Chapter 5
Kalman Filtering over Wireless Fading Channels

Yasamin Mostofi and Alireza Ghaffarkhah

Abstract In this chapter, we consider the estimation of a dynamical system over a wireless fading channel using a Kalman filter. We develop a framework for understanding the impact of stochastic communication noise, packet drop and the knowledge available on the link qualities on Kalman filtering over fading channels. We consider three cases of “full knowledge”, “no knowledge” and “partial knowledge”, based on the knowledge available on the communication quality. We characterize the dynamics of these scenarios and establish the necessary and sufficient conditions to ensure stability. We then propose new ways of optimizing the packet drop in order to minimize the average estimation error variance of the Kalman filter. We consider both adaptive and non-adaptive optimization of the packet drop. Our results show that considerable performance improvement can be achieved by using the proposed framework.

5.1 Introduction

The unprecedented growth of sensing, communication and computation, in the past few years, has created the possibility of exploring and controlling the environment in ways not possible before. A wide range of sensor network applications have emerged, from environmental monitoring, emergency response, smart homes and factories to surveillance, security and military applications [1, 2]. The vision of a multi-agent mobile network cooperatively learning and adapting in harsh unknown environments to achieve a common goal is closer than ever. In such networks, each node has limited capabilities and needs to cooperate with others in order to achieve the task. Therefore, there can be cases that an agent senses a parameter of interest and sends its measurement wirelessly to a remote node, which will be in charge of estimation and possibly producing a control command. Wireless communication then plays a key role in the overall performance of such cooperative networks as

Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87113, USA e-mail: {ymostofi,alinem}@ece.unm.edu.

the agents share sensor measurements and receive control commands over wireless links.

In a realistic communication setting, such as an urban area or indoor environment, Line-Of-Sight (LOS) communication may not be possible due to the existence of several objects that can attenuate, reflect, diffract or block the transmitted signal. The received signal power typically experiences considerable variations and can change drastically in even a small distance. Fig. 5.1 (left and solid dark curve of right), for instance, shows examples of channel measurements in the Electrical and Computer Engineering (ECE) building at the University of New Mexico. It can be seen that channel can change drastically with a small movement of an agent. In general, communication between sensor nodes can be degraded due to factors such as fading, shadowing or distance-dependent path loss [3,4]. As a result, the received Signal to Noise Ratio of some of the receptions can be too low resulting in a packet drop. Furthermore, poor link quality can result in bit flips and therefore the packets that are kept are not necessarily free of error. As a result, sensing and control can not be considered independent of communication issues. An integrative approach is needed, where sensing, communication and control issues are jointly considered in the design of these systems.

Considering the impact of communication on networked estimation and control is an emerging research area, which has received considerable attention in the past few years. Among the uncertainties introduced by communication, impact of quantization on networked estimation and control has been studied extensively [5, 6]. Stabilization of linear dynamical systems in the presence of quantization noise has been characterized [7–9] and trade-offs between information rate and convergence time is formulated [10]. To address the inadequacy of the classical definition of capacity for networked control applications, anytime capacity was introduced and utilized for stabilization of linear systems [11]. Disturbance rejection and the corresponding required extra rate in these systems were considered in [12]. Nilsson studied the impact of random delays on stability and optimal control of networked control systems [13].

In estimation and control over a wireless link, poor link qualities are the main performance degradation factors. Along this line, Micheli et al. investigated the impact of packet loss on estimation by considering random sampling of a dynamical system [14]. This is followed by the work of Sinopoli et al. which derived bounds for the maximum tolerable probability of packet loss to maintain stability [15]. Several other papers have appeared since then that consider the impact of packet loss or random delays on networked control systems [16–21].

The current framework in the literature, however, does not consider the impact of stochastic communication errors on estimation and control over wireless links. Transmission over a wireless fading channel, as shown in Fig. 5.1, can result in a time-varying packet drop. Furthermore, the packets that are kept are not necessarily free of noise. The variance of this noise is also stochastic due to its dependency on the Signal to Noise Ratio. Therefore, we need to consider the impact of both packet drop and stochastic communication noise on the performance. Moreover, the knowledge on the communication reception quality may not be fully available to the estimator, resulting in different forms of Kalman filtering. In [22,23], we considered the impact of communication errors on estimation over wireless links. We showed that dropping all the erroneous packets may not be the optimum strategy for estimation of a rapidly-changing dynamical system over a wireless link. In this chapter, we extend our previous work and characterize the impact of both packet drop and communication errors on Kalman filtering over wireless fading channels. We are interested in characterizing the impact of the knowledge available on the link qualities on Kalman filtering over wireless channels. We consider three cases of "full knowledge", "no knowledge" and "partial knowledge", based on the knowledge available on the communication quality. We characterize the dynamics of these scenarios and establish necessary and sufficient conditions to ensure stability. We then show how to optimize the packet drop in the physical layer in order to minimize the average estimation error variance of the Kalman filter. Our approach is truly integrative as the available knowledge on link qualities is utilized in the estimator and current quality of estimation is used in some of our proposed packet drop designs.

We conclude this section with an overview of the chapter. In Section 5.2, we formulate the problem and summarize key features of a transmission over a wireless fading channel. In Section 5.3, we consider and characterize three possible Kalman filtering scenarios, depending on the knowledge available on the communication error variance in the estimator. In Section 5.4, we then optimize the packet drop mechanism. We show how dropping all the erroneous packets (or keeping them all) may not be the optimum strategy and that packet drop should be designed and adapted based on the knowledge available on the link and estimation quality as well as the dynamics of the system under estimation. We conclude in Section 5.5.

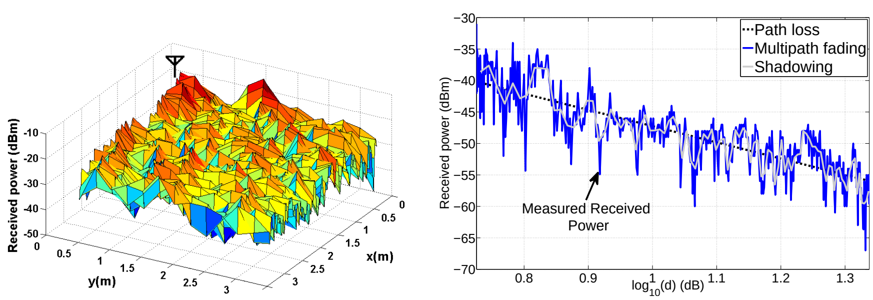

Fig. 5.1: Channel measurement (left) in the Cooperative Network Lab and (right) along a hallway in the basement of the ECE building at the University of New Mexico.

5.2 System Model

Consider a mobile sensor observing a system with a linear dynamics as follows:

$$x[k+1] = Ax[k] + w[k] \qquad \text{and} \qquad y[k] = Cx[k] + v[k], \tag{5.1}$$

where $x[k] \in \mathbb{R}^N$ and $y[k] \in \mathbb{R}^M$ represent the state and observation respectively. $w[k] \in \mathbb{R}^N$ and $v[k] \in \mathbb{R}^M$ represent zero-mean Gaussian process and observation noise vectors with variances of $Q \succcurlyeq 0$ and $R_s \succ 0$ respectively. In this chapter, we take $M = N$ and C invertible to focus on the impact of time-varying fading channels and the resulting communication error on Kalman filtering. We are interested in estimating unstable dynamical systems. Therefore matrix A has one or more of its eigenvalues outside the unit circle. It should, however, be noted that the proposed optimization framework of this chapter is also applicable to the case of a stable A. Furthermore, we take the pair $(A, Q^{1/2})$ to be controllable. The sensor then transmits its observation over a **time-varying fading channel** to a remote node, which is in charge of estimation. Fig. 5.2 shows the high-level schematic of the considered problem. In this chapter, we consider and optimize such problems, i.e. Kalman filtering in the presence of packet drop and stochastic communication noise.

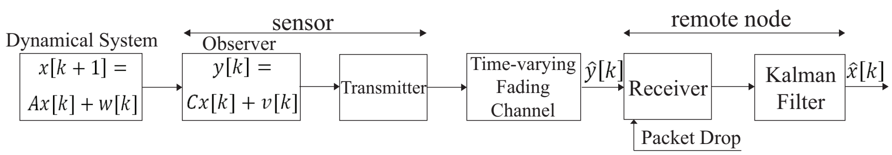

Fig. 5.2: A schematic of estimation over a wireless fading channel.

5.2.1 An Overview of Wireless Communications [3, 4, 24]

In this section we briefly describe the key factors needed for modeling the impact of a time-varying fading channel on Kalman filtering.

5.2.1.1 Received Signal To Noise Ratio (SNR)

A fundamental parameter that characterizes the performance of a communication channel is the received Signal to Noise Ratio. Received Signal to Noise Ratio (SNR) is defined as the ratio of the received signal power divided by the receiver thermal noise power. SNR is a key factor in determining whether a received packet will be kept and used in the receiving node or not. Furthermore, it determines how well the received bits and, as a result, the received measurement can be retrieved. Let $\Upsilon[k]$ represent the instantaneous received Signal to Noise Ratio at the k^{th} transmission. We will have $\Upsilon[k] = \frac{|h[k]|^2 \sigma_s^2}{\sigma_T^2}$, where $h[k] \in \mathbb{C}$ represents the time-varying coefficient of the baseband equivalent channel during the transmission of $x[k]$, $\sigma_s^2 = \mathbb{E}\left(|s|^2\right)$ is the transmitted signal power and $\sigma_T^2 = \mathbb{E}\left(|n_{\text{thermal}}|^2\right)$ is the power of the receiver thermal noise. As the sensor moves, the remote node will experience different channels and therefore different SNRs.

5.2.1.2 Probabilistic Characterization of SNR

In wireless communications, it is common to model the channel (and as a result the received SNR) probabilistically, with the goal of capturing its underlying dynamics. The utilized probabilistic models are the results of analyzing several empirical data over the years. In general, a communication channel between two nodes can be modeled as a multi-scale dynamical system with three major dynamics: *multi-path fading*, *shadowing* and *path loss*. Fig. 5.1 (right) shows the received signal power across a route in the basement of ECE building at UNM [24]. The three main dynamics of the received signal power are marked on the figure. When a wireless transmission occurs, replicas of the transmitted signal will arrive at the receiver due to phenomena such as reflection and scattering. Different replicas can be added constructively or destructively depending on the phase terms of individual ones. As a result, with a small movement of a node, the phase terms can change drastically, resulting in the rapid variations of the channel. Such rapid variations are referred to as multipath fading and can be seen from Fig. 5.1 (right, solid dark curve). The higher the number of reflectors and scatterers in the environment, the more severe small-scale variations could be. By spatially averaging the received signal locally and over distances that channel can still be considered stationary, a slower dynamic emerges, which is called shadowing (light gray curve of Fig. 5.1 right). Shadowing is the result of the transmitted signal being possibly blocked by a number of obstacles before reaching the receiver. Empirical data has shown lognormal to be a good match for the distribution of shadowing. Finally, by averaging over the variations of shadowing, a distance-dependent trend is seen, which is, in dB, an affine function of the log of the distance between the transmitter and receiver.

Since SNR is proportional to the received signal power, it has similar dynamics. Therefore, if the movement of the transmitting node is confined to a small area, then $\Upsilon[k]$ can be considered a stationary stochastic process. On the other hand, if the node is moving fast and over larger distances during the estimation process, then the non-stationary nature of the channel should also be considered by taking into account shadowing. Finally, for movements over even larger areas, non-stationary behavior of shadowing should also be considered by taking into account the underlying path loss trend. In this chapter, we consider the multipath fading, i.e. we model the SNR as a stochastic but stationary process. Our results are easily extendable to the cases of non-stationary channels.

5.2.1.3 Distribution of Fading

In this chapter, we use the term fading to refer to multipath fading. Over small enough distances where channel (or equivalently SNR) can be considered stationary, it can be mathematically shown that Rayleigh distribution is a good match for the distribution of the square root of SNR if there is no Line Of Sight (LOS) path while Rician provides a better match if an LOS exists. A more general distribution is Nakagami [25], which has the following pdf for $\Upsilon_{\mathrm{sqrt}} \triangleq \sqrt{\Upsilon}$:

$$\chi(\Upsilon_{\text{sqrt}}) = \frac{2m^m \Upsilon_{\text{sqrt}}^{2m-1}}{\Gamma(m)\Upsilon_{\text{ave}}^m} \exp\left(\frac{-m\Upsilon_{\text{sqrt}}^2}{\Upsilon_{\text{ave}}}\right) \tag{5.2}$$

for $m \geq 0.5$, where m is the fading parameter, Υ_{ave} denotes the average of SNR and $\Gamma(.)$ is the Gamma function. If $m = 1$, this distribution becomes Rayleigh whereas for $m = \frac{(m'+1)^2}{2m'+1}$, it is reduced to a Rician distribution with parameter m'. These distributions also match several empirical data. In this chapter, we do not make any assumption on the probability distribution of Υ. Only when we want to provide an example, we will take Υ to be exponentially distributed (i.e. the square root is Rayleigh), which is a common model for outdoor fading channels with no Line-of-Sight path. We also take the SNR to be uncorrelated from one transmission to the next.

5.2.1.4 Stochastic Communication Noise Variance

The sensor node of Fig. 5.2 quantizes the observation, $y[k]$, transforms it into a packet of bits and transmits it over a fading channel. The remote node will receive a noisy version of the transmitted data due to bit flip. Let $\hat{y}[k]$ represent the received signal, as shown in Fig. 5.2. $\hat{y}[k]$ is what the second node assumes the k^{th} transmitted observation was. Let $n[k]$ represent the difference between the transmitted observation and the received one:

$$n[k] = \hat{y}[k] - y[k]. \tag{5.3}$$

We refer to $n[k]$ as communication error (or communication noise) throughout the chapter. This is the error that is caused by poor link quality and the resulting flipping of some of the transmitted bits during the k^{th} transmission.[1] The variance of $n[k]$ at the k^{th} transmission is given by

$$R_c[k] = \mathbb{E}\big(n[k]n^T[k]|\Upsilon[k]\big) = \sigma_c^2\big(\Upsilon[k]\big)I, \tag{5.4}$$

where I represents the identity matrix throughout this chapter and $\sigma_c^2\big(\Upsilon[k]\big)$ is a non-increasing function of SNR that depends on the transmitter and receiver design principles, such as modulation and coding, as well as the transmission environment. It can be seen that the communication noise has a stochastic variance as the variance is a function of SNR. Throughout the chapter, we use $\mathbb{E}(z)$ to denote average of the variable z. Note that we took the communication error in the transmission of different elements of vector $y[k]$ to be uncorrelated due to fast fading. This assumption is mainly utilized in Section 5.4 for the optimization of packet drop. To keep our analysis general, in this chapter we do not make any assumption on the form of $\sigma_c^2\big(\Upsilon[k]\big)$ as a function of $\Upsilon[k]$. By considering both the communication and

[1] Note that for transmissions over a fading channel, the impact of quantization is typically negligible as compared to the communication errors. Therefore, in this chapter, we ignore the impact of quantization as it has also been heavily explored in the context of networked control systems.

observation errors, we have,

$$\hat{y}[k] = Cx[k] + \underbrace{v[k] + n[k]}_{r[k]}, \tag{5.5}$$

where $r[k]$ represents the overall error that enters the estimation process at the k^{th} time step and has the variance of

$$R[k] = \mathbb{E}\big(r[k]r^T[k]|\Upsilon[k]\big) = R_s + R_c[k]. \tag{5.6}$$

$R[k]$ is therefore stochastic due to its dependency on the SNR. Since our emphasis is on the communication noise, throughout the chapter, we refer to $R[k]$ as the "overall communication noise variance".

5.2.1.5 Packet Drop Probability

Let $\mu[k]$ indicate if the receiver drops the k^{th} packet, i.e. $\mu[k] = 1$ means that the k^{th} packet is dropped while $\mu[k] = 0$ denotes otherwise. $\mu[k]$ can also be represented as a function of $\Upsilon[k]$: $\mu[k] = G\big(\Upsilon[k]\big)$. It should be noted that the receiver may not decide on dropping packets directly based on the instantaneous received Signal to Noise Ratio. However, since any other utilized measure is a function of $\Upsilon[k]$, we find it useful to express μ as a function of this fundamental parameter. Experimental results have shown G to be well approximated as follows [26]:

$$\mu[k] = \begin{cases} 0 & \Upsilon[k] \geq \Upsilon_T \\ 1 & \text{else} \end{cases} \tag{5.7}$$

This means that the receiver keeps those packets with the received instantaneous Signal to Noise Ratio above a designated threshold Υ_T. In this chapter, we show how the threshold Υ_T can be optimized in order to control the amount of information loss and communication error that enters the estimation process.

5.2.2 Estimation using a Kalman Filter

The remote node estimates the state based on the received observations using a Kalman filter [27]. Let $\hat{x}[k] = \hat{x}[k|k-1]$ represent the estimate of $x[k]$ using all the received observations up to and including time $k-1$. Then $P[k]$ represents the corresponding estimation error variance given $\Upsilon[0], \cdots, \Upsilon[k-1]$:

$$P[k] = P[k\,|\,k-1] = \mathbb{E}\left[\Big(x[k] - \hat{x}[k]\Big)\Big(x[k] - \hat{x}[k]\Big)^T \,\Big|\, \Upsilon[0], \cdots, \Upsilon[k-1]\right]. \tag{5.8}$$

As can be seen, Eq. 5.8 is different from the traditional form of Kalman Filter since $P[k]$ is a stochastic function due to its dependency on $\Upsilon[0], \Upsilon[1], \cdots, \Upsilon[k-1]$.

Therefore, to obtain $\mathbb{E}\big(P[k]\big)$, $P[k]$ should be averaged over the joint distribution of $\Upsilon[0],\Upsilon[1],\cdots,\Upsilon[k-1]$.

5.2.3 Impact of Packet Drop and Stochastic Communication Error on Kalman Filtering

A transmission over a fading channel can result in the dropping of some of the packets. Furthermore, the packets that are kept may not be free of noise, depending on the packet drop threshold Υ_T. It is possible to increase the threshold such that the packets that are kept are free of error. However, this increases packet drop probability and can result in instability or poor performance. Therefore, we need to consider the impact of both packet drop and communication noise on the performance. The communication noise variance is also stochastic due to its dependency on SNR. As a result, Kalman filtering over a fading channel will not have its traditional form.

Furthermore, Kalman filtering is done in the application layer whereas the physical layer is in charge of wireless communication. An estimate of SNR is typically available in the physical layer. Such an estimate can be translated to an assessment of the communication noise variance if a mathematical characterization for σ_c^2 exists. However, this knowledge may or may not be available in the application layer, depending on the system design. There can also be cases where an exact assessment of the communication noise variance and the resulting reception quality is only partially available in the physical layer. As a result, full knowledge of the communication noise variance may not be available at the estimator. Depending on the available knowledge on the communication noise variance, the Kalman filter will have different forms. In this chapter, we are interested in understanding the impact of the knowledge available on the link qualities on Kalman filtering over wireless channels. We consider three cases of “full knowledge”, “no knowledge” and “partial knowledge”, based on the knowledge available on the communication quality in the estimator. In the next section, we characterize the dynamics of these scenarios and establish necessary and sufficient conditions to ensure stability. Then, in Section 5.4, we show how to optimize packet drop threshold in the physical layer in order to minimize the average estimation error variance of the Kalman filter.

5.3 Stability Analysis

Consider the traditional form of Kalman filtering with no packet drop and full knowledge on the reception quality. We have, $\hat{x}[k+1]=A\hat{x}[k]+K_f[k]\big(\hat{y}[k]-C\hat{x}[k]\big)$ and $P[k+1]=Q+\big(A-K_f[k]C\big)P[k]\big(A-K_f[k]C\big)^T+K_f[k]R[k]K_f^T[k]$, where $K_f[k]=AP[k]C^T\big(CP[k]C^T+R[k]\big)^{-1}$ is the optimal gain [27]. However, as discussed in the previous section, the received packet is dropped if $\Upsilon[k]<\Upsilon_T[k]$. Furthermore, full knowledge on the variance of the stochastic communication noise might not be available. Then, the traditional form of Kalman filter needs to be modified to reflect packet drop and the available information at the estimator. Let $0 \preceq \tilde{R}[k] \preceq R[k]$ denote the part of $R[k]$ that is known to the estimator. $\tilde{R}[k]$ can be considered what the estimator perceives $R[k]$ to be. Then, $\tilde{R}[k]$ is used to calculate the filtering gain which, when considered jointly with the packet drop, results in the following recursion for $P[k]$:

$$P[k+1] = \mu[k]AP[k]A^T + Q + \big(1-\mu[k]\big)\Big[\big(A-\tilde{K}_f[k]C\big)P[k]\big(A-\tilde{K}_f[k]C\big)^T + \tilde{K}_f[k]R[k]\tilde{K}_f^T[k]\Big], \quad (5.9)$$

where $\tilde{K}_f[k] = AP[k]C^T\big(CP[k]C^T + \tilde{R}[k]\big)^{-1}$ and $\mu[k]$ is defined in Eq. 5.7. We assume that the initial error variance, $P[0]$, is positive definite. Then, using the fact that $(A, Q^{1/2})$ is controllable, one can easily verify that $P[k] \succ 0$ for $k \geq 0$, which we utilize later in our derivations. By inserting $\tilde{K}_f$ in Eq. 5.9, we obtain

$$\begin{aligned} P[k+1] =& AP[k]A^T + Q - \big(1-\mu[k]\big)AP[k]C^T\big(CP[k]C^T + \tilde{R}[k]\big)^{-1} \\ &\times \big(CP[k]C^T + 2\tilde{R}[k] - R[k]\big)\big(CP[k]C^T + \tilde{R}[k]\big)^{-1}CP[k]A^T. \end{aligned} \quad (5.10)$$

In this part, we consider three different cases, based on the knowledge available on the link quality at the estimator:

1. Estimator has no knowledge on the reception quality: $\tilde{R}[k] = 0$,
2. Estimator has full knowledge on the reception quality: $\tilde{R}[k] = R[k]$,
3. Estimator has partial knowledge on the reception quality: $0 \prec \tilde{R}[k] \prec R[k]$.

In the following, we derive the necessary and sufficient condition for ensuring the stability of these cases and show them to have the same stability condition. Before doing so, we introduce a few basic definitions and lemmas that will be used throughout the chapter. Note that all the variables used in this chapter are real.

Definition 5.1. We consider the estimation process stable as long as the average estimation error variance, $\mathbb{E}\big(P[k]\big)$, stays bounded.

Definition 5.2. Let $f : \mathbb{R}^{N\times N} \to \mathbb{R}^{N'\times N'}$ represent a symmetric matrix-valued function. Then f is non-decreasing with respect to its symmetric input matrix variable if $\Pi_1 \succeq \Pi_2 \;\Rightarrow\; f\big(\Pi_1\big) \succeq f\big(\Pi_2\big)$, for any symmetric Π_1 and $\Pi_2 \in \mathbb{R}^{N\times N}$. Similarly, f is increasing if $\Pi_1 \succ \Pi_2 \;\Rightarrow\; f\big(\Pi_1\big) \succ f\big(\Pi_2\big)$.

Definition 5.3. Let f represent a symmetric matrix-valued function, $f : \mathbb{R}^{N\times N} \to \mathbb{R}^{N'\times N'}$. Function f is convex with respect to matrix inequality if $f(\theta\Pi_1 + (1-\theta)\Pi_2) \preceq \theta f(\Pi_1) + (1-\theta)f(\Pi_2)$, for arbitrary Π_1 and $\Pi_2 \in \mathbb{R}^{N\times N}$ and $\theta \in [0,1]$. Similarly, f is concave if $-f$ is convex. See [28] for more details.

Lemma 5.1. *We have the following:*

1. *Let $\Pi_1 \in \mathbb{R}^{N\times N}$ and $\Pi_2 \in \mathbb{R}^{N\times N}$ represent two symmetric positive definite matrices. Then $\Pi_1 \preceq \Pi_2$ if and only if $\Pi_1^{-1} \succeq \Pi_2^{-1}$.*

2. *If* $\Pi_1 \preceq \Pi_2$ *for symmetric* Π_1 *and* Π_2*, then* $\Psi\Pi_1\Psi^T \preceq \Psi\Pi_2\Psi^T$ *for an arbitrary* $\Psi \in \mathbb{R}^{N'\times N}$.
3. *Let* $f : \mathbb{R}^{N\times N} \to \mathbb{R}^{N'\times N'}$ *represent a symmetric function. Let* $\Pi_1 \in \mathbb{R}^{N\times N}$ *be symmetric. If* f *is an increasing (non-decreasing) function of* Π_1*, then* $f\big(\Pi_1+\Pi_2\big)$ *is also increasing (non-decreasing) with respect to* Π_1 *for a constant* Π_2.
4. *If* $f : \mathbb{R}^{N\times N} \to \mathbb{R}^{N'\times N'}$ *is symmetric and a non-decreasing function of symmetric* Π_1*, then* $\Psi f\big(\Pi_1\big)\Psi^T$ *is also non-decreasing as a function of* Π_1 *and for an arbitrary* $\Psi \in \mathbb{R}^{N''\times N'}$.
5. *Let* $f : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ *represent matrix inverse function:* $f(\Pi_1) = \Pi_1^{-1}$ *for a symmetric positive definite* Π_1*. Then* f *is a convex function of* Π_1.
6. *If* $f : \mathbb{R}^{N\times N} \to \mathbb{R}^{N'\times N'}$ *is symmetric and a convex function of* Π_1*, then* $f\big(\Pi_1+\Pi_2\big)$ *is also convex with respect to* Π_1 *and for a constant* Π_2.
7. *If* $f : \mathbb{R}^{N\times N} \to \mathbb{R}^{N'\times N'}$ *is symmetric and a convex function of* Π_1*, then* $\Psi f\big(\Pi_1\big)\Psi^T$ *is also convex for an arbitrary matrix* $\Psi \in \mathbb{R}^{N''\times N'}$.
8. *Let* $f(\Pi_1) = \Pi_1\big(\Pi_2+\Pi_1\big)^{-1}\Pi_1$*, where* Π_1 *and* Π_2 *are symmetric positive definite matrices.* f *is a convex function of* Π_1.

Proof. see [29, 30]. □

Lemma 5.2. *Consider the following Lyapunov equation with* Δ *Hermitian:* $\Sigma = \Pi\Sigma\Pi^T + \Delta$*. Then the following holds:*

1. *If* Π *is a stable matrix (spectral radius less than one),* Σ *will be unique and Hermitian and can be expressed as follows:*

$$\Sigma = \sum_{i=0}^{\infty} \Pi^i \Delta \big(\Pi^T\big)^i. \tag{5.11}$$

2. *If* $\big(\Pi, \Delta^{1/2}\big)$ *is controllable and* $\Delta \succeq 0$*, then* Σ *will be Hermitian, unique and positive definite iff* Π *is stable.*

Proof. see [27]. □

5.3.1 Noise-free Case

Consider the case where $\tilde{R}[k] = R[k] = 0$ if a packet is kept. We refer to this scenario as the noise-free case. Comparison with the dynamics of this case will help us characterize the dynamics of the three aforementioned scenarios. In this case, Eq. 5.10 is simplified to $P[k+1] = \mu[k]AP[k]A^T + Q \triangleq \Phi_{\text{perf}}(P[k], \mu[k])$, with the following average dynamics:

$$\mathbb{E}\big(P[k+1]\big) = \mu_{\text{ave}}\big(\Upsilon_T\big)A\mathbb{E}\big(P[k]\big)A^T + Q, \tag{5.12}$$

where $\mu_{\text{ave}}\big(\Upsilon_T\big)$ is the average probability of packet drop:

$$\mu_{\text{ave}}(\Upsilon_T) = \mathbb{E}\big(\mu[k]\big) = \int_0^{\Upsilon_T} \chi(\Upsilon)d\Upsilon, \tag{5.13}$$

with χ representing the probability density function of Υ. Let $\rho_{\max}(A)$ represent the spectral radius of matrix A. The average dynamical system of Eq. 5.12 is stable iff

$$\mu_{\text{ave}}(\Upsilon_T) < \rho_{\max}^{-2}(A). \tag{5.14}$$

5.3.2 Estimator has no knowledge on the reception quality

From Eq. 5.10, we have the following for the case where no information on the reception quality is available at the estimator:

$$\begin{aligned} P[k+1] =& AP[k]A^T + Q - \big(1-\mu[k]\big)AP[k]C^T\big(CP[k]C^T\big)^{-1} \\ & \times \big(CP[k]C^T - R[k]\big)\big(CP[k]C^T\big)^{-1}CP[k]A^T \\ =& \mu[k]AP[k]A^T + Q + (1-\mu[k])AC^{-1}R[k]\big(C^T\big)^{-1}A^T \\ \triangleq& \Phi_{\text{no}}\big(P[k],\mu[k],R[k]\big). \end{aligned} \tag{5.15}$$

The average of $P[k]$ is then given by

$$\mathbb{E}\big(P[k+1]\big) = \mu_{\text{ave}}\big(\Upsilon_T\big)A\mathbb{E}\big(P[k]\big)A^T + Q + AC^{-1}R_{\text{ave}}\big(\Upsilon_T\big)\big(C^T\big)^{-1}A^T, \tag{5.16}$$

where $\mu_{\text{ave}}\big(\Upsilon_T\big)$ is defined in Eq. 5.13 and $R_{\text{ave}}\big(\Upsilon_T\big)$ is the average noise variance that enters the estimation process:

$$R_{\text{ave}}(\Upsilon_T) = \mathbb{E}\big(R[k]\big) = \Big(1-\mu_{\text{ave}}\big(\Upsilon_T\big)\Big)R_s + \left[\int_{\Upsilon_T}^{\infty} \sigma_c^2(\Upsilon)\chi(\Upsilon)d\Upsilon\right]I \tag{5.17}$$

with I denoting the identity matrix of the appropriate size.

Lemma 5.3. *Let $\rho_{\max}(A)$ represent the spectral radius of matrix A. The average dynamical system of Eq. 5.16 is stable if and only if $\mu_{ave}(\Upsilon_T) < \rho_{\max}^{-2}(A)$ or equivalently $\Upsilon_T < \Upsilon_{T,c}$, where $\Upsilon_{T,c}$ is the unique solution of the following equation:*

$$\int_0^{\Upsilon_{T,c}} \chi(\Upsilon)d\Upsilon = \rho_{\max}^{-2}(A). \tag{5.18}$$

Proof. See [27]. □

Note that $\Upsilon_{T,c}$ is often referred to as the *critical threshold*. In the next sections, we use the dynamics of the "noise-free" and "no knowledge" cases in order to understand the dynamics of the "full" and "partial" knowledge scenarios.

5.3.3 Estimator has full knowledge on the reception quality

Consider Eq. 5.10 with $\tilde{R}[k] = R[k]$. We have

$$\begin{aligned} P[k+1] =& AP[k]A^T + Q - \big(1-\mu[k]\big)AP[k]C^T\big(CP[k]C^T + R[k]\big)^{-1}CP[k]A^T \\ =& Q + AP[k]A^T - AP[k]\left(P[k] + \Pi_z[k]\right)^{-1}P[k]A^T \\ =& Q + A\Pi_z[k]A^T - A\Pi_z[k]\big(P[k] + \Pi_z[k]\big)^{-1}\Pi_z[k]A^T \\ \triangleq& \Phi_{\text{full}}\big(P[k],\mu[k],R[k]\big), \end{aligned} \tag{5.19}$$

where $\Sigma_z[k] = \begin{cases} R[k] & \mu[k]=0 \\ \infty & \text{otherwise} \end{cases}$, $\Pi_z[k] = C^{-1}\Sigma_z[k]\left(C^T\right)^{-1}$ and the third line is written using matrix inversion lemma [29]. The following set of lemmas relate the dynamics of this case to those of the "no knowledge" and "noise free" cases.

Lemma 5.4. *Consider any symmetric P_1, P_2 and P_3 such that $P_1 \succeq P_2 \succeq P_3 \succ 0$. For any $0 \le \mu \le 1$ and $R \succ 0$, we have*

$$\Phi_{no}\big(P_1,\mu,R\big) \succeq \Phi_{full}\big(P_2,\mu,R\big) \succeq \Phi_{perf}\big(P_3,\mu\big). \tag{5.20}$$

Proof. For any $P \succ 0$, we have $\big(CPC^T + R\big)^{-1} = \big(CPC^T\big)^{-1} - \big(CPC^T\big)^{-1}\Big[\big(CPC^T\big)^{-1} + R^{-1}\Big]^{-1}\big(CPC^T\big)^{-1}$. Therefore using Lemma 5.1,

$$\begin{gathered} R \succeq \Big[\big(CPC^T\big)^{-1} + R^{-1}\Big]^{-1} \Rightarrow \\ \big(CPC^T\big)^{-1}R\big(CPC^T\big)^{-1} \succeq \big(CPC^T\big)^{-1}\Big[\big(CPC^T\big)^{-1} + R^{-1}\Big]^{-1}\big(CPC^T\big)^{-1} \Rightarrow \\ \big(CPC^T\big)^{-1}\big(CPC^T - R\big)\big(CPC^T\big)^{-1} \preceq \big(CPC^T + R\big)^{-1} \Rightarrow \\ \Phi_{\text{no}}\big(P,\mu,R\big) \succeq \Phi_{\text{full}}\big(P,\mu,R\big). \end{gathered} \tag{5.21}$$

Furthermore, for any $\tilde{K}_f$, $\big(A - \tilde{K}_f C\big)P\big(A - \tilde{K}_f C\big)^T \succeq 0$ and $\tilde{K}_f R \tilde{K}_f^T \succeq 0$. Then from Eq. 5.9,

$$\Phi_{\text{full}}\big(P,\mu,R\big) \succeq \mu APA^T + Q = \Phi_{\text{perf}}\big(P,\mu\big). \tag{5.22}$$

From the third line of Eq. 5.19, it can be seen that $\Phi_{\text{full}}\big(P,\mu,R\big)$ is a non-decreasing functions of $P \succ 0$. Moreover, $\Phi_{\text{perf}}\big(P,\mu\big)$ is a non-decreasing function of $P \succ 0$. Then for $P_1 \succeq P_2 \succeq P_3 \succ 0$, we obtain

$$\Phi_{\text{no}}\big(P_1,\mu,R\big) \succeq \Phi_{\text{full}}\big(P_1,\mu,R\big) \succeq \Phi_{\text{full}}\big(P_2,\mu,R\big) \succeq \Phi_{\text{perf}}\big(P_2,\mu\big) \succeq \Phi_{\text{perf}}\big(P_3,\mu\big). \tag{5.23}$$

□

Lemma 5.5. *Let $P_1[k+1] = \Phi_{no}\big(P_1[k],\mu[k],R[k]\big)$, $P_2[k+1] = \Phi_{full}\big(P_2[k],\mu[k],R[k]\big)$ and $P_3[k+1] = \Phi_{perf}\big(P_3[k],\mu[k]\big)$. Then starting from the same initial error covari-*

ance P_0, we have

$$\mathbb{E}\big(P_1[k]\big) \succeq \mathbb{E}\big(P_2[k]\big) \succeq \mathbb{E}\big(P_3[k]\big), \qquad \text{for } k \geq 0. \tag{5.24}$$

Proof. At $k=0$, $P_1[0]=P_2[0]=P_3[0]=P_0$. Assume that at time k, $P_1[k] \succeq P_2[k] \succeq P_3[k]$. Then, from Lemma 5.4 we have,

$$\begin{aligned} \Phi_{\text{no}}\big(P_1[k],\mu[k],R[k]\big) \succeq \Phi_{\text{full}}\big(P_2[k],\mu[k],R[k]\big) \succeq \Phi_{\text{perf}}\big(P_3[k],\mu[k]\big) \Rightarrow \\ P_1[k+1] \succeq P_2[k+1] \succeq P_3[k+1]. \end{aligned} \tag{5.25}$$

This proves that starting from the same P_0, $P_1[k] \succeq P_2[k] \succeq P_3[k]$ for $k \geq 0$ (by using mathematical induction). Furthermore, at any time k, $P_1[k]$, $P_2[k]$ and $P_3[k]$ can be expressed as functions of P_0 and the sequence $\Upsilon[0],\cdots,\Upsilon[k-1]$:
$P_1[k] = \Lambda_{\text{no}}\big(P_0,\Upsilon[0],\cdots,\Upsilon[k-1]\big)$, $P_2[k] = \Lambda_{\text{full}}\big(P_0,\Upsilon[0],\cdots,\Upsilon[k-1]\big)$ and $P_3[k] = \Lambda_{\text{perf}}\big(P_0,\Upsilon[0],\cdots,\Upsilon[k-1]\big)$. Therefore

$$\mathbb{E}\big(P_1[k]\big) = \int_0^\infty \cdots \int_0^\infty \Lambda_{\text{no}}\big(P_0,\Upsilon[0],\cdots,\Upsilon[k-1]\big)\left[\prod_{j=0}^{k-1}\chi(\Upsilon_j)\right] d\Upsilon[0]\cdots d\Upsilon[k-1]. \tag{5.26}$$

Similar expressions can be written for $\mathbb{E}\big(P_2[k]\big)$ and $\mathbb{E}\big(P_3[k]\big)$. Since $\chi(\Upsilon_j)$ is non-negative for $j=0,\cdots,k-1$ and $\Lambda_{\text{no}}\big(P_0,\Upsilon[0],\cdots,\Upsilon[k-1]\big) \succeq \Lambda_{\text{full}}\big(P_0,\Upsilon[0],\cdots,\Upsilon[k-1]\big) \succeq \Lambda_{\text{perf}}\big(P_0,\Upsilon[0],\cdots,\Upsilon[k-1]\big)$, we have $\mathbb{E}\big(P_1[k]\big) \succeq \mathbb{E}\big(P_2[k]\big) \succeq \mathbb{E}\big(P_3[k]\big)$. □

The next theorem shows that the stability region of the full knowledge and no knowledge cases are the same.

Theorem 5.1. *The dynamical system of Eq. 5.19 is stable if and only if $\mu_{ave}\big(\Upsilon_T\big) < \rho_{\max}^{-2}(A)$ or equivalently $\Upsilon_T < \Upsilon_{T,c}$, where $\rho_{\max}(A)$ and $\Upsilon_{T,c}$ are as defined in Lemma 5.3.*

Proof. From Lemma 5.5, we know that the average estimation error variance of the full-knowledge case is upper bounded by that of the no-knowledge case and lower bounded by that of the noise-free case. Therefore, from Lemma 5.3 and Eq. 5.14 for the stability of the noise-free case, we can see that $\mu_{\text{ave}}\big(\Upsilon_T\big) < \rho_{\max}^{-2}(A)$ is the necessary and sufficient condition for ensuring the stability of the case of full knowledge. □

It is also possible to deduce the same final result by directly relating the average of the dynamics of the full-knowledge case to those of noise-free and no-knowledge cases. We show this alternative approach in the appendix.

5.3.4 Estimator has partial knowledge on the reception quality

Consider Eq. 5.10 with $0 \prec \tilde{R}[k] \prec R[k]$:

$$\begin{aligned} P[k+1] =& AP[k]A^T + Q - \big(1-\mu[k]\big)AP[k]C^T\big(CP[k]C^T + \tilde{R}[k]\big)^{-1} \\ & \times \big(CP[k]C^T + 2\tilde{R}[k] - R[k]\big)\big(CP[k]C^T + \tilde{R}[k]\big)^{-1}CP[k]A^T \\ \triangleq & \Phi_{\text{par}}\big(P[k],\mu[k],R[k],\tilde{R}[k]\big). \end{aligned} \tag{5.27}$$

To facilitate mathematical derivations of the partial-knowledge case, we assume diagonal $\tilde{R}$ and R in this part: $\tilde{R}[k] = \tilde{\sigma}_n^2\big(\Upsilon[k]\big)I$ and $R[k] = \sigma_n^2\big(\Upsilon[k]\big)I$.

Lemma 5.6. *Assume $\Pi \succ 0$ and $0 \leq \xi_2 \leq \xi_1 \leq \xi_{\max}$. Then $\big(\Pi+\xi_1 I\big)^{-1}\big(\Pi+2\xi_1 I - \xi_{\max}I\big)\big(\Pi+\xi_1 I\big)^{-1} \succeq \big(\Pi+\xi_2 I\big)^{-1}\big(\Pi+2\xi_2 I-\xi_{\max}I\big)\big(\Pi+\xi_2 I\big)^{-1}$.*

Proof. Define $\xi_{\text{diff}} \triangleq \xi_1 - \xi_2 \geq 0$ and $\tilde{\Pi} \triangleq \Pi + \xi_2 I$. Since $\xi_2 - \xi_{\max} \leq 0$ and $\xi_1 + \xi_2 - 2\xi_{\max} \leq 0$, we have

$$\begin{gathered} \tilde{\Pi}^3 + (\xi_2 - \xi_{\max} + 2\xi_{\text{diff}})\tilde{\Pi}^2 \succeq \\ \tilde{\Pi}^3 + (\xi_2 - \xi_{\max} + 2\xi_{\text{diff}})\tilde{\Pi}^2 + \xi_{\text{diff}}^2(\xi_2-\xi_{\max})I + \xi_{\text{diff}}\underbrace{(2\xi_2 - 2\xi_{\max} + \xi_{\text{diff}})}_{\xi_1+\xi_2-2\xi_{\max}}\tilde{\Pi} = \\ \big(\tilde{\Pi}^2 + \xi_{\text{diff}}\tilde{\Pi}\big)\big(\tilde{\Pi}^{-1} + (\xi_2-\xi_{\max})\tilde{\Pi}^{-2}\big)\big(\tilde{\Pi}^2 + \xi_{\text{diff}}\tilde{\Pi}\big) \end{gathered} \tag{5.28}$$

or equivalently $\tilde{\Pi}\big(\tilde{\Pi} + (\xi_2 + 2\xi_{\text{diff}} - \xi_{\max})I\big)\tilde{\Pi} \succeq \tilde{\Pi}\big(\tilde{\Pi}+\xi_{\text{diff}}I\big)\tilde{\Pi}^{-1}\big(\tilde{\Pi} + (\xi_2 - \xi_{\max})I\big)\tilde{\Pi}^{-1}\big(\tilde{\Pi}+\xi_{\text{diff}}I\big)\tilde{\Pi}$. Let $\Psi = \big(\Pi+\xi_1 I\big)^{-1}\big(\Pi+\xi_2 I\big)^{-1}$. Then, using Lemma 5.1, we have,

$$\begin{gathered} \Psi\tilde{\Pi}\big(\tilde{\Pi} + (\xi_2 + 2\xi_{\text{diff}} - \xi_{\max})I\big)\tilde{\Pi}\Psi^T \succeq \\ \Psi\tilde{\Pi}\big(\tilde{\Pi}+\xi_{\text{diff}}I\big)\tilde{\Pi}^{-1}\big(\tilde{\Pi}+(\xi_2-\xi_{\max})I\big)\tilde{\Pi}^{-1}\big(\tilde{\Pi}+\xi_{\text{diff}}I\big)\tilde{\Pi}\Psi^T, \end{gathered} \tag{5.29}$$

which can be easily verified to be the same as the desired inequality. □

Lemma 5.7. *Consider $0 \leq \mu \leq 1, \tilde{R} = \tilde{\sigma}_n^2 I, R = \sigma_n^2 I$ and $P \succ 0$. Then, $\Phi_{par}\big(P,\mu,R,\tilde{R}\big)$ is a non-increasing function of $\tilde{\sigma}_n^2 \in [0,\sigma_n^2]$.*

Proof. For $\tilde{R} = \tilde{\sigma}_n^2 I$ and $R = \sigma_n^2 I$, $\Phi_{\text{par}}\big(P,\mu,R,\tilde{R}\big)$ can be written as

$$\begin{aligned} \Phi_{\text{par}}\big(P,\mu,R,\tilde{R}\big) =& APA^T + Q - (1-\mu)APC^T\big(CPC^T + \tilde{\sigma}_n^2 I\big)^{-1} \\ & \times \big(CPC^T + 2\tilde{\sigma}_n^2 I - \sigma_n^2 I\big)\big(CPC^T + \tilde{\sigma}_n^2 I\big)^{-1}CPA^T. \end{aligned} \tag{5.30}$$

Then Lemma 5.6 implies that $\big(CPC^T + \tilde{\sigma}_n^2 I\big)^{-1}\big(CPC^T + 2\tilde{\sigma}_n^2 I - \sigma_n^2 I\big)\big(CPC^T + \tilde{\sigma}_n^2 I\big)^{-1}$ is a non-decreasing function of $\tilde{\sigma}_n^2$ and as a result $\Phi_{\text{par}}\big(P,\mu,R,\tilde{R}\big)$ is a non-increasing function of $\tilde{\sigma}_n^2$. □

Lemma 5.7 indicates that as the partial knowledge increases ($\frac{\tilde{\sigma}_n^2}{\sigma_n^2} \to 1$), the error covariance $P[k+1]$ decreases, given the same $P[k]$. Also one can easily confirm, using Lemma 5.7, that for diagonal R and $\tilde{R}$:

$$\Phi_{\text{no}}\big(P,\mu,R\big) \succeq \Phi_{\text{par}}\big(P,\mu,R,\tilde{R}\big) \succeq \Phi_{\text{full}}\big(P,\mu,R\big) \succeq \Phi_{\text{perf}}\big(P,\mu\big). \tag{5.31}$$

Since characterizing the necessary and sufficient stability condition for the partial case is considerably challenging, we next make an approximation by assuming that $\tilde{R}$ is small. This approximation will then help us establish that $\Phi_{\text{par}}\big(P,\mu,R,\tilde{R}\big)$ is non-decreasing as a function of $P \succ 0$, which will then help us characterize the stability region.

Lemma 5.8. *Let $\rho_{min}(\Pi)$ represent the minimum eigenvalue of symmetric $\Pi \succ 0$. If $\Pi \succ 0$, $\xi_2 \geq 0$ and $0 \leq \xi_1 \ll \rho_{min}(\Pi)$, then we have the following approximation by only keeping the first-order terms: $\big(\Pi+\xi_1 I\big)^{-1}\big(\Pi+2\xi_1 I-\xi_2 I\big)\big(\Pi+\xi_1 I\big)^{-1} \approx \Pi^{-1}\Big(\Pi-\xi_2 I+2\xi_1\xi_2\Pi^{-1}\Big)\Pi^{-1}$.*

Proof. We have $\big(\Pi+\xi_1 I\big)\big(\Pi^{-1}-\xi_1\Pi^{-2}\big) = I-\xi_1^2\Pi^{-2}$. Therefore, for $0 \leq \xi_1 \ll \rho_{\min}(\Pi)$, we have $\|\xi_1^2\Pi^{-2}\| \ll 1$, which results in the following approximations by only keeping the first-order terms as a function of ξ_1:

$$\big(\Pi+\xi_1 I\big)^{-1} \approx \Pi^{-1}-\xi_1\Pi^{-2} \text{ and } \big(\Pi+\xi_1 I\big)^{-2} \approx \Pi^{-2}-2\xi_1\Pi^{-3}. \tag{5.32}$$

Then, we have $\big(\Pi+\xi_1 I\big)^{-1}\big(\Pi+2\xi_1 I-\xi_2 I\big)\big(\Pi+\xi_1 I\big)^{-1} = \big(\Pi+\xi_1 I\big)^{-1}-(\xi_2-\xi_1)\big(\Pi+\xi_1 I\big)^{-2}$. But

$$\begin{aligned}\big(\Pi+\xi_1 I\big)^{-1}-(\xi_2-\xi_1)\big(\Pi+\xi_1 I\big)^{-2} &\approx \Pi^{-1}-\xi_1\Pi^{-2}-(\xi_2-\xi_1)\big(\Pi^{-2}-2\xi_1\Pi^{-3}\big)\\ &= \Pi^{-1}\Big(\underbrace{I-2\xi_1^2\Pi^{-2}}_{\approx I}-\xi_2\Pi^{-1}+2\xi_1\xi_2\Pi^{-2}\Big)\\ &\approx \Pi^{-1}-\xi_2\Pi^{-2}+2\xi_1\xi_2\Pi^{-3}\\ &= \Pi^{-1}\Big(\Pi-\xi_2 I+2\xi_1\xi_2\Pi^{-1}\Big)\Pi^{-1}.\end{aligned} \tag{5.33}$$

□

Lemma 5.9. *Let $\tilde{R} = \tilde{\sigma}_n^2 I$, $R = \sigma_n^2 I$ and assume that $\tilde{\sigma}_n^2 \ll \rho_{min}\big(CPC^T\big)$. Then, the first-order approximation of $\Phi_{par}\big(P,\mu,R,\tilde{R}\big)$ is a non-decreasing function of $P \succ 0$.*

Proof. By using Lemma 5.8, we can approximate $\Phi_{\text{par}}\big(P,\mu,R,\tilde{R}\big)$ as follows

$$\begin{aligned}\Phi_{\text{par}}\big(P,\mu,R,\tilde{R}\big) \approx{}& \mu APA^T+Q+(1-\mu)\sigma_n^2 A\big(C^TC\big)^{-1}A^T\\ &-2(1-\mu)\tilde{\sigma}_n^2\sigma_n^2 A\big(C^TC\big)^{-1}P^{-1}\big(C^TC\big)^{-1}A^T,\end{aligned} \tag{5.34}$$

which is a non-decreasing function of P. □

Lemma 5.10. *Let $\tilde{R}[k] = \tilde{\sigma}_n^2\big(\Upsilon[k]\big)I$, $R[k] = \sigma_n^2\big(\Upsilon[k]\big)I$ and $\tilde{\sigma}_n^2\big(\Upsilon[k]\big) \ll \rho_{min}\big(CP[k]C^T\big)$ for all k. Let $P_1[k+1] = \Phi_{no}\big(P_1[k],\mu[k],R[k]\big)$, $P_2[k+1] = \Phi_{par}\big(P_2[k],\mu[k],R[k],\tilde{R}[k]\big)$,*

$P_3[k+1] = \Phi_{full}\big(P_3[k], \mu[k], R[k]\big)$ and $P_4[k+1] = \Phi_{perf}\big(P_4[k], \mu[k]\big)$. Then starting from the same initial P_0, we have

$$\mathbb{E}\big(P_1[k]\big) \succeq \mathbb{E}\big(P_2[k]\big) \succeq \mathbb{E}\big(P_3[k]\big) \succeq \mathbb{E}\big(P_4[k]\big), \qquad k \geq 0. \tag{5.35}$$

Proof. From Eq. 5.31, we have the following for diagonal $R[k]$ and $\tilde{R}[k]$ and any $P[k] \succ 0$:

$$\begin{aligned} &\Phi_{\text{no}}\big(P[k], \mu[k], R[k]\big) \succeq \Phi_{\text{par}}\big(P[k], \mu[k], R[k], \tilde{R}[k]\big) \succeq \\ &\Phi_{\text{full}}\big(P[k], \mu[k], R[k]\big) \succeq \Phi_{\text{perf}}\big(P[k], \mu[k]\big). \end{aligned} \tag{5.36}$$

For $\tilde{\sigma}_n^2\big(\Upsilon[k]\big) \ll \rho_{\min}\big(CP[k]C^T\big)$, Φ_{par} is a non-decreasing function of P according to Lemma 5.9. Therefore, we have the following for $P_1[k] \succeq P_2[k] \succeq P_3[k] \succeq P_4[k] \succ 0$,

$$\begin{aligned} &\Phi_{\text{no}}\big(P_1[k], \mu[k], R[k]\big) \succeq \Phi_{\text{par}}\big(P_1[k], \mu[k], R[k], \tilde{R}[k]\big) \succeq \\ &\Phi_{\text{par}}\big(P_2[k], \mu[k], R[k], \tilde{R}[k]\big) \succeq \Phi_{\text{full}}\big(P_2[k], \mu[k], R[k]\big) \succeq \\ &\Phi_{\text{full}}\big(P_3[k], \mu[k], R[k]\big) \succeq \Phi_{\text{perf}}\big(P_3[k], \mu[k]\big) \succeq \Phi_{\text{perf}}\big(P_4[k], \mu[k]\big). \end{aligned} \tag{5.37}$$

Then, Eq. 5.35 can be easily proved similar to Lemma 5.5. □

The next theorem shows that, under the assumptions we made, the stability region of the partial-knowledge case is the same as the full knowledge and no knowledge cases.

Theorem 5.2. *Assume $\tilde{R}[k] = \tilde{\sigma}_n^2\big(\Upsilon[k]\big)I$, $R[k] = \sigma_n^2\big(\Upsilon[k]\big)I$ and for all k we have $\tilde{\sigma}_n^2\big(\Upsilon[k]\big) \ll \rho_{min}\big(CP[k]C^T\big)$. Then the dynamical system of Eq. 5.27 is stable if and only if $\mu_{ave}\big(\Upsilon_T\big) < \rho_{\max}^{-2}(A)$ or equivalently $\Upsilon_T < \Upsilon_{T,c}$, where $\rho_{\max}(A)$ and $\Upsilon_{T,c}$ are as defined in Lemma 5.3.*

Proof. By utilizing Lemma 5.10, this can be proved similar to Theorem 5.1. □

Fig. 5.3 shows the norm of the asymptotic average estimation error variance for the three cases of full knowledge, partial knowledge and no knowledge, and as a function of Υ_T. For this case, SNR is exponentially distributed with the average of 30dB, $\sigma_n^2(\Upsilon) = 0.50 + 533.3 \times \Omega\big(\sqrt{\Upsilon}\big)$, where $\Omega(d) = \frac{1}{\sqrt{2\pi}}\int_d^\infty e^{-t^2/2}dt$ for an arbitrary d and $\tilde{\sigma}_n^2(\Upsilon) = 0.5$. This is the variance of the communication noise for a binary modulation system that utilizes gray coding [31]. The following parameters are chosen for this example: $A = \begin{pmatrix} 2 & 0.3 & 0.45 \\ 0.4 & 0.2 & 0.5 \\ 1.5 & 0.6 & 0.34 \end{pmatrix}$, $Q = 0.001I$ and $C = 2I$, which results in $\Upsilon_{T,c} = 22.53$ dB. As can be seen, all the three curves have the same critical stability point, indicated by $\Upsilon_{T,c}$. As expected, the more is known about the link qualities, the better the performance is. The figure also shows that not all $\Upsilon_T < \Upsilon_{T,c}$ result in the same performance and that there is an optimum threshold that minimizes the average estimation error variance. Υ_T impacts both packet drop and the amount of communication noise that enters the estimation process. Therefore, optimizing it can properly control the overall performance, as we shall show in the next section.

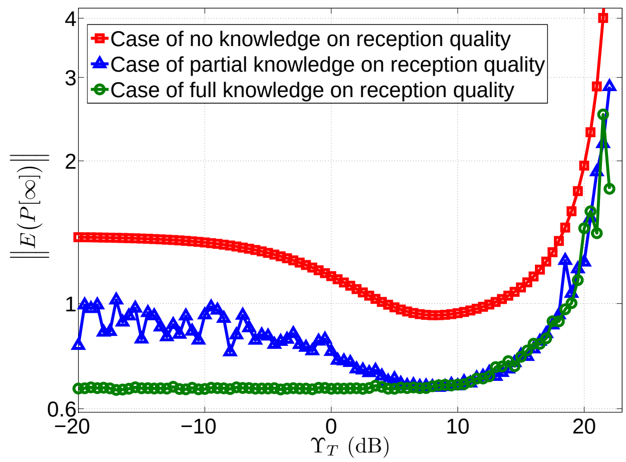

Fig. 5.3: Impact of the knowledge available on the link qualities on the performance and stability of Kalman filtering over fading channels.

5.4 Optimization of Packet Drop in the physical layer

In this part, we consider the optimization of packet drop in the physical layer. In Eq. 5.10, the threshold Υ_T impacts both the amount of information loss ($\mu[k]$) and the quality of those packets that are kept ($R[k]$). If this threshold is chosen very high, the information loss rate could become considerably high, depending on the quality of the link. However, once a packet is kept, the communication error that enters the estimation process will be small. On the other hand, for low thresholds, information loss rate will be lower at the cost of more noisy receptions. In this section, we are interested in characterizing the optimum threshold for the two cases that the estimator has "full knowledge" and "no knowledge" on the communication quality. Since the optimization of the threshold occurs in the physical layer, in this section we assume that the physical layer has full knowledge of SNR and the corresponding communication error variance function. For instance, for the case where the estimator has no knowledge on the reception quality, the physical layer would still have link quality information, based on which it will optimize the threshold. However, the information on the link quality is not transferred to the application layer, which is in charge of estimation. In other words, the physical layer controls the impact of the communication links on the estimation process. For the case where the estimator has full knowledge on the reception quality, on the other hand, the physical layer constantly passes on the information on the link qualities to the estimator. Therefore, both the physical and application layers utilize this knowledge. The framework of this section can then be extended to other scenarios, for instance to the case where the physical layer only has partial knowledge on the link qualities (such as an upper bound). While in the previous section, we took the threshold to be constant during the estimation process, in this section we consider both adaptive (time-varying) and non-adaptive thresholding. In the adaptive case, we show how the threshold can be optimized at every time step in order to minimize the average estimation error variance of the next time. For the non-adaptive thresholding, we find the optimum threshold that minimizes the asymptotic average estimation error variance. There are interesting tradeoffs between the two approaches, as discussed in this section.

5.4.1 Estimator has no knowledge on the reception quality

In this part we consider the case where the estimator has no information on the reception quality. We consider both scenarios with fixed and adaptive thresholding. From Eq. 5.15, we have $P[k+1]=\mu[k]AP[k]A^T+Q+(1-\mu[k])AC^{-1}R[k]\left(C^T\right)^{-1}A^T$. Both $\mu[k]$ and $R[k]$ are functions of $\Upsilon_T[k]$, which is taken to be time-varying to allow for adaptive thresholding. Since the stability analysis of the previous section was carried out assuming that Υ_T is time-invariant, we need to first establish the stability of the time-varying case. For the derivations of this section, we take $R[k]$ to be diagonal: $R[k]=\sigma_n^2(\Upsilon[k])I$. We have,

$$P[k+1]=\mu[k]AP[k]A^T+Q+\underbrace{(1-\mu[k])\sigma_n^2\big(\Upsilon[k]\big)}_{\beta[k]}A\big(C^TC\big)^{-1}A^T, \tag{5.38}$$

with the following solution as a function of the initial variance P_0:

$$\begin{aligned}P[k]=\left(\prod_{i=0}^{k-1}\mu[i]\right)A^kP_0\big(A^T\big)^k+\\ \sum_{i=0}^{k-1}\left(\prod_{j=i+1}^{k-1}\mu[j]\right)A^{k-i-1}\Big(Q+\beta[i]A\big(C^TC\big)^{-1}A^T\Big)\big(A^T\big)^{k-i-1}.\end{aligned} \tag{5.39}$$

At the k^{th} time step, $\mu[k]$ and $\beta[k]$ are functions of the instantaneous received SNR, $\Upsilon[k]$. Since $\Upsilon[k_1]$ and $\Upsilon[k_2]$, for $k_1\neq k_2$, are independent, $\mu[k_1]$ and $\mu[k_2]$ as well as $\mu[k_1]$ and $\beta[k_2]$ are also independent for $k_1\neq k_2$. Therefore,

$$\begin{aligned}\mathbb{E}\Big(P[k]\Big)=\left(\prod_{i=0}^{k-1}\mu_{\text{ave}}\big(\Upsilon_T[i]\big)\right)A^kP_0\big(A^T\big)^k+\\ \sum_{i=0}^{k-1}\left(\prod_{j=i+1}^{k-1}\mu_{\text{ave}}\big(\Upsilon_T[j]\big)\right)A^{k-i-1}\Big(Q+\sigma_{n,\text{ave}}^2\big(\Upsilon_T[i]\big)A\big(C^TC\big)^{-1}A^T\Big)\big(A^T\big)^{k-i-1},\end{aligned} \tag{5.40}$$

where $\mu_{\text{ave}}\big(\Upsilon_T[k]\big)$ and $\sigma_{n,\text{ave}}^2\big(\Upsilon_T[k]\big)$ represent time-varying average probability of packet loss (spatial averaging over fading) and average noise variance that entered the estimation process respectively:

$$\mu_{\text{ave}}\big(\Upsilon_T[k]\big)=\mathbb{E}\big(\mu[k]\big)=\int_0^{\Upsilon_T[k]}\chi(\Upsilon)d\Upsilon \tag{5.41}$$

and

$$\sigma^2_{n,\text{ave}}\big(\Upsilon_T[k]\big) = \mathbb{E}\big(\beta[k]\big) = \int_{\Upsilon_T[k]}^{\infty} \sigma_n^2(\Upsilon)\chi(\Upsilon)d\Upsilon, \tag{5.42}$$

where χ represents the probability density function of Υ. Then Eq. 5.40 is the solution to the following average dynamical system:

$$\mathbb{E}\big(P[k+1]\big) = \mu_{\text{ave}}\big(\Upsilon_T[k]\big)A\mathbb{E}\big(P[k]\big)A^T + Q + \sigma^2_{n,\text{ave}}\big(\Upsilon_T[k]\big)A(C^TC)^{-1}A^T. \tag{5.43}$$

It can be seen that for a time-varying $\Upsilon_T[k]$, we have a time-varying average dynamical system, for which we have to ensure stability.

Lemma 5.11. *Let $\rho_{\max}(A)$ denote the spectral radius of matrix A. If for an arbitrary small $\varepsilon > 0$, $\mu_{ave}(\Upsilon_T[k]) \le \underbrace{(1-\varepsilon)\rho_{\max}^{-2}(A)}_{\mu_{ave,max}}$, then the average dynamical system of Eq. 5.43 will be stable.*

Proof. Let $\sigma^2_{n,\text{ave,max}} = \max_{\Upsilon_T[k]} \ \sigma^2_{n,\text{ave}}\big(\Upsilon_T[k]\big) = \int_0^\infty \sigma_n^2\big(\Upsilon\big)\chi(\Upsilon)d\Upsilon$. Then,

$$\mathbb{E}\big(P[k+1]\big) \preceq \mu_{\text{max,ave}}A\mathbb{E}\big(P[k]\big)A^T + Q + \sigma^2_{n,\text{ave,max}}A\big(C^TC\big)^{-1}A^T. \tag{5.44}$$

Starting from the same initial condition, $\mathbb{E}\big(P[k]\big)$ is upper-bounded by the solution of $\mathbb{E}\big(P_m[k+1]\big) = \mu_{\text{ave,max}}A\mathbb{E}\big(P_m[k]\big)A^T + Q + \sigma^2_{n,\text{ave,max}}A\big(C^TC\big)^{-1}A^T$, which is stable [27]. This shows that $\mu_{\text{ave}}\big(\Upsilon_T[k]\big) \le \mu_{\text{ave,max}}$ is a sufficient condition for ensuring the stability of Eq. 5.43. □

Note that $\mu_{\text{ave}}(\Upsilon_T[k]) \le \mu_{\text{ave,max}}$ is equivalent to $\Upsilon_T[k] \le \Upsilon^{\varepsilon}_{T,c}$, where $\Upsilon^{\varepsilon}_{T,c}$ is such that $\int_0^{\Upsilon^{\varepsilon}_{T,c}} \chi(\Upsilon)d\Upsilon = (1-\varepsilon)\rho_{\max}^{-2}(A)$ for an unstable A. Next, we develop a foundation for the optimization of the threshold. Fig. 5.4 shows examples of μ_{ave} and $\sigma^2_{n,\text{ave}}$ as a function of the threshold. It can be seen that the optimum threshold should properly control and balance the amount of information loss and communication error that enters the estimation process, as we shall characterize in this section.

5.4.1.1 Adaptive Thresholding

In this part, we consider the case where $\Upsilon_T[k]$ is time varying and is optimized at every time step based on the current estimation error variance. Consider Eq. 5.38. We have the following conditional average for the estimation error variance at time $k+1$:

$$\mathbb{E}\Big(P[k+1]\ \Big|\ P[k]\Big) = AP[k]A^T + Q + \int_{\Upsilon_T[k]}^{\infty}\Big[\sigma_n^2\big(\Upsilon\big)A\big(C^TC\big)^{-1}A^T - AP[k]A^T\Big]\chi(\Upsilon)d\Upsilon. \tag{5.45}$$

Then at the k^{th} time step, our goal is to find $\Upsilon_T[k]$ such that $\mathbb{E}\Big(P[k+1]\ \Big|\ P[k]\Big)$ is minimized.

Minimization of Trace:

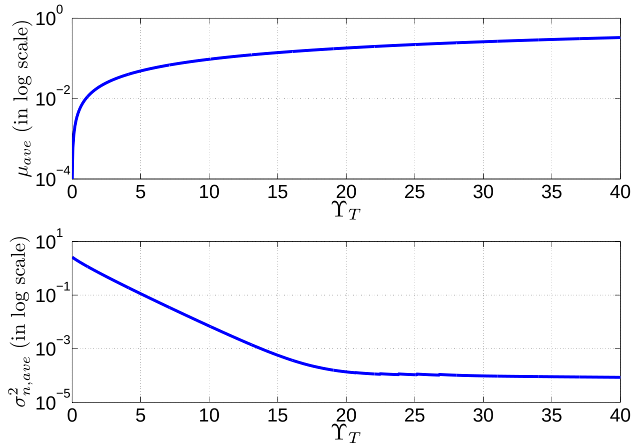

Fig. 5.4: Examples of the average probability of packet drop and average overall communication noise variance as a function of Υ_T. As the threshold increases, μ_{ave} increases while communication noise variance $\sigma^2_{n,\text{ave}}$ decreases.

Consider minimizing $\text{tr}\Big(\mathbb{E}\Big(P[k+1]\,\Big|\,P[k]\Big)\Big)$, where tr(.) denotes the trace of the argument. We have

$$\text{tr}\Big(\mathbb{E}\Big(P[k+1]\,\Big|\,P[k]\Big)\Big) = \text{tr}\Big(AP[k]A^T + Q\Big) + \int_{\Upsilon_T[k]}^{\infty}\Big[\sigma_n^2(\Upsilon)\,\text{tr}\Big(A\big(C^TC\big)^{-1}A^T\Big) - \text{tr}\Big(AP[k]A^T\Big)\Big]\chi(\Upsilon)d\Upsilon, \tag{5.46}$$

and the following optimization problem:

$$\Upsilon^*_{T,\text{trace,adp}}[k] = \arg\min_{\Upsilon_T[k]} \int_{\Upsilon_T[k]}^{\infty} \underbrace{\Big[\sigma_n^2(\Upsilon)\,\text{tr}\Big(A\big(C^TC\big)^{-1}A^T\Big) - \text{tr}\Big(AP[k]A^T\Big)\Big]}_{\Theta(\Upsilon,P[k])}\chi(\Upsilon)d\Upsilon,$$
$$\text{subject to} \quad 0 \le \Upsilon_T[k] \le \Upsilon^{\varepsilon}_{T,c} \tag{5.47}$$

with $\Upsilon^*_{T,\text{trace,adp}}[k]$ denoting the optimum threshold at time step k, when minimizing the trace in the adaptive case.

Note that $\sigma_n^2(\Upsilon)$ and, as a result $\Theta(\Upsilon,P[k])$, is a non-increasing function of Υ. Furthermore, $\chi(\Upsilon)$ is always non-negative, which results in $\int_{\Upsilon_T[k]}^{\infty}\Theta(\Upsilon,P[k])\chi(\Upsilon)d\Upsilon$ minimized when $\Theta(\Upsilon,P[k])$ changes sign. Thus, three different cases can happen:

1. $\lim_{\Upsilon\to 0}\Theta(\Upsilon,P[k]) \ge 0$ and $\lim_{\Upsilon\to\infty}\Theta(\Upsilon,P[k]) \ge 0$: In this case, $\Theta(\Upsilon,P[k])$ is non-negative as a function of Υ. Therefore, $\Upsilon^*_{T,\text{trace,adp}}[k] = \Upsilon^{\varepsilon}_{T,c}$. One possible scenario that results in non-negative $\Theta(\Upsilon,P[k])$ is when $\text{tr}\Big(AP[k]A^T\Big)$ is considerably low.

2. $\lim_{\Upsilon\to 0}\Theta\big(\Upsilon,P[k]\big)\le 0$ and $\lim_{\Upsilon\to\infty}\Theta\big(\Upsilon,P[k]\big)<0$: In this case $\Theta\big(\Upsilon,P[k]\big)$ is negative as a function of Υ. Therefore, $\Upsilon^*_{T,\text{trace,adp}}[k]=0$, i.e. the next packet is kept independent of its quality. For instance, if $\operatorname{tr}\Big(AP[k]A^T\Big)$ is considerably high, it can result in a negative $\Theta\big(\Upsilon,P[k]\big)$.
3. $\lim_{\Upsilon\to 0}\Theta\big(\Upsilon,P[k]\big)>0$ and $\lim_{\Upsilon\to\infty}\Theta\big(\Upsilon,P[k]\big)<0$: In this case the optimum threshold is $\Upsilon^*_{T,\text{trace,adp}}[k]=\min\big\{\Upsilon^{\varepsilon}_{T,c},\gamma^*_{T,\text{trace,adp}}\big\}$ where $\gamma^*_{T,\text{trace,adp}}$ is the unique solution to the following equation:

$$\Theta\big(\gamma^*_{T,\text{trace,adp}},P[k]\big)=0\;\Rightarrow\;\sigma_n^2\big(\gamma^*_{T,\text{trace,adp}}\big)\operatorname{tr}\Big(A\big(C^TC\big)^{-1}A^T\Big)=\operatorname{tr}\Big(AP[k]A^T\Big). \tag{5.48}$$

Minimization of Determinant:
Consider Eq. 5.45, which can be written as:

$$\mathbb{E}\Big(P[k+1]\;\Big|\;P[k]\Big)=\mu_{\text{ave}}(\Upsilon_T[k])AP[k]A^T+\sigma^2_{n,\text{ave}}(\Upsilon_T[k])A\big(C^TC\big)^{-1}A^T+Q. \tag{5.49}$$

Lemma 5.12. *Consider the case where Q is negligible in Eq. 5.49. The optimum adaptive threshold, $\Upsilon^*_{T,det,adp}[k]$, that minimizes the determinant of $\mathbb{E}\Big(P[k+1]\;\Big|\;P[k]\Big)$ is then the solution to the following optimization problem:*

$$\begin{aligned}&\Upsilon^*_{T,det,adp}[k]=\operatorname{argmin}_{\Upsilon_T[k]}\;\prod_i\Big(\mu_{ave}(\Upsilon_T[k])+\zeta_i\sigma^2_{n,ave}(\Upsilon_T[k])\Big)\\&\text{subject to}\quad 0\le\Upsilon_T[k]\le\Upsilon^{\varepsilon}_{T,c},\end{aligned} \tag{5.50}$$

where $\Xi[k]=P^{-1/2}[k](C^TC)^{-1}P^{-1/2}[k]$ and ζ_is for $1\le i\le N$ are the eigenvalues of $\Xi[k]$.

Proof. Since $P[k]$ is positive definite, there exists a unique positive definite $P^{1/2}[k]$ such that $P[k]=P^{1/2}[k]P^{1/2}[k]$. We then have the following for negligible Q:

$$\mathbb{E}\Big(P[k+1]\;\Big|\;P[k]\Big)=AP^{1/2}[k]\Big(\mu_{\text{ave}}\big(\Upsilon_T[k]\big)I+\sigma^2_{n,\text{ave}}\big(\Upsilon_T[k]\big)\Xi[k]\Big)P^{1/2}[k]A^T, \tag{5.51}$$

where $\Xi[k]$ is as denoted earlier in the lemma with the following diagonalization $\Xi[k]=F\zeta_{\text{diag}}F^T$, with $FF^T=I$ and $\zeta_{\text{diag}}=\operatorname{diag}\{\zeta_1,\zeta_2,\ldots,\zeta_N\}$. Then,

$$\begin{aligned}\det\bigg(\mathbb{E}\Big(P[k+1]\;\Big|\;P[k]\Big)\bigg)&=\det\big(AP[k]A^T\big)\det\Big(\mu_{\text{ave}}\big(\Upsilon_T[k]\big)I+\sigma^2_{n,\text{ave}}\big(\Upsilon_T[k]\big)\zeta_{\text{diag}}\Big)\\&=\det\big(AP[k]A^T\big)\prod_i\Big(\mu_{\text{ave}}\big(\Upsilon_T[k]\big)+\zeta_i\sigma^2_{n,\text{ave}}\big(\Upsilon_T[k]\big)\Big).\end{aligned} \tag{5.52}$$

Therefore, $\Upsilon^*_{T,\text{det,adp}}[k]$ is the solution to the optimization problem of Eq. 5.50. □

We know that the optimum threshold is either at the boundaries of the feasible set ($\Upsilon_T[k] = 0$ and $\Upsilon_T[k] = \Upsilon^{\varepsilon}_{T,c}$) or is where the derivative is zero: $\frac{\partial \det\left(\mathbb{E}\left(P[k+1] \,\middle|\, P[k]\right)\right)}{\partial \Upsilon_T[k]} = 0 \Rightarrow \sum_i \frac{\chi(\Upsilon_T[k])\left(1-\zeta_i\sigma_n^2\left(\Upsilon_T[k]\right)\right)}{\mu_{\text{ave}}\left(\Upsilon_T[k]\right)+\zeta_i\sigma^2_{n,\text{ave}}\left(\Upsilon_T[k]\right)} = 0$. Among these points, the one that minimizes Eq. 5.52 is the optimum threshold. Therefore in practice, based on the given parameters such as the shape of the overall communication noise variance (σ_n^2) and the pdf of SNR, the optimum threshold can be identified for this case. Fig. 5.5 shows the

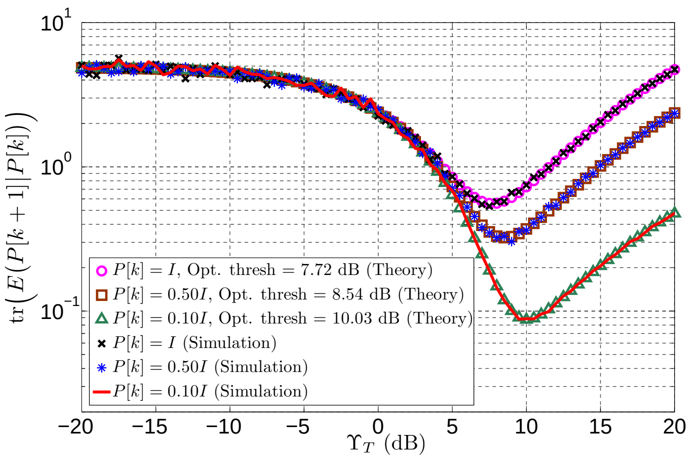

Fig. 5.5: Optimization of packet drop through one-step adaptive thresholding for three different $P[k]$s and average SNR of 20dB. This is for the case where knowledge of reception quality is not available at the estimator.

performance of adaptive thresholding at time $k+1$ given three different $P[k]$s. For this example, Υ is taken to have an exponential distribution with the average of 20dB and the overall communication noise variance is taken as follows: $\sigma_n^2(\Upsilon) = 1.27 \times 10^{-4} + 533.3 \times \Omega(\sqrt{\Upsilon})$, where $\Omega(d) = \frac{1}{\sqrt{2\pi}} \int_d^{\infty} e^{-t^2/2} dt$ for an arbitrary d. This is the variance of the communication noise for a binary modulation system that utilizes gray coding [31] and corresponds to 10 bits per sample and quantization step size of 0.0391. The rest of the parameters are the same as for Fig. 5.3. The optimum thresholds can be seen from the figure. Furthermore, the theoretical curves are confirmed with simulation results. As can be seen, the higher $\mathrm{tr}(AP[k]A^T)$ is (which corresponds to a higher $P[k]$ for this example), the lower the threshold will be. This can also be seen from the solution of the optimization problem of Eq. 5.47. For instance, in Eq. 5.48, the higher $\mathrm{tr}(AP[k]A^T)$ is, the lower the optimum threshold will be. This also makes intuitive sense as for higher $P[k]$s, the overall noise variance is more tolerable and the threshold should be lowered to avoid more information loss.

5.4.1.2 Non-Adaptive Thresholding through Minimization of $\mathbb{E}\big(P[\infty]\big)$

In this part, we consider the case where a non-adaptive threshold is used throughout the estimation process. We show how to optimize this threshold such that $\mathbb{E}\big(P[\infty]\big)$ is minimized [22]. We consider minimization of both norm and determinant of $\mathbb{E}\big(P[\infty]\big)$. The derivations can be similarly carried out to minimize the trace of $\mathbb{E}\big(P[\infty]\big)$. Consider Eq. 5.43. In this case, Υ_T will be time invariant. Then, using Lemma 5.2, we will have the following expression for $\mathbb{E}\big(P[\infty]\big)$:

$$\mathbb{E}\big(P[\infty]\big) = \sum_{i=0}^{\infty} \mu_{\text{ave}}^{i}(\Upsilon_T) A^i Q \big(A^T\big)^i + \sigma_{n,\text{ave}}^2(\Upsilon_T) \sum_{i=0}^{\infty} \mu_{\text{ave}}^{i}(\Upsilon_T) A^{i+1} (C^T C)^{-1} \big(A^T\big)^{i+1}. \tag{5.53}$$

Let $\Upsilon_{T,\text{norm,fixed}}^{*}$ represent the optimum way of dropping packets which will minimize the spectral norm of the asymptotic average estimation error variance for a non-adaptive threshold: $\Upsilon_{T,\text{norm,fixed}}^{*} = \text{argmin} \,\big\|\mathbb{E}\big(P[\infty]\big)\big\|$, for $\Upsilon_T < \Upsilon_{T,c}$. Also, let $\Upsilon_{T,\text{det,fixed}}^{*}$ represent the optimum way of dropping packets which will minimize the determinant of the asymptotic average estimation error variance: $\Upsilon_{T,\text{det,fixed}}^{*} = \text{argmin}\,\det\Big(\mathbb{E}\big(P[\infty]\big)\Big)$, for $\Upsilon_T < \Upsilon_{T,c}$. Since we need to find an exact expression for the norm and determinant of $\mathbb{E}\big(P[\infty]\big)$, we need to make a few simplifying assumptions in this part. More specifically, we assume that $C = \varsigma I$ and $Q = qI$. We furthermore take $A = A_s$, where A_s is a symmetric matrix, i.e. $A_s = A_s^T$. Under these assumptions, Eq. 5.53 can be rewritten as

$$\mathbb{E}\big(P[\infty]\big) = \varsigma^{-2}\sigma_{n,\text{ave}}^2(\Upsilon_T) \sum_{i=0}^{\infty} \mu_{\text{ave}}^{i}(\Upsilon_T) \big(A_s\big)^{2i+2} + q \sum_{i=0}^{\infty} \mu_{\text{ave}}^{i}(\Upsilon_T) \big(A_s\big)^{2i}. \tag{5.54}$$

Theorem 5.3. *(Balance of Information Loss & Communication Noise) Consider the average system dynamics of Eq. 5.54 with an unstable A_s. Then $\Upsilon_{T,norm,fixed}^{*}$ will be as follows:*

$$\Upsilon_{T,norm,fixed}^{*} = \begin{cases} \gamma_{T,norm,fixed}^{*} & \gamma_{T,norm,fixed}^{*} \geq 0 \\ 0 & \textit{otherwise} \end{cases} \tag{5.55}$$

where $\gamma_{T,norm,fixed}^{*}$ is the unique solution to the following equation:

$$\underbrace{\mu_{ave}(\gamma_{T,norm,fixed}^{*})}_{\textit{information loss}} + \underbrace{\sigma_{n,norm}^2(\gamma_{T,norm,fixed}^{*})}_{\textit{overall comm. noise}} + \frac{\varsigma^2 q}{\rho_{\max}^2 \sigma_n^2(\Upsilon = \gamma_{T,norm,fixed}^{*})} = \rho_{\max}^{-2}, \tag{5.56}$$

with $\sigma_{n,norm}^2$ denoting the normalized average overall communication noise variance: $\sigma_{n,norm}^2(\gamma_{T,norm,fixed}^{*}) = \frac{\sigma_{n,ave}^2(\gamma_{T,norm,fixed}^{*})}{\sigma_n^2(\Upsilon=\gamma_{T,norm,fixed}^{*})}$. Furthermore, $\Upsilon_{T,det,fixed}^{*}$ is determined as follows:

$$\Upsilon_{T,det,fixed}^{*} = \begin{cases} \gamma_{T,det,fixed}^{*} & \gamma_{T,det,fixed}^{*} \geq 0 \\ 0 & \textit{otherwise} \end{cases} \tag{5.57}$$

where $\gamma^*_{T,det,fixed}$ is the unique solution to the following equation:

$$\sum_{i=1}^{N} \frac{\rho_i^2}{1-\rho_i^2 \mu_{ave}(\gamma^*_{T,det,fixed})} = \sum_{i=1}^{N} \frac{1}{\sigma^2_{n,norm}(\gamma^*_{T,det,fixed}) + \frac{q\varsigma^2}{\sigma_n^2(\Upsilon=\gamma^*_{T,det,fixed})\rho_i^2}}, \tag{5.58}$$

with ρ_1, ρ_2,...,ρ_N representing the ordered eigenvalues of matrix A: $|\rho_1| \geq |\rho_2| \geq \ldots \geq |\rho_N|$ and $\rho_{\max} = |\rho_1|$.

Proof. Consider the diagonalization of matrix A_s: $A_s = LDL^T$, where $L^T L = I$ and $D = \text{diag}\{\rho_1, \rho_2, \ldots, \rho_N\}$. The following can be easily confirmed from Eq. 5.54: $\mathbb{E}(P[\infty]) = L\text{diag}\left\{\frac{q+\varsigma^{-2}\rho_1^2\sigma^2_{n,\text{ave}}(\Upsilon_T)}{1-\rho_1^2\mu_{\text{ave}}(\Upsilon_T)}, \ldots, \frac{q+\varsigma^{-2}\rho_N^2\sigma^2_{n,\text{ave}}(\Upsilon_T)}{1-\rho_N^2\mu_{\text{ave}}(\Upsilon_T)}\right\}L^T$, which results in

$$\left\|\mathbb{E}(P[\infty])\right\| = \frac{q+\varsigma^{-2}\rho_1^2\sigma^2_{n,\text{ave}}(\Upsilon_T)}{1-\rho_1^2\mu_{\text{ave}}(\Upsilon_T)}. \tag{5.59}$$

Let $\gamma^*_{T,\text{norm,fixed}}$ represent any solution to Eq. 5.56. It can be easily verified that $\frac{\partial\|\mathbb{E}(P[\infty])\|}{\partial\Upsilon_T}$ is only zero at $\gamma^*_{T,\text{norm,fixed}}$. Next we show that Eq. 5.56 has a unique solution. Assume that Eq. 5.56 has two solutions: $\gamma^*_{T,\text{norm,fixed},1}$ and $\gamma^*_{T,\text{norm,fixed},2} > \gamma^*_{T,\text{norm,fixed},1}$. Since σ_n^2 is a non-increasing function of Υ, we will have the following:

$$\begin{aligned}
&\mu_{\text{ave}}(\gamma^*_{T,\text{norm,fixed},1}) + \sigma^2_{n,\text{norm}}(\gamma^*_{T,\text{norm,fixed},1}) + \frac{\varsigma^2 q}{\rho^2_{\max}\sigma_n^2(\gamma^*_{T,\text{norm,fixed},1})} - \\
&\left[\mu_{\text{ave}}(\gamma^*_{T,\text{norm,fixed},2}) + \sigma^2_{n,\text{norm}}(\gamma^*_{T,\text{norm,fixed},2}) + \frac{\varsigma^2 q}{\rho^2_{\max}\sigma_n^2(\gamma^*_{T,\text{norm,fixed},2})}\right] = \\
&\underbrace{\int_{\gamma^*_{T,\text{norm,fixed},2}}^{\gamma^*_{T,\text{norm,fixed},1}} \chi(\Upsilon)d\Upsilon + \int_{\gamma^*_{T,\text{norm,fixed},1}}^{\gamma^*_{T,\text{norm,fixed},2}} \frac{\sigma_n^2(\Upsilon)\chi(\Upsilon)}{\sigma_n^2(\Upsilon=\gamma^*_{T,\text{norm,fixed},1})}d\Upsilon}_{<0} + \\
&\underbrace{\left(\frac{1}{\sigma_n^2(\Upsilon=\gamma^*_{T,\text{norm,fixed},1})} - \frac{1}{\sigma_n^2(\Upsilon=\gamma^*_{T,\text{norm,fixed},2})}\right)\int_{\gamma^*_{T,\text{norm,fixed},2}}^{\infty} \sigma_n^2(\Upsilon)\chi(\Upsilon)d\Upsilon}_{<0} + \\
&\underbrace{\frac{\varsigma^2 q}{\rho^2_{\max}}\left(\frac{1}{\sigma_n^2(\Upsilon=\gamma^*_{T,\text{norm,fixed},1})} - \frac{1}{\sigma_n^2(\Upsilon=\gamma^*_{T,\text{norm,fixed},2})}\right)}_{<0} < 0.
\end{aligned} \tag{5.60}$$

Therefore[2], $\gamma^*_{T,\text{norm,fixed},1} = \gamma^*_{T,\text{norm,fixed},2}$. Let $\Upsilon_{T,c}$ be the critical stability threshold defined in the previous section: $1-\rho^2_{\max}\mu_{\text{ave}}(\Upsilon_{T,c}) = 0$. We have $\gamma^*_{T,\text{norm,fixed}} < \Upsilon_{T,c}$. Consider those cases where there exists a non-negative solution to Eq. 5.56. Then using the fact that $\lim_{\Upsilon_T \to \Upsilon_{T,c}} \mathbb{E}(P[\infty]) \to \infty$ shows that $\gamma^*_{T,\text{norm,fixed}}$ corresponds to the unique minimum of $\left\|\mathbb{E}(P[\infty])\right\|$, i.e. $\Upsilon^*_{T,\text{norm,fixed}} = \gamma^*_{T,\text{norm,fixed}}$. If the process noise is the dominant noise, compared to the communication noise, there may be no

[2] Note that $\frac{\partial\sigma_n^2(\Upsilon)}{\partial\Upsilon}$ is taken to be zero only asymptotically.

positive solution to Eq. 5.56. It can be easily seen that, in such cases, $\left\|\mathbb{E}\left(P[\infty]\right)\right\|$ will be an increasing function for $\Upsilon_T \geq 0$, resulting in $\Upsilon^*_{T,\text{norm,fixed}} = 0$.

Next we will find $\Upsilon^*_{T,\text{det,fixed}}$. We will have, $\det\left(\mathbb{E}(P[\infty])\right) = \prod_{i=1}^{N} \frac{\rho_i^2 \varsigma^{-2}\sigma^2_{n,\text{ave}}(\Upsilon_T)+q}{1-\rho_i^2\mu_{\text{ave}}(\Upsilon_T)}$. It can be easily confirmed that

$$\frac{\partial \det\ \mathbb{E}(P[\infty])}{\partial \Upsilon_T} = \chi(\Upsilon_T)\frac{\prod_{i=1}^{N}\left(\varsigma^{-2}\sigma^2_{n,\text{ave}}(\Upsilon_T)\rho_i^2+q\right)}{\prod_{i=1}^{N}\left(1-\rho_i^2\mu_{\text{ave}}(\Upsilon_T)\right)}\left[\sum_{j=1}^{N}\frac{\rho_j^2}{1-\rho_j^2\mu_{\text{ave}}(\Upsilon_T)} - \sum_{j=1}^{N}\frac{\varsigma^{-2}\sigma_n^2(\Upsilon=\Upsilon_T)\rho_j^2}{\varsigma^{-2}\sigma^2_{n,\text{ave}}(\Upsilon_T)\rho_j^2+q}\right]. \tag{5.61}$$

Therefore, $\frac{\partial \det\left(\mathbb{E}\left(P[\infty]\right)\right)}{\partial \Upsilon_T}\Big|_{\Upsilon_T=\Upsilon^*_{T,\text{det,fixed}}} = 0$ will result in Eq. 5.58.

In a similar manner, it can be easily confirmed that Eq. 5.58 has a unique solution and that $\Upsilon^*_{T,\text{det,fixed}}$ corresponds to the global minimum of the determinant of the asymptotic average estimation error variance. □

Theorem 5.3 shows that in this case, the optimum way of dropping packets is the one that provides a balance between information loss (μ_{ave}) and overall communication noise ($\sigma^2_{n,\text{ave}}$). Eq. 5.56 (and Eq. 5.58) may not have a positive solution if process noise is the dominant noise compared to the overall communication noise (the third term on the left hand side of Eq. 5.56, for instance, will then get considerably high values). In such cases, the receiver should keep all the packets as communication noise is not the bottleneck. However, as long as process noise is not the dominant noise, the optimum way of dropping packets is the one that provides a balance between information loss and communication noise as indicated by Eq. 5.56 (and Eq. 5.58). Fig. 5.6 shows the asymptotic average estimation error variance as a function

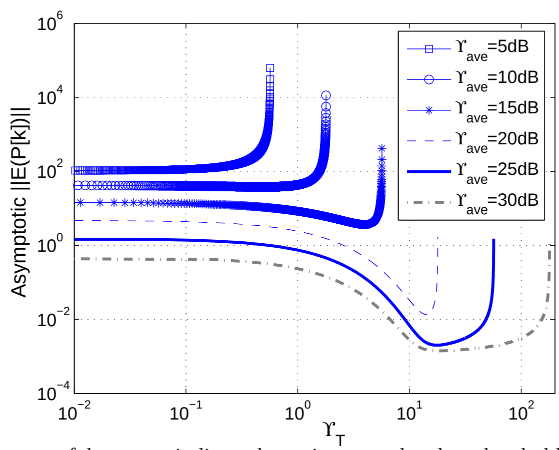

Fig. 5.6: Minimums of the curves indicate the optimum packet drop threshold for the non-adaptive case, when the knowledge of reception quality is not available at the estimator.

of a non-adaptive threshold, for different average SNRs and for the same parameters of Fig. 5.5. It can be seen that if Υ_T is too low, estimation performance degrades due to excessive communication noise. On the other hand, having Υ_T too high will result in the loss of information, which will degrade the performance. The optimum Υ_T (as predicted by Theorem 5.3) provides the necessary balance between loss of information and communication noise, reaching the minimums of the estimation error curves. As Υ_T increases, the estimation will approach the instability regions, predicted by Eq. 5.18 due to high information loss. Note that while Eq. 5.56 is derived for symmetric A matrices, Fig. 5.6 is plotted for a non-symmetric one. Yet the minima of the curves satisfy Eq. 5.56. This suggests that a similar expression could be valid for the general case.

Fig. 5.7 (left) shows the trace of the average estimation error variance as a function of time and for both adaptive and non-adaptive approaches. The average SNR is 18dB for this figure and the rest of the parameters are the same as in Fig. 5.5. Performance for the cases of $\Upsilon_T = 0$ and $\Upsilon_T = \Upsilon_{T,c}^{\varepsilon}$ are also plotted for comparison, which show that they perform poorly in this case. Fig. 5.7 (right) shows the corresponding thresholds for this figure. It can be seen that the adaptive approach chooses lower thresholds (on average) at the beginning. However, as the estimation error variance decreases, it can afford to choose higher thresholds. As compared with the non-adaptive approach, it can be seen that the adaptive approach performs better, as expected. By monitoring the current estimation error variance constantly, the adaptive approach can optimize the threshold better. However, in order to do so, it (the physical layer) requires constant knowledge of the estimation error variance from the application layer. Furthermore, the optimum threshold for the adaptive case can take values close to the critical threshold more often in order to optimize the performance. However, this comes at the cost of decreasing the stability margin. Thus, there are interesting tradeoffs between the two approaches. Fig. 5.8 shows similar curves for the case of average SNR of 25dB. Similar trends can be seen.

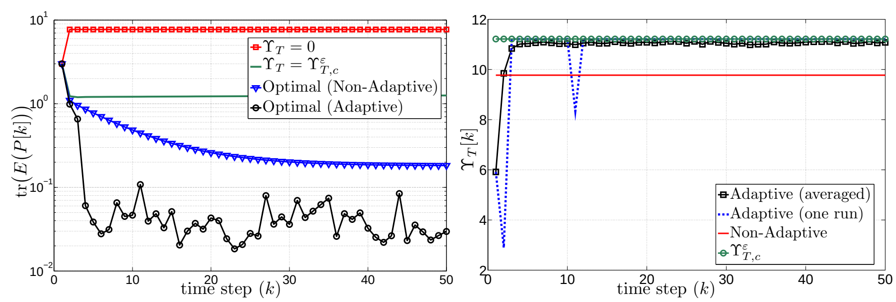

Fig. 5.7: (left) Packet drop threshold optimization for the case where the estimator has no knowledge of the reception quality and $\Upsilon_{\text{ave}} = 18\text{dB}$. (right) The corresponding thresholds for the left figure.

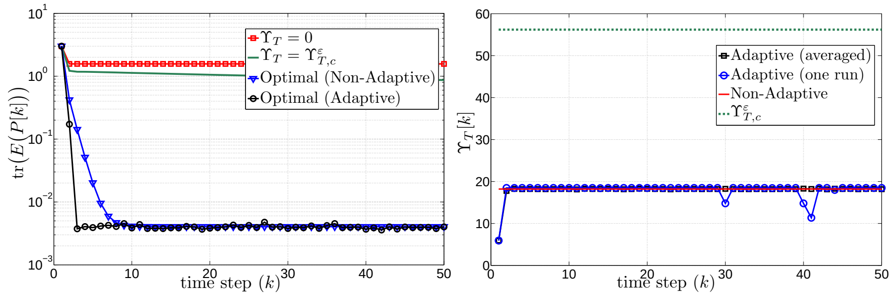

Fig. 5.8: (left) Packet drop threshold optimization for the case where the estimator has no knowledge of the reception quality and $\Upsilon_{\text{ave}} = 25$dB. (right) The corresponding thresholds for the left figure.

5.4.2 Estimator has full knowledge on the reception quality

In this part, we consider the case where the full knowledge of reception quality (R) is available at the estimator. Consider Eq. 5.19 (second and third lines). First, we characterize the one-step optimization of packet drop. Let $\Upsilon_{T_1}[k]$ and $\Upsilon_{T_2}[k]$ represent two possible thresholds at time step k, where $\Upsilon_{T_1}[k] < \Upsilon_{T_2}[k]$. Note that $R(\Upsilon[k]) = R_s + \sigma_c^2(\Upsilon[k])I$. Then, given $P[k]$, it can be easily confirm (using Eq. 5.19) that if $\Upsilon_{T_1}[k] < \Upsilon[k] < \Upsilon_{T_2}[k]$, then $P_1[k+1] \preceq P_2[k+1]$. Otherwise, $P_1[k+1] = P_2[k+1]$. Therefore, for any SNR at time k, we have $P_1[k+1] \preceq P_2[k+1]$. This means that the design with a smaller threshold will lower the next step estimation error variance. Therefore, the optimum threshold will be zero: $\Upsilon_T^{**} = 0$. Next, we consider non-adaptive optimization of Υ_T.

Theorem 5.4. *Consider Eq. 5.19. Keeping all the packets, i.e.* $\Upsilon_T = 0$, *will minimize the estimation error covariance at any time and therefore its average too.*

Proof. Let P_1 and P_2 represent estimation error covariance matrices of two estimators using fixed thresholds of Υ_{T1} and Υ_{T2}, where $\Upsilon_{T1} < \Upsilon_{T2}$. Then, $\Pi_{z,1}[k] \preceq \Pi_{z,2}[k]$. Assume that $P_1[0] = P_2[0]$. It is easy to see that $P_1[1] \preceq P_2[1]$ for any $\Upsilon[0]$. Consider the case where $P_1[k] \preceq P_2[k]$. Then, we have

$$\begin{aligned} P_1[k+1] \preceq & \quad AP_1[k]A^T + Q - AP_1[k]\,(P_1[k] + \Pi_{z,2}[k])^{-1} P_1[k]A^T \\ & = Q + A\Pi_{z,2}[k]A^T - A\Pi_{z,2}[k](P_1[k] + \Pi_{z,2}[k])^{-1}\Pi_{z,2}[k]A^T \\ & \preceq Q + A\Pi_{z,2}[k]A^T - A\Pi_{z,2}[k](P_2[k] + \Pi_{z,2}[k])^{-1}\Pi_{z,2}[k]A^T \\ & \preceq \qquad P_2[k+1]. \end{aligned} \tag{5.62}$$

Therefore, $\Upsilon_T = 0$ minimizes the estimation error variance at any time as well as its average. □

Fig. 5.9 shows the performance of threshold optimization for both cases where the estimator has “full knowledge” and “no knowledge” on the reception quality. It can be seen that the case with full knowledge (optimum threshold is zero in this case) performs the best, as expected. It can also be seen that the performance of the adaptive approach, when channel knowledge is not available at the estimator, is close to the case of full knowledge.

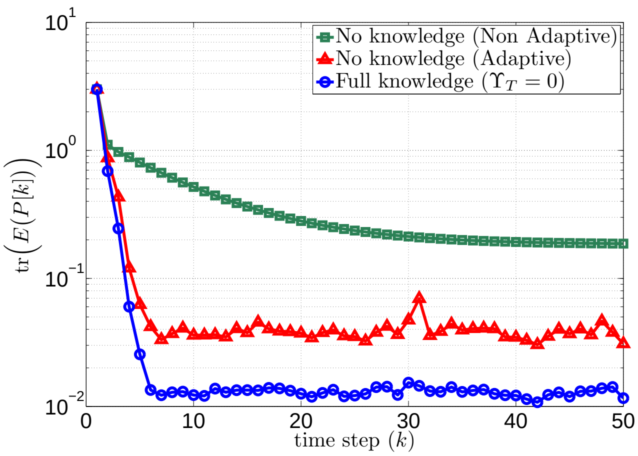

Fig. 5.9: Packet drop threshold optimization for both cases where the estimator has "full knowledge" and "no knowledge" on the reception quality and $\Upsilon_{\text{ave}} = 18$ dB.

5.5 Conclusions and Further Extensions

In this chapter, we considered the impact of both stochastic communication noise and packet drop on the estimation of a dynamical system over a wireless fading channel. We characterized the impact of the knowledge available on the link qualities, by considering the stability and performance of three cases of Kalman filtering with "full knowledge", "no knowledge" and "partial knowledge". We then proposed adaptive and non-adaptive ways of optimizing the packet drop in order to minimize the average estimation error variance of the Kalman filter. Our results showed that considerable performance can be gained by using our proposed optimization framework. There are several ways of extending our results, as we discussed throughout the chapter. We assumed an invertible C matrix. Furthermore, the derivations of the partial-knowledge case were carried out assuming that the known part of stochastic noise variance is small. Our proposed optimization framework of Section 5.4 can also be extended to scenarios where the physical layer only has partial information on the link quality, such as an upper bound. Mathematically characterizing the performance of such cases and properly optimizing the threshold are among possible extensions of this work. Finally, we assumed stochastic stationary channels in this chapter. Our framework can be easily extended to the case of non-stationary channels.

5.6 Appendix

Let $P_1[k]$ and $P_2[k]$ represent the estimation error covariance matrices of the noise-free and full knowledge cases respectively. From Eq. 5.22, we know that $\mathbb{E}\big(P_2[k+1]\big) \succeq \mu_{\text{ave}}\big(\Upsilon_T\big)A\mathbb{E}\big(P_2[k]\big)A^T + Q$. Therefore, we can easily establish that $\mathbb{E}\big(P_2[k]\big) \succeq \mathbb{E}\big(P_1[k]\big) \Rightarrow \mathbb{E}\big(P_2[k+1]\big) \succeq \mathbb{E}\big(P_1[k+1]\big)$. Next, we compare the average dynamics of the full knowledge case with that of no knowledge. Let $P_1[k]$ and $P_2[k]$ represent the estimation error covariance matrices of full knowledge and no knowledge cases respectively. We have,

$$\mathbb{E}\big(P_1[k+1]\big|P_1[k]\big) = \\ \big(1-\mu_{\text{ave}}\big(\Upsilon_T\big)\big)\mathbb{E}\big(P_1[k+1]\big|P_1[k],\Upsilon[k] > \Upsilon_T\big) + \mu_{\text{ave}}\big(\Upsilon_T\big)\mathbb{E}\big(P_1[k+1]\big|P_1[k],\Upsilon[k] \leq \Upsilon_T\big). \tag{5.63}$$

Using Lemma 5.1, it can be easily confirmed that $P_1[k+1]$ is a concave function of $\Sigma_z\big(\Upsilon[k]\big)$ in Eq. 5.19. Therefore, using conditional Jensen's inequality, we have, $\mathbb{E}\big(P_1[k+1]\big|P_1[k],\Upsilon[k] > \Upsilon_T\big) \preceq AP_1[k]A^T + Q - f\big(P_1[k]\big)$, where

$$f\big(P_1[k]\big) = AP_1[k]\Big(P_1[k] + C^{-1}\mathbb{E}\big(\Sigma_z[k]\big|\Upsilon[k] > \Upsilon_T\big)C^{-T}\Big)^{-1}P_1[k]A^T, \tag{5.64}$$

and $\mathbb{E}\big(P_1[k+1]\big|P_1[k]\big) \preceq AP_1[k]A^T + Q - \big(1-\mu_{\text{ave}}\big(\Upsilon_T\big)\big)f\big(P_1[k]\big)$. It can be seen, using Lemma 5.1, that f is a convex function of $P_1[k]$. Therefore by applying Jensen's inequality, $\mathbb{E}\big(P_1[k+1]\big) \preceq A\mathbb{E}\big(P_1[k]\big)A^T + Q - \big(1-\mu_{\text{ave}}\big(\Upsilon_T\big)\big)f\big(\mathbb{E}\big(P_1[k]\big)\big)$. By noting that $\mathbb{E}\big(\Sigma_z\big(\Upsilon[k]\big)\big|\Upsilon[k] > \Upsilon_T\big) = \frac{R_{\text{ave}}(\Upsilon_T)}{1-\mu_{\text{ave}}(\Upsilon_T)}$, it can be confirmed, after a few lines of derivations using Eq. 5.16, that $\mathbb{E}\big(P_2[k]\big) \succeq \mathbb{E}\big(P_1[k]\big) \quad \Rightarrow \quad \mathbb{E}\big(P_2[k+1]\big) \succeq \mathbb{E}\big(P_1[k+1]\big)$. Then, the stability condition for the full-knowledge case can be established in a similar manner to Theorem 5.1.

Acknowledgements This work is supported in part by NSF award # 0812338.

References

1. C. Chong and S. Kumar, "Sensor networks: evolution, opportunities and challenges," *Proceedings of the IEEE*, vol. 91, pp. 1247–1256, Aug. 2003.
2. B. Sinopoli, C. Sharp, L. Schenato, S. Schaffert, and S. Sastry, "Distributed control applications within sensor networks," *Proceedings of the IEEE*, vol. 91, pp. 1235–1246, Aug. 2003.
3. W. Jakes, *Microwave Mobile Communications*. IEEE Press, 1974.
4. T. S. Rappaport, *Wireless Communications, Principles and Practice*. Prentice-Hall, July 1999.
5. W. Wong and R. Brockett, "Systems with Finite Communication Bandwidth Constraints–Part I: State Estimation Problems," *IEEE Transactions on Automatic Control*, vol. 42, Sept. 1997.
6. W. Wong and R. Brockett, "Systems with Finite Communication Bandwidth Constraints–II: Stabilization with Limited Information Feedback," *IEEE Transactions on Automatic Control*, vol. 44, May 1999.
7. S. Tatikonda, *Control under Communication Constraints*. PhD thesis, MIT, 2000.
8. J. Hespanha, A. Ortega, and L. Vasudevan, "Towards the control of linear systems with minimum bit-rate," in *In Proc. of the Int. Symp. on the Mathematical Theory of Networks and Syst.*, vol. 1, Aug. 2002.

9. N. Elia and S. Mitter, “Stabilization of linear systems with limited information,” *IEEE Transactions on Automatic Control*, vol. 46, Sept. 2001.
10. F. Faganini and S. Zampieri, “Stability analysis and synthesis for scalar linear systems with a quantized feedback,” *IEEE Transactions on Automatic Control*, vol. 48, Sept. 2003.
11. A. Sahai and S. Mitter, “The Necessity and Sufficiency of Anytime Capacity for Stabilization of a Linear System over a Noisy Communication Link, Part I: Scalar Systems,” *IEEE Transactions on Information Theory*, vol. 52, pp. 3369–3395, Aug. 2006.
12. N. Martins and M. Dahleh, “Fundamental limitations of performance in the presence of finite capacity feedback,” in *Proceedings of 2005 American Control Conference*, vol. 1, pp. 79–86, June 2005.
13. J. Nilsson, *Real-time Control Systems with Delays*. PhD thesis, Department of Automatic Control, Lund Institute of Technology, 1998.
14. M. Micheli and M. I. Jordan, “Random sampling of a continuous-time dynamical systems,” in *Proceedings of 15th International Symposium on the Mathematical Theory of Networks and Systems (MTNS)*, (University of Notre Dame, South Bend, Indiana), 2002.
15. B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, “Kalman filtering with intermittent observations,” in *Proceedings of the 42nd IEEE Conference on Decision and Control*, 2003.
16. X. Liu and A. Goldsmith, “Kalman filtering with partial observation losses,” in *Proceedings of the 43rd IEEE Conf. Decision and Control*, vol. 4, (Atlantis, Bahamas), pp. 4180–4186, Dec. 2004.
17. L. A. Montestruque and P. Antsaklis, “Stability of model-based networked control systems with time-varying transmission times,” *IEEE Transactions on Automatic Control*, vol. 49, pp. 1562–1572, Sept. 2004.
18. M. Chow, Z. Sun, and H. Li, “Optimal stabilizing gain selection for networked control systems with time delays and packet losses,” *IEEE Transactions on Control Systems Technology*, vol. 17, pp. 1154–1162, Sept. 2009.
19. O. C. Imer, S. Yuksel, and T. Basar, “Optimal control of LTI systems over communication networks,” *Automatica*, vol. 42, pp. 1429–1440, Sept. 2006.
20. L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, “Foundations of control and estimation over lossy networks,” *Proceedings of the IEEE*, vol. 95, pp. 163–187, Jan. 2007.
21. J. P. Hespanha, P. Naghshtabrizi, and Y. Xu, “A survey of recent results in networked control systems,” *Proceedings of the IEEE*, vol. 95, pp. 138–162, Jan. 2007.
22. Y. Mostofi and R. Murray, “To Drop or Not to Drop: Design Principles for Kalman Filtering over Wireless Fading Channels,” *IEEE Transactions on Automatic Control*, vol. 54, pp. 376–381, February 2009.
23. Y. Mostofi and R. Murray, “Kalman filtering over wireless fading channels - How to handle packet drop,” *International Journal of Robust and Nonlinear Control*, vol. 19, pp. 1993–2015, February 2009.
24. Y. Mostofi, A. Gonzales-Ruiz, A. Ghaffarkhah, and D. Li, “Characterization and Modeling of Wireless Channels for Networked Robotic and Control Systems - A Comprehensive Overview,” in *Proceedings of 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, (St. Louis, MO), October 2009.
25. A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
26. D. Son, B. Krishnamachari, and J. Heidemann, “Experimental Analysis of Concurrent Packet Transmissions in Low-Power Wireless Networks,” USC-ISI Technical Report, ISI-TR-2005-609, November 2005.
27. T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall information and system sciences series, 2000.
28. S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
29. R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1999.
30. A. W. Marshal and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. Academic Press: New York, 1979.
31. Y. Mostofi and R. Murray, “Effect of time-varying fading channels on the control performance of a mobile sensor node,” in *Proceedings of IEEE 1st International Conference on Sensor and Adhoc Communications and Networks (SECON)*, (Santa Clara, California), Oct. 2004.