\left(\text{one jump in } [t, t+\tau)\right) = \lambda(e)\tau,$$

$$\Pr(\text{more than one jump in } [t, t+\tau)) = o(\tau),$$

where $\lim_{\tau\to 0} \frac{o(\tau)}{\tau} = 0$. Therefore,

$$\begin{aligned}
&\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e] \\
&= (1 - \lambda(e)\tau)\,\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{no jump}] \\
&\quad + \lambda(e)\tau\,\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{one jump}] \\
&\quad + \mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{more than one jump}]o(\tau),
\end{aligned}$$

where $\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{more than one jumps}]$ is finite. We conclude that

$$\begin{aligned}
(\mathcal{L}V)(e) &= \lim_{\tau\to 0} \frac{(1 - \lambda(e)\tau)\,\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{no jump}]}{\tau} \\
&\qquad + \frac{\lambda(e)\tau\,\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{one jump}] - V(e)}{\tau} \\
&= \lim_{\tau\to 0} \frac{\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{no jump}] - V(e)}{\tau} \\
&\qquad + \lim_{\tau\to 0} \lambda(e)\,\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{one jump}] \\
&\qquad - \lambda(e)\,\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{no jump}] \\
&= \lim_{\tau\to 0} \frac{\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{no jump}] - V(e)}{\tau} \\
&\qquad + \lambda(e) \int_{\mathcal{M}} V(v)\,\mathrm{d}\mu(v) - \lambda(e)V(e).
\end{aligned}$$

We complete the proof by noting that,

$$\begin{aligned}
&\lim_{\tau\to 0} \frac{\mathrm{E}[V(\mathbf{e}(t+\tau))|\mathbf{e}(t) = e,\ \text{no jump}] - V(e)}{\tau} \\
&= \frac{\partial V(e)}{\partial e} \cdot Ae + \frac{1}{2}\,\mathrm{trace}\left[\sigma' \frac{\partial^2 V(e)}{\partial e^2}\sigma\right]
\end{aligned}$$

which is the generator of the diffusion process without jumps [38]. ■

The analysis approach is a Lyapunov method in stochastic settings. Letting $x = \mathbf{e}(t)$ in (3.4) and taking expectation, one obtains

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}[V(\mathbf{e}(t))] = \mathrm{E}[(\mathcal{L}V)(\mathbf{e}(t))], \tag{3.8}$$

from which the stochastic stability properties of the process $\mathbf{e}(t)$ will be deduced by appropriate choice of the Lyapunov function V.

Chapter 4

Communication schedulers

A communication scheduler is a mechanism that determines whether or not a node sends data to the network, given information available. The mechanism can be either deterministic or stochastic. For a deterministic communication scheduler, the decision to send data is a deterministic function of the given information set. For a stochastic communication scheduler, the decision is based on outcomes of a certain random process.

Our main interest is in stochastic communication schedulers. In particular, we use counting processes such as the Poisson process to model communication schedulers. We use *jump-diffusion processes* to model NCSs embedded with stochastic communication schedulers and apply a stochastic analysis method that resembles Lyapunov's methods. The method enables us to analyze a large class of stochastic communication schedulers and various types of networks (*e.g.*, networks with delay and networks that drop data packets). The resulting jump-diffusion processes are of theoretical interest on their own in that the jump in-

tensities are coupled with system dynamics.

We explore communication schedulers via *estimation over networks* problems. The chapter is organized as follows. In Section 4.1, we give basic definitions and motivate stochastic communication schedulers. Sections 4.2 and 4.3 contain the main results on stochastic communications for full state measurement plants and partially observed plants, respectively. Section 4.4 briefly discusses deterministic communication schedulers. Section 4.5 gives some simulation results.

4.1 Definitions and Motivation

In this section we give some basic definitions and discuss the merits of stochastic communication schedulers.

4.1.1 Definitions

We mainly consider a lossless network with a fixed delay τ. Unless otherwise stated, we simply call it a *τ-delayed network*, or a *delay-free network* if $\tau = 0$. The definitions for stability and observability are similar to the ones in Chapter 2.

Definition 4.1 (Stability in the m^{th} moment) *A random process $\mathbf{x}(t)$ has initial distribution $\mathbf{x}(0) = \mathbf{x}_0$ such that $\mathrm{E}[|\mathbf{x}_0|^m] < \delta$, for some $\delta > 0$. The process $\mathbf{x}(t)$ is stable in the m^{th} moment, if for any $T > 0$, there exists $\epsilon > 0$, such that for any $t > T$, we have*

$$\mathrm{E}[|\mathbf{x}(t)|^m] < \epsilon, \tag{4.1}$$

where $|\cdot|$ *stands for the Euclidean norm. It is unstable in the* m^{th} *moment, if for any* $\epsilon > 0$*, there exists* $t > 0$ *such that*

$$\mathrm{E}[|\mathbf{x}(t)|^m] > \epsilon.$$

Definition 4.2 (Observability in the m^{th} moment) *A system with the state* $\mathbf{x}(t)$ *is observable in the* m^{th} *moment over a network if there exists an encoder-estimator pair such that* $\mathbf{x}(t) - \hat{\mathbf{x}}(t)$ *is stable in the* m^{th} *moment, where* $\hat{\mathbf{x}}(t)$ *is the state of the estimator.*

The long-term average is a natural criterion in communication network applications. We define a measure of communication cost as follows:

Definition 4.3 (Packet rate) *The* packet rate *of a node is the asymptotic rate of messages that are sent by the node,* i.e.*,*

$$\mathcal{R} := \lim_{k\to\infty} \mathbf{E}\Big[\frac{k}{\mathbf{t}_k}\Big],$$

where $\mathbf{t}_k$ *is the time when the node sends the* k^{th} *packet.*

This definition is in contrast to the logarithm-based concepts such as *entropy rate* or *channel capacity* [10]. The definition is appealing in the study of packet-based communication networks. The channel capacity (the ability to deliver data packets reliably in unit time) can be defined in parallel to the foregoing. By finite network utilization, we mean that a node sends data with a finite packet rate $\mathcal{R}$.

4.1.2 Why stochastic communication schedulers?

Communication scheduling methods can be either *time-driven* or *event-driven*. The physics of time-driven and event-driven models has been studied extensively, most recently by Åström *et al.* [53]. Periodic communication is a simple and widely-used scheduling method. Periodic communication is generally not optimal to reduce network utilization because a node does not always need to send data, especially when the receiver has good knowledge of it. The communication schedulers we propose are event-driven.

In *stochastic communication schedulers*, the *probability* of a node sending a message is a function of the current estimation error. In the core of the stochastic communication schedulers lie *counting processes*, which were defined in Chapter 3. In essence, a node's communication decision is triggered by a counting process whose "incremental rate" depends on the system dynamics. Stochastic communication schedulers have both practical and theoretical advantages, of which we list a few:

- Stochastic communication schedulers can be incorporated in existing communication protocols and therefore improve medium access algorithms. For example, in the p-consistent carrier sense multiple access (CSMA) protocols, the parameter p indicates the probability of packet sending when a node detects a collision. An algorithm can be designed so that each node may dynamically choose its own parameter p.

- Stochastic communication schedulers are manageable in analysis. With stochastic communication schedulers, the resulting dynamics evolves according to stochastic differential equations with resets triggered by counting

processes. Analysis tools are available from the statistics literature, *e.g.*, [3, 22].

- Stochastic schedulers consist of a rich class of communication decision making mechanisms. Deterministic communication schedulers, discussed in Section 4.4, can be viewed as their limiting cases.

- The stability results for stochastic communication schedulers are applicable to periodic communications and to networks that drop data packets (see Section 4.3.6).

4.2 Remote estimation with local state measurement

In this section, we consider the remote estimation with full local state measurement over a network with fixed time delay τ (see Figure 4.1). We analyze both delay-free and τ-fixed delay systems, and we consider various stochastic communication schedulers.

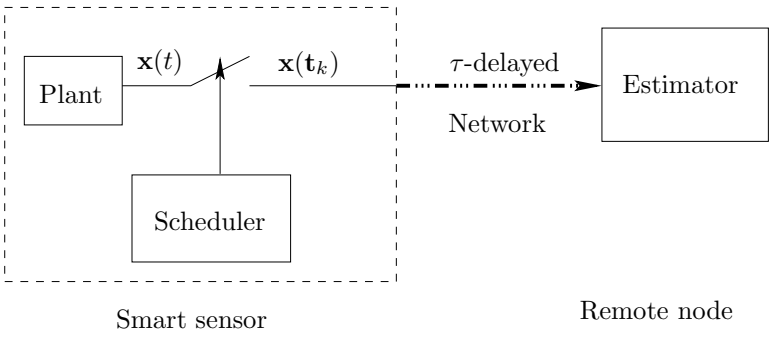


Figure 4.1. Remote estimation with local state measurement.

4.2.1 A remote estimation scheme

The continuous-time LTI system under consideration is

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{w}(t), \tag{4.2}$$

where the matrix A has at least one pole in the right half plane, and the zero-mean Gaussian white noise $\mathbf{w}(t)$ has covariance matrix $\sigma\sigma' \in \mathbb{R}^{n\times n}$. The state $\mathbf{x}(t)$ is measured locally, and it is then sent to a remote node, where the minimum variance estimator is constructed as

$$\dot{\hat{\mathbf{x}}}(t) = A\hat{\mathbf{x}}(t). \tag{4.3}$$

This remote estimation problem arises in many practical scenarios. The networked control problem considered in [60], in which spatially distributed nodes coordinate through message broadcasting, boils down to the estimation of (4.2).

The local plant needs to send its own state to the remote site. It is the *communication scheduler*'s responsibility to determine when data transmission should take place. To this effect, it builds an *internal estimate* of its own state. The difference between this internal state estimate and the measured state provides a criterion to judge the quality of the remote estimate.

Denote by $\{\mathbf{t}_k\}_{k\in\mathbb{N}}$ the times when the plant sends its measured state $\mathbf{x}(\mathbf{t}_k)$ to the remote estimator. Due to factors such as noise and quantization errors during transmission, the data received is not exactly what was sent, *i.e.*,

$$\mathbf{y}(\mathbf{t}_k) = \mathbf{x}(\mathbf{t}_k) + \mathbf{v}(k), \tag{4.4}$$

where the process $\mathbf{v}(k)$ is assumed to be stationary with known probability distribution $\mu(\cdot)$ on a finite support $\mathcal{M}$. In addition, $\mathbf{v}(k)$ is zero-mean and independent of $\mathbf{w}$. From now on, $\mathbf{v}(k)$ is abbreviated as $\mathbf{v}$ due to stationarity.

The remote estimation at time t is based on all information received up to time t, which is τ-delayed, *i.e.*,

$$\big\{\mathbf{y}(\mathbf{t}_j) : \mathbf{t}_j \leq t - \tau\big\}. \tag{4.5}$$

The corresponding minimum-variance estimate is given by a Kalman filter for (4.2) with discrete measurements [30]. For simplicity we shall assume that the measurement noise $\mathbf{v}(k)$ is negligible, in which case the filter takes a particularly simple form because one does not need to propagate the covariance matrix and the optimal estimator is given by the following open-loop "computational model,"

$$\begin{aligned} \dot{\hat{\mathbf{x}}}(t) &= A\hat{\mathbf{x}}(t), \\ \hat{\mathbf{x}}(t) &= \mathbf{z}(t), \quad \text{when } t = \mathbf{t}_k + \tau, \end{aligned} \tag{4.6}$$

where

$$\mathbf{z}(\mathbf{t}_k + \tau) := \exp(A\tau)\mathbf{y}(\mathbf{t}_k). \tag{4.7}$$

4.2.2 Error dynamics

Denote by $\mathbf{N}(t)$ a right-continuous function that is constant almost everywhere except at the times $\mathbf{t}_k$ ($k \in \mathbb{N}$) when it is increased by 1. An increment in $\mathbf{N}(t)$ signifies that the local plant sends a data packet, and vice versa. We rewrite the estimator equation (4.6) together with its jumps (4.7) in a compact form:

$$\mathrm{d}\hat{\mathbf{x}}(t) = A\hat{\mathbf{x}}(t)\mathrm{d}t + \big(\mathbf{z}(k) - \hat{\mathbf{x}}(t^-)\big)\mathrm{d}\mathbf{N}(t - \tau). \tag{4.8}$$

Remark 4.1 *The process $\hat{\mathbf{x}}(t)$ is also right-continuous. At time $t = \mathbf{t}_k + \tau$ ($k \in \mathbb{N}$) $\mathbf{N}(t - \tau)$ has an increment of 1, and therefore $\hat{\mathbf{x}}(t)$ has a jump of size*

$\mathbf{z}(k) - \hat{\mathbf{x}}(t^-)$, *which means*

$$\hat{\mathbf{x}}(\mathbf{t}_k + \tau) = \lim_{t \uparrow \mathbf{t}_k + \tau} \hat{\mathbf{x}}(t) + \lim_{t \uparrow \mathbf{t}_k + \tau} (\mathbf{z}(k) - \hat{\mathbf{x}}(t)) = \mathbf{z}(k).$$

From (4.2) and (4.8), we can verify that the estimation error,

$$\hat{\mathbf{e}} := \mathbf{x} - \hat{\mathbf{x}},$$

satisfies the following jump-diffusion process (see Chapter 3):

$$\mathrm{d}\hat{\mathbf{e}}(t) = A\hat{\mathbf{e}}(t)\mathrm{d}t + \mathrm{d}\mathbf{w}(t) + \big(\boldsymbol{\eta}(t) - \hat{\mathbf{e}}(t^-)\big)\mathrm{d}\mathbf{N}(t - \tau), \tag{4.9}$$

where

$$\boldsymbol{\eta}(t) = \exp\{A\tau\}\mathbf{v} + \int_{t-\tau}^{t} \exp\{A(t-r)\}\mathrm{d}\mathbf{w}(r), \tag{4.10}$$

in which $\mathbf{v}$ is the error term as in (4.4), and integral is defined in the Itô sense [41, 26]. It is straightforward to show that the stochastic moments of $\boldsymbol{\eta}(t)$ are finite for any delay of τ.

An optimal communication scheduling problem is solved in [61] for discrete-time systems. It was shown that the optimal communication decision is a function of the estimation error associated with an additional estimate of its local state $\mathbf{x}(t)$ that should be updated in a delay-free fashion right after data is transmitted, *i.e.*, without waiting τ time units in the discrete update (4.7), even though the network has delays. This *delay-free estimate* $\hat{\mathbf{x}}^0$ is constructed inside the local node as follows

$$\mathrm{d}\hat{\mathbf{x}}^0(t) = A\hat{\mathbf{x}}^0(t)\mathrm{d}t + \big(\mathbf{y}(t) - \hat{\mathbf{x}}^0(t^-)\big)\mathrm{d}\mathbf{N}(t), \tag{4.11}$$

in which $\mathbf{N}(t)$ is determined by the local node's communication scheduling decision. Inspired by the results in [61], we consider communication schedulers that base their decisions on the delay-free estimation error,

$$\mathbf{e} := \mathbf{x} - \hat{\mathbf{x}}^0,$$

whose dynamics is given by

$$\mathrm{d}\mathbf{e}(t) = A\mathbf{e}(t)\mathrm{d}t + \mathrm{d}\mathbf{w}(t) + \big(\mathbf{v} - \mathbf{e}(t^-)\big)\mathrm{d}\mathbf{N}(t). \tag{4.12}$$

For delay-free networks, Eq. (4.9) becomes (4.12). With a stochastic communication scheduler, the plant sends data at an average rate that depends on the current value of the internal estimation error $\mathbf{e}(t^-)$, as in (4.12). To this effect, we define $\mathbf{N}(t)$ to be a counting process with the jump intensity function

$$\lambda(\mathbf{e}^-), \tag{4.13}$$

where $\mathbf{e}^-$ is a shorthand for $\mathbf{e}(t^-)$. We will use this convention whenever the context allows in the following sections. We start by considering the delay-free case (4.12).

4.2.3 Uncontrolled communications

Take a Poisson process $\mathbf{N}(t)$ with constant rate γ as the counting process in (4.12). The corresponding packet rate is $\mathcal{R} = \gamma$. With this communication scheduling method, data packets are sent randomly in an open-loop fashion. Therefore, we call it *uncontrolled communication*. This model can also be used to model a network that drops data packets, which will be discussed in Section 4.3.

The following theorem is the main result of this section.

Theorem 4.1 *Let* $\mathbf{e}(t)$ *be the jump-diffusion process defined by* (4.12)*, with* $\lambda(\mathbf{e}^-) = \gamma$*. Then we have the following statements:*

1. *If* $\gamma > \max\{\Re[\mathrm{eig}(A)]\}$*, then* $\mathrm{E}[\mathbf{e}(t)]$ *converges to zero exponentially fast.*

2. *If* $\gamma > m \max\{\Re[\mathrm{eig}(A)]\}$*, then* $\mathbf{e}(t)$ *is stable in the* m^{th} *moment, for each* $m \geq 2$*.*

3. *If* $\gamma < m \max\{\Re[\mathrm{eig}(A)]\}$*, then* $\mathbf{e}(t)$ *is unstable in the* m^{th} *moment, for each* $m \geq 2$*.*

4. *If* $\gamma > 2 \max\{\Re[\mathrm{eig}(A)]\}$*, and there exist* $P, Q \in \mathbb{R}^{n\times n}$ *positive definite and* $c > 0$ *such that*

$$P\Big(A - \frac{\gamma}{2}I\Big) + \Big(A - \frac{\gamma}{2}I\Big)' P \leq -Q, \qquad\qquad Q \geq cP,$$

then $\mathbf{e}(t)$ *is stable in the second moment and*

$$\lim_{t\to\infty} \mathrm{E}[\mathbf{e}'(t) P \mathbf{e}(t)] \leq \frac{\gamma\rho^2 + \theta}{c},$$

where $\rho^2 := \int_{\mathcal{M}} v' P v \, \mathrm{d}\mu(v)$*, and* $\theta := \mathrm{trace}(\sigma' P \sigma)$*.*

Remark 4.2 *We observe that* $m \max\{\Re[\mathrm{eig}(A)]\}$ *is a* tight bound *for the* m^{th} *moment stability,* i.e.*, for any* $\epsilon > 0$*, there exists* $\gamma > m \max\{\Re[\mathrm{eig}(A)]\} - \epsilon$ *such that* $\mathbf{e}(t)$ *is unstable. Note that the bound only depends on the least stable mode of matrix* A*.*

Remark 4.3 *Whatever the Poisson rate* γ *is, the estimation error is unstable in the* m^{th} *moment for sufficiently large* m*. Then what should* m *be in practice?*

There is no clear answer to this question. For any $m > 0$, moment stability implies almost sure *stability. The latter means stability with probability one, which is practically interesting in many applications. Stability in the moment of higher order means the process's probability distribution is less "heavy-tailed",* i.e., *the density function decays faster. Means square stability ($m = 2$) is frequently used in applications.*

We need the following lemma to prove this theorem. The applications of this lemma will recur in what follows.

Lemma 4.1 *Given a scalar random variable* $\mathbf{x}$ *that is nonnegative with probability one, a positive constant* δ*, and* $k > \ell > 0$*, we have*

$$\mathrm{E}[\mathbf{x}^k] \geq \delta^\ell \, \mathrm{E}[\mathbf{x}^{k-\ell}] - \delta^k.$$

Proof.[Lemma 4.1] Suppose $\mathbf{x}$ has probability distribution $\mu(\mathbf{x})$. For every $\delta > 0$, the following inequalities hold

$$\begin{aligned}
\mathrm{E}[\mathbf{x}^k] &\geq \int_{x \geq \delta} x^k \, \mathrm{d}\mu(x) \geq \delta^\ell \int_{x \geq \delta} x^{k-\ell} \, \mathrm{d}\mu(x) \\
&= \delta^\ell \Big(\int_{x \geq 0} x^{k-\ell} \, \mathrm{d}\mu(x) - \int_{x < \delta} x^{k-\ell} \, \mathrm{d}\mu(x) \Big) \\
&\geq \delta^\ell (\mathrm{E}[\mathbf{x}^{k-\ell}] - \delta^{k-\ell}).
\end{aligned}$$

■

Proof.[Theorem 4.1] To prove statement 1, take[1] $V(e) = e$. Because $\int_{\mathcal{M}} V(v) \mathrm{d}\mu(v) = 0$, Eq. (3.7) becomes

$$\mathcal{L}V(e) = Ae - \gamma e.$$

[1]Here V is not a positive function as required in the definition of a *generator*. But the derivation does not change for this smooth function.

Therefore, Eq. (3.8) takes the form of:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}[\mathbf{e}(t)] = (A - \gamma I)\,\mathrm{E}[\mathbf{e}(t)],$$

which leads to statement 1 because $A - \gamma I$ is Hurwitz, *i.e.*, all the eigenvalues are in the left half plane.

To prove statement 2, let $P, Q > 0$ and $c > 0$ such that

$$P\Big(A - \frac{\gamma}{m}I\Big) + \Big(A - \frac{\gamma}{m}I\Big)' P \leq -Q, \qquad\qquad Q \geq cP.$$

Such matrices exist as long as $\gamma > m \max\{\Re[eig(A)]\}$. For $m \geq 2$, define

$$V(e) := (e'Pe)^{\frac{m}{2}}. \tag{4.14}$$

We infer from (3.7) that

$$\begin{aligned}
\mathcal{L}V(e) &= \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\, e'(PA + A'P)e + \lambda(e)\rho^m - \lambda(e)V(e) \\
&\quad + m(\frac{m}{2} - 1)(e'Pe)^{\frac{m}{2}-2}\, e'P\sigma\sigma'Pe + \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\theta \\
&= \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\, e'[P(A - \frac{\gamma}{m}I) + (A - \frac{\gamma}{m}I)'P]e + \gamma\,\rho^m \\
&\quad + m(\frac{m}{2} - 1)(e'Pe)^{\frac{m}{2}-2}\, e'P\sigma\sigma'Pe + \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\theta \\
&\leq -\frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\, e'Qe + \gamma\,\rho^m \\
&\quad + m(\frac{m}{2} - 1)(e'Pe)^{\frac{m}{2}-2}\, e'P\sigma\sigma'Pe + \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\theta \\
&\leq -\frac{cm}{2}V(e) + \gamma\,\rho^m + \frac{m}{2}\big(2c_2(\frac{m}{2} - 1) + \theta\big)(e'Pe)^{\frac{m}{2}-1},
\end{aligned}$$

where $\rho^m := \int (v'Pv)^{\frac{m}{2}}\mathrm{d}\mu(v)$ and $c_2 > 0$ is such that $P\sigma\sigma'P \leq c_2 P$. From this and (3.8), we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}[V(\mathbf{e})] \leq -c\frac{m}{2}\,\mathrm{E}[V(\mathbf{e})] + \gamma\,\rho^m + \big(c_2 m(\frac{m}{2} - 1) + \frac{m\theta}{2}\big)\,\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}-1}].$$

Given some $\delta > 0$, we appeal to Lemma 4.1 to deduce that

$$\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}-1}] \leq \frac{1}{\delta}\mathrm{E}[V(\mathbf{e})] + \delta^{\frac{m}{2}-1},$$

and therefore,

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{E}[V(\mathbf{e})] \\ &\leq -\frac{cm}{2}\mathrm{E}[V(\mathbf{e})] + \gamma\rho^m + \big(c_2 m(\frac{m}{2}-1) + \frac{m\theta}{2}\big)\Big(\frac{1}{\delta}\mathrm{E}[V(\mathbf{e})] + \delta^{\frac{m}{2}-1}\Big) \\ &= -\frac{m}{2}\Big(c - \frac{c_2 m - 2c_2 + \theta}{\delta}\Big)\mathrm{E}[V(\mathbf{e})] + \gamma\rho^m + \frac{m}{2}\delta^{\frac{m}{2}-1}\big(2c_2(\frac{m}{2}-1) + \theta\big). \end{aligned} \tag{4.15}$$

For sufficiently large δ, $c - \frac{c_2 m - 2c_2 + \theta}{\delta} > 0$ and $\mathrm{E}[V(\mathbf{e})]$ is uniformly bounded.

To prove statement 3, we decompose the error dynamics (4.12) using the Jordan form,

$$TAT^{-1} = J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{bmatrix},$$

where J_i is a Jordan block, $1 \leq i \leq r$, and $T \in \mathbb{R}^{n\times n}$ is invertible. Denote $\bar{\mathbf{e}} = T\mathbf{e}$. From (4.12), we have

$$\mathrm{d}\bar{\mathbf{e}}(t) = J\bar{\mathbf{e}}(t)\mathrm{d}t + T\mathrm{d}\mathbf{w}(t) + (T\mathbf{v} - \bar{\mathbf{e}}(t))\mathrm{d}\mathbf{N}(t). \tag{4.16}$$

Partition $\bar{\mathbf{e}}$ according to the structure of J, *i.e.*, $\bar{\mathbf{e}} = \begin{bmatrix} \bar{\mathbf{e}}_1 \\ \vdots \\ \bar{\mathbf{e}}_r \end{bmatrix}$. For $1 \leq i \leq r$,

$$\mathrm{d}\bar{\mathbf{e}}_i(t) = J_i\bar{\mathbf{e}}_i(t)\mathrm{d}t + T_{i-}\mathrm{d}\mathbf{w}(t) + (T_{i-}\mathbf{v} - \bar{\mathbf{e}}_i(t))\mathrm{d}\mathbf{N}(t), \tag{4.17}$$

where T_{i-} is the i^{th} row block of T as in the matrix J. Because T is invertible, $T_{i-}\mathrm{d}\mathbf{w}$ is another non-zero Gaussian white noise with covariance matrix $T_{i-}\sigma\sigma' T_{i-}'$. Hence (4.16) is bounded in the m^{th} moment if and only if (4.17) is, for any $1 \leq i \leq r$. We can complete the proof in the same way as that of statement 2.

To prove statement 4, we rewrite (4.15) for $m = 2$ and obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}[V(\mathbf{e})] \leq -\Big(c - \frac{\theta}{\delta}\Big)\,\mathrm{E}[V(\mathbf{e})] + \gamma\rho^2 + \theta.$$

Applying the Comparison Lemma [24], we infer that,

$$\lim_{t\to\infty} \mathrm{E}[V(\mathbf{e}(t))] \leq \frac{\gamma\rho^2 + \theta}{c - \frac{\theta}{\delta}},$$

which completes the proof by making $\delta \to \infty$. ■

4.2.4 Controlled communications

We now consider intensity functions that depend on the current estimation error in a closed-loop fashion. We call such communication scheduling methods *controlled communications* [62]. The rationale is that a "larger" estimation error should more rapidly lead to a message exchange. We consider the intensity function of the form:

$$\lambda(\mathbf{e}^-) = (\mathbf{e}^{-\prime} P \mathbf{e}^-)^k \tag{4.18}$$

where P is some positive definite matrix and k a positive integer.

Theorem 4.2 *Let* $\mathbf{e}(t)$ *be the jump-diffusion process defined by* (4.12) *with jump intensity* (4.18)*. For every* $k > 0$*, the estimation error* $\mathbf{e}(t)$ *is stable in the* m^{th} *moment,* $m \geq 2$ *and the packet rate* $\mathcal{R}$ *is finite.*

Proof.[Theorem 4.2] Choose c_1 sufficiently large such that $A - \frac{c_1}{2}I$ is asymptotically stable. Then there exists a matrix $P > 0$ such that

$$P\Big(A - \frac{c_1}{2}I\Big) + \big(A - \frac{c_1}{2}I\big)P < 0,$$

i.e., $PA + A'P < c_1 P$. Choose $c_2 > 0$ so that $P\sigma\sigma' P \le c_2 P$.

We start by proving that the m^{th} moment of $\mathbf{e}(t)$ is bounded for $m > 2k$. Let V be as in (4.14), $\rho^m := \int_{\mathcal{M}} (v'Pv)^{\frac{m}{2}} d\mu(v)$, and $\theta := \operatorname{trace}(\sigma' P \sigma)$. From (3.7), we obtain

$$\begin{aligned}
\mathcal{L}V(e) &= \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\, e'(PA + A'P)e + \lambda(e)\rho^m - \lambda(e)V(e) \\
&\quad + \frac{m}{2}(\frac{m}{2}-1)(e'Pe)^{\frac{m}{2}-2}\, e'P\sigma\sigma' Pe + \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\theta \\
&= \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\, e'(PA + A'P)e + \rho^m (e'Pe)^k - (e'Pe)^{\frac{m}{2}+k} \\
&\quad + m(\frac{m}{2}-1)(e'Pe)^{\frac{m}{2}-2}\, e'P\sigma\sigma' Pe + \frac{m}{2}(e'Pe)^{\frac{m}{2}-1}\theta \\
&\le c_1 \frac{m}{2}(e'Pe)^{\frac{m}{2}} + \rho^m (e'Pe)^k - (e'Pe)^{\frac{m}{2}+k} \\
&\quad + \big(c_2 m(\frac{m}{2}-1) + \frac{m\theta}{2}\big)(e'Pe)^{\frac{m}{2}-1},
\end{aligned}$$

which together with (3.8) yields

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}[V(\mathbf{e})] &\le c_1 \frac{m}{2}\,\mathrm{E}[V(\mathbf{e})] + \rho^m\,\mathrm{E}[(\mathbf{e}'P\mathbf{e})^k] - \mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}+k}] \\
&\quad + \big(c_2 m(\frac{m}{2}-1) + \frac{m\theta}{2}\big)\,\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}-1}].
\end{aligned} \tag{4.19}$$

For any $\delta_1, \delta_2, \delta_3 > 0$, Lemma 4.1 yields

$$\begin{aligned}
\mathrm{E}[(\mathbf{e}'P\mathbf{e})^k] &\le \frac{\mathrm{E}[V(\mathbf{e})]}{\delta_1^{\frac{m}{2}-k}} + \delta_1^k, \\
\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}+k}] &\ge \delta_2^k\,\mathrm{E}[V(\mathbf{e})] - \delta_2^{\frac{m}{2}+k}, \\
\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}-1}] &\le \frac{1}{\delta_3}\,\mathrm{E}[V(\mathbf{e})] + \delta_3^{\frac{m}{2}-1}.
\end{aligned}$$

Consequently, Eq. (4.19) becomes

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}[V(\mathbf{e})] &\le c_1\frac{m}{2}\,\mathrm{E}[V(\mathbf{e})] + \rho^m\Big(\frac{\mathrm{E}[V(\mathbf{e})]}{\delta_1^{\frac{m}{2}-k}} + \delta_1^k\Big) - (\delta_2^k\,\mathrm{E}[V(\mathbf{e})] - \delta_2^{\frac{m}{2}+k}) \\ &\quad + \frac{m}{2}\big(2c_2(\frac{m}{2}-1)+\theta\big)\Big(\frac{\mathrm{E}[V(\mathbf{e})]}{\delta_3} + \delta_3^{\frac{m}{2}-1}\Big) \\ &\le \big(\frac{c_1 m}{2} + \frac{\rho^m}{\delta_1^{\frac{m}{2}-k}} - \delta_2^k + m\frac{c_2(\frac{m}{2}-1)+\theta/2}{\delta_3}\big)\,\mathrm{E}[V(\mathbf{e})] \\ &\quad + \rho^m\delta_1^k + \delta_2^{\frac{m}{2}+k} + m\big(c_2(\frac{m}{2}-1)+\frac{\theta}{2}\big)\delta_3^{\frac{m}{2}-1}.\end{aligned}$$

For sufficiently large δ_2,

$$c_1\frac{m}{2} + \frac{\rho^m}{\delta_1^{\frac{m}{2}-k}} - \delta_2^k + \frac{m\big(c_2(\frac{m}{2}-1)+\frac{\theta}{2}\big)}{\delta_3} < 0,$$

and the boundedness of $\mathrm{E}[V(\mathbf{e})]$ follows. Consequently $\mathbf{e}(t)$ is stable in the m^{th} moment.

To prove the stability of $\mathbf{e}(t)$ in the m^{th} moment for $m \le 2k$, we use Lemma 4.1 to bound

$$\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}}] \le \frac{\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{k+\frac{1}{2}}]}{\delta_4^{k+\frac{1}{2}-\frac{m}{2}}} + \delta_4^k, \tag{4.20}$$

where $\delta_4 > 0$. From the last case we know that $\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{k+\frac{1}{2}}]$ is bounded (because $2(k+\frac{1}{2}) > 2k$), we conclude that $\mathrm{E}[(\mathbf{e}'P\mathbf{e})^{\frac{m}{2}}]$ is also bounded for $m \le 2k$. From (4.18), since $\mathrm{E}[(\mathbf{e}'P\mathbf{e})^k]$ is finite, so is the packet rate $\mathcal{R}$. ■

4.2.5 τ-delayed networks

We now extend the stochastic stability results to networks with τ time units delay. For constant intensity functions, since the estimation error process in (4.9) is driven by a constant intensity Poisson process $\mathbf{N}(t-\tau)$, Theorem 4.1 holds and we have the following corollary.

Corollary 4.1 *For any $\tau > 0$, let $\mathbf{e}(t)$ and $\hat{\mathbf{e}}(t)$ be the jump-diffusion processes defined by* (4.12) *and* (4.9)*, respectively, in which the counting process $\mathbf{N}(t)$ is a Poisson process with rate γ. Then we have the following statements:*

1. *If $\gamma > m \max\{\Re[\mathrm{eig}(A)]\}$, then both $\mathbf{e}(t)$ and $\hat{\mathbf{e}}(t)$ are stable in the m^{th} moment, for each $m \geq 2$.*

2. *If $\gamma < m \max\{\Re[\mathrm{eig}(A)]\}$, then neither $\mathbf{e}(t)$ nor $\hat{\mathbf{e}}(t)$ is stable in the m^{th} moment, for each $m \geq 2$.*

For state-dependent intensity functions like (4.18), a result analogous to Theorem 4.2 is stated as follows.

Theorem 4.3 *For any $\tau > 0$, let $\mathbf{e}(t)$ and $\hat{\mathbf{e}}(t)$ be the jump-diffusion processes defined by* (4.12) *and* (4.9)*, respectively, in which the counting process $\mathbf{N}(t)$ has jump intensity function* (4.18)*. Then both $\mathbf{e}(t)$ and $\hat{\mathbf{e}}(t)$ are stable in the m^{th} moment, $m \geq 2$, and the packet rate $\mathcal{R}$ is finite.*

Proof.[Theorem 4.3] Since the process $\mathbf{e}(t)$ is not affected by delay, from Theorem 4.2, $\mathbf{e}(t)$ is stable in all moments and the packet rate $\mathcal{R}$ is finite. It remains to prove the stability of the estimation error $\hat{\mathbf{e}}(t)$. We consider the function $V : \mathbb{R}^n \to \mathbb{R}$ defined in (4.14) and show that for an arbitrary time t, $\mathrm{E}\left[V\big(\hat{\mathbf{e}}(t+\tau)\big)\right]$ is bounded.

For a given time $t \geq 0$, let $\mathbf{s}(t)$ be the time at which the communication scheduler sends the last measurement before or at t, *i.e.*,

$$\mathbf{s}(t) := \max\big\{r \leq t : d\mathbf{N}(r) > 0\big\}.$$

The random variable $\mathbf{s}(t)$ is a *stopping time* (see Appendix A), which is independent of any event after time t. Since no data is sent during the time interval $(\mathbf{s}(t), t]$, the processes $\mathbf{N}(t)$ and $\mathbf{e}(t)$ in (4.12) have no jumps on this interval. Consequently, because of the network delay of τ, the remote node does not receive any data on $(\mathbf{s}(t)+\tau, t+\tau]$ and during this interval the process $\hat{\mathbf{e}}$ in (4.9) has no jumps.

At time $\mathbf{s}(t)+\tau$, $\hat{\mathbf{e}}$ is reset to $\boldsymbol{\eta}(\mathbf{s}(t)+\tau)$ defined by (4.10), and therefore

$$\hat{\mathbf{e}}(\mathbf{s}(t)+\tau) = \exp\{A\tau\}\mathbf{v} + \int_{\mathbf{s}(t)}^{\mathbf{s}(t)+\tau} \exp\{A(\mathbf{s}(t)+\tau-r)\}\mathrm{d}\mathbf{w}(r). \tag{4.21}$$

On the other hand, since the process $\mathbf{e}$ is reset to $\mathbf{v}$ at time $\mathbf{s}(t)$ and this process has no jumps on $(\mathbf{s}(t), t]$, we conclude from (4.12) that,

$$\mathbf{e}(t) = \exp\{A(t-\mathbf{s}(t))\}\mathbf{v} + \int_{\mathbf{s}(t)}^{t} \exp\{A(t-r)\}\mathrm{d}\mathbf{w}(r),$$

which is equivalent to

$$\mathbf{v} = \exp\{A(-t+\mathbf{s}(t))\}\mathbf{e}(t) - \int_{\mathbf{s}(t)}^{t} \exp\{A(\mathbf{s}(t)-r)\}\mathrm{d}\mathbf{w}(r).$$

We then plug $\mathbf{v}$ in (4.21) to get

$$\hat{\mathbf{e}}(\mathbf{s}(t)+\tau) = \exp\{A(\tau+\mathbf{s}(t)-t)\}\mathbf{e}(t) + \int_{t}^{\mathbf{s}(t)+\tau} \exp\{A(\mathbf{s}(t)+\tau-r)\}\mathrm{d}\mathbf{w}(r).$$

Moreover, since the process $\hat{\mathbf{e}}$ has no jumps on $(\mathbf{s}(t)+\tau, t+\tau]$, we conclude from (4.9) that

$$\hat{\mathbf{e}}(t+\tau) = \exp\{A(t-\mathbf{s}(t))\}\hat{\mathbf{e}}(\mathbf{s}(t)+\tau) + \int_{\mathbf{s}(t)+\tau}^{t+\tau} \exp\{A(t+\tau-r)\}\mathrm{d}\mathbf{w}(r).$$

We combine the two equations above to obtain

$$\hat{\mathbf{e}}(t+\tau) = \exp\{A\tau\}\mathbf{e}(t) + \int_{t}^{t+\tau} \exp\{A(t+\tau-r)\}\mathrm{d}\mathbf{w}(r).$$

Since $V(a+b) \leq 2^{2m}V(a) + 2^{2m}V(b)$, for any $a, b \in \mathbb{R}^n$, we conclude that

$$V(\hat{\mathbf{e}}(t+\tau)) \leq 2^{2m}V\big(\exp\{A\tau\}\mathbf{e}(t)\big) + 2^{2m}V\Big(\int_t^{t+\tau} \exp\{A(t+\tau-r)\}\mathrm{d}\mathbf{w}(r)\Big).$$

Since the integral $\int_t^{t+\tau} \exp\{A(t+\tau-r)\}\mathrm{d}\mathbf{w}(r)$ has Gaussian distribution, independent of $\mathbf{s}(t)$, we conclude that there exist finite constants c_5 and c_6 such that,

$$\mathrm{E}\left[V(\hat{\mathbf{e}}(t+\tau))\right] \leq c_5\,\mathrm{E}\left[V\big(\mathbf{e}(t)\big)\right] + c_6.$$

Then the boundedness of $\mathrm{E}\left[V\big(\hat{\mathbf{e}}(t+\tau)\big)\right]$ follows from the already established boundedness of $\mathrm{E}[V(\mathbf{e}(t))]$ (Corollary 4.1). ■

4.3 Remote estimation with local output measurement

We now consider the continuous-time LTI plant with partial observations

$$\begin{aligned} \dot{\mathbf{x}}(t) &= A\mathbf{x}(t) + \mathbf{w}(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{v}(t), \end{aligned} \tag{4.22}$$

where $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{l\times n}$, and $(C,\ A)$ is an observable pair. The Gaussian white noise $\mathbf{v}(t) \in \mathbb{R}^l$ and disturbance $\mathbf{w}(t) \in \mathbb{R}^n$ are mutually independent and zero-mean with positive definite covariance matrices $\Sigma_w \in \mathbb{R}^{n\times n}$ and $\Sigma_v \in \mathbb{R}^{l\times l}$.

We use an estimator scheme as shown in Figure 4.2, in which the local node has a *Kalman filter*. The Kalman filter state is sampled and sent to a remote estimator. We model the times at which data arrive at the remote estimator

with a counting process, as defined in Chapter 3. The communication schedulers work similar to those in Section 4.2. However, the construct of the controlled stochastic scheduler relies on the Kalman filter state $\tilde{\mathbf{x}}$ instead of the measured state $\mathbf{x}(t)$. This estimation scheme is motivated by the burgeon of *smart sensors*, which incorporate both computation and communication units.

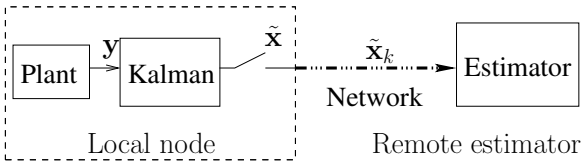


Figure 4.2. A remote estimation scheme: local Kalman filter state is sent.

This scheme is in contrast to the one in Figure 4.3, in which "raw measurements" are sent to a time-varying Kalman filter (TVKF) [30, 49].

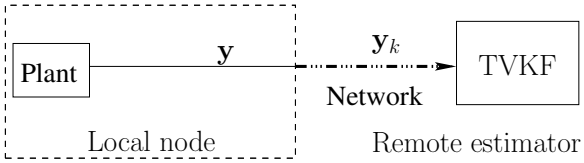


Figure 4.3. A remote estimate scheme: measurements are sent.

In this section we consider a broader class of communication schedulers and networks. We start with a communication-estimation scheme and its error dynamics, and then give stability conditions for uncontrolled schedulers as well as controlled stochastic schedulers. We also explore saturated intensity functions for controlled communications. We apply the results to periodic communication and networks that drop data packets. At the end of the section, we compare several aspects of the estimation schemes in Figures 4.2 and 4.3.

4.3.1 A communication-estimation scheme

The estimation scheme consists of a local node, a remote estimator, and a network, as shown in Figure 4.4. The local node has a Kalman filter and a communication scheduler. The Kalman filter state $\tilde{\mathbf{x}}(t)$ is computed locally in a continuous fashion (or with a short sampling period) and sent to the remote estimator at times determined by the communication scheduler. The decision on when to send data is based on how well the remote estimate $\hat{\mathbf{x}}$ matches the current Kalman filter state $\tilde{\mathbf{x}}$. To implement this idea, the local node keeps an identical copy of the remote estimate $\hat{\mathbf{x}}$. For simplicity of notations, we denote by $\hat{\mathbf{x}}$ the copy of $\hat{\mathbf{x}}$ on the local node.

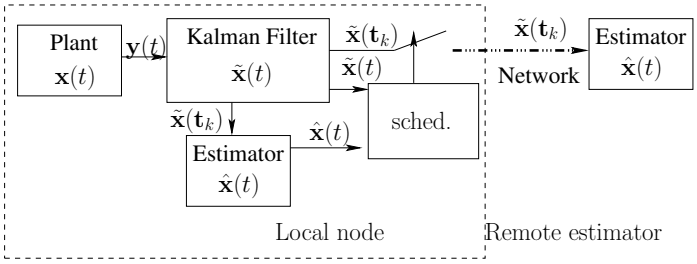


Figure 4.4. The detailed structure of the estimation scheme.

The Kalman filter for plant (4.22) is

$$\dot{\tilde{\mathbf{x}}} = A\tilde{\mathbf{x}} + L(\mathbf{y} - C\tilde{\mathbf{x}}), \tag{4.23}$$

where $L \in \mathbb{R}^{n\times l}$ makes $A - LC$ Hurwitz. The estimator, which has two identical copies on both the local node and the remote site, propagates the estimate open-loop most of the time and resets itself to the Kalman filter state $\tilde{\mathbf{x}}$ when data is

received,

$$\dot{\hat{\mathbf{x}}} = A\hat{\mathbf{x}} \tag{4.24}$$
$$\hat{\mathbf{x}}(\mathbf{t}_k) = \tilde{\mathbf{x}}(\mathbf{t}_k^-) - \mathbf{z}(k),$$

where $\mathbf{t}_k$ ($k \in \mathbb{N}$) indicates instants of data arrivals, and $\mathbf{z}(k)$ models the quantization errors. We assume that $\mathbf{z}(k)$ has probability distribution μ_k such that all finite moments are bounded, *i.e.*,

$$\mathrm{E}\left[|\mathbf{z}(k)|^m\right] < \Delta_z(m) < \infty, \quad m \geq 1.$$

We assume *i.i.d.* quantization errors with probability distribution function $\mu(z)$ on a finite support $\mathcal{M}$, and we replace $\mathbf{z}(k)$ by $\mathbf{z}$ when the context allows. The construction above implicitly requires the following assumption:

Assumption 4.1

1. *The delay is negligible,* i.e., $\tau = 0$.
2. *In case of lossy networks, an instantaneous acknowledgment of the packet arrival is available to the local node.*

4.3.2 Error dynamics

The estimation error $\hat{\mathbf{e}}$, the propagation error $\tilde{\mathbf{e}}$, and the Kalman filtering error $\boldsymbol{\xi}$ are defined as follows,

$$\hat{\mathbf{e}} := \mathbf{x} - \hat{\mathbf{x}}, \quad \tilde{\mathbf{e}} := \tilde{\mathbf{x}} - \hat{\mathbf{x}}, \quad \boldsymbol{\xi} := \mathbf{x} - \tilde{\mathbf{x}} \tag{4.25}$$

of which only $\tilde{\mathbf{e}}$ is known to the local node.

From (4.25), the estimation error $\hat{\mathbf{e}}(t)$ is stable in the m^{th} moment if both $\mathrm{E}\left[|\tilde{\mathbf{e}}(t)|^m\right]$ and $\mathrm{E}\left[|\boldsymbol{\xi}(t)|^m\right]$ are bounded. From (4.22)-(4.24), we conclude that the error systems of $\tilde{\mathbf{e}}(t)$ and $\boldsymbol{\xi}(t)$ have the following dynamics,

$$\dot{\tilde{\mathbf{e}}} = A\tilde{\mathbf{e}} + LC\boldsymbol{\xi} + L\mathbf{v}, \quad \tilde{\mathbf{e}}(\mathbf{t}_k) = \mathbf{z}, \tag{4.26a}$$

$$\dot{\boldsymbol{\xi}} = (A - LC)\boldsymbol{\xi} - L\mathbf{v} + \mathbf{w}. \tag{4.26b}$$

We regard $\tilde{\mathbf{e}}(t)$ in (4.26a) as a process driven by $\boldsymbol{\xi}(t)$ as given in (4.26b). Eq. (4.26a) can be written as a *jump-diffusion process*,

$$\mathrm{d}\tilde{\mathbf{e}} = A\tilde{\mathbf{e}}\,\mathrm{d}t + LC\mathrm{d}\boldsymbol{\xi} + L\,\mathrm{d}\mathbf{v} + \int_{\mathcal{M}} \left(\mathbf{z}(k) - \tilde{\mathbf{e}}(t^-)\right) \mathrm{d}\mathbf{N}(\mathrm{d}t, \mathrm{d}z). \tag{4.27}$$

Since $A - LC$ is Hurwitz, we have that $\boldsymbol{\xi}$ has all finite-order moments bounded, *i.e.*,

$$\mathrm{E}\left[|\boldsymbol{\xi}(t)|^m\right] \leq \Delta_\xi(m), \quad \forall\ t \geq 0. \tag{4.28}$$

In proving the theorems, we need an alternative representation of (4.27). If we stack the states, $\mathbf{e} := \left[\begin{smallmatrix} \tilde{\mathbf{e}} \\ \boldsymbol{\xi} \end{smallmatrix}\right]$, then

$$\dot{\mathbf{e}} = \bar{A}\mathbf{e} + \bar{\mathbf{w}}, \quad \tilde{\mathbf{e}}(\mathbf{t}_k) = \mathbf{z}, \tag{4.29}$$

where

$$\bar{A} := \left[\begin{smallmatrix} A & LC \\ 0 & A-LC \end{smallmatrix}\right],$$

and $\bar{\mathbf{w}} \in \mathbb{R}^{2n}$ has a covariance matrix

$$\Sigma = \left[\begin{smallmatrix} L\Sigma_v L' & L\Sigma_v L' \\ L\Sigma_v L' & L\Sigma_v L' + \Sigma_w \end{smallmatrix}\right] \geq 0.$$

The process (4.29) can also be written as a jump-diffusion equation,

$$\mathrm{d}\mathbf{e}(t) = \bar{A}\mathbf{e}\,\mathrm{d}t + \mathrm{d}\bar{\mathbf{w}} + \int \left(\left[\begin{smallmatrix} \mathbf{z}(k) \\ 0 \end{smallmatrix}\right] - \left[\begin{smallmatrix} I & 0 \\ 0 & 0 \end{smallmatrix}\right] \mathbf{e}(t)\right)\, \mathrm{d}\mathbf{N}(\mathrm{d}t, \mathrm{d}z). \tag{4.30}$$

4.3.3 Uncontrolled communications

We give a tight bound[2] on the Poisson rate for stochastic moment stability in uncontrolled communications. Consider a communication scheduler that is driven by a Poisson process with rate $\gamma > 0$.

Theorem 4.4 *Let the estimation error* $\hat{\mathbf{e}}$ *be defined as in* (4.22)-(4.25)*, in which the time sequence* $\mathbf{t}_k$ *(*$k \in \mathbb{N}$*) is generated by a Poisson process with rate* $\gamma > 0$*. For any* $m \geq 2$*,*

1. *if* $\gamma > m \max\{\Re[\mathrm{eig}(A)]\}$*, then* $\hat{\mathbf{e}}(t)$ *is stable in the* m^{th} *moment.*

2. *if* $\gamma < m \max\{\Re[\mathrm{eig}(A)]\}$*, then* $\hat{\mathbf{e}}(t)$ *is unstable in the* m^{th} *moment.*

Proof.[Theorem 4.4] Consider the jump-diffusion equation (4.30), in which $\mathbf{e} := \left[\begin{smallmatrix} \tilde{\mathbf{e}} \\ \boldsymbol{\xi} \end{smallmatrix}\right]$. The rest of the proof runs parallel to that of Theorem 4.1. ■

4.3.4 Controlled communications

We give a sufficient stability condition for controlled communications. Consider the stochastic communication scheduler that is driven by a counting process $\mathbf{N}(t)$ with the jump intensity that depends on the measured error dynamics. One choice is

$$\lambda(\tilde{\mathbf{e}}^-) = (\tilde{\mathbf{e}}^{-\prime} P \tilde{\mathbf{e}}^-)^k, \tag{4.31}$$

where $P \in \mathbb{R}^{n \times n}$ is positive definite and $k \in \mathbb{R}^+$. We will show in Section 4.5 that for the same packet rates, such communication schedules lead to smaller error variances than those with uncontrolled communication schedulers.

[2]See Remark 4.2 for the meaning of *tight*.

Theorem 4.5 *Let the estimation error* $\hat{\mathbf{e}}$ *be defined as in* (4.22)-(4.25)*, in which the time sequence* $\mathbf{t}_k$ *(*$k \in \mathbb{N}$*) is generated by a counting process with jump intensity* $\lambda(\tilde{\mathbf{e}}^-)$ *given by* (4.31)*. The estimation error* $\hat{\mathbf{e}}(t)$ *is stable in the* m^{th} *moment, for all* $m \geq 2$*.*

Proof.[Theorem 4.5] The proof runs parallel to the ones for uncontrolled communication schedulers. See Appendix B.1. ■

4.3.5 Controlled communications: saturated intensities

If the intensity function (4.31) is "saturated", but the saturation level is given by Theorem 4.4, the moment stability is maintained.

Theorem 4.6 *Let the estimation error* $\hat{\mathbf{e}}$ *be defined as in* (4.22)-(4.25)*, in which the time sequence* $\mathbf{t}_k$ *(*$k \in \mathbb{N}$*) is generated by an integer-valued process with jump intensity,*

$$\lambda(\tilde{\mathbf{e}}^-) = \min\{\gamma, (\tilde{\mathbf{e}}^{-\prime} P \tilde{\mathbf{e}}^-)^k\},$$

where $P > 0$*,* $\gamma > m \max\{\Re[\mathrm{eig}(A)]\}$*,* $m \geq 2$*, and* $k \in \mathbb{N}$*. Then* $\hat{\mathbf{e}}(t)$ *is stable in the* m^{th} *moment.*

The proof of Theorem 4.6 runs parallel to that of Theorem 4.4 and Theorem 4.1. Note that the intensity function can be written as $\lambda(\tilde{e}^-) = \gamma - b(\tilde{e}^-)$, where

$$b(\tilde{e}^-) = \begin{cases} \gamma - (\tilde{e}^{-\prime} P \tilde{e}^-)^k, & \text{if} \quad (\tilde{e}^{-\prime} P \tilde{e}^-)^k < \gamma, \\ 0, & \text{otherwise,} \end{cases}$$

and $\mathrm{E}\left[b(\tilde{\mathbf{e}}^-)V(\tilde{\mathbf{e}}^-)\right]$ is uniformly bounded. The details are omitted.

4.3.6 Periodic communication over lossy networks

We consider periodic communication schedulers in which data packets are sent to the network with period T_s but get lost with probability p.

Theorem 4.7 *If the data loss probability p satisfies*

$$p < \exp(-mT_s \max\{\Re[\mathrm{eig}(A)]\}),$$

the estimation error $\hat{\mathbf{e}}(t)$ is stable in the m^{th} moment, for all $m \geq 2$.

To prove the result, we apply Theorem 4.4. Subtleties arise from the fact that the Poisson process does not necessarily jump at the sampling instants, and there may be more than one jump in one sampling period.

Proof.[Theorem 4.7] We start from the process $\mathbf{e}(t)$ in (4.29) with constant Poisson rate γ. Define a continuous-time jump-diffusion process $\mathbf{e}_p = \begin{bmatrix} \tilde{\mathbf{e}}_p \\ \boldsymbol{\xi}_p \end{bmatrix}$. For $k \in \mathbb{N}$, we have

$$\begin{aligned} \dot{\mathbf{e}}_p &= \bar{A}\mathbf{e}_p + \bar{\mathbf{w}} \\ \tilde{\mathbf{e}}_p(kT_s) &= \mathbf{z}, \quad \text{if } \tilde{\mathbf{e}}(t) \text{ has jumps in } \big((k-1)T_s, kT_s\big], \end{aligned} \tag{4.32}$$

i.e., $\mathbf{e}_p$ is driven by the same disturbance/noise $\bar{\mathbf{w}}$ as $\mathbf{e}$ is, but has a jump on any instant kT_s if and only if $\mathbf{e}$ has jump(s) in the interval $\big((k-1)T_s, kT_s\big]$. By construction, the process $\mathbf{e}_p(t)$ is the error system for a periodic sending scheme with period T_s and data loss probability

$$p = \exp(-\gamma T_s), \tag{4.33}$$

which is the probability that the process $\mathbf{e}(t)$ has no jumps on the interval $\big((k-1)T_s, kT_s\big]$.

If the process $\mathbf{e}(t)$ in (4.29) has bounded the m^{th} moment, $t \geq 0$, so does $\mathbf{e}_p(t)$ in (4.32). A proof can be adapted from the proof of Theorem 4.3.

When $p < \exp(-mT_s \max\{\Re[\text{eig}(A)]\})$, Eq. (4.33) leads to

$$\gamma = -\frac{\ln p}{T_s} > m \max\{\Re[\text{eig}(A)]\},$$

from which $\mathbf{e}(t)$ in (4.29) is stable in the m^{th} moment (Theorem 4.4). Hence $\mathbf{e}_p(t)$ in (4.32) is also stable in the m^{th} moment. This completes the proof. ■

4.3.7 Discussions

We compare the stability conditions between the discrete-time estimation schemes in which the Kalman filter states are sent (Figure 4.2) and the schemes in which the raw measurements are sent (Figure 4.3).

To this effect, we interpret Theorem 4.7 in the discrete-time domain. Consider the discrete-time LTI plant,

$$\begin{aligned} \mathbf{x}_{t+1} &= A_d \mathbf{x}_t + \mathbf{w}_t \\ \mathbf{y}_t &= C_d \mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{4.34}$$

where the matrices are defined similar to (4.22), and $\mathbf{w}_t$ and $\mathbf{v}_t$ are *i.i.d.* Gaussian random variables.

Let (4.34) be a zero-order-hold discretization of (4.22) with sampling time T_s, *i.e.*, $A_d = \exp(AT_s)$ and $C_d = C$. For simplicity, suppose there is no quantization error, *i.e.*, $\mathbf{z} = 0$ in (4.32). When the node sends the Kalman filter state (Fig 4.2) with a data loss probability $p = \exp(-\gamma T_s)$, the discrete-time estimator has the same dynamics as the process $\mathbf{e}_p$ defined in (4.32) but sampled at times tT_s

($t \in \mathbb{N}$). From Theorem 4.7, we conclude that the m^{th} moment of the discrete-time estimation error is bounded provided that the data loss probability p satisfies

$$p < \frac{1}{(\max\{|\operatorname{eig}(A_d)|\})^m}. \tag{4.35}$$

In [49], the authors consider an estimation scheme in which raw measurements are sent. They show the existence of a critical data loss probability p_c, beyond which a transition to an unbounded state error covariance occurs. This critical value is bounded by $\underline{p} \leq p_c \leq \bar{p}$, where $\underline{p}$ and $\bar{p}$ are solved by LMIs and the latter has a closed form,

$$\bar{p} = \frac{1}{(\max\{|\operatorname{eig}(A_d)|\})^2}.$$

The lower bound $\underline{p}$ depends on A_d and C_d. They show that $p_c < \bar{p}$ in general, and $p_c = \bar{p}$ only in special cases, *e.g.*, if C_d is invertible, or if A_d has one unstable eigenvalue.

To achieve the second moment stability, the remote estimation structure in Figure 4.2 tolerates a data loss probability, which is in general higher than the structure in Figure 4.3. The bound in (4.35) depends only on the matrix A_d. Eq. (4.35) also provides stability conditions for higher order moments. However, the scheme in Figure 4.2 requires Kalman filtering on the local node.

We compare the packet rate requirements when schemes in Figure 4.2 and Figure 4.3 are used to estimate the plant in Example 2.1.

Example 4.1 *Consider the estimation of the following LTI plant*

$$\mathbf{x}_{t+1} = \begin{bmatrix} 2 & 0 \\ 0 & \sqrt{2} \end{bmatrix} \mathbf{x}_t + \mathbf{w}_t, \qquad \mathbf{y}_t = \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{x}_t + \mathbf{v}_t.$$

where $\mathbf{w}_t$ *and* $\mathbf{v}_t$ *are Gaussian random variables. The network has data packet loss probability* $p \in [0, 1)$*, and the network delay is* $\tau = 1$*.*

1. *The plant sends* $\tilde{\mathbf{x}}_t$ *at time* t*, and it arrives at the remote estimator at time* $t+1$ *with probability* $1-p$ *(see Figure 4.2). The estimation error variance is bounded if and only if*

$$p < 0.2500.$$

2. *The plant sends* $\mathbf{y}_t$ *at time* t*, and it arrives at the TVKF at time* $t+1$ *with probability* $1-p$ *(see Figure 4.3). To have bounded estimation variance using the TVKF scheme, the dropout probability must be such that* $p < p_c$*, where* p_c *is some value bounded by*

$$0.1249 \leq p_c \leq 0.2500,$$

in which the lower bound is solved by LMIs as proposed by [49].

We observe that former imposes less stringent requirements on packet transmission reliability.

4.4 Deterministic communication schedulers

We now consider communication schedulers that utilize deterministic rules but restrict our attention to delay-free networks. While the estimation error dynamics is still represented by (4.12), the increments of the process $\mathbf{N}(t)$ are determined by the process $\mathbf{e}(t)$ in a deterministic fashion. The communication scheduler

monitors a radially unbounded, positive, and continuous *communication index* $S : \mathbb{R}^n \to \mathbb{R}^+$ and forces a node to broadcast its state when $S(\mathbf{e}) \geq 1$. In particular, a message broadcast occurs at time $\mathbf{t}_k$ when $\lim_{t \uparrow t_k} S\big(\mathbf{e}(t)\big) \geq 1$. To avoid chattering, the post-reset value $\mathbf{v}(k)$ should satisfy $S\big(\mathbf{v}(k)\big) < 1$ with probability one. Again, we assume $\mathbf{v}(k)$ has finite support $\mathcal{M}$ and is stationary. Denote $\mathbf{v}(k)$ by $\mathbf{v}$ as context allows. Such resetting guarantees that $\mathbf{e}(t)$ is bounded, since

$$\mathbf{e}(t) \in \mathcal{D} := \{e \in \mathbb{R}^n | S(e) \leq 1\}, \qquad \forall t \geq 0, \tag{4.36}$$

with probability one.

Now we determine the packet rate $\mathcal{R}$. Suppose that a message exchange occurs at time $\mathbf{t}_{k-1}$, when $\mathbf{e}(\mathbf{t}_{k-1})$ is reset to $\mathbf{v}$. From $\mathbf{t}_{k-1}$ to the next reset time $\mathbf{t}_k$, $\mathbf{e}(t)$ is a pure diffusion process,

$$\dot{\mathbf{e}} = A\mathbf{e} + \dot{\mathbf{w}}. \tag{4.37}$$

Given $\mathbf{v} = v$, define $\mathbf{T}_k(v)$ to be the inter-communication time, *i.e.*,

$$\mathbf{T}_k(v) = \inf\{t - \mathbf{t}_{k-1} \geq 0 : \mathbf{e}(t) \in \partial\mathcal{D}, \mathbf{e}(\mathbf{t}_{k-1}) = v\},$$

where $\mathbf{e}(t)$ is governed by (4.37) for $t \geq \mathbf{t}_{k-1}$ and $\partial\mathcal{D}$ denotes the boundary of $\mathcal{D}$. The random variable $\mathbf{T}_k(v)$ is known as the *first exit time* of $\mathbf{e}(t)$ from $\mathcal{D}$. It is in general difficult to obtain the distribution of $\mathbf{T}_k(\mathbf{v})$ in a closed form, but its expected value can be obtained from Dynkin's equation. In particular, defining $g(v) := \mathrm{E}[\mathbf{T}_k(v)]$, it is known that $g(v)$ is a solution to the following boundary value problem:

$$\begin{aligned} &\frac{\partial g(v)}{\partial v} \cdot Av + \frac{1}{2}\,\mathrm{trace}\left[\sigma' \frac{\partial^2 g(v)}{\partial v^2}\sigma\right] = -1, \\ &\forall\, v \in \mathcal{D}, \quad g(v) = 0 \quad \forall\, v \in \partial\mathcal{D}, \end{aligned} \tag{4.38}$$

where $\frac{\partial g(v)}{\partial v}$ and $\frac{\partial^2 g(v)}{\partial v^2}$ denote the gradient vector and the Hessian matrix, respectively [44]. Define $\mathbf{T}_k := \mathbf{t}_k - \mathbf{t}_{k-1}$ to be the *intercommunication time* between the $(k-1)^{th}$ and the k^{th} messages. Given $g(v)$, we have

$$\mathrm{E}[\mathbf{T}_k] = \mathrm{E}[g(\mathbf{v})] = \int_{\mathcal{M}} g(v) d\mu(v).$$

Since $\mathbf{T}_k$ is *i.i.d.*, it is straightforward to show that the packet rate

$$\mathcal{R} = \lim_{k\to\infty} \mathbf{E}\Big[\frac{k}{\sum_{i=1}^{k} \mathbf{T}_k}\Big] = \frac{1}{\mathbf{E}[\mathbf{T}_k]},$$

from which we have

$$\mathcal{R} = \frac{1}{\int_{\mathcal{M}} g(v) \mathrm{d}\mu(v)}.$$

In practice, (4.38) needs to be solved numerically. Since $\mathcal{D}$ is compact, Eq. (4.36) provides a bound for $\mathbf{e}(t)$ and consequently on its statistical moments. To obtain tighter bounds one can use Kolmogorov's forward equation with appropriate boundary conditions to compute the probability density function of the error $\mathbf{e}(t)$. However, this method is computationally intensive for higher-order systems. In special cases, we have closed-form solutions.

Example 4.2 *For a one-dimensional system with $A = 0$ and $\mathcal{D} := [-L, L]$, we obtain*

$$g(z) = \frac{L^2 - z^2}{\sigma^2}, \quad z \in \mathcal{D}.$$

Assuming that the quantization error $\mathbf{v}$ is uniformly distributed on $[-\frac{q}{2}, \frac{q}{2}]$ (for some $q \geq 0$), we obtain the following packet rate,

$$\mathcal{R} = \frac{12L^2 - q^2}{12\Sigma_w}.$$

4.5 Simulations

In this section we validate the theoretical results through Monte Carlo simulations. All the simulations are done in Matlab/Simulink. The counting process $\mathbf{N}(t)$ is realized by a single binomial test. Specifically, for a fixed time step h, if a binomial test characterized by a success probability $p = 1 - e^{-h\lambda(\mathbf{e}(kh))}$ succeeds, a message exchange is triggered at time $t_k := k\,h$ ($k \in \mathbb{N}$). Convergence results for similar procedures can be found in [14] and references therein.

4.5.1 Rate-variance curves

To study the trade-off between packet rates and estimation error variances, we consider the remote state estimator of a first order unstable plant,

$$\dot{\mathbf{x}}(t) = \mathbf{x}(t) + \mathbf{w},$$

which corresponds to a jump-diffusion process defined by (4.12) with $A = 1$ and $\sigma = 1$. The simulation time for these results is 1000 seconds. The system's instability presents additional challenges to the estimation.

Figure 4.5 (left) depicts the trade-off between the packet rate and the estimation error variance for four different communication schedulers: periodic, uncontrolled stochastic scheduling, controlled stochastic scheduling with quadratic intensities, and deterministic scheduling with quadratic communication indexes. The curves are obtained by varying the parameters that define these schedulers. For a given packet rate, the uncontrolled stochastic scheduling (constant intensity) results in the largest error variances, whereas the deterministic scheduling results in the smallest. The packet rate obtained with the stochastic scheduling

for the quadratic $\lambda(e)$ is much smaller than the upper bound provided by (4.20), which can be shown to be 1.

Figure 4.5 (right) provides a comparison between deterministic and stochastic schedulers. The deterministic schedulers have a communication index in the form of $S(e) := \frac{e^2}{\Delta} \leq 1$, and the different points on the curve are generated by changing Δ. The stochastic schedulers have intensities of the form $\lambda(e) = (\frac{e^2}{\Delta})^k$, where Δ is a positive parameter and $k \in \{1, 2, 3, 4, 5\}$. For large k, $\lambda(e)$ essentially provides a barrier at $e^2 = \Delta$, which acts as the bound in the deterministic schedulers. Therefore it is not surprising to see that as k increases, the stochastic schedulers converge to the deterministic schedulers. As proved for discrete systems, the deterministic curve provides optimal trade-off between communication cost and control performance [61].

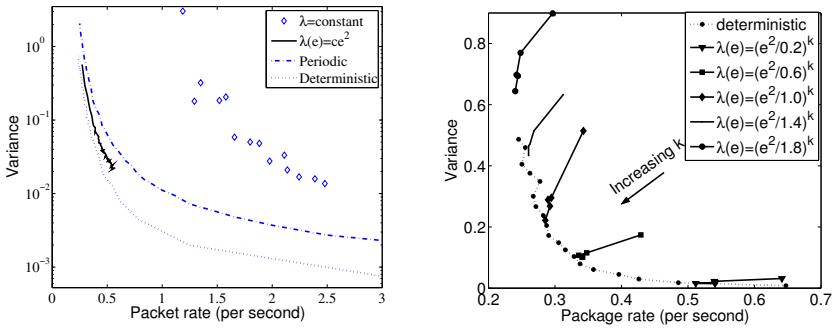


Figure 4.5. Packet rate vs. estimation error variance.

4.5.2 Remote estimation with local output measurement

A second-order plant example with two unstable poles in [49] is recaptured in continuous settings with $T_s = 0.1s$, in which

$$A = \left[\begin{smallmatrix} 2.23 & 0 \\ 8.52 & 0.95 \end{smallmatrix}\right], \quad C = \left[\begin{smallmatrix} 1 & 1 \end{smallmatrix}\right],$$

and the disturbance and noise covariance matrices are

$$\Sigma_w = \left[\begin{smallmatrix} 200 & 0 \\ 0 & 200 \end{smallmatrix}\right], \quad \Sigma_v = 25,$$

respectively. The Kalman filter gain is $L = \left[\begin{smallmatrix} 31.88 \\ 8.36 \end{smallmatrix}\right]$. Figure 4.6 is the variance of 5000-sample Monte Carlo simulation for different constant Poisson rates γ, which are chosen around $\Gamma := 2\max\{\Re[\text{eig}(A)]\} \approx 4.46$. For $\gamma = 0.2\Gamma$ the error variance blows up, and for $\gamma = 1.8\Gamma$, it is well below 1000. For $\gamma = 1.1\Gamma$, the variance is still bounded, but not as clear-cut. The curve for $\gamma = 0.9\Gamma$ has a trend of going up. We conjecture that limited samples in the simulation cause statistical variations.

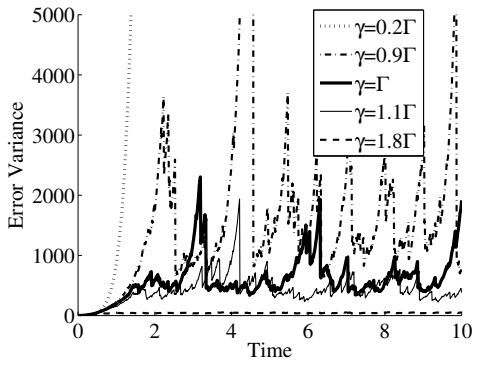


Figure 4.6. Estimation error variances under different Poisson rates.

The controlled communication scheduler chooses the intensity function $\lambda(\tilde{e}^-) = \tilde{e}^{-\prime} P \tilde{e}^-$. Figure 4.7 gives the variance and packet rate for different P. For

$P = 0.2I_{2\times 2}$, the corresponding rate is 5.25, which is roughly the same as that of constant $\gamma = 1.1\Gamma$. The former has a variance around 5 (Figure 4.7), while the latter has 427 (Figure 4.6). The controlled communication schedule does reduce error variance given communication constraints.

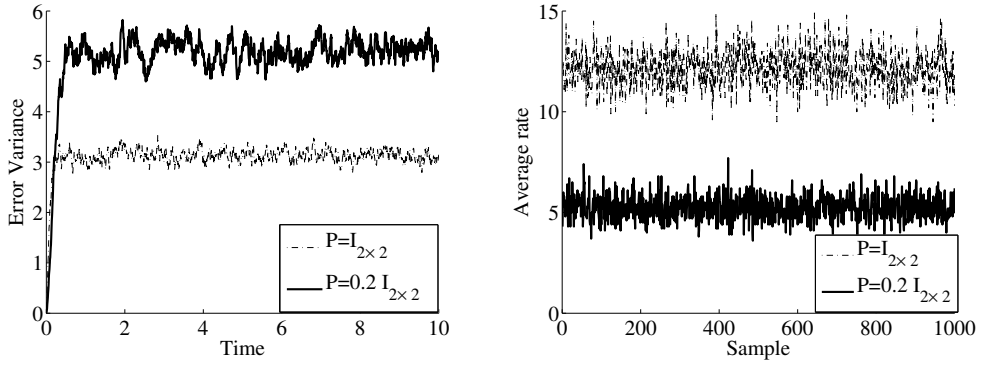


Figure 4.7. Variances (left) and packet rates (right) with jump intensity $\tilde{e}^{-\prime} P \tilde{e}^{-}$.

4.6 Summary

Stochastic communication schedulers are proposed for estimation over networks. Using tools from jump-diffusion processes, we investigate conditions under which such schedulers guarantee stochastic moment stability of estimation errors. We consider plants with full or partial observations and networks with or without delays. We also apply the stochastic communication scheduling results to periodic communication and lossy networks. Monte Carlo simulations demonstrate that these communication schedulers can save communication resources over periodic schemes. Deterministic communication schedulers are briefly discussed, which are solved as a first exit time problem via Dynkin's equation.

Chapter 5

Optimal communication schedulers

In Chapter 4, we introduced various communication scheduling methods and analyzed their effects on estimation over networks. In this chapter, we turn our attention to performance and pursue optimal communication schedulers.

We propose a communication-estimation scheme that includes a local node (plant), a remote node (estimator), and a network. The objective is to increase the estimation performance and to reduce the communication load. We pose a long-term average cost optimization problem and address three questions:

1. *What* does the local node send to the estimator through the network?
2. *When* does the local node send data?
3. *How* does the remote node make the best use of data available?

The emphasis is on the second issue, *i.e.*, *how to schedule data communication.*

We call this an *optimal communication scheduling* problem.

We consider the discrete-time domain to avoid the technicalities that arise in continuous-time jump diffusion processes. We consider stochastic communication schedulers for which, at each time instant, a probability $u \in [0, 1]$ is assigned to whether or not data should be sent. These scheduling methods also include deterministic rules as special cases for which $u \in \{0, 1\}$.

For networks that do not drop data packets (lossless), we develop an algorithm to calculate the optimal communication scheduler. For lossy networks, our analysis provides both upper and lower bounds for the optimal solution.

This chapter is organized as follows. In Section 5.1, we discuss the application of optimal communication schedulers in *medium access control* (MAC) design and introduce an optimization criterion. In Section 5.2, we formulate the problem of LTI estimation over delayed networks and detail the communication-estimation scheme. In Section 5.4, we optimize the communication schedulers through dynamic programming. The average cost optimality problem is of interest on its own because it deals with a Borel state-space and an unbounded per-stage cost. In Section 5.5, we give numerical examples.

5.1 Motivation

In this section, we first discuss the so-called p-persistent MAC protocols [54] to motivate *stochastic communication schedulers*. Stochastic communication schedulers can be integrated into communication protocols to improve the communication channel efficiency in NCS applications. Then we discuss a cost criterion

that includes a quadratic term on estimation errors and a linear term on communications.

5.1.1 Stochastic communication scheduling methods

In the p-persistent method, when a collision happens at time t, a node sends data with probability u, $0 \leq u \leq 1$, at time $t+1$. We use the symbol u to emphasize that the data sending probability is a control variable. A larger u implies a more aggressive algorithm. The value of u is chosen so that the network is efficiently used and the probability of repeated collisions is low. The introduction of randomness allows each node to act independently, which is critical when the network scale is large and coordination between nodes is expensive. When the parameter u is properly chosen, the network achieves better communication performance, *e.g.*, less frequent collisions and a larger data throughput.

The parameter u can be chosen dynamically. Generally, we may consider a feasible set

$$u \in [p_{min}, p_{max}], \quad 0 \leq p_{min} < p_{max} \leq 1. \tag{5.1}$$

The lower bound, p_{min}, is chosen such that the network is effectively utilized. The upper bound, p_{max}, ensures that the probability of packet collisions among different nodes is below a certain level. The parameters p_{min} and p_{max} depend on the number of nodes, network bandwidth, and system dynamics. We use $u \in [0, 1]$, but the results can be extended to the general case (5.1) with minimal modifications, as discussed later in Remark 5.2.

5.1.2 Cost criterion

We propose a cost function that *quadratically* penalizes the estimation error and *linearly* penalizes data communication.

At time t, the communication scheduler decides whether or not to send a data packet, based on the information available up to time t. The decision is signified by the value of $\mathbf{a}_t$ ($\mathbf{a}_t = 1$ if sending, and $\mathbf{a}_t = 0$ otherwise).

We consider an estimation problem, in which the estimation error is denoted by $\mathbf{e}_t \in \mathbb{R}^n$. Define the long-term average cost as

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N} \mathbf{e}_t' Q \mathbf{e}_t + \lambda \mathbf{a}_t\right],$$

where $Q \in \mathbb{R}^{n\times n}$ is positive definite and $\|Q\|_2 = 1$, and $\lambda > 0$. The cost criterion consists of two parts, namely, the weighted average estimation error variance,

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{k=0}^{N-1} \mathbf{e}_t' Q \mathbf{e}_t\right],$$

and the average packet rate,

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N} \mathbf{a}_t\right].$$

The weight λ penalizes data communication. The larger λ becomes, the more "expensive" data communication is. This formulation is a form of *multicriterion optimization*. By changing Q and λ, we can also solve optimization problems with hard constraints (*e.g.* the weighted average estimation error variance is below a certain level) [6, pp 54].

5.2 Problem statement

The problem is to estimate the state, $\mathbf{x}_t$, of a given discrete-time LTI plant over a delayed network,

$$\begin{aligned} \mathbf{x}_{t+1} &= A\mathbf{x}_t + \mathbf{w}_t \\ \mathbf{y}_t &= C\mathbf{x}_t + \mathbf{v}_t, \end{aligned} \tag{5.2}$$

where $\mathbf{w}_t$ and $\mathbf{v}_t$ are zero-mean Gaussian variables with covariance matrices Σ_w and Σ_v, respectively. The initial condition of $\mathbf{x}_t$ has a Gaussian distribution. The plant, including its computational and communication units, is also referred to as the *local node*. The remote estimator, including its communication and computational units, is also referred to as the *remote node*. The local node needs to decide *when to send data* and *what to send*. The remote node needs to make the best use of the received data to reconstruct the state of (5.2). We assume the following:

Assumption 5.1 *Computation time is negligible compared to communication time.*

Assumption 5.2 *If the local node sends data at t, the packet gets lost with probability p,* i.i.d.*, but at time $t+1$, the local node gets the acknowledgment of whether or not the packet will be received.*

Assumption 5.3 *It takes the remote node $\tau \geq 1$ time steps to receive any data packet.*

Assumption 5.2 holds for some MAC mechanisms, in which acknowledgments are "received" right away and can be used at the next time period. For example,

in CSMA/CD, a transmitting node monitors its own transmission. It detects a collision by watching if there is excess current above what it is generating. The fixed length of delay in Assumption 5.3 streamlines the main results.

5.2.1 Information patterns

We consider the information patterns of the local node and that of the remote estimator (see Figure 5.1).

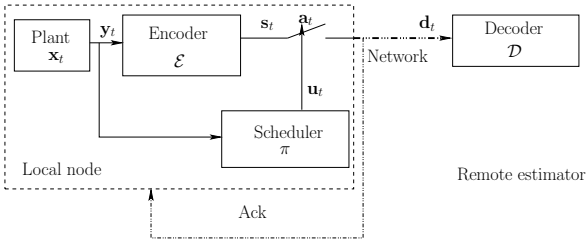


Figure 5.1. Information patterns.

The local node

The local node has access to all measurements, $\mathcal{Y}(t) := \{\mathbf{y}_j : j \leq t\}$. It sends data to the network in forms of *packets*. The node decides whether or not to send data packet $\mathbf{s}_t$ at time t. This decision is signified by the value of $\mathbf{a}_t$:

$$\mathbf{a}_t = \begin{cases} 1, & \text{if } \mathbf{s}_t \text{ is sent at } t, \\ 0, & \text{otherwise,} \end{cases}$$

and we denote by $\Theta_a(t) := \{\mathbf{a}_j : j \leq t\}$ the history of $\mathbf{a}_t$. We denote by $\boldsymbol{\nu}_t$ whether or not the remote node receives the data packet $\mathbf{s}_{t-\tau}$ at time t, *i.e.*,

$$\boldsymbol{\nu}_t = \begin{cases} 1, & \text{if } \mathbf{s}_{t-\tau} \text{ is received at } t, \\ 0, & \text{otherwise,} \end{cases}$$

and denote by $\Theta_\nu(t) := \{\boldsymbol{\nu}_j : j \leq t\}$ the history of $\boldsymbol{\nu}_t$. By Assumptions 5.2 and 5.3, at time t, the local node has the knowledge of $\Theta_\nu(t-1+\tau)$ [1]. Therefore the local node's information set at time t is,

$$\mathcal{I}_s(t) = \{\mathcal{Y}(t), \Theta_a(t-1), \Theta_v(t-1+\tau)\}.$$

The decision on whether or not to send data at time t is an outcome of a random draw, *i.e.*,

$$\mathbf{a}_t = \begin{cases} 1, & \text{with probability } \mathbf{u}_t, \\ 0, & \text{with probability } 1-\mathbf{u}_t, \end{cases} \tag{5.3}$$

where the *control* $\mathbf{u}_t$ is determined by a communication scheduling function π, *i.e.*,

$$\mathbf{u}_t = \pi(\mathcal{I}_s(t)). \tag{5.4}$$

Each packet $\mathbf{s}_t$ encodes information in the form of

$$\mathbf{s}_t = \mathcal{E}(\mathcal{I}_s(t)), \tag{5.5}$$

in which the notation $\mathcal{E}$, understood as an *encoder*, is a slight abuse of its counterpart in information theory literature [10].

We define the *packet rate* as the long-term average data *sending* frequency,

$$\mathcal{R} := \limsup_{N\to\infty} \mathrm{E}\left[\frac{\sum_{j=1}^{N} \mathbf{a}_j}{N}\right].$$

[1] The latest packet in question is $\mathbf{s}_{t-1}$.

One of our objectives is to reduce $\mathcal{R}$ so as to improve the communication performance. As mentioned in Chapter 2, the channel performance is adversely affected when the packet rate is too high.

The remote estimator

The local node sends data packets to the network, and the remote estimator receives some of them after a fixed time delay of τ. The remote estimator's information set at time t is

$$\mathcal{I}_r(t) := \{(\boldsymbol{\nu}_j, \mathbf{d}_j) : j \leq t\},$$

where $\mathbf{d}_t$ is the packet that the estimator receives at time t. From Assumption 5.3, there is a fixed delay τ and therefore

$$\mathbf{d}_t = \begin{cases} \mathbf{s}_{t-\tau}, & \text{if} \quad \boldsymbol{\nu}_t = 1, \\ \varnothing, & \text{if} \quad \boldsymbol{\nu}_t = 0. \end{cases}$$

The estimator reconstructs the state of (5.2) by an estimator of the form

$$\hat{\mathbf{x}}_t = \mathcal{D}(\mathcal{I}_r(t)), \tag{5.6}$$

in which $\mathcal{D}$ is understood as a *decoder*.

5.2.2 Problem formulation

To evaluate the long-term performance of a communication-estimation scheme, we use a long-term average cost criterion that penalizes both the estimation error,

$$\hat{\mathbf{e}}_t := \mathbf{x}_t - \hat{\mathbf{x}}_t \tag{5.7}$$

and the packet rate $\mathcal{R}$.

Problem 5.1 *Estimate the state of the plant* (5.2) *over networks under Assumptions 5.1-5.3, with communication scheduling function* (5.4)*, encoder* (5.5)*, and decoder* (5.6)*, so as to minimize the long-term average cost function,*

$$J_1(\pi, \mathcal{E}, \mathcal{D}) = \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} \hat{\mathbf{e}}_t' Q \hat{\mathbf{e}}_t + \lambda \mathbf{a}_t \mid \pi, \mathcal{E}, \mathcal{D}\right]. \tag{5.8}$$

The communication, encoding, and decoding policies that achieve the minimum in (5.8) *are denoted by* π^**,* $\mathcal{E}^*$*, and* $\mathcal{D}^*$*, respectively. The corresponding minima is denoted by* $J_1^* = J(\pi^*, \mathcal{E}^*, \mathcal{D}^*)$*.*

In the next sections, we reformulate the problem in simpler forms without compromising the optimal performance. We will take advantage of the Markovian and Gaussian properties of the plant (5.2).

5.2.3 Communication-estimation structure

In this section, we propose a special structure for π, $\mathcal{D}$, and $\mathcal{E}$, and show that they do not lead to larger optimal cost values. The structure is depicted in Figure 5.2.

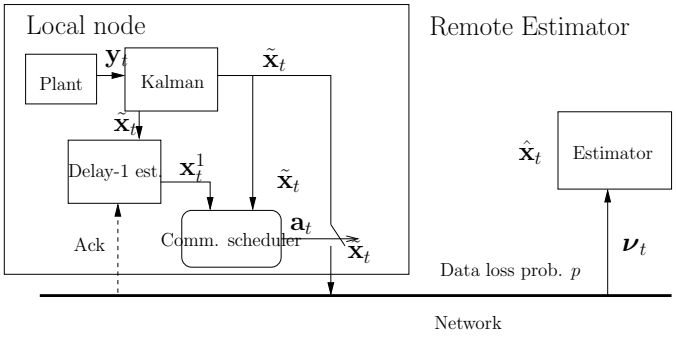


Figure 5.2. A structured solution to Problem 5.1.

The local Kalman filter

The encoder $\mathcal{E}$ encodes a local Kalman filter state at time t into a packet $\mathbf{s}_t$. The local Kalman estimator for plant (5.2) is based on the complete measurement history, independent of network delays:

$$\tilde{\mathbf{x}}_{t+1} = A\tilde{\mathbf{x}}_t + L(\mathbf{y}_{t+1} - CA\tilde{\mathbf{x}}_t), \tag{5.9}$$

where $L \in \mathbb{R}^{n\times l}$ makes $A - LCA$ Hurwitz. Define the estimation error,

$$\tilde{\mathbf{e}}_t := \mathbf{x}_t - \tilde{\mathbf{x}}_t.$$

The estimator (5.9) results in an error covariance matrix $\Sigma_L = \mathrm{E}[\tilde{\mathbf{e}}_t\tilde{\mathbf{e}}_t']$. We have that

$$\tilde{\mathbf{e}}_{t+1} = (A - LCA)\tilde{\mathbf{e}}_t - L\left(C\mathbf{w}_t + \mathbf{v}_{t+1}\right) + \mathbf{w}_t. \tag{5.10}$$

The local node sends the Kalman filter state $\tilde{\mathbf{x}}_t$ to the network if the communication scheduler decides to do so at time t, which is to be specified later.

The remote estimator

The remote estimator receives the data packet containing the Kalman filter state $\tilde{\mathbf{x}}_t$ after a delay of τ with probability $1-p$, where p the probability that the data packet gets lost.

The decoder $\mathcal{D}$ is an unbiased estimator, which calculates $\hat{\mathbf{x}}_t$ using data received by time t according to

$$\hat{\mathbf{x}}_t = \begin{cases} A\hat{\mathbf{x}}_{t-1}, & \text{if } \boldsymbol{\nu}_t = 0, \\ A^\tau\tilde{\mathbf{x}}_{t-\tau}, & \text{if } \boldsymbol{\nu}_t = 1. \end{cases} \tag{5.11}$$

The local estimator

The communication policy π is related to the so-called 1-delay estimator, constructed as if the network only has a 1-step delay:

$$\hat{\mathbf{x}}_t^1 = \begin{cases} A\hat{\mathbf{x}}_{t-1}^1, & \text{if } \boldsymbol{\nu}_t = 0, \\ A\tilde{\mathbf{x}}_{t-1}, & \text{if } \boldsymbol{\nu}_t = 1, \end{cases} \tag{5.12}$$

which is feasible, because by Assumption 5.2 the local node has the acknowledgment by time t of whether $\tilde{\mathbf{x}}_{t-1}$ will be received or not. Define the 1-delay estimation error $\mathbf{e}_t^1$ as

$$\mathbf{e}_t^1 := \tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_t^1, \tag{5.13}$$

which will be used by the communication scheduler.

During the derivation of the main result, we need to use the k-delay estimators, $1 < k \leq \tau$, defined as follows,

$$\hat{\mathbf{x}}_t^k = \begin{cases} A\hat{\mathbf{x}}_{t-1}^k, & \text{if } \boldsymbol{\nu}_t = 0, \\ A^k\tilde{\mathbf{x}}_{t-k}, & \text{if } \boldsymbol{\nu}_t = 1, \end{cases} \tag{5.14}$$

which are not physically implemented in the communication-estimation scheme (see Figure 5.2).

The stochastic communication scheduler

The local node decides whether or not to send data according to (5.3), *i.e.*, sending data with probability $\mathbf{u}_t$. However, the communication scheduling function (5.4) is determined by the 1-delay estimation error $\mathbf{e}_t^1$, *i.e.*,

$$\mathbf{u}_t = \gamma_t(\mathbf{e}_t^1),$$

where the function $\gamma_t : \mathbb{R}^n \to [0,1]$. A policy π is a sequence of communication scheduling functions, *i.e.*,

$$\pi := \{\gamma_0, \gamma_1, \cdots\}. \tag{5.15}$$

5.2.4 Structured reformulation

We impose the communication-estimation structure in Figure 5.2 to Problem 5.1:

Problem 5.2 *Given the communication-estimation structure in Figure 5.2, with Kalman filter* (5.9)*, remote estimator* (5.11)*, and 1-delay estimator* (5.12)*, find the optimal communication policy* π *as in* (5.15) *to minimize the long-term average cost function*

$$J_2(\pi) = \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} \hat{\mathbf{e}}_t' Q \hat{\mathbf{e}}_t + \lambda \mathbf{u}_t | \pi\right], \quad \hat{\mathbf{e}}_t \in \mathbb{R}^n. \tag{5.16}$$

The minimizing communication policy is denoted by π^**, and the corresponding minimum long-term average cost is denoted by* $J_2^* = J(\pi^*)$*.*

Lemma 5.1 *Problem 5.2 solves Problem 5.1 without loss of optimality,* i.e.*,*

$$J_1^* = J_2^*,$$

where J_1^* *and* J_2^* *are minimum costs obtained by solving Problem 5.1 and Problem 5.2, respectively.*

Proof.[Lemma 5.1] For the purpose of state estimation with a second moment criterion, $\tilde{\mathbf{x}}_t$ contains as much information as $\mathcal{Y}(t)$ does [52]. Therefore, sending

at time t $\tilde{\mathbf{x}}_t$ does not lose optimality. At time t, the most information that the remote node can expect is $\mathcal{Y}(t-j)$, or $\tilde{\mathbf{x}}_{t-j}$ $(j \geq \tau)$, which is the most recently received packet. Then the estimator (5.11) is the minimum variance estimator. Given *what to send*, and *how to reconstruct*, the optimality of the communication policy of the form (5.15) is proved [61] in a similar setting. It will be shown later in this chapter. ■

5.2.5 Error dynamics

We analyze the error dynamics for the communication-estimation scheme in Figure 5.2. During the derivation, we need the definition of the k-delay estimation error $\hat{\mathbf{e}}_t^k$,

$$\hat{\mathbf{e}}_t^k := \mathbf{x}_t - \hat{\mathbf{x}}_t^k,$$

where $\hat{\mathbf{x}}_t^k$ is the state of the k-delay estimator defined in (5.14). For the special case of $k = \tau$, $\hat{\mathbf{e}}_t^\tau$ is abbreviated as $\hat{\mathbf{e}}_t$ as in (5.7). The lemma below is central to converting Problem 5.2 to a more tractable one.

Lemma 5.2 *1. The k-delay estimator* (5.14) *satisfies*

$$\hat{\mathbf{x}}_t^k = A\hat{\mathbf{x}}_{t-1}^{k-1}.$$

2. The k-delay and $(k+1)$-delay estimation errors satisfy

$$\hat{\mathbf{e}}_{t+1}^{k+1} = A\hat{\mathbf{e}}_t^k + \mathbf{w}_t.$$

3. The τ-delay estimation error satisfies

$$\hat{\mathbf{e}}^{\tau}_{t+\tau} = A^{\tau}\tilde{\mathbf{e}}_t + \sum_{s=0}^{\tau-1} A^{\tau-s-1}\mathbf{w}_{t+s} + \begin{cases} A^{\tau}\mathbf{e}^1_t, & w.p. \quad 1-(1-p)\mathbf{u}_t, \\ 0, & w.p. \quad (1-p)\mathbf{u}_t. \end{cases} \tag{5.17}$$

4. The probability density transition function from $\mathbf{e}^1_t$ to $\mathbf{e}^1_{t+1}$ satisfies

$$\mathbf{e}^1_{t+1} = \begin{cases} A\mathbf{e}^1_t + L(CA\tilde{\mathbf{e}}_t + C\mathbf{w}_t + \mathbf{v}_{t+1}), & w.p. \quad 1-(1-p)\mathbf{u}_t, \\ L(CA\tilde{\mathbf{e}}_t + C\mathbf{w}_t + \mathbf{v}_{t+1}), & w.p. \quad (1-p)\mathbf{u}_t. \end{cases} \tag{5.18}$$

Proof. The proof is mostly technical. See Appendix B.2.1. ■

We note that at time t, the local node knows $\mathbf{e}^1_t$, which is the difference between the Kalman filter state $\tilde{\mathbf{x}}_t$ and the 1-delay estimator state $\hat{\mathbf{x}}^1_t$. We also note that $\tilde{\mathbf{e}}_t$, $\mathbf{w}_t$, and $\mathbf{v}_{t+1}$ are mutually independent Gaussian variables, as seen in (5.10). Let the probability density function (*pdf*) of $L(CA\tilde{\mathbf{e}}_t + C\mathbf{w}_t + \mathbf{v}_{t+1})$ be f, *i.e.*,

$$f(w) := \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big\{ -\frac{w'\Sigma^{-1}w}{2}\Big\}, \qquad \forall w \in \mathbb{R}^n, \tag{5.19}$$

where the covariance matrix,

$$\Sigma = L(CA\Sigma_L A'C' + C\Sigma_w C' + \Sigma_v)L'.$$

Remark 5.1 *The Kalman gain L may not be full rank (or even a square matrix). If the covariance matrix Σ is not full rank, the corresponding Gaussian random process is* degenerated*, which is equivalent to a Gaussian variable with a lower dimension. All the arguments hold if (A, L) is controllable, in which we consider n-step transition instead of 1-step transition [15]. We omit the details for this case.*

5.2.6 Optimal communication problem

From now on we drop the superscript 1 in $\mathbf{e}_t^1$ for brevity.

Lemma 5.3 *Define*

$$\rho^2 := \operatorname{trace}\left(A'^{\tau} Q A^{\tau} \Sigma_L\right) + \sum_{s=0}^{\tau-1} \operatorname{trace}\left(A'^{\tau-s-1} Q A^{\tau-s-1} \Sigma_w\right).$$

Then the criterion (5.16) *can be written as*

$$J_\pi(e) = \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} c(\mathbf{e}_t, \mathbf{u}_t) | \pi, \mathbf{e}_0 = e\right],$$

where

$$c(e,u) = \left(1-(1-p)u\right) e' A'^{\tau} Q A^{\tau} e + \lambda u + \rho^2. \tag{5.20}$$

Proof. The proof is technical. See Appendix B.2.2. ■

For this reason, $c(\cdot,\cdot)$ is referred to as the *per-stage cost* function of the optimization Problem 5.2. We define the last optimization problem as follows:

Problem 5.3 *Consider the stochastic discrete dynamic system* (5.18)*, which has probability density transition kernel*

$$P(dy|e,u) = \big[(1-p)uf(y) + \left(1-(1-p)u\right) f(y-Ae)\big]dy. \tag{5.21}$$

Find a policy $\pi = \{\gamma_0, \gamma_1, \cdots\}$*, where* $\gamma_t : \mathbb{R}^n \mapsto [0,1]$*, that minimizes the cost function*

$$J_\pi(e) = \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} c\left(\mathbf{e}_t, \gamma_t(\mathbf{e}_t)\right) | \pi, \mathbf{e}_0 = e\right],$$

where the per-stage cost c *is defined in* (5.20)*.*

Remark 5.2 *We can modify the problem statement slightly so that the feasible policy set is* $[0, p_{max}]$ $(0 < p_{max} < 1)$ *by replacing* $(1-p)$ *with* $p_{max}(1-p)$ *in* (5.21). *As we will see, changing the feasible set from* $[0, p_{max}]$ *to* $[p_{min}, p_{max}]$ $(0 < p_{min} < 1)$ *does not cause any additional difficulty.*

The optimal cost function is defined as,

$$J^*(e) = \min_{\pi \in \Pi} J_\pi(e), \quad e \in \mathbb{R}^n, \tag{5.22}$$

where Π is a feasible set. A policy that achieves the optimal cost is denoted by π^*. A policy π in the form of $(\gamma, \gamma, \cdots)$ is called stationary, in which γ is also referred to as a *policy*. The corresponding cost function is denoted by $J_\gamma(e)$. A stationary policy that achieves the optimal cost (5.22) is denoted by γ^*.

We will solve the optimal communication Problem 5.3 for networks that do not drop packets, in Section 5.3. For lossy networks, we give both the upper and the lower bounds for the optimal solution, in Section 5.4.

5.3 Lossless networks

We state the main results on optimal communication schedulers for lossless networks in the first part of this section, which involves an average cost optimality equality (ACOE). In the second part, we give an iterative procedure to solve the ACOE and prove that the procedure converges exponentially.

5.3.1 Main results

When the channel is lossless ($p = 0$), we can solve the optimal communication problem exactly. We consider an admissible policy set Π_δ, consisting of policies π with the properties that

$$\gamma_t(e) \in \begin{cases} \{0\}, & \text{if} \quad e = 0, \\ [0,1], & \text{if} \quad 0 < e' A'^{\tau} Q A^{\tau} e < \delta, \\ \{1\}, & \text{if} \quad e' A'^{\tau} Q A^{\tau} e \geq \delta, \end{cases} \tag{5.23}$$

where δ is some large positive constant.

Remark 5.3 *The restriction of admissible policy set* (5.23) *is for technical reasons. For the admissible policy set* (5.23)*, if* $e' A'^{\tau} Q A^{\tau} e \geq \delta$ (δ *very large*)*, the communication is triggered* ($\gamma_t(e) = 1$)*. This leads to a type of* ergodicity *property, which is discussed in the proof of Lemma 5.4. The admissible set becomes less restrictive when* δ *is larger. We conjecture that there exists* $\Delta > 0$ *such that for any* $\delta_1 > \delta_2 > \Delta$*,*

$$J^*_{\delta_1} = J^*_{\delta_2}, \quad \pi^*_{\delta_1} = \pi^*_{\delta_2},$$

where $J^*_{\delta_i}$ *and* $\pi^*_{\delta_i}$*,* $i \in \{1, 2\}$*, are the optimal cost and optimal policy, respectively. This was observed in numerical calculation and proved for one-dimensional systems. But it remains to be proved for general cases.*

Let $h : \mathbb{R}^n \to \mathbb{R}$ be a function bounded below. Define the following average cost optimality equality (ACOE),

$$Th(e) = h(e) + J,$$

where the dynamic programming operator T is defined as

$$Th(e) = \min_{u=\gamma(e)\in\Pi_\delta(e)} (1-u)e'A'^{\tau}QA^{\tau}e + \lambda u + \rho^2 + \int_{\mathbb{R}^n} h(y)[uf(y) + (1-u)f(y-Ae)]\mathrm{d}y. \tag{5.24}$$

We have the following theorem for lossless networks:

Theorem 5.1 *1. There exist a unique constant J and a unique bounded function $h : \mathbb{R}^n \to \mathbb{R}$ such that*

$$h(e) + J = (Th)(e), \quad \forall e \in \mathbb{R}^n \tag{5.25}$$

with $h(0) = 0$.

2. The optimal average cost

$$J = \rho^2 + \mathrm{E}[h(\mathbf{w})]. \tag{5.26}$$

3. The optimal communication policy is deterministic, stationary, and given by

$$\gamma^*(e) = \begin{cases} 0, & e'A'^{\tau}QA^{\tau}e < \delta,\ \mathrm{E}[h(Ae+\mathbf{w})] - \mathrm{E}[h(\mathbf{w})] < \lambda - e'A'^{\tau}QA^{\tau}e, \\ 1, & \textit{otherwise.} \end{cases}$$

Remark 5.4 *The function $h(e)$ can be seen as a* cost potential *[4]. One may choose to pay $h(e)$ to "stop" the process now and continue paying an average cost of J per time step in the future. Alternatively, one can play according to the optimal policy $\gamma^*(e)$ and pay a per-stage cost given by* (5.20)*. The two strategies result in the same average cost.*

The first part of Theorem 5.1 is proved by span-contraction arguments. We start by introducing two technical lemmas. Lemma 5.4 is on the *ergodicity* of the state transition kernel $P(\cdot|e,u)$. Lemma 5.5 states that T is a contraction mapping, which allows the use of a fixed-point theorem.

Let $(\mathbb{R}^n, \mathcal{B})$ be a measurable space and μ be a signed measure defined on the sigma algebra $\mathcal{B}$ [42]. Define the *total variation* of μ as,

$$\|\mu\|_{TV} := \sup_{B\in\mathcal{B}} \mu(B) - \inf_{B\in\mathcal{B}} \mu(B).$$

Lemma 5.4 (Ergodic condition) *There exists a constant $\beta \in (0,1)$ such that*

$$\|P(\cdot|e,u) - P(\cdot|\tilde{e},\tilde{u})\|_{TV} \leq 2\beta,$$

for every $e, \tilde{e} \in \mathbb{R}^n$, $u = \gamma(e) \in \Pi_\delta(e)$, $\tilde{u} = \gamma(\tilde{e}) \in \Pi_\delta(\tilde{e})$.

Proof.[Lemma 5.4] The signed measure $P(\cdot|e,u) - P(\cdot|\tilde{e},\tilde{u})$ satisfies,

$$\sup_B [P(B|e,u) - P(B|\tilde{e},\tilde{u})] + \inf_B [P(B|e,u) - P(B|\tilde{e},\tilde{u})] = 0,$$

from which one concludes that

$$\begin{aligned}\|P(\cdot|e,u) - P(\cdot|\tilde{e},\tilde{u})\|_{TV} &= 2\sup_B [P(B|e,u) - P(B|\tilde{e},\tilde{u})] \\ &= \int_{\mathbb{R}^n} |p(y|e,u) - p(y|\tilde{e},\tilde{u})| \mathrm{d}y. \qquad (5.27)\end{aligned}$$

The second equality above follows from Scheffe's Theorem [12], and p is the transition probability density function of $P(\cdot|e,u)$ in (5.21). From (5.27) we have

$$\begin{aligned}\|P(\cdot|e,u) - P(\cdot|\tilde{e},\tilde{u})\|_{TV} &= \int_{\mathbb{R}^n} p(y|e,u) + p(y|\tilde{e},\tilde{u}) - 2\min\{p(y|e,u), p(y|\tilde{e},\tilde{u})\} \mathrm{d}y \\ &= 2 - 2\int_{\mathbb{R}^n} \min\{p(y|e,u), p(y|\tilde{e},\tilde{u})\} \mathrm{d}y. \qquad (5.28)\end{aligned}$$

Because $u = \gamma(e) \in \Pi_\delta(e)$ and $\tilde{u} = \gamma(\tilde{e}) \in \Pi_\delta(\tilde{e})$,

$$p(y|e,u) = (1-p)f(y) + pf(y - Ae) \qquad \text{if} \quad e'A'^{\tau}QA^{\tau}e > \delta$$
$$p(y|\tilde{e},\tilde{u}) = (1-p)f(y) + pf(y - A\tilde{e}) \qquad \text{if} \quad \tilde{e}'A'^{\tau}QA^{\tau}\tilde{e} > \delta.$$

We consider three cases separately for (5.28),

$$\begin{aligned}
&\|P(\cdot|e,u) - P(\cdot|\tilde{e},\tilde{u})\|_{TV} \\
&\overset{(a)}{=} 2 - 2 \begin{cases}
\int_{\mathbb{R}^n} \min\{(1-p)f(y) + pf(y-Ae), (1-p)f(y) + pf(y-Ae)\}\mathrm{d}y, \\
\qquad \text{if} \quad e'A'^{\tau}QA^{\tau}e \geq \delta,\ \tilde{e}'A'^{\tau}QA^{\tau}\tilde{e} \geq \delta \\
\int_{\mathbb{R}^n} \min\{(1-p)uf(y) + (1-(1-p)u)f(y-Ae), f(y)\}\mathrm{d}y, \\
\qquad \text{if} \quad e'A'^{\tau}QA^{\tau}e < \delta,\ \tilde{e}'A'^{\tau}QA^{\tau}\tilde{e} \geq \delta \\
\int_{\mathbb{R}^n} \min\{(1-p)uf(y) + (1-(1-p)u)f(y-Ae), \\
\qquad\quad (1-p)\tilde{u}f(y) + (1-(1-p)\tilde{u})f(y-A\tilde{e})\}\mathrm{d}y, \\
\qquad \text{if} \quad e'A'^{\tau}QA^{\tau}e < \delta,\ \tilde{e}'A'^{\tau}QA^{\tau}\tilde{e} < \delta
\end{cases} \\
&\overset{(b)}{\leq} 2 - 2 \begin{cases}
1, & \text{if} \quad e'A'^{\tau}QA^{\tau}e \geq \delta, \tilde{e}'A'^{\tau}QA^{\tau}\tilde{e} \geq \delta \\
\int_{\mathbb{R}^n} \min\{f(y-Ae), f(y)\}\mathrm{d}y, & \text{if} \quad e'A'^{\tau}QA^{\tau}e < \delta,\ \tilde{e}'A'^{\tau}QA^{\tau}\tilde{e} \geq \delta \\
\int_{\mathbb{R}^n} \min\{f(y), f(y-Ae)\}\mathrm{d}y, & \text{if} \quad e'A'^{\tau}QA^{\tau}e < \delta, \tilde{e}'A'^{\tau}QA^{\tau}\tilde{e} < \delta
\end{cases} \\
&\overset{(c)}{\leq} 2 - 2\int_{\mathbb{R}^n} \min\{f(y), f(y-Ae)\}\mathrm{d}y, \quad \forall\ e'A'^{\tau}QA^{\tau}e < \delta,
\end{aligned}$$

where (a) is from (5.27), (b) is because $(1-p)uf(y) + (1-(1-p)u)f(y-Ae)$ is a linear combination of $f(y)$ and $f(y-Ae)$, and (c) is because

$$\int_{\mathbb{R}^n} \min\{f(y), f(y-Ae)\}\mathrm{d}y \leq 1.$$

For $e'A'^{\tau}QA^{\tau}e < \delta$, there exists $\epsilon > 0$ and a measurable set B with Lebesgue measure $\mu(B) > 0$, such that for all $y \in B$, $f(y) > \epsilon$, $f(y-Ae) > \epsilon$, and

$\epsilon\mu(B) < 1$. Therefore,

$$\|P(\cdot|e,u) - P(\cdot|\tilde{e},\tilde{u})\|_{TV} \leq 2 - 2\int_B \min\{f(y), f(y-Ae)\}\mathrm{d}y \leq 2 - 2\epsilon\mu(B),$$

which completes the proof by setting $\beta = 1 - \epsilon\mu(B)$. Note that the feasible set (5.23) circumvents the difficulty by forcing $u = 1$ when e is large. ■

The *span semi-norm* of a bounded function $h : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\mathrm{span}\{h\} := \sup_y h(y) - \inf_y h(y).$$

Lemma 5.5 *The dynamic programming operator T in* (5.24) *is a span-contraction,* i.e., *for any bounded functions $g(\cdot)$ and $h(\cdot)$, $\mathrm{span}\{Tg - Th\} \leq \beta\,\mathrm{span}\{g - h\}$, where $\beta \in [0, 1)$ is given by Lemma 5.4.*

Proof. This result follows directly from Lemma 3.5 in [15, pp 59]. ■

With the last two lemmas, we are now ready to prove Theorem 5.1:

Proof.[Theorem 5.1] From Lemma 5.5 we conclude that the operator T is a span-contraction and therefore, by Banach's Fixed-Point Theorem, there exists a bounded function $h(\cdot)$ that is a fixed-point in the span sense, *i.e.*, $\mathrm{span}\{(Th)(e) - h(e)\} = 0$. To complete the proof for part one, we choose

$$h(e) := h(e) - h(0).$$

By setting $e = 0$ in (5.25) and using (5.24), we obtain (5.26). It remains to prove that J is the minimum per-stage cost and γ^* is the optimal communication logic. The proof is similar to the one in [5, pp 191], but we are dealing with a Borel state space.

Let $\pi = \{\gamma_0, \gamma_1, \cdots\} \in \Pi_\delta$, and N be a positive integer. From (5.25),

$$T_{\gamma_{N-1}} h \geq h + J.$$

Apply $T_{\gamma_{N-2}}$ to both sides of the equation above,

$$\begin{aligned} T_{\gamma_{N-2}} T_{\gamma_{N-1}} h &\geq T_{\gamma_{N-2}}(h + J) \\ &= T_{\gamma_{N-2}} h + J \\ &\geq h + 2J. \end{aligned}$$

Continuing this procedure, we have

$$T_{\gamma_0} \cdots T_{\gamma_{N-1}} h(e) \geq h(e) + NJ,$$

the right side of which equals the N-stage cost corresponding to initial state e, policy π, and the terminal cost function h, *i.e.*,

$$\mathrm{E}\left[h(\mathbf{e}_N) + \sum_{t=0}^{N-1} c(\mathbf{e}_t, \mathbf{u}_t(\mathbf{e}_t))|\mathbf{e}_0 = e, \pi\right] = T_{\gamma_0} \cdots T_{\gamma_{N-1}} h(e).$$

Then we have

$$\frac{1}{N}\,\mathrm{E}\left[h(\mathbf{e}_N) + \sum_{t=0}^{N-1} c(\mathbf{e}_t, \mathbf{u}_t(\mathbf{e}_t))|\mathbf{e}_0 = e, \pi\right] \geq \frac{1}{N} h(e) + J.$$

Since h is bounded, letting $N \to \infty$, we obtain

$$J_\pi(e) \geq J,$$

which means J is the minima. When $\pi = \gamma^*$, equalities hold in all the inequalities above. We conclude that γ^* minimizes J_π, *i.e.*,

$$J_{\gamma^*} = J.$$

■

5.3.2 Value iteration

Eq. (5.25) can be solved efficiently using a *value iteration algorithm* defined as follows:

1. $h^0(e) = 0$, $\forall e \in \mathbb{R}^n$;

2. The sequences of functions $h^k : \mathbb{R}^n \to \mathbb{R}$ and $J^k : \mathbb{R}^n \to \mathbb{R}$ ($\forall\ k \geq 0$) are defined by the following iteration:

$$\begin{aligned} \bar{h}^{k+1} &:= Th^k, \\ J^{k+1} &:= \bar{h}^{k+1} - h^k, \\ h^{k+1}(e) &:= \bar{h}^{k+1}(e) - \bar{h}^{k+1}(0), \quad \forall e \in \mathbb{R}^n. \end{aligned} \tag{5.29}$$

The value iteration scheme is exponentially convergent, as stated in the following theorem:

Theorem 5.2 *The sequence (J^k, h^k) converges exponentially fast to the pair (J, h). In particular,*

$$\|h^k - h\|_\infty \leq \beta^k \cdot \text{span}\{h\}, \tag{5.30}$$

$$\|J^k - J\|_\infty \leq 3\beta^{k-1} \cdot \text{span}\{h\}, \tag{5.31}$$

where $\|\cdot\|_\infty$ denotes the sup-norm of a function.

Proof.[Theorem 5.2] We start by noting for any given function $g : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned} \|g\|_\infty &\leq \|g - g(x)\|_\infty + |g(x)| \\ &\leq \text{span}\{g\} + |g(x)|, \qquad \forall x \in \mathbb{R}^n. \end{aligned} \tag{5.32}$$

From (5.25), $\operatorname{span}\{h\} = \operatorname{span}\{Th\}$. Therefore, we have that

$$\begin{aligned}\operatorname{span}\{h^k - h\} &= \operatorname{span}\{Th^{k-1} - Th\} \leq \beta \operatorname{span}\{h^{k-1} - h\} \\ &\leq \beta^k \operatorname{span}\{h^0 - h\} = \beta^k \operatorname{span}\{h\},\end{aligned}$$

from which the inequality (5.30) follows using (5.32) and the fact that $h^k(0) = h(0) = 0$. To prove (5.31), we use the fact that

$$\begin{aligned}\operatorname{span}\{J^k - J\} &= \operatorname{span}\{Th^{k-1} - h^{k-1} - \big(Th - h\big)\} \\ &= \operatorname{span}\{Th^{k-1} - Th - \big(h^{k-1} - h\big)\} \\ &\leq \operatorname{span}\{Th^{k-1} - Th\} + \operatorname{span}\{h^{k-1} - h\} \\ &\leq (\beta^k + \beta^{k-1}) \operatorname{span}\{h\}. \end{aligned} \tag{5.33}$$

Moreover, we conclude from the definitions of J^k and T, and (5.26) that

$$\begin{aligned}J^k(0) &= (Th^{k-1})(0) - h^{k-1}(0) \\ &= \rho^2 + \mathrm{E}[h^{k-1}(\mathbf{w})] = J + \mathrm{E}[h^{k-1}(\mathbf{w}) - h(\mathbf{w})],\end{aligned}$$

and therefore,

$$|J^k(0) - J| = |\,\mathrm{E}[h^{k-1}(\mathbf{w}) - h(\mathbf{w})]| \leq \beta^{k-1} \operatorname{span}\{h(\cdot)\},$$

from which the inequality (5.31) follows, together with (5.32) and (5.33). ■

5.4 Erasure networks

In this section, we state the main results for Problem 5.3, in which the network drops data packets with probability p, by providing both an upper bound and a lower bound to the optimal solution. We obtain the upper bound by solving an

average cost optimality inequality (ACOI), while getting the lower bound from an average cost optimality equality (ACOE).

Let $h : \mathbb{R}^n \to \mathbb{R}$ be a function bounded below. Define the following ACOI,

$$T_1 h(e) \leq h(e) + J, \tag{5.34}$$

where the dynamic programming operator T_1 is defined by,

$$T_1 h(e) = \min_{u=\gamma(e)\in\Pi_\delta(e)} (1-(1-p)u)e' A'^{\tau} Q A^{\tau} e + \lambda u + \rho^2 \tag{5.35}$$

$$+ \int_{\mathbb{R}^n} h(y)[(1-p)uf(y) + (1-(1-p)u)f(y-Ae)]\mathrm{d}y. \tag{5.36}$$

We define the following ACOE,

$$T_2 h(e) = h(e) + J, \tag{5.37}$$

where the dynamic programming operator T_2 is defined by,

$$T_2 h(e) = \min_{u=\gamma(e)\in\Pi_\delta(e)} (1-u)e' A'^{\tau} Q A^{\tau} e + \lambda u + \rho^2 \tag{5.38}$$

$$+ \int_{\mathbb{R}^n} h(y)[(1-p)uf(y) + (1-(1-p)u)f(y-Ae)]\mathrm{d}y.$$

5.4.1 Main results

The following results provide bounds for the optimal communication Problem 5.3:

Theorem 5.3 *If the packet dropout probability p satisfies*

$$p < \frac{1}{\max |\operatorname{eig}(A)|^{2\tau}},$$

the optimal communication Problem 5.3 has the following upper and lower bounds:

1. *The ACOI* (5.34) *has a solution* $\left(\bar{h}, \bar{J}\right)$*, where*

$$\bar{h}(e) = e'Ke,$$
$$\bar{J} = \lambda + \rho^2 + \text{trace}(K\Sigma), \tag{5.39}$$

in which $K = 0$ *if* $p = 0$*, and* K *is the unique positive definite solution to the discrete Lyapunov equation,*

$$pA'^{\tau}KA^{\tau} - K + pA'^{\tau}QA^{\tau} = 0, \tag{5.40}$$

if $0 < p < \frac{1}{\max|\,\text{eig}(A)|^{2\tau}}$*. The optimal cost function* J^* *is upper bounded by*

$$J^* \leq \bar{J}. \tag{5.41}$$

2. *The ACOE* (5.37) *has the unique solution* $(\underline{h}(e), \underline{J})$*. The optimal cost function* J^* *is lower bounded by*

$$J^* \geq \underline{J}. \tag{5.42}$$

Proof.[Theorem 5.3] We prove the upper bound in (5.41) in Section 5.4.2 and the lower bound (5.42) in Section 5.4.3. ■

5.4.2 Upper bound

Proof.[Theorem 5.3 Part 1] It is well known that the Lyapunov equation (5.40) has a positive definite solution K if $\sqrt{p}A^{\tau}$ is stable, *i.e.*,

$$p < \frac{1}{\max|\,\text{eig}(A)|^{2\tau}}.$$

For the *pdf* $f(\cdot)$ as in (5.19), we have that

$$\int_{\mathbb{R}^n} y'Kyf(y)\mathrm{d}y = \mathrm{trace}(K\Sigma),$$
$$\int_{\mathbb{R}^n} y'Kyf(y-Ae)\mathrm{d}y = \mathrm{trace}(K\Sigma) + e'A'KAe.$$

Now we verify that (5.39) solves the ACOI (5.34) by showing

$$\begin{aligned}
&T_1\bar{h}(e)\\
&= T_1(e'Ke)\\
&= \min_{u=\gamma(e)\in\Pi_\delta(e)} (1-(1-p)u)\, e'A'^{\tau}QA^{\tau}e + \lambda u + \rho^2 + (1-p)u\,\mathrm{trace}(K\Sigma)\\
&\qquad\qquad + (1-(1-p)u)\,(\mathrm{trace}(K\Sigma) + e'A'^{\tau}KA^{\tau}e)\\
&= \min_{u=\gamma(e)\in\Pi_\delta(e)} (\lambda - (1-p)e'A'^{\tau}(Q+K)A^{\tau}e)u\\
&\qquad\qquad + e'A'^{\tau}QA^{\tau}e + \rho^2 + \mathrm{trace}(K\Sigma) + e'A'^{\tau}KA^{\tau}e.
\end{aligned}$$

We choose the parameter δ in (5.23) so that if $e'A'^{\tau}QA^{\tau}e \geq \delta$,

$$e'Ke \geq \lambda\frac{p}{1-p}.$$

This choice of δ ensures that the policies are in the admissible set (5.23) for both cases that we will discuss next. We consider two cases:

- If $e'Ke \geq \lambda\frac{p}{1-p}$, from (5.40), we have $(1-p)e'A'^{\tau}(Q+K)A^{\tau}e \geq \lambda$, and the

minimizing policy is $u = \gamma(e) = 1$. We have that

$$\begin{aligned} T_1\bar{h}(e) &= \lambda - (1-p)e'A'^{\tau}QA^{\tau}e - (1-p)e'A'^{\tau}KA^{\tau}e + e'A'^{\tau}QA^{\tau}e \\ &\quad + \rho^2 + \mathrm{trace}(K\Sigma) + e'A'^{\tau}KA^{\tau}e \\ &= e'A'^{\tau}\left(Q + K - (1-p)Q - (1-p)K\right)A^{\tau}e + \lambda + \rho^2 + \mathrm{trace}(K\Sigma) \\ &= pe'A'^{\tau}(Q+K)A^{\tau}e + \lambda + \rho^2 + \mathrm{trace}(K\Sigma) \\ &= e'A'^{\tau}KA^{\tau}e + \lambda + \rho^2 + \mathrm{trace}(K\Sigma) \\ &=: \bar{h}(e) + \bar{J}. \end{aligned}$$

- If $e'Ke < \lambda\frac{p}{1-p}$, from (5.40), we have $(1-p)e'A'^{\tau}(Q+K)A^{\tau}e < \lambda$, and the minimizing policy is $u = \gamma(e) = 0$. We have that

$$\begin{aligned} &T_1\bar{h}(e) \\ &= e'A'^{\tau}QA^{\tau}e + \rho^2 + \mathrm{trace}(K\Sigma) + e'A'^{\tau}KA^{\tau}e \\ &= pe'A'^{\tau}(Q+K)A^{\tau}e + \lambda + \rho^2 + \mathrm{trace}(K\Sigma) + (1-p)e'A'^{\tau}(Q+K)A^{\tau}e - \lambda \\ &=: \bar{h}(e) + \bar{J} + (1-p)e'A'^{\tau}(Q+K)A^{\tau}e - \lambda \\ &\leq \bar{h}(e) + \bar{J}. \end{aligned}$$

Now we prove (5.41). Let $\bar{u} = \bar{\gamma}(e)$ be minimizing in (5.35) for $\bar{h}(e)$, *i.e.*,

$$T_1\bar{h}(e) = c(e,\bar{u}) + \int_{\mathbb{R}^n} \bar{h}(y)P(\mathrm{d}y|e,\bar{u}),$$

where

$$c(e,\bar{u}) := (1-(1-p)\bar{u})e'A'^{\tau}QA^{\tau}e + \lambda\bar{u} + \rho^2,$$

and

$$P(\mathrm{d}y|e,\bar{u}) := [(1-p)\bar{u}f(y) + (1-(1-p)\bar{u})f(y-Ae)]\mathrm{d}y.$$

Because $c(e,u) \geq 0$, $\forall e, u$, we have that

$$\begin{aligned}\int_{\mathbb{R}^n} \bar{h}(y) P(\mathrm{d}y|e,\bar{\gamma}) &= T_1\bar{h}(e) - c(e,\bar{u}) \\ &\leq T_1\bar{h}(e) \\ &\leq \bar{h} + \bar{J},\end{aligned}$$

from which we have, for any $N > 0$,

$$\mathrm{E}[\bar{h}(\mathbf{e}_N)|\mathbf{e}_0 = e, \bar{\gamma}] \leq \bar{h}(e) + N\bar{J} < \infty.$$

From the equation above, we have that

$$\begin{aligned}\mathrm{E}\left[\sum_{t=0}^{N-1} c(\mathbf{e}_t, u(\mathbf{e}_t))|\mathbf{e}_0 = e, \bar{\gamma}\right] &\leq \bar{h}(e) + N\bar{J} - \mathrm{E}[\bar{h}(\mathbf{e}_N)] \\ &\leq \bar{h}(e) + N\bar{J},\end{aligned}$$

in which the last inequality is because $\bar{h}(e) \geq 0$. The proof of (5.41) is completed by dividing both sides of the inequality above by N and letting $N \to \infty$. ■

5.4.3 Lower bound

Now we prove the last part of Theorem 5.3, which provides a lower bound for the optimal solution.

Proof.[Theorem 5.3 Part 2] The per-stage cost and the state probability transition kernel for the dynamic programming operator T_2, defined in (5.38), are

$$c(e,u) = (1-u)e'A'^{\tau}QA^{\tau}e + \lambda u + \rho^2,$$

and

$$P(dy|e,u) = \big[(1-p)uf(y) + (1-(1-p)u)\, f(y - Ae)\big]dy,$$

respectively. Following the proof of Lemma 5.4, we can show that the probability transition kernel P satisfies the ergodic condition. Therefore (5.37) has a unique solution, denoted by $(\underline{h}(e), \underline{J})$. From the definitions of T_1 and T_2, we know that

$$T_1 h(e) \geq T_2 h(e), \quad \forall\ h(\cdot), e \in \mathbb{R}^n.$$

Following the proof of Theorem 5.1, we have

$$T_1^N \underline{h}(e) \geq T_2^N \underline{h}(e) \geq \underline{h}(e) + N\underline{J}, \quad N > 0,$$

the right side of which equals the N-stage cost corresponding to initial state e, policy π, and the terminal cost function $\underline{h}$, *i.e.*,

$$\mathrm{E}\left[\underline{h}(\mathbf{e}_N) + \sum_{t=0}^{N-1} c(\mathbf{e}_t, \mathbf{u}_t(\mathbf{e}_t))|\mathbf{e}_0 = e, \pi\right] = T_1^N \underline{h}(e).$$

Then we have

$$\frac{1}{N}\mathrm{E}\left[\underline{h}(\mathbf{e}_N) + \sum_{t=0}^{N-1} c(\mathbf{e}_t, \mathbf{u}_t(\mathbf{e}_t))|\mathbf{e}_0 = e, \pi\right] \geq \frac{1}{N}\underline{h}(e) + \underline{J}.$$

Since $\underline{h}$ is bounded, letting $N \to \infty$, we obtain,

$$J_\pi(e) \geq \underline{J},$$

from which we conclude that $J^* \geq \underline{J}$. ■

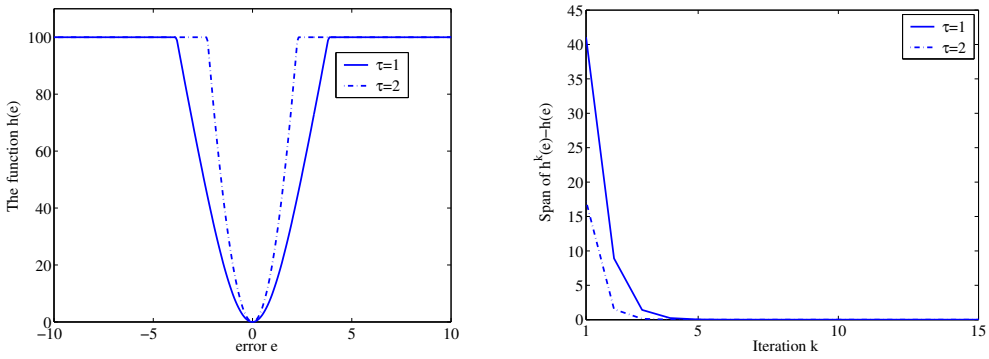
5.5 Numerical examples

In this section, we apply Theorems 5.1-5.3 numerically to a one-dimensional plant (5.2) with $A = 2$, $C = 1$, $\Sigma_w = 1$, and $\Sigma_v = 1$. In the per-stage cost function (5.20), $Q = 1$, and $\lambda = 100$. In the admissible policy set (5.23), $\delta = 10$.

For a lossless network ($p = 0$), Theorem 5.1 gives an exact solution $(h(e), J^*)$ and the optimal communication scheduling method. When $\tau = 1$, we obtain $J^* = 69.73$; and when $\tau = 2$, we obtain $J^* = 140.24$. The functions $h(\cdot)$ is plotted in Figure 5.3(a), which shows that $h(\cdot)$, as a "cost potential" (cf. Remark 5.4), is larger for $\tau = 2$ than for $\tau = 1$. The optimal communication policy

$$\gamma(e) = \begin{cases} 1 & \text{if } |e| > 2.25 \\ 0 & \text{otherwise,} \end{cases}$$

when $\tau = 1$, and the threshold become 3.80 (instead of 2.25) if $\tau = 2$. This shows that our choice of $\delta = 10$ for the admissible policy set (5.23) is not binding. Figure 5.3(b) shows that the iterative method is convergent exponentially.



(a) (b)

Figure 5.3. Numerical solution for a lossless network. In (a) Solutions of $h(e)$ to (5.25). (b) The iterative scheme converges exponentially.

For a network that drops packets with probability $p = 0.05$ and a delay $\tau = 2$, the Lyapunov equation (5.40) holds with $K = 4$. The lower and upper bounds of the optimal cost function are $\underline{J} = 141.06$ and $\bar{J} = 222.37$, respectively. Figure 5.5 shows function $\underline{h}(e)$ (solution to the ACOE (5.37)) and the function $\bar{h}(e) = e'Ke$.

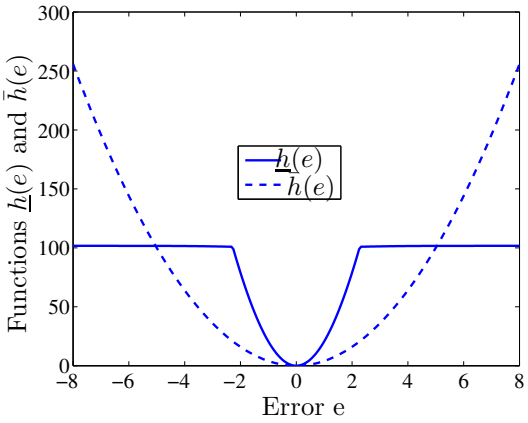


Figure 5.4. The functions $\underline{h}(e)$ and $\bar{h}(e)$ as in Theorem 5.3 for a network that drops packets with probability $p = 0.05$.

5.6 Summary

The chapter is on the design of optimal communication scheduling methods for the remote estimation of discrete-time LTI systems. The objective is to find optimal communication scheduler to minimize the estimation error variance and the average packet rate. We started from stochastic communication schedulers and formulated a structured communication-estimation scheme. For networks that do not drop packets, we solved the optimal communication scheduling problem via dynamic programming. The optimal scheduling methods turns out to be deterministic. For networks that drop packets, we obtained both upper and lower bounds to the optimal solution.

Chapter 6

Summary and future work

The dissertation work focuses on data communication scheduling methods for estimation over networks. A communication scheduler is a mechanism that determines whether or not a node sends data to the network given the information pattern available. We focused on stochastic communication schedulers, in which decisions to send data are based on a counting process (for continuous-time systems) or a random draw (for discrete-time systems). We proposed, analyzed, and optimized communication schedulers in the context of estimation over networks.

We gave a unified answer to both the bit rate and the packet rate requirements. We observed that the estimator is insensitive to packet size once the packet is of a certain size. This observation motivates us to measure communications in terms of *packets*, instead of *bits*.

We proposed stochastic communication schedulers for LTI systems with full or partial state measurements and pursued two objectives: observability and finite network resource usage. We used counting processes to model the schedulers,

jump-diffusion processes to model the dynamics and solve stability problems via stochastic analysis based on infinitesimal generators. We analyzed a large class of stochastic communication schedulers and various types of networks (*e.g.*, networks with delay and networks with packet dropouts). The resulting jump diffusion processes are of theoretical interest on their own in that the jump intensities are dynamically coupled with system dynamics.

We pursued optimal communication schedulers. For networks with no packet dropouts, we solved the optimal communication problem exactly. For general networks with a packet dropout probability p, we obtained both an upper bound and a lower bound of the optimal solution. The upper bound was obtained by solving an average cost optimality inequality (ACOI), and the lower bound was solved by an average cost optimality equality (ACOE). These schedulers can be dynamically implemented in medium access control (MAC) designs to make the communication protocols suitable for NCS applications.

Communication scheduling is an important topic for networked control systems. Future research directions include:

1. The optimal communication scheduler problem for erasure networks is yet to be solved. Average cost optimality equations such as (5.25) and inequalities such as (5.34) may still play a central role. But the results in the dissertation need to be extended in such cases.

2. Robustness issues need to be addressed for applications. Throughout the dissertation, we assumed that the network has mechanisms to detect transmission errors, upon which optimal estimation schemes were developed. What if transmission errors are not detected occasionally? We conjecture

that the optimal estimator on the remote node should take the form of a Kalman filter to make it less sensitive to transmission errors. The analysis needs to be extended to such cases.

3. Communication schedulers are designed for the remote estimation of LTI systems. It remains open for nonlinear systems in most cases. It is conjectured that the Lipschitz constant plays a critical role, because it can be used as a bound for how fast the state of a system can grow [28].

4. It remains to apply and experimentally test the stochastic communication scheduling algorithms in networked control systems.

Bibliography

[1] P. K. Andersen, Ø. Borgan, R. G. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer, New York, 1993.

[2] P. Antsaklis and J. Baillieul. Guest editorial special issue on networked control systems. *IEEE Trans. on Automat. Contr.*, 49(9):1421–1423, September 2004.

[3] David Applebaum. *Lévy Processes and Stochastic Calculus*, volume 93 of *Cambridge studies in advanced mathematics*. Cambridge, UK, 2004.

[4] A. Arapostathis, V.S. Borkar, E. Fernandez-Gaucherand, M.K. Ghosh, and S.I. Marcus. Discrete-time controlled markov processes with average cost criterion: A survey. *SIAM J. Contr. Optimization*, 31(2):282–344, March 1993.

[5] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 2, pages 191–192. Athena Scientific, Belmont, Mass, 1995.

[6] Stephen P. Boyd and Craig H. Barratt. *Linear Controller Design: Limits of Performance*. Prentice-Hall information and system sciences series. Prentice-Hall, 1991.

[7] F. Bullo and D. Liberzon. Quantized control via locational optimization. *IEEE Trans. on Automat. Contr.*, 51(1), January 2006.

[8] Xiaoming Chen and Geok-Soon Hong. A simulation study of the predictive p-persistent CSMA protocol. In *Proc. of the 35th Annual Simulation Symposium*, pages 345–351, April 2002.

[9] Sinem Coleri, Anuj Puri, and Pravin Varaiya. Power efficient system for sensor networks. In *Proc. of the Eighth IEEE International Symposium on Computers and Communication*, volume 2, pages 837–842, 2003.

[10] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, Inc, USA, 1991.

[11] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer-Verlag, 1988.

[12] Luc Devroye and Laszlo Gyorfi. *Nonparametric Density Estimation: The L_1 View*. Wiley, New York, 1985.

[13] Nicola Elia and Sanjoy K. Mitter. Stabilization of linear systems with limited information. *IEEE Trans. on Automat. Contr.*, 46(9):1384–1400, September 2001.

[14] Paul Glasserman and Nocolas Merener. Numerical solution of jump-diffusion LIBOR market models. *Finance and Stochastics*, 7:1–27, 2003.

[15] Onésimo Hernández-Lerma. *Adaptive Markov Control Processes*, volume 79 of *Applied mathematical sciences*. Springer-Verlag, New York, 1989.

[16] João Pedro Hespanha, Margaret L. McLaughlin, and Gaurav Sukhatme. Haptic collaboration over the Internet. In *Proc. of the 5th Phantom Users Group Workshop*, October 2000.

[17] João Pedro Hespanha, Payam Naghshtabrizi, and Yonggang Xu. A survey of recent results in networked control systems. To appear in the Proceedings of IEEE, July 2005.

[18] João Pedro Hespanha, Antonio Ortega, and Lavanya Vasudevan. Towards the control of linear systems with minimum bit-rate. In *Proc. of the Int. Symp. on the Mathematical Theory of Networks and Syst.*, August 2002.

[19] João Pedro Hespanha and Andrew Teel. Stochastic impulsive systems driven by renewal processes. Submitted for publication, December 2005.

[20] K. Hikichi, H. Morino, I. Arimoto, K. Sezaki, and Y. Yasuda. The evaluation of delay jitter for haptics collaboration over the Internet. In *IEEE Global Telecomm. Conf. (GLOBECOM)*, volume 2, pages 1492 – 1496, November 2002.

[21] Hideaki Ishii and Bruce Francis. Stabilizing a linear system by switching control with dwell time. *IEEE Trans. on Automat. Contr.*, 47(12):1962–1973, December 2002.

[22] Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *A series of comprehensive studies in mathematics*. Springer-Verlag, Berlin, 1987.

[23] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, Upper Saddle River, New Jersey, 2000.

[24] Hassan K. Khalil. *Nonlinear Systems*. Prentice Hall, Upper Saddle River, NJ, 1996.

[25] F. Kozin. A survey of stability of stochastic systems. *Automatica*, 5:95–112, 1969.

[26] Harold Kushner. *Heavy Traffic Analysis of Controlled Queueing and Communication Networks*, volume 47 of *Applications of Mathematics*. Springer, Berlin, 2001.

[27] Keyong Li and John Baillieul. Robust quantization for digital finite communication bandwidth (DFCB) control. *IEEE Trans. on Automat. Contr.*, 49(9):1573–1584, September 2004.

[28] Daniel Liberzon and João Pedro Hespanha. Stabilization of nonlinear systems with limited information feedback. *IEEE Trans. on Automat. Contr.*, 50(6):910–915, April 2004.

[29] Nuno C. Martins and Munther A. Dahleh. Feedback control in the presence of noisy channels: "Bode-like" fundamental limitations of performance. Submitted for publication, October 2004.

[30] A. Matveev and A. Savkin. The problem of state estimation via asynchronous communication channels with irregular transmission times. *IEEE Trans. on Automat. Contr.*, 48(4):670–676, April 2003.

[31] Cai Meng, Tianmiao Wang, Wusheng Chou, Sheng Luan, Yuru Zhang, and Zengmin Tian. Remote surgery case: Robot-assisted teleneurosurgery. In *IEEE Int. Conf. on Robot. and Auto. (ICRA'04)*, volume 1, pages 819–823, April 2004.

[32] R. M. Murray, K. J. Åström, S. P. Boyd, R. W. Brockett, and G. Stein. Control in an information rich world. *IEEE Contr. Syst. Mag.*, 23(2):20–33, April 2003.

[33] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran. Topological feedback entropy and nonlinear stabilization. *IEEE Trans. on Automat. Contr.*, 49(9):1585–1597, September 2004.

[34] Girish N. Nair and Robin J. Evans. Exponential stabilisability of finite-dimensional linear systems with limited data rates. *Automatica*, 39(4):585–593, April 2003.

[35] Girish N. Nair and Robin J. Evans. Stabilizability of stochastic linear systems with finite feedback data rates. *SIAM J. Contr. Optimization*, 43(2):413–436, 2004.

[36] D. Nesić and A. Teel. Input-output stability properties of networked control systems. *IEEE Trans. on Automat. Contr.*, 49(10):1650–1667, October 2004.

[37] Petter Ogren, Edward Fiorelli, and Naomi Ehrich Leonard. Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment. *IEEE Trans. on Automat. Contr.*, 49(8):1292–1302, August 2004.

[38] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 2000.

[39] Bernt Øksendal and Agnes Sulem. *Applied Stochastic Control of Jump Diffusions*. Springer, Berlin, 2005.

[40] Vu N. Phat, Jianming Jiang, Andrey V. Savkin, and Ian R. Petersen. Robust stabilization of linear uncertain discrete-time systems via a limited capacity communication channel. *Syst. & Contr. Lett.*, 53(5):347–360, December 2004.

[41] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion.* Springer, Berlin, 1999.

[42] Walter Rudin. *Real and Complex analysis.* McGraw-Hill, 3 edition, 1987.

[43] A. Sahai. *Anytime Information Theory.* PhD thesis, MIT, Cambridge, MA, 2001.

[44] Zeev Schuss. *Theory and Applications of Stochastic Differential Equations.* Wiley Series in Probability and mathematical statistics. John Wiley and Sons, 1980.

[45] Pete Seiler and Raja Sengupta. Analysis of communication losses in vehicle control problems. In *Proc. of the 2001 Amer. Contr. Conf.*, pages 1491–1496, June 2001.

[46] Pete Seiler and Raja Sengupta. An H_∞ approach to networked control. *IEEE Trans. on Automat. Contr.*, 50(3):356–364, March 2005.

[47] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, July 1948.

[48] S. Shirmohammadi and N. Ho Woo. Evaluating decorators for haptic collaboration over the Internet. In *The 3rd IEEE Int. Workshop on Haptic, Audio and Visual Env. and their Applic.*, pages 105–109, October 2004.

[49] Bruno Sinopoli, Luca Schenato, Missimo Franceschetti, Kameshwar Poolla, Michael I. Jordan, and Shankar S. Sastry. Kalman filtering with intermittent observations. *IEEE Trans. on Automat. Contr.*, 49(9):1453–1464, September 2004.

[50] Craig Smith and Peter Seiler. Estimation with lossy measurements: Jump estimators for jump systems. *IEEE Trans. on Automat. Contr.*, 48(12):2163–2171, December 2003.

[51] Roy Smith and Fred Y. Hadaegh. Parallel estimators and communication in spacecraft formations. In *Proc. of the 16th World Congress of Int. Federation of Automat. Contr.*, July 2005.

[52] K. J. Åström. *Introduction to Stochastic Control Theory.* Academic Press, 1970.

[53] K. J. Åström and B. M. Bernhardsson. Comparison of Riemann and Lebesgue sampling for first order stochastic systems. In *Proc. of the 41st Conf. on Decision and Contr.*, volume 2, pages 2011–2016, December 2002.

[54] Andrew Tanenbaum. *Computer Networks.* Prentice Hall, 3rd edition, 1996.

[55] Gongguo Tang and Lei Guo. Adaptive stabilization of unknown linear systems under communication constraints. Submitted to the Chinese Control Conference (in Chinese), March 2005.

[56] S. Tatikonda and S. Mitter. Control over noisy channels. *IEEE Trans. on Automat. Contr.*, 49(7):1196–1201, July 2004.

[57] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Trans. on Automat. Contr.*, 49(7):1056–1068, July 2004.

[58] Wikipedia. `http://www.wikipedia.org/`.

[59] Wing Shing Wong and Roger W. Brockett. Systems with finite communication bandwidth constraints–II: Stabilization with limited information feedback. *IEEE Trans. on Automat. Contr.*, 44(5), May 1999.

[60] Yonggang Xu and João Pedro Hespanha. Communication logics for networked control systems. In *Proc. of the 2004 Amer. Contr. Conf.*, volume 1, pages 572–577, June 2004.

[61] Yonggang Xu and João Pedro Hespanha. Optimal communication logics for networked control systems. In *Proc. of the 43rd Conf. on Decision and Contr.*, pages 3527–3532, December 2004.

[62] Yonggang Xu and João Pedro Hespanha. Estimation under uncontrolled and controlled communications in networked control systems. In *Proc. of the 44th Conf. on Decision and Contr.*, pages 842–847, December 2005.

[63] J. K. Yook, D. M. Tilbury, and N. R. Soparkar. Trading computation for bandwidth: Reducing communication in distributed control systems using state estimators. *IEEE Trans. Contr. Syst. Technol*, 10(4):503–518, July 2002.

Appendix A

Concepts in stochastic processes

In this section, we summarize some concepts related to stochastic processes, especially jump diffusion processes. More details can be found in texts on stochastic processes, *e.g.* [38, 22].

Definition A.1 (Probability space) *Let $\mathcal{F}$ be a σ-field on the space Ω, and $(\Omega, \mathcal{F})$ be a measurable space. The triplet $(\Omega, \mathcal{F}, P)$ is called a* probability space*, where P is a probability measure on $(\Omega, \mathcal{F})$,* i.e.*, $P : \mathcal{F} \to [0, 1]$ such that*

- *$P(\varnothing) = 0$ and $P(\Omega) = 1$;*
- *If F_1, F_2, $\cdots \in \mathcal{F}$ and $\{F_i\}_{i=1}^{\infty}$ is disjoint, $P\left(\cup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P\left(F_i\right)$.*

The subsets F of Ω that belong to $\mathcal{F}$ are called $\mathcal{F}$-measurable sets.

Definition A.2 (Stochastic basis) *A stochastic basis, denoted by $(\Omega, \mathcal{F}, \mathbf{F}, P)$, is a probability space $(\Omega, \mathcal{F}, P)$ equipped with a* filtration $\mathbf{F} = (\mathcal{F}(t))_{t \in \mathbb{R}^+}$*, where*

filtration means increasing and right-continuous family of sub-σ-fields of $\mathcal{F}$, i.e., *$\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$ and $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$.*

Roughly speaking, a σ-algebra defines the set of events that can be measured, or discriminated, and a filtration is the changes in the set of events that can be measured, or the information available at time t.

Definition A.3 (Random process) *A random process is a family $(\mathbf{x}(t))_{t\in\mathbb{R}^+}$ of mappings from Ω into some set E,* i.e., *$\mathbf{x}(\omega, t) : \Omega \times \mathbb{R}^+ \to E$, or $\mathbf{x}_t(\omega) : \Omega \times \mathbb{R}^+ \to E$.*

We only consider $\mathbb{R}^n$-valued random processes.

Definition A.4 (Random measure) *A random measure on $\mathbb{R}_+ \times E$ is a family $\mu = (\mu(w; dt, dx) : w \in \Omega)$ of nonnegative measures on $(\mathbb{R}_+ \times E, \mathcal{R}_+ \times \mathcal{E})$ satisfying $\mu(w; \{0\} \times E) = 0$ identically.*

Definition A.5 (Adapted process) *A random process $\mathbf{x}$ is adapted to the filtration $\mathbf{F}$ if $\mathbf{x}(t)$ is $\mathcal{F}_t$-measurable for every $t \in \mathbb{R}^+$.*

Definition A.6 (Stopping time) *A function $\tau : \Omega \to [0, \infty]$ is called an $\mathcal{F}_t$-stopping time if*

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t, \quad \forall t \geq 0.$$

That is to say, given the information set $\mathcal{F}_t$, one can tell whether $\tau \leq t$ or not.

Definition A.7 (Predictable) *A* predictable *σ-field is a σ-field on $\Omega \times \mathbb{R}_+$ that is generated by all left-continuous adapted process.*

Remark A.1 *A càdlàg process $\mathbf{x}(t)$ is not a predictable process, because it is not left-continuous. However, $\mathbf{y}(t) := \mathbf{x}(t^-)$ is.*

Appendix B

Technical proofs

We prove a few technical results for Chapter 4 in Section B.1 and for Chapter 5 in Section B.2.

B.1 Proof of theorems in Chapter 4

The proof of Theorem 4.5 is similar to that of Theorem 4.2.

Proof.[Theorem 4.5] Since the jump intensity $\lambda(\cdot)$ in (4.31) is dependent on $\tilde{\mathbf{e}}(t^-)$ and it is known that $\xi(t)$ is stochastically bounded, to simplify the proof, take $V := (\tilde{e}' P \tilde{e})^{\frac{m}{2}}$, $P \in \mathbb{R}^{n\times n}$ positive definite. From Lemma 3.1, the generator for the jump diffusion process (4.26a) is given by

$$\begin{aligned}\mathcal{L}V(\tilde{e}) =& \frac{\partial V(\tilde{e})}{\partial \tilde{e}} \cdot \big(A\tilde{e} + LC\xi\big) + \frac{1}{2}\operatorname{trace}\left(\Sigma_{Lv}\frac{\partial^2 V(\tilde{e})}{\partial \tilde{e}^2}\right) \\ &+ \lambda(\tilde{e})\Big(\int V(z)\,\mathrm{d}\mu(z) - V(\tilde{e})\Big),\end{aligned}$$

where $\frac{\partial V(\tilde{e})}{\partial \tilde{e}}$ and $\frac{\partial^2 V(\tilde{e})}{\partial \tilde{e}^2}$ denote the gradient and Hessian matrix of V, respectively,

$\Sigma_{Lv} := L\Sigma_v L'$, and ξ is considered external as given by (4.26b).

For positive definite matrix P, $\exists c_5, c_6 > 0$ such that

$$PA + A'P \leq c_5 P,$$

$$P\Sigma_{Lv}P \leq c_6 P,$$

and $\exists\, c_7 > 0$ such that

$$\tilde{e}'(PLC + C'L'P)\xi \leq c_7(\tilde{e}'P\tilde{e})^{\frac{1}{2}}(\xi'\xi)^{\frac{1}{2}},$$

for any $\tilde{e}$ and $\xi \in \mathbb{R}^n$. Now consider the generator,

$$\begin{aligned}
\mathcal{L}V(\tilde{e}) &= \frac{m}{2}(\tilde{e}'P\tilde{e})^{\frac{m}{2}-1}\,\tilde{e}'(PA + A'P)\tilde{e} + \frac{m}{2}(\tilde{e}'P\tilde{e})^{\frac{m}{2}-1}\,\tilde{e}'(PLC + C'L'P)\xi \\
&\quad + (\tilde{e}'P\tilde{e})^k\ \mathrm{E}^{\mathbf{z}}\left[V(\mathbf{z})\right] + 2\frac{m}{2}(\frac{m}{2}-1)(\tilde{e}'P\tilde{e})^{\frac{m}{2}-2}\,\tilde{e}'P\Sigma_{Lv}P\tilde{e} \\
&\quad + \frac{m}{2}(\tilde{e}'P\tilde{e})^{\frac{m}{2}-1}\,\mathrm{trace}\,(\Sigma_{Lv}P) - (\tilde{e}'P\tilde{e})^{\frac{m}{2}+k} \\
&\leq c_5\frac{m}{2}V(\tilde{e}) + c_7\frac{m}{2}(\tilde{e}'P\tilde{e})^{\frac{m}{2}-\frac{1}{2}}(\xi'\xi)^{\frac{1}{2}} + (\tilde{e}'P\tilde{e})^k\ \mathrm{E}^{\mathbf{z}}\left[V(\mathbf{z})\right] \\
&\quad + \frac{m}{2}\big(2c_6(\frac{m}{2}-1) + \mathrm{trace}\,(\Sigma_{Lv}P)\,\big)(\tilde{e}'P\tilde{e})^{\frac{m}{2}-1} - (\tilde{e}'P\tilde{e})^{\frac{m}{2}+k}.
\end{aligned}$$

Consider the case in which $\frac{m}{2} > k$. Take the expectation of the generator w.r.t. $\tilde{\mathbf{e}}$ and ξ. For any δ_2, δ_3, and $\delta_4 > 0$,

$$\begin{aligned}
\mathrm{E}\left[(\tilde{\mathbf{e}}'P\tilde{\mathbf{e}})^k\right] &\leq \frac{\mathrm{E}\left[V(\tilde{\mathbf{e}})\right]}{\delta_2^{\frac{m}{2}-k}} + \delta_2^k, \\
\mathrm{E}\left[(\tilde{\mathbf{e}}'P\tilde{\mathbf{e}})^{\frac{m}{2}+k}\right] &\geq \delta_3^k\,\mathrm{E}\left[V(\tilde{\mathbf{e}})\right] - \delta_3^{\frac{m}{2}+k}, \\
\mathrm{E}\left[(\tilde{\mathbf{e}}'P\tilde{\mathbf{e}})^{\frac{m}{2}-1}\right] &\leq \frac{1}{\delta_4}\,\mathrm{E}\left[V(\tilde{\mathbf{e}})\right] + \delta_4^{\frac{m}{2}-1}.
\end{aligned}$$

By Hölder's Inequality,

$$\mathrm{E}^{\tilde{\mathbf{e}},\xi}\left[(\tilde{e}'P\tilde{e})^{\frac{m}{2}-\frac{1}{2}}(\xi'\xi)^{\frac{1}{2}}\right] \leq \left(\mathrm{E}^{\tilde{\mathbf{e}}}\left[(\tilde{e}'P\tilde{e})^{\frac{m}{2}}\right]\right)^{1-\frac{1}{m}}\left(\mathrm{E}^{\xi}\left[(\xi'\xi)^{\frac{m}{2}}\right]\right)^{\frac{1}{m}}.$$

Using Lemma 4.1, for any $\delta_5 > 0$, we obtain

$$\left(\mathrm{E}^{\tilde{\mathbf{e}}} \left[(\tilde{e}' P \tilde{e})^{\frac{m}{2}} \right] \right)^{1-\frac{1}{m}} = \mathrm{E}[V(\tilde{\mathbf{e}})]^{1-\frac{1}{m}} \leq \frac{1}{\delta_5^{\frac{1}{m}}} \mathrm{E}\left[V(\tilde{\mathbf{e}})\right] + \delta_5^{1-\frac{1}{m}}.$$

Therefore,

$$\mathrm{E}^{\tilde{\mathbf{e}},\xi} \left[(\tilde{e}' P \tilde{e})^{\frac{m}{2}-\frac{1}{2}} (\xi' \xi)^{\frac{1}{2}} \right] \leq (\Delta_\xi(m))^{\frac{1}{m}} \left(\frac{1}{\delta_5^{\frac{1}{m}}} \mathrm{E}\left[V(\tilde{\mathbf{e}})\right] + \delta_5^{1-\frac{1}{m}} \right),$$

where $\Delta_\xi(m)$ is as in (4.28). Then boundedness follows using the same argument as in Theorem 4.4 by choosing δ_3 large enough.

For $k \geq \frac{m}{2}$, using Lemma 4.1 again, for any $\delta_6 > 0$, we obtain

$$\mathrm{E}\left[(\tilde{e}' P \tilde{e})^{\frac{m}{2}} \right] \leq \frac{\mathrm{E}\left[(\tilde{e}' P \tilde{e})^{k+1} \right]}{\delta_6^{k+1-\frac{m}{2}}} + \delta_6^{\frac{m}{2}},$$

which leads to the boundedness of $\mathrm{E}\left[(\tilde{e}' P \tilde{e})^{\frac{m}{2}} \right]$, as the boundedness of $\mathrm{E}\left[(\tilde{e}' P \tilde{e})^{k+1} \right]$ is already established in the previous case. ■

B.2 Proof of theorems in Chapter 5

B.2.1 Proof of Lemma 5.2

Proof.[Lemma 5.2] Following the derivation in [61],

$$\begin{aligned}
\hat{\mathbf{x}}_{t+1}^{k+1} &= \begin{cases} A\hat{\mathbf{x}}_t^{k+1}, & \text{if } \boldsymbol{\nu}_{t+1} = 0 \\ A^{k+1}\tilde{\mathbf{x}}_{t-k}, & \text{if } \boldsymbol{\nu}_{t+1} = 1 \end{cases} \\
&= \begin{cases} A\hat{\mathbf{x}}_t^{k}, & \text{if } \boldsymbol{\nu}_{t+1} = 0 \\ A^{k+1}\tilde{\mathbf{x}}_{t-k}, & \text{if } \boldsymbol{\nu}_{t+1} = 1 \end{cases} \\
&= A\hat{\mathbf{x}}_t^{k},
\end{aligned}$$

which proves part one. Part two follows by taking account of (5.2),

$$\begin{aligned}\hat{\mathbf{e}}_{t+1}^{k+1} &= A\mathbf{x}_t + \mathbf{w}_t - A\hat{\mathbf{x}}_t^k \\ &= A\hat{\mathbf{e}}_t^k + \mathbf{w}_t.\end{aligned}$$

Propagate the equation above from $t+1$ to $t+\tau$,

$$\hat{\mathbf{e}}_{t+\tau}^{\tau} = A^{\tau-1}\hat{\mathbf{e}}_{t+1}^1 + \sum_{s=1}^{\tau-1} A^{\tau-s-1}\mathbf{w}_{t+s},$$

in which

$$\begin{aligned}\hat{\mathbf{e}}_{t+1}^1 &= \mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1}^1 \\ &= A\mathbf{x}_t + \mathbf{w}_t - \begin{cases} A\hat{\mathbf{x}}_t^1, & \text{if} \quad \boldsymbol{\nu}_{t+1} = 0 \\ A\tilde{\mathbf{x}}_t, & \text{if} \quad \boldsymbol{\nu}_{t+1} = 1 \end{cases} \\ &= \begin{cases} A(\mathbf{x}_t - \tilde{\mathbf{x}}_t) + A(\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_t^1) + \mathbf{w}_t, & \text{if} \quad \boldsymbol{\nu}_{t+1} = 0 \\ A(\mathbf{x}_t - \tilde{\mathbf{x}}_t) + \mathbf{w}_t, & \text{if} \quad \boldsymbol{\nu}_{t+1} = 0. \end{cases}\end{aligned}$$

Combining the two equations above,

$$\begin{aligned}\hat{\mathbf{e}}_{t+\tau}^{\tau} &= A^{\tau}(\mathbf{x}_t - \tilde{\mathbf{x}}_t) + \sum_{s=0}^{\tau-1} A^{\tau-s-1}\mathbf{w}_{t+s} + \begin{cases} A^{\tau}(\tilde{\mathbf{x}}_t - \hat{\mathbf{x}}_t^1), & \text{if} \quad \boldsymbol{\nu}_{t+1} = 0 \\ 0, & \text{if} \quad \boldsymbol{\nu}_{t+1} = 1 \end{cases} \\ &= A^{\tau}\tilde{\mathbf{e}}_t + \sum_{s=0}^{\tau-1} A^{\tau-s-1}\mathbf{w}_{t+s} + \begin{cases} A^{\tau}\mathbf{e}_t^1, & w.p. \quad 1-(1-p)\mathbf{u}_t \\ 0, & w.p. \quad (1-p)\mathbf{u}_t, \end{cases}\end{aligned}$$

which proves the third part of the lemma. For the last part, we have that

$$
\begin{aligned}
\mathbf{e}_{t+1}^1 &\overset{(a)}{=} \tilde{\mathbf{x}}_{t+1} - \hat{\mathbf{x}}_{t+1}^1 \\
&\overset{(b)}{=} A\tilde{\mathbf{x}}_t + L(\mathbf{y}_{t+1} - CA\tilde{\mathbf{x}}_t) - \begin{cases} A\hat{\mathbf{x}}_t^1, & \text{if } \boldsymbol{\nu}_{t+1} = 0 \\ A\tilde{\mathbf{x}}_t, & \text{if } \boldsymbol{\nu}_{t+1} = 1 \end{cases} \\
&\overset{(c)}{=} A\tilde{\mathbf{x}}_t + L(CA\mathbf{x}_t + C\mathbf{w}_t + \mathbf{v}_{t+1} - CA\tilde{\mathbf{x}}_t) - \begin{cases} A\hat{\mathbf{x}}_t^1, & \text{if } \boldsymbol{\nu}_{t+1} = 0 \\ A\tilde{\mathbf{x}}_t, & \text{if } \boldsymbol{\nu}_{t+1} = 1 \end{cases} \\
&\overset{(d)}{=} \begin{cases} A\mathbf{e}_t^1 + L(CA\tilde{\mathbf{e}}_t + C\mathbf{w}_t + \mathbf{v}_{t+1}), & \text{if } \boldsymbol{\nu}_{t+1} = 0 \\ L(CA\tilde{\mathbf{e}}_t + C\mathbf{w}_t + \mathbf{v}_{t+1}), & \text{if } \boldsymbol{\nu}_{t+1} = 1 \end{cases} \\
&\overset{(e)}{=} \begin{cases} A\mathbf{e}_t^1 + L(CA\tilde{\mathbf{e}}_t + C\mathbf{w}_t + \mathbf{v}_{t+1}), & w.p. \quad 1 - (1-p)\mathbf{u}_t \\ L(CA\tilde{\mathbf{e}}_t + C\mathbf{w}_t + \mathbf{v}_{t+1}), & w.p. \quad (1-p)\mathbf{u}_t, \end{cases}
\end{aligned}
$$

in which (a) is by definition, (b) is from the Kalman filter (5.9), (c) is from the plant dynamics (5.2), (d) is by definitions, and (e) is from Assumptions 5.2 and 5.3. ■

B.2.2 Proof of Lemma 5.3

Proof.[Lemma 5.3]

$$
J_{\pi_t}(e) = \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} c(\mathbf{e}_t, \mathbf{u}_t) | \pi_t, \mathbf{e}_0 = e\right],
$$

From (5.16),

$$
\begin{aligned}
J_{\pi_t}(e) &\overset{(a)}{=} \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} \hat{\mathbf{e}}_t'^{\tau} Q \hat{\mathbf{e}}_t^{\tau} + \lambda \mathbf{a}_t | \pi_t, \mathbf{e}_0 = e\right] \\
&\overset{(b)}{=} \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} \hat{\mathbf{e}}_{t+\tau}'^{\tau} Q \hat{\mathbf{e}}_{t+\tau}^{\tau} + \lambda \mathbf{a}_t | \pi_t, \mathbf{e}_0 = e\right] \\
&\quad + \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{\tau} \hat{\mathbf{e}}_t'^{\tau} Q \hat{\mathbf{e}}_t^{\tau} | \pi_t, \mathbf{e}_0 = e\right] \\
&\overset{(c)}{=} \limsup_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{t=1}^{N-1} c(\mathbf{e}_t^1, \mathbf{u}_t) | \pi_t, \mathbf{e}_0 = e\right],
\end{aligned}
$$

where (b) is because $\mathrm{E}\left[\hat{\mathbf{e}}_t'^{\tau} Q \hat{\mathbf{e}}_t^{\tau}\right]$ is finite, and (c) is by (5.17) and the definitions of ρ^2 and $c(e, u)$. ■