Performance of Estimation and Detection Algorithms in Wireless Networks

by

Alex Seak Chon Leong

Submitted in total fulfilment of
the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Electronic Engineering
The University of Melbourne
Australia

Dec 2007

The University of Melbourne

Australia

Abstract

Performance of Estimation and Detection Algorithms in Wireless Networks

by Alex Seak Chon Leong

This thesis focuses on techniques for analyzing the performance of estimation and detection algorithms under conditions which could be encountered in wireless networks, with emphasis on wireless sensor networks. These include phenomena such as measurement losses, fading channels, measurement delays and power constraints.

We first look at the hidden Markov model (HMM) filter with random measurement losses. The loss process is governed by another Markov chain. In the two-state case we derive analytical expressions to compute the probability of error. In the multi-state case we derive approximations that are valid at high signal-to-noise ratio (SNR). Relationships between the error probability and parameters of the loss process are investigated.

We then consider the problem of detecting two-state Markov chains in noise, under the Neyman-Pearson formulation. Our measure of performance here is the error exponent, and we give methods for computing this, firstly when channels are time-invariant, and then for time-varying fading channels. We also characterize the behaviour of the error exponent at high SNR.

We will look at the fixed lag Kalman smoother with random measurement losses. We investigate both the notion of estimator stability via expectation of the error covariance, and a probabilistic constraint on the error covariance. A comparison with the Kalman filter where lost measurements are retransmitted is made.

Finally we consider the distributed estimation of scalar linear systems using multiple sensors under the analog forwarding scheme. We study the asymptotic behaviour of the steady state error covariance as the number of sensors increases.

We formulate optimization problems to minimize the sum power subject to error covariance constraints, and to minimize the error covariance subject to sum power constraints. We compare between the performance of multi-access and orthogonal access schemes, and for fading channels the effects of various levels of channel state information (CSI).

This is to certify that

(i) the thesis comprises only my original work,

(ii) due acknowledgement has been made in the text to all other material used,

(iii) the thesis is less than 100,000 words in length, exclusive of table, maps, bibliographies, appendices and footnotes.

Signature_____

Date_____

Acknowledgments

I would like to thank my supervisors Professor Subhrakanti Dey and Associate Professor Jamie Evans for their guidance and endless source of ideas and suggestions, which have contributed greatly to the contents of this thesis.

I wish to thank the ARC Special Research Center for Ultra-Broadband Information Networks (CUBIN) for providing a wonderful working environment, and allowing me to meet many new people.

Financial support was provided by an Australian Postgraduate Award, and a postgraduate studentship supported by CUBIN. I am also grateful for the funding provided by my supervisors for the last few months of candidature.

Finally, I'd like to thank my friends and family for their encouragement, support and patience.

List of Publications

Journal Papers

1. A. S. Leong, S. Dey, and J. S. Evans. Probability of error analysis for hidden Markov model filtering with random packet loss. *IEEE Trans. Signal Processing*, 55(3):809–821, March 2007.

2. A. S. Leong, S. Dey, and J. S. Evans. Error exponents for Neyman-Pearson detection of Markov chains in noise. *IEEE Trans. Signal Processing*, 55(10):5097–5103, October 2007.

3. A. S. Leong, S. Dey, and J. S. Evans. On Kalman smoothing with random packet loss. *IEEE Trans. Signal Processing*. To appear.

4. A. S. Leong, S. Dey, and J. S. Evans. Power efficient state estimation using multiple sensors. In preparation.

Conference Papers

1. A. S. Leong, J. S. Evans, and S. Dey. Power control and multiuser diversity in multiple access channels with two time-scale fading. In *Proc. WiOpt'05*, pages 86–95, Trentino, Italy, April 2005.

2. A. S. Leong, S. Dey, and J. S. Evans. Error exponents for Neyman-Pearson detection of Markov chains in noise. In *Proc. Information, Decision and Control*, pages 94–99, Adelaide, Australia, February 2007.

Contents

List of Figures

List of Tables

List of Acronyms

CDMA	Code Division Multiple Access
CSMA	Carrier Sense Multiple Access
CSI	Channel State Information
FDMA	Frequency Division Multiple Access
HMM	Hidden Markov Model
i.i.d.	independent and identically distributed
KKT	Karush-Kuhn-Tucker
MAP	Maximum a posteriori
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
PAM	Pulse Amplitude Modulation
p.d.f.	probability density function
SNR	Signal-to-Noise Ratio
TDMA	Time Division Multiple Access

Chapter 1

Introduction

In this chapter we will briefly describe the basic concepts related to the topics in this thesis. We first provide an overview of wireless sensor networks, fading channels, and measurement losses and delays in wireless networks. We will then give a summary of the various detection and estimation algorithms studied in the thesis, and mention some of their main properties. Asymptotic concepts will be utilized in some chapters, and the notation which we will use in the thesis will be defined. We conclude the chapter with an outline of the thesis and its main contributions.

1.1 Wireless Sensor Networks

Wireless sensor networks are collections of sensors which can communicate with each other or to a central node/base station via wireless links. These sensors are equipped with sensing, computing and communications capabilities. Current and future advances in hardware technologies have allowed these sensors to become small and low cost. Potential applications of wireless sensor networks are many, and include environmental and infrastructure monitoring, healthcare, home automation, target tracking and surveillance. Recent years have seen a huge interest in wireless sensor networks by researchers in many different disciplines such as electrical engineering, computer science, physics and mathematics; some surveys of the field include (Akyildiz *et al.*, 2002; Chong & Kumar, 2003). However many challenges remain before their full benefits can be realised.

For instance, sensor nodes can be deployed either in predetermined locations or randomly, and due to their low cost we could have many of them. However the sensors often run on batteries which are difficult/costly to replace, so that low

power and energy consumption is of paramount importance. Another reason why utilising sensor networks effectively is challenging is because the sensor nodes have to communicate with each other through unreliable wireless links.

This thesis will focus on techniques for analyzing the performance of signal processing algorithms under conditions which could be encountered in wireless networks, in particular wireless sensor networks. These include channel fading, measurement losses, measurement delays, and the presence of power constraints. While these issues are prevalent in the communications and networking communities, they have received much less attention in other areas such as radar and control. Knowledge of the achievable performance and fundamental limitations of particular algorithms will be useful both for engineers who wish to apply these algorithms to wireless networks, as well as motivating the design of new algorithms which may be more suited to these environments. In the past, underlying assumptions such as all (noisy) measurements being received at the receiver perfectly without any delay, have often been used in the literature. However, the increasing use of wireless technologies in practical systems means that these assumptions may no longer be appropriate. This thesis will involve the performance analysis of a few algorithms of interest where such assumptions are relaxed.

1.1.1 Fading channels

Wireless channels are inherently time-varying. The quality of the channel is determined by factors such as distance, line of sight, channel bandwidths and frequencies used, and transmission data rates. Fading refers to the random fluctuations in the amplitude and phase of a signal due to interference between different versions of the transmitted signal.

The effect of a fading channel on a signal is often modelled as follows. Suppose a message s_k is transmitted through a wireless channel. Then the received message y_k is regarded as

$$y_k = h_k s_k + n_k$$

where n_k is a noise term and h_k is the channel gain, which models the effect of

random fluctuations on the amplitude and phase of the signal. Various assumptions on h_k can be made. One can give h_k various different probability distributions, such as Rayleigh when there is no line of sight, or Rician when there is a line of sight, or others such as Nakagami. Depending on the situation, h_k will either change at every time instant or can be assumed to be constant for a number of time steps. Even when the channels are time-varying, the successive instances may be either independent or correlated.

A survey of results on fading channels from an information theoretic viewpoint is (Biglieri *et al.*, 1998). One issue which can affect the performance of wireless systems dramatically is the question of how much we know about h_k. This is referred to as channel state information (CSI) in the literature. If the transmitter has knowledge of the actual values of h_k we say that we have full transmitter CSI, and if the receiver has this knowledge we say that we have full receiver CSI. CSI is often obtained/estimated by sending some training symbols and taking measurements, and it may not be feasible for both the transmitter and receiver to have full CSI, depending on the situation. When the transmitter has full CSI it can do power control to improve performance, such as increasing capacity or reliability. Even if one doesn't have full CSI, if the distribution of h_k is known then we can still hope to gain some improvement over not assuming anything about h_k at all, though not as great as when there is full CSI. From a signal processing perspective, it is important to be able to characterize the performance of estimation and detection algorithms in the presence of fading under various different assumptions on the availability of CSI.

1.1.2 Measurement Losses and Delays

Another phenomenon which can occur in wireless networks that can reduce performance is measurement loss. Suppose the measurements are digitized and grouped together in packets which are then sent over the wireless channel. Then due to collisions the entire packet may be lost entirely (Sinopoli *et al.*, 2004). Depending on the fading channel quality, we might also decide not to use any corrupted packets that arrive (Mostofi & Murray, 2005).

Delays in receiving measurements can occur if lost packets have to be retransmitted. The use of protocols such as Carrier Sense Multiple Access (CSMA), which involve transmitters checking for existing activity in the channel and deferring transmission if the channel is busy, can also result in delays.

The presence and absence of measurements will also affect the performance of signal processing algorithms, many of which were originally designed assuming measurements to always be available with negligible delay. Being able to analyze and characterize the performance of algorithms when there are random packet losses and/or delays is thus also of interest when doing signal processing in wireless networks.

1.2 Signal Detection

General references on hypothesis testing and detection include (Poor, 1994; Kay, 1998; Casella & Berger, 2002). In statistics a fundamental method of inference is that of *hypothesis testing*, where we have a number of different hypotheses and we want to determine which of them is true. To do this one makes a number of observations (assumed to all be drawn from the same hypothesis) and on the basis of these observations comes to a decision using a hypothesis test. Here we concentrate on binary hypothesis testing where there are two hypotheses H_0 and H_1. H_0 is often called the null hypothesis and H_1 the alternative hypothesis. We also restrict ourselves to *simple* hypotheses where each of the two hypotheses can be specified by a single probability distribution. For multiple hypothesis testing and composite hypothesis testing the reader is referred to the references above.

An example of a simple binary hypothesis testing problem is a signal detection problem where we have the hypotheses

$$\begin{aligned} H_0 &: y_k = w_k \\ H_1 &: y_k = s_k + w_k \end{aligned} \tag{1.1}$$

for $k = 1, \ldots, n$, i.e. we take n observations before making a decision. Here H_0

represents the situation where our measurements y_k consists of random noise w_k, and H_1 represents the case where there is a signal s_k immersed in noise w_k.[1] The signal s_k may either be a known signal, or a random signal where we have some knowledge of its distribution. The situation (1.1) is often used in radar where s_k represents (the signal generated by) a target, so that H_0 would be the case where a target is not present and H_1 the case where a target is present. Some other applications include communications e.g. receiver detecting whether signals have been sent to it, seismology e.g. detection of earthquakes, and astronomy e.g. detecting signals from pulsars.

Two quantities which are of interest in binary hypothesis testing are the probabilities of making the following two types of errors: 1) deciding H_1 when H_0 is actually true, known as a Type I error in the statistics literature, and 2) deciding H_0 when H_1 is actually true, known as a Type II error. Alternative notation in the radar literature which is perhaps more illuminating, is to call a Type I error a *false alarm*, and a Type II error a *missed detection*. This is due to the fact that H_0 is commonly used to represent the absence of a target, and H_1 the presence of a target. One way of designing hypothesis tests is to use rules which reduce the probabilities of these two types of errors from occurring. However in general there is a trade-off involved in that we cannot reduce both at the same time. Neyman-Pearson detection and Bayesian detection are two commonly used criteria which will be described below.

As a preliminary let

$$L(y_1, \ldots, y_n) \equiv \frac{p_1(y_1, \ldots, y_n)}{p_0(y_1, \ldots, y_n)}$$

be the likelihood ratio, where p_1 and p_0 are the likelihoods (densities) under H_1 and H_0 respectively.

[1] In this thesis we will not be too particular with the distinction between a random variable/process and its realization.

1.2.1 Neyman-Pearson detection

The Neyman-Pearson criterion, first formalized by J. Neyman and E. Pearson in (Neyman & Pearson, 1933), is to minimize the probability of missed detection P_M subject to the constraint that the probability of false alarm P_{FA} satisfies $P_{FA} \leq \alpha$, where α is usually chosen to be small.[2] It recognises the fact that P_{FA} and P_M are not necessarily of equal importance. For instance, in some applications false alarms may be highly undesirable, and we want to minimise the probability of missed detection while guaranteeing that P_{FA} is sufficiently small.

The Neyman-Pearson lemma states that if the null and alternative hypotheses are both simple, then the optimal test which satisfies the Neyman-Pearson criterion is to decide H_1 if $L(y_1, \ldots, y_n) > \gamma$, otherwise decide H_0, where γ is determined such that $P_{FA} \leq \alpha$ is satisfied. Thus the optimal test involves forming the likelihood ratio and seeing if it exceeds a certain threshold γ. This threshold can be determined analytically in some simple situations, though in general one has to resort to simulations or bounds.

1.2.2 Bayesian detection

When one has prior information $P(H_0)$ and $P(H_1)$ on the probability of each hypothesis occurring, then Bayesian detection is another commonly used alternative. Suppose we assign certain costs to error and/or detection events, e.g. let C_{ij} be the cost of deciding H_i when H_j is true. Then the Bayes risk is defined as

$$R = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} P(\text{decide } H_i | H_j) P(H_j).$$

For example, if we assign costs such that $C_{00} = C_{11} = 0$ and $C_{01} = C_{10} = 1$, then the Bayes risk becomes

$$P(\text{decide } H_0 | H_1) P(H_1) + P(\text{decide } H_1 | H_0) P(H_0)$$

[2] Often the solution will be such that $P_{FA} = \alpha$ is achieved, though in complicated situations one might be satisfied as long as $P_{FA} \leq \alpha$ is achieved (Casella & Berger, 2002, p.385), hence the use of the inequality.

which can be regarded as the average probability of error.

It is known (Poor, 1994) that the hypothesis test which minimizes the Bayes risk is to decide H_1 if $L(y_1, \ldots, y_n) > \gamma$, otherwise decide H_0, where the threshold γ is now

$$\gamma = \frac{(C_{10} - C_{00})P(H_0)}{(C_{01} - C_{11})P(H_1)},$$

and it is assumed that $C_{10} > C_{00}$ and $C_{01} > C_{11}$. So like the Neyman-Pearson test, the Bayesian test also compares the likelihood ratio against a threshold, though in general the values of the two thresholds will be different. For the case when $C_{00} = C_{11} = 0$ and $C_{01} = C_{10} = 1$, this reduces to deciding H_1 when

$$\frac{p_1(y_1, \ldots, y_n)}{p_0(y_1, \ldots, y_n)} > \frac{P(H_0)}{P(H_1)},$$

which can be easily shown to be equivalent to deciding H_1 when

$$P(H_1|y_1, \ldots, y_n) > P(H_0|y_1, \ldots, y_n)$$

and deciding H_0 otherwise. This is the well-known result that maximum a posteriori (MAP) detection minimizes the probability of error.

1.3 Signal Estimation

In signal estimation[3] we are interested in estimating a process $\{x_k\}$ from measurements $\{y_k\}$ which are related to $\{x_k\}$ in some way. For example, we may have detected that a target is present, and we now wish to track it. Specifically, at time k we want to estimate x_k, given measurements $y_1, \ldots, y_m$. Often this estimate is chosen to be the one with minimum mean squared error (MMSE), i.e. the estimate $\hat{x}_k$ which minimizes

$$\mathbb{E}\left[||X_k - \hat{X}_k||^2 | Y_1, \ldots, Y_m\right].$$

[3] Also referred to as state estimation in some contexts, such as hidden Markov models and linear systems.

In the case where $m < k$, this is referred to as prediction, for $m = k$ as filtering and for $m > k$ as smoothing.[4] From probability theory it is well-known (e.g. (Shiryaev, 1996, p.237)) that the optimal solution is the conditional expectation

$$\hat{X}_k = \mathbb{E}\left[X_k | Y_1, \ldots, Y_m\right].$$

Now if one knows exactly the distribution of $p(x_k|y_1, \ldots, y_m)$, then the problem is essentially solved, e.g. perform a numerical integration to get $\mathbb{E}\left[X_k|Y_1, \ldots, Y_m\right]$. However finding $p(x_k|y_1, \ldots, y_m)$ is in general very difficult. The subject of nonlinear filtering deals with these issues, and some recent developments for approximately solving these problems include particle filters (Arulampalam *et al.*, 2002) and unscented filters (Julier & Uhlmann, 2004).

Two commonly used models for describing signals that are considered in this thesis are hidden Markov models and linear systems. Hidden Markov model filters (see Section 1.4) are a type of nonlinear filter which can be found exactly because of its finite state space (assuming the underlying Markov chain is finite state), so that we can represent the distribution of $p(x_k|y_1, \ldots, y_m)$ by a probability mass function. This is in contrast to a (continuous) probability density which in general requires an infinite number of parameters to specify[5], so that the resulting filter is often infinite dimensional.

However, recall that the Gaussian distribution is a continuous distribution which *can* be completely specified by two parameters, its mean and variance. Furthermore, linear transformations of Gaussian random variables remain Gaussian. Then in the case where we have linear systems with additive Gaussian noise, the optimal filter will turn out to be finite dimensional, and its solution admits a simple recursive form known as the Kalman filter (see Section 1.5).

[4] In the literature the term "filtering" is also sometimes used in broad terms to refer to all three of these situations.

[5] In mathematics it is well-known that many function spaces are infinite dimensional, even though the continuous distributions commonly encountered in undergraduate courses might give the impression that all probability density functions (p.d.f.'s) can be specified by a few parameters.

1.4 Hidden Markov Models

References on hidden Markov models (HMMs) include (Rabiner, 1989; Ephraim & Merhav, 2002; Elliott *et al.*, 1995). Let $\{X_k\}$ be a finite state Markov chain (for definiteness let it have M states) with transition probability matrix A, i.e. the entries a_{ij} of A represent

$$a_{ij} = P(X_{k+1} = j | X_k = i),$$

and initial probability distribution π_0. Let Y_k be a random function of X_k (though not depending on previous values $X_{k-1}, X_{k-2}, \dots$), with

$$P(Y_k = y_k | X_k = i) \equiv b_i(y_k) \tag{1.2}$$

known. $\{Y_k\}$ is then conditionally independent given the Markov chain $\{X_k\}$. The term "hidden" refers to the situation where one can only observe $\{Y_k\}$ even though the Markov chain $\{X_k\}$ is the underlying process of interest, and one then attempts to make inferences on X_k based on Y_k. A hidden Markov model is often specified as the tuple (A, B, π_0) where B denotes the collection of all b_i's in (1.2). Though it will not be used in this thesis, we note that hidden Markov models can also be written in a state space form reminiscent of (1.4)-(1.5) in the next section, however what one might regard as the process noise term is now no longer independent Gaussian noise, but a martingale difference sequence (Elliott *et al.*, 1995).

HMMs were first studied in the 1960s by L. Baum and T. Petrie (Baum & Petrie, 1966), who referred to them as "probabilistic functions of Markov chains". In the ensuing years, many applications of HMMs have been found, including speech recognition, radar, communications, and biology, see (Rabiner, 1989; Ephraim & Merhav, 2002) and the references therein. As an example, a two-state Markov chain could be used to model binary data where one of the states corresponds to a 1 and the other 0, with the transition probabilities of the Markov chain representing the probability of moving from 1 to 0, or 0 to 1. Some reasons for the popularity of HMMs is its flexibility in modelling nonlinear phenomena and its amenability to

implementation due to the finite dimensionality of the HMM filter.[6]

1.4.1 HMM Filter

Recall from Section 1.3 (also MAP detection in Section 1.2.2) that to perform optimal estimation, one is often required to know the distribution of $p(x_k|y_1,\ldots,y_m)$. For the hidden Markov models considered in this thesis, where we assume that $\{X_k\}$ is a finite state Markov chain, these are just the conditional probabilities

$$P(X_k = i|y_1,\ldots,y_m), i = 1,\ldots,M$$

and the collection of these forms a probability mass function. Since the number of variables is finite, the filter is finite dimensional and can be derived relatively easily. Here we present the formulas for the filtering case, HMM prediction and smoothing can be found in the general references (Rabiner, 1989; Ephraim & Merhav, 2002). Define

$$P(X_k = i|y_1,\ldots,y_k) \equiv \Pi^i_{k|k}$$

and let $\Pi_{k|k}$ be the column vector with $\Pi^i_{k|k}$ as its entries. Then the conditional probabilty estimates can be computed recursively as follows:

$$\Pi_{k+1|k+1} = \frac{B_{y_{k+1}} A^T \Pi_{k|k}}{\begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} B_{y_{k+1}} A^T \Pi_{k|k}} \tag{1.3}$$

where

$$B_{y_{k+1}} = \begin{bmatrix} b_1(y_{k+1}) & 0 & \ldots & 0 \\ 0 & b_2(y_{k+1}) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & b_M(y_{k+1}) \end{bmatrix}.$$

Hence the HMM filter admits a recursive form, and is a nonlinear function of the measurements y_k in general, c.f. the Kalman filter of Section 1.5.

[6] This is true for *finite state* Markov chains. However the notion of HMMs can also be extended to Markov chains with countable state spaces or continuous state Markov processes.

1.4.2 State Estimation of HMMs

Given observations $Y_1, \ldots, Y_k$, we wish to find the "best" estimate of X_k based on these observations. The maximum a posteriori (MAP) rule

$$\hat{X}_k = \arg \max_{j=1,\ldots,M} \Pi_{k|k}^j$$

gives the estimate $\hat{X}_k$ that minimizes the probability of error $P(\hat{X}_k \neq X_k)$. If the states correspond to e.g. points in Euclidean space, then mean squared error (MSE) criteria can also been considered, see (Golubev, 2000).

Another related problem is to determine the most likely sequence $X_1, \ldots, X_k$ given $Y_1, \ldots, Y_k$ which can be solved by using the Viterbi algorithm (Rabiner, 1989; Ephraim & Merhav, 2002).

1.5 Kalman Filtering

General references for this section include (Anderson & Moore, 1979) and (Kailath *et al.*, 2000).

1.5.1 Linear systems

Suppose the process $\{x_k\}$ evolves according to

$$x_{k+1} = A_k x_k + w_k \tag{1.4}$$

and the measurements $\{y_k\}$ are given by

$$y_k = C_k x_k + v_k, \tag{1.5}$$

where A_k and C_k are matrices, and $\{w_k\}$ and $\{v_k\}$ are mutually independent[7] white[8] Gaussian noise processes with zero mean and covariance matrices $Q_k \geq 0$ and $R_k > 0$ respectively. We refer to x_k as the state and the problem of estimating x_k as a state estimation problem. The initial state x_0 is assumed to be Gaussian with known mean and covariance matrix. We call w_k the process noise and v_k the measurement noise.

Linear systems have many applications in engineering. The use of linear systems in modelling is ubiquitous in control theory. In radar, the motion of a target can also be modelled using linear systems. Other situations which have been modelled using linear systems include variations in daily temperatures and the concentration of chemicals (air pollution) in the atmosphere.

1.5.2 Kalman filter

A recursive solution to the filtering problem for the linear system (1.4)-(1.5) was given by R. Kalman in (Kalman, 1960).[9] We first introduce some notation. Let

$$\hat{x}_{k|m} \equiv \mathbb{E}\left[x_k | y_0, \ldots, y_m\right]$$

and

$$P_{k|m} \equiv \mathbb{E}\left[(x_k - \hat{x}_{k|m})(x_k - \hat{x}_{k|m})^T | y_0, \ldots, y_m\right]$$

[7]The results can also be derived for more general linear systems such as correlation between the process and measurement noises, see (Anderson & Moore, 1979).

[8]Here $\{w_k\}$ being white means that w_k and w_l are independent random variables for $k \neq l$. Similarly for $\{v_k\}$.

[9]The continuous time version of the problem was later solved by R. Kalman and R. Bucy in (Kalman & Bucy, 1961), and is known as the Kalman-Bucy filter.

represent the state estimates and error covariance matrices of these estimates. Then the Kalman filtering equations can be written as:[10]

$$\begin{aligned}
\hat{x}_{k+1|k} &= A_k \hat{x}_{k|k} \\
P_{k+1|k} &= A_k P_{k|k} A_k^T + Q_k \\
\hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + P_{k+1|k} C_{k+1}^T (C_{k+1} P_{k+1|k} C_{k+1}^T + R_{k+1})^{-1} (y_{k+1} - C_{k+1} \hat{x}_{k+1|k}) \\
P_{k+1|k+1} &= P_{k+1|k} - P_{k+1|k} C_{k+1}^T (C_{k+1} P_{k+1|k} C_{k+1}^T + R_{k+1})^{-1} C_{k+1} P_{k+1|k}.
\end{aligned} \tag{1.6}$$

The first and second equations of (1.6) are commonly referred to as the time update equations, the third and fourth equations of (1.6) the measurement update equations.

We note that the Kalman filter equations compute the state estimates and corresponding error covariances recursively. It also computes both the filtered estimate $\hat{x}_{k|k}$ and the one-step ahead predicted estimate $\hat{x}_{k+1|k}$. The Kalman filter is a linear filter since the measurements y_k are used in a linear fashion to compute the estimates. Another property of the Kalman filter is that the recursions for the error covariances do not actually depend on the measurements, so can be precomputed (assuming A_k, C_k, R_k, Q_k are known beforehand).[11]

Now suppose A_k, C_k, R_k, Q_k are constant, and we drop the subscript k on these quantities. Then in many cases the error covariance $P_{k+1|k}$ converges to a limit P_∞ as $k \to \infty$. For stable systems (i.e. the eigenvalues of the matrix A are all less than 1 in magnitude), this is guaranteed and the steady state error covariance P_∞ can be found as the solution of the equation

$$P_\infty = A[P_\infty - P_\infty C^T (C P_\infty C^T + R)^{-1} C P_\infty] A^T + Q. \tag{1.7}$$

In general there is no closed form solution for P_∞,[12] though various routines can be used to solve for P_∞ numerically, such as those found in Matlab. For unstable

[10] Other equivalent ways of writing these equations are possible, see (Anderson & Moore, 1979; Kailath *et al.*, 2000).

[11] When there is random measurement loss as in Chapter 4 this is no longer the case as the error covariances will then depend on the values of the loss process.

[12] In the scalar case however closed form solutions can be found, see e.g. Chapter 5.

systems, P_∞ can still exist (as a bounded quantity) and be found by solving the same equation (1.7), but extra assumptions are needed. One such set of assumptions is that the system is detectable and stabilizable (Anderson & Moore, 1979).

1.5.3 Kalman smoothers

The original derivations in (Kalman, 1960) allowed one to recursively compute both the filtered and one-step ahead predicted estimates and their corresponding covariances. Prediction for further instances $\hat{x}_{k+N|k}, N > 1$ can also be derived easily from these equations, see (Anderson & Moore, 1979). However, it was not immediate how to extend these results to smoothing, and it was left to later researchers such as (Rauch, 1963; Fraser & Potter, 1969) to derive the various smoothers; both (Anderson & Moore, 1979) and (Kailath *et al.*, 2000) discuss some of the history of these results.

There are in fact three different types of smoothers commonly used. The *fixed-point smoother* estimates x_j for some fixed j, while measurements $y_1, \ldots, y_j, y_{j+1}, \ldots$ are taken. This smoother is used when there is some important time instance at which one wants to estimate the state accurately as more and more measurements come in, e.g. an initial time instance.

The *fixed-lag smoother* estimates x_{k-N} for some fixed lag N from measurements $y_1, \ldots, y_k$, which is used to provide on-line smoothing of data as it is received, though introducing a delay of N.

The *fixed-interval smoother* considers a fixed interval e.g. $[1, 2, \ldots, M]$, and estimates x_i for $i = 1, \ldots, M$ based on $y_1, \ldots, y_M$. This type of smoother is used when one can do off-line processing of a set of measurements and we want the best estimate at each time instance based on the entire set of measurements.

1.6 Asymptotic Notation

Suppose we wish to study the asymptotic behaviour of a function $f(x)$ near a point of interest x_0. For instance, if we are interested in the large x behaviour we can take x_0 to be infinity, and for small x behaviour we can take x_0 to be zero. Asymptotic

notation can be used to describe how "close" one function is to another in a mathematically precise manner. Below we will fix the notation that will be used in later chapters, particularly Chapters 3 and 5, noting that other authors sometimes use the same symbols with slightly different meanings. For further reference see (Olver, 1974).

1) We say that $f(x) = O(g(x))$ if there is a constant M such that

$$\left|\frac{f(x)}{g(x)}\right| \leq M$$

for all x in some neighbourhood of x_0.

2) We say that $f(x) = o(g(x))$ if

$$\frac{f(x)}{g(x)} \to 0$$

as $x \to x_0$.

3) We say that $f(x) \sim g(x)$ if

$$\frac{f(x)}{g(x)} \to 1$$

as $x \to x_0$. One thing to note is that $f(x) \sim g(x)$ does not necessarily mean that $f(x) \to g(x)$ as $x \to x_0$. For instance, let $f(x) = x^2 + x$ and $g(x) = x^2$. Then $f(x) \sim g(x)$ for $x \to \infty$, but the difference $f(x) - g(x) = x$ is unbounded as $x \to \infty$.

1.7 Outline and Contributions of Thesis

This thesis deals with the performance analysis of estimation and detection algorithms operating in wireless environments. While assumptions such as those of all measurements being received without delay may have been satisfactory in the past, the increasing reliance on wireless technologies means that consideration of phenomena such as measurement losses, fading, delays and power constraints will be of increasing importance.

The algorithms whose performance we will analyze include the HMM filter,

Neyman-Pearson detector, and the Kalman filter and smoother. Different notions of performance are used (depending on the algorithm), such as the probability of error, the error exponent, steady state error covariances, and notions of stability for time-varying error covariances. The main contributions of this thesis are contained in Chapters 2 to 5. Chapters 2 and 3 deal with algorithms related to hidden Markov models, while Chapters 4 and 5 deal with algorithms related to linear systems and Kalman filtering. A brief description of each of these chapters follows.

In Chapter 2 we look at the performance of the HMM filter, with random packet losses governed by another Markov chain. Here we are interested in the probability of error, specifically methods for computing it analytically/numerically. In the case where only measurement losses occur but no noise is added, we derive an expression for the probability of error which is given in terms of an infinite series. We also prove some monotonicity properties that the parameters of the loss process will have on the probability of error. In the case where there are both losses and additive Gaussian noise we further specialise into two-state and multi-state Markov chain situations. For two-state Markov chains we present a numerical method for computing the probability of error, and its numerical properties are analyzed. For multi-state Markov chains, we give an approximation for the probability of error which is valid at high signal to noise ratio (SNR). The derivation of the approximation is based on methods used in the noiseless case. Numerical comparisons demonstrating the accuracy of our formulae with simulation results are presented throughout.

In Chapter 3 we consider the detection of two-state Markov chains in noise and fading under the Neyman-Pearson criterion. The measure of performance we are interested in here is the error exponent, which represents the rate at which the probability of missed detection P_M decays for a fixed constraint on the probability of false alarm P_{FA}, as the number of samples taken goes to infinity. We first present a numerical method for computing the error exponent in the case of time-invariant channels. The method is similar to the numerical method for computing the probability of error for the noisy two-state case of Chapter 2, even though the two problems at first appear to be quite different. We also determine the asymptotic form of the error exponent at high SNR, which is shown to scale *linearly* with the

SNR, independently of the values of the Markov chain transition probabilities. We then extend our numerical method for computing the error exponent to the case of fading channels without CSI. The asymptotic form at high SNR is determined, and is now shown to scale *logarithmically* with the SNR, in contrast to the time-invariant channel case.

In Chapter 4 we study the Kalman fixed lag smoother with random packet loss (mainly i.i.d. Bernoulli here), and whether the additional delay introduced by smoothing brings us any advantages over the Kalman filter. Two different performance criteria will be considered, one being a notion of estimator stability via boundedness of the expectation of the error covariance (Sinopoli *et al.*, 2004), and the other a probabilistic notion of performance (Shi *et al.*, 2005) which deals with questions such as the probability of the error covariance being less than a certain value at a given time. The equations for the fixed lag smoother with random packet loss are first derived. We look at the stability of the smoother in terms of the boundedness of the expectation of the error covariance. We show that for unstable systems with random losses, Kalman smoothing does not provide any advantage over Kalman filtering in terms of stability of the estimator, since the critical thresholds for the arrival rate guaranteeing stability turn out to be the same in both cases. However, if instead we consider a probabilistic notion of performance then we find that Kalman smoothing does perform better than Kalman filtering. We also analyze two simple strategies for retransmitting lost measurements, and we find that in both cases, their distributions turn out to be the same as the Kalman filter (without retransmission), so provides no advantage using both notions of performance, but with the disadvantage of introducing an additional delay.

In Chapter 5 we address the problem of estimation of scalar linear systems with multiple sensors, using a scheme known as analog forwarding, where the different sensors make measurements of the scalar state, and transmits directly (uncoded but with a possible scaling) to the fusion center for further processing. The situation turns out to be just another linear system, so that optimal state estimation can be achieved using a Kalman filter. Error covariances will be the measure of performance in this chapter. In the case where the system is stable and the channels are time-

invariant, there will be a steady state error covariance. We will consider both a multi-access scheme and an orthogonal access scheme for the sensors to transmit their measurements to the fusion center, and make comparisons between them. We will show that in many situations, as the number of sensors M goes to infinity, the steady state error covariance converges to the process noise variance at the rate $1/M$. Moreover, we find that in the multi-access case this can be achieved even if the total power used by all sensors is bounded, though this is not possible in the orthogonal case. Power efficiency will also be another consideration in this chapter. We formulate and solve two optimization problems, one being to minimize the sum power subject to an error covariance constraint, the other being to minimize the error covariance subject to a sum power constraint. Both of these problems turn out to be convex and hence can be solved efficiently. We then switch to time-varying fading channels. In the case where CSI is available, there is no notion of a steady state error covariance, so we look at the error covariance at each individual time step and perform optimization at each time step. When we don't have CSI but have channel statistics, we will derive the best *linear* estimator whose equations turn out to be similar to the Kalman filter, and study its performance.

Chapter 2

Hidden Markov Model Filtering with Random Packet Loss

2.1 Introduction

This chapter studies the probability of error for maximum a posteriori (MAP) estimation of hidden Markov models, where measurements can be either lost or received according to another Markov process. Estimation with lossy measurements was considered for linear systems in (Sinopoli *et al.*, 2004), for the Kalman filtering problem with losses modelled by an independent and identically distributed (i.i.d.) Bernoulli process.[1] They showed that for an unstable system, there exists a threshold such that if the probability of reception exceeds this threshold then the expected value (with respect to the loss process) of the error covariance matrix (which is a random quantity due to random losses) will be bounded, but if the probability of reception is lower than this threshold then the error covariance diverges. In a slightly different context, (Tiwari *et al.*, 2005) (also see (Huang & Dey, 2007)) extends these results to Markovian loss processes, which allows modelling of more "bursty" types of errors. Estimation with Markovian packet losses was also studied in (Smith & Seiler, 2003), and sub-optimal estimators were derived which can be used to provide upper bounds on the estimation errors of the optimal estimator.

The purpose of this chapter is to use some of the ideas relating to lossy measurement processes, but to apply it to the different problem of state estimation for Markov chains. For HMM estimation problems the state space is often finite, thus notions of estimation stability as in (Sinopoli *et al.*, 2004) are not appropriate.

[1] The extension of some of these results to Kalman smoothing is the subject of Chapter 4.

Instead here we will use the probability of estimation error as our measure of performance, also see (Wonham, 1965; Golubev, 2000) for mean square error criterions. Obtaining analytical expressions for the error performance associated with filtering for HMMs (even without any loss process) is a difficult problem however, where few general results are known. Some results for two-state Markov chains in the continuous time case may be found in (Wonham, 1965). In discrete time, asymptotic formulae for "slow" Markov chains or Markov chains with "rare transitions" with finite state space were obtained in (Khasminskii & Zeitouni, 1996) for the probability of error, and (Golubev, 2000) for a mean square error criterion. This was later extended to HMM smoothing for Markov chains with rare transitions in (Shue *et al.*, 2000). A general characterization of the error probability for the two-state hidden Markov model in discrete time was derived in (Shue *et al.*, 2001), and a numerical method to calculate it was proposed. However the problem for general multi-state Markov chains is still open.

The organisation of the chapter is as follows. We will first study the simpler problem with observation losses but no noise in Section 2.3. Analytical expressions for the error probabilities will be derived and some special cases presented in 2.3.1 and 2.3.2 respectively. Some relationships between the error probability and the parameters of the Markovian loss process are established in 2.3.3 and numerical studies presented in 2.3.4. Multi-state Markov chains are considered in 2.3.5. We will then shift our attention to the more general HMM problem with noise in Section 2.4. In 2.4.1 and 2.4.2 we will characterize the error probabilities for the two-state Markov chain, though using quite different methods. Section 2.4.3 will present some numerical studies for the noisy case. It is more difficult here to prove properties similar to the ones in the noiseless case, and some conjectures which will require further investigation are stated. Section 2.4.4 studies the situation when the signal is i.i.d. In 2.4.5 and 2.4.6, high SNR approximations for the two-state and multi-state Markov chains respectively are derived and numerical comparisons made.

2.2 Model and notational conventions

The main model of interest is

$$Y_k = \gamma(Z_k)h(X_k) + v_k. \tag{2.1}$$

Here $\{Y_k\}$ is the observation process, $\{v_k\}$ is the noise process which will be i.i.d. and $N(0, \sigma^2)$.[2] $\{X_k\}$ and $\{Z_k\}$ are homogeneous two-state Markov chains[3] which are assumed to be independent of each other, with $h(1) = -1, h(2) = 1, \gamma(1) = 0, \gamma(2) = 1$. $h(X_k)$ can be interpreted as the signal that we wish to estimate, and $\gamma(Z_k)$ as the correlated loss process, with the correlation modelled by a Markov chain. We will assume that $\gamma(Z_k)$ is known to us at each time instant. In Sections 2.3.5 and 2.4.6 we will look at the situation where $\{X_k\}$ is a multi-state Markov chain.

Originally, in the paper (Leong *et al.*, 2007) we followed (Shue *et al.*, 2001) (also notably in the monograph (Elliott *et al.*, 1995)) and used the conventions $a_{ij} = P(X_{k+1} = i|X_k = j)$ and $g_{ij} = P(Z_{k+1} = i|Z_k = j)$ for the transition probabilities. However, to maintain consistency with the rest of the thesis, in this chapter we will use instead the more common conventions

$$a_{ij} = P(X_{k+1} = j|X_k = i) \text{ and } g_{ij} = P(Z_{k+1} = j|Z_k = i)$$

for the transition probability matrices $A = (a_{ij})$ and $G = (g_{ij})$. Assuming that $a_{21}, a_{12}, g_{21}, g_{12} \neq 0$, unique stationary probabilities will then exist and are given by $P(X_k = 1) = \frac{a_{21}}{a_{12}+a_{21}}$, $P(X_k = 2) = \frac{a_{12}}{a_{12}+a_{21}}$, $P(Z_k = 1) = \frac{g_{21}}{g_{12}+g_{21}}$, $P(Z_k = 2) = \frac{g_{12}}{g_{12}+g_{21}}$.[4]

Denote the conditional probability vector for the HMM filter by $\Pi_{k|k}$, with the i-th entry being $\Pi^i_{k|k} = P(X_k = i|Y_0 = y_0, \ldots, Y_k = y_k, Z_0 = z_0, \ldots, Z_k = z_k)$. The MAP estimate of X_k, which is well-known to minimize the probability of error, is

[2] Other noise types such as noise with state-dependent variances are possible, but some derivations will be more complicated.

[3] Some work on the case where $\{X_k\}$ is a multi-state Markov chain is considered in Sections 2.3.5 and 2.4.6.

[4] Strictly speaking, this is true if the initial state of the Markov chain has the same distribution as the stationary distribution, otherwise this holds only in the limit as $k \to \infty$.

for the two-state case

$$\hat{X}_k = \begin{cases} 1 & , \quad \Pi^1_{k|k} > \Pi^2_{k|k} \\ 2 & , \quad \text{otherwise.} \end{cases}$$

2.3 Noiseless case

We first study the simpler version of the model (2.1)

$$Y_k = \gamma(Z_k)h(X_k)$$

which does not have the noise term v_k. The probability of estimation error that we derive will be given in terms of an infinite series, which is a more explicit form than that which will be derived for the noisy case in Section 2.4.1. The noiseless case considered here is quite suitable for the noisy situation at high SNR, since (roughly speaking) at high SNR the errors due to the packet loss process tends to dominate the errors due to the noise term, see for example the discussion at the end of Section 2.4.5. Indeed, the derivation in Section 2.4.5 of an approximation for the error probability at high SNR will be based on some of the techniques of this section.

2.3.1 Derivation of probability of error

For this simple noiseless model, whenever there is no packet loss, i.e. $Z_k = 2$, the estimate (of $h(X_k)$) will be the same as the measurement. The probability vectors $\Pi_{k|k}$ are therefore updated as:

$$\Pi_{k+1|k+1} = \begin{cases} A^T\Pi_{k|k} & , \quad \gamma(Z_{k+1}) = 0 \\ \begin{bmatrix} 1 & 0 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} = h(1) \\ \begin{bmatrix} 0 & 1 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} = h(2). \end{cases}$$

So whenever there is no packet loss, the probability vector will "reset" to either $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$ or $\begin{bmatrix} 0 & 1 \end{bmatrix}^T$, a fact we will exploit in our derivation of the probability of

error. When a packet is received, no errors will be made, so

$$P(\text{Error}) = P(\text{Error}, Z_k = 1) + P(\text{Error}, Z_k = 2) = P(\text{Error}, Z_k = 1).$$

This can be further split up as follows:

$$\begin{aligned} P(\text{Error}, Z_k = 1) &= P(\text{Error}, Z_{k-1} = 1, Z_k = 1) + P(\text{Error}, Z_{k-1} = 2, Z_k = 1) \\ &= P(\text{Error}, Z_{k-2} = 1, Z_{k-1} = 1, Z_k = 1) \\ &\quad + P(\text{Error}, Z_{k-2} = 2, Z_{k-1} = 1, Z_k = 1) \\ &\quad + P(\text{Error}, Z_{k-1} = 2, Z_k = 1) \\ &\vdots \\ &= \sum_{n=1}^{\infty} p(n) \end{aligned}$$

where $p(n) \equiv P(\text{Error}, Z_{k-n} = 2, Z_{k-n+1} = 1, Z_{k-n+2} = 1, \ldots, Z_k = 1)$.

Expressions for each term $p(n)$ can be derived. Define $\Pi_{1,n} \equiv (A^n)^T \begin{bmatrix} 1 & 0 \end{bmatrix}^T$ and $\Pi_{2,n} \equiv (A^n)^T \begin{bmatrix} 0 & 1 \end{bmatrix}^T$, with $\Pi_{i,n}^j$ representing the j-th element of $\Pi_{i,n}$. For brevity, also let $Z \equiv (Z_{k-n} = 2, Z_{k-n+1} = 1, Z_{k-n+2} = 1, \ldots, Z_k = 1)$. Then

$$\begin{aligned} p(n) &= P(Z, X_k = 1, \arg\max_j \Pi_{1,n}^j \neq 1, X_{k-n} = 1) \\ &\quad + P(Z, X_k = 2, \arg\max_j \Pi_{1,n}^j \neq 2, X_{k-n} = 1) \\ &\quad + P(Z, X_k = 1, \arg\max_j \Pi_{2,n}^j \neq 1, X_{k-n} = 2) \\ &\quad + P(Z, X_k = 2, \arg\max_j \Pi_{2,n}^j \neq 2, X_{k-n} = 2) \\ &= \frac{g_{12}}{g_{12} + g_{21}} g_{21}(1 - g_{12})^{n-1} \times \\ &\quad \Big[P(X_{k-n} = 1) a_{11}^{(n)} I_{\text{argmax}_j \Pi_{1,n}^j \neq 1} + P(X_{k-n} = 1) a_{12}^{(n)} I_{\text{argmax}_j \Pi_{1,n}^j \neq 2} \\ &\quad + P(X_{k-n} = 2) a_{21}^{(n)} I_{\text{argmax}_j \Pi_{2,n}^j \neq 1} + P(X_{k-n} = 2) a_{22}^{(n)} I_{\text{argmax}_j \Pi_{2,n}^j \neq 2} \Big] \\ &= \frac{g_{12}}{g_{12} + g_{21}} g_{21}(1 - g_{12})^{n-1} \sum_{r,s=1}^{2} P(X_{k-n} = r) a_{rs}^{(n)} I_{\text{argmax}_j \Pi_{r,n}^j \neq s} \end{aligned} \tag{2.2}$$

where I is the indicator function and $a_{ij}^{(n)}$ is the (i, j)-th entry of the matrix A^n. For

a more explicit expression for $p(n)$, note that

$$A^n = \begin{bmatrix} \frac{a_{21}+a_{12}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} & \frac{a_{12}-a_{12}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} \\ \frac{a_{21}-a_{21}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} & \frac{a_{12}+a_{21}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} \end{bmatrix} \tag{2.3}$$

which may be verified using induction. Then

$$\Pi_{1,n} = \begin{bmatrix} \frac{a_{21}+a_{12}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} \\ \frac{a_{12}-a_{12}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} \end{bmatrix}$$

and

$$\Pi_{2,n} = \begin{bmatrix} \frac{a_{21}-a_{21}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} \\ \frac{a_{12}+a_{21}(1-a_{12}-a_{21})^n}{a_{12}+a_{21}} \end{bmatrix}.$$

Also define

$$\begin{aligned} q_n^1 &\equiv \frac{a_{21} - a_{12} + 2a_{12}(1 - a_{12} - a_{21})^n}{a_{12} + a_{21}} \\ q_n^2 &\equiv \frac{a_{21} - a_{12} - 2a_{21}(1 - a_{12} - a_{21})^n}{a_{12} + a_{21}}. \end{aligned} \tag{2.4}$$

Note now that $\text{argmax}_j \Pi_{1,n}^j = 1$ is equivalent to $q_n^1 > 0$, $\text{argmax}_j \Pi_{1,n}^j = 2$ is equivalent to $q_n^1 \leq 0$, $\text{argmax}_j \Pi_{2,n}^j = 1$ is equivalent to $q_n^2 > 0$, and $\text{argmax}_j \Pi_{2,n}^j = 2$ is equivalent to $q_n^2 \leq 0$. Then it is easily shown that (2.2) can also be written in the form

$$p(n) = \begin{cases} \frac{g_{12}}{g_{12}+g_{21}} g_{21}(1-g_{12})^{n-1} \left[\frac{a_{21}}{a_{12}+a_{21}} a_{11}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{21}^{(n)} \right] & , \quad q_n^1 \leq 0, q_n^2 \leq 0 \\ \frac{g_{12}}{g_{12}+g_{21}} g_{21}(1-g_{12})^{n-1} \left[\frac{a_{21}}{a_{12}+a_{21}} a_{11}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{22}^{(n)} \right] & , \quad q_n^1 \leq 0, q_n^2 > 0 \\ \frac{g_{12}}{g_{12}+g_{21}} g_{21}(1-g_{12})^{n-1} \left[\frac{a_{21}}{a_{12}+a_{21}} a_{12}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{21}^{(n)} \right] & , \quad q_n^1 > 0, q_n^2 \leq 0 \\ \frac{g_{12}}{g_{12}+g_{21}} g_{21}(1-g_{12})^{n-1} \left[\frac{a_{21}}{a_{12}+a_{21}} a_{12}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{22}^{(n)} \right] & , \quad q_n^1 > 0, q_n^2 > 0. \end{cases} \tag{2.5}$$

Hence

$$P(\text{Error}) = \sum_{n=1}^{\infty} p(n)$$

where $p(n)$ is given by (2.5), $a_{ij}^{(n)}$ is the (i,j)-th entry of A^n in (2.3), and q_n^1 and q_n^2 are given by (2.4). Numerical computation of such infinite series can be easily

Table 2.1: Simulation and analytical comparison of noiseless error probabilities

a_{12}	a_{21}	g_{12}	g_{21}	simulation	analytical
0.51	0.67	0.06	0.30	0.3591	0.3596
0.31	0.34	0.65	0.97	0.2210	0.2208
0.43	0.91	0.18	0.38	0.2178	0.2178
0.74	0.20	0.69	0.59	0.0982	0.0981
0.04	0.09	0.55	0.69	0.0507	0.0506

handled using computer algebra software such as Mathematica.

In Table 2.1 we compare the derived expression with simulation results for a selection of different parameter values. We set the distribution of the initial states of the Markov chains to be equal to the stationary distributions (though by the exponential forgetting property of the HMM filter (Le Gland & Mevel, 2000; Shue *et al.*, 1998) the effect of the initial state should not have a major effect for long runs), and then run Monte Carlo simulations of the filtering updates to obtain the probability of error. The simulations results were averaged over 10 runs, each run of length 100000. It may be seen that, not suprisingly, there is very close agreement at all values considered.

2.3.2 Special cases

In certain cases, the expression for the error probability can be further simplified. We present two examples.

(i) If $a_{12} = a_{21} < 0.5$ so that A is symmetric, then $q_n^1 > 0$ and $q_n^2 \leq 0$ always, and

$$\begin{aligned} P(\text{Error}) &= \sum_{n=1}^{\infty} \frac{1}{2} \frac{g_{12}g_{21}}{g_{12}+g_{21}} (1-g_{12})^{n-1} \left[1-(1-2a_{21})^n\right] \\ &= \frac{g_{21}a_{21}}{(g_{12}+g_{21})(2a_{21}+g_{12}-2a_{21}g_{12})}. \end{aligned}$$

(ii) Suppose the signal is i.i.d., i.e. $a_{12} = 1-a_{21}$. If $a_{21} \leq 0.5$, then $q_n^1 \leq 0, q_n^2 \leq 0, \forall n$, and

$$P(\text{Error}) = \sum_{n=1}^{\infty} \frac{g_{12}g_{21}}{g_{12}+g_{21}} (1-g_{12})^{n-1} a_{21} = \frac{a_{21}g_{21}}{g_{12}+g_{21}}.$$

If $a_{21} > 0.5$, then $q_n^1 > 0, q_n^2 > 0, \forall n$, and

$$P(\text{Error}) = \sum_{n=1}^{\infty} \frac{g_{12}g_{21}}{g_{12}+g_{21}}(1-g_{12})^{n-1}(1-a_{21}) = \frac{(1-a_{21})g_{21}}{g_{12}+g_{21}}.$$

Figure 2.5 of Section 2.3.4 contains some simulation results for this example. The linear dependence on the stationary probability $\frac{g_{21}}{g_{12}+g_{21}}$ when the data is i.i.d. also holds in the noisy case, see Section 2.4.4.

2.3.3 Theoretical properties

We now demonstrate some relationships between the probability of error and the parameters of the packet loss process. Proofs of Theorems 2.3.1 - 2.3.3 may be found in the appendix to this chapter, Section 2.6.

Theorem 2.3.1. *For fixed A and g_{12}, the probability of error is monotonically increasing in g_{21}.*

Theorem 2.3.2. *For fixed A and g_{21}, the probability of error is monotonically decreasing in g_{12}.*

Theorem 2.3.1 states that the error probability increases as g_{21} increases, when all other parameters are fixed. Intuitively this is reasonable, since g_{21} is the probability that the next packet is lost given that the current packet has been received, so we are more likely to drop packets and do worse at estimation when this parameter is increased. Theorem 2.3.2 is also quite intuitive, as g_{12} is the probability that the next packet will be received correctly given that the current packet has been lost, so increasing this parameter should improve our estimation performance.

For the third result, let p_0 be the stationary probability that a measurement is not received, i.e. $p_0 = P(Z_k = 1) = \frac{g_{21}}{g_{12}+g_{21}}$. Theorem 2.3.3 shows that in general p_0 alone does not uniquely determine the error probabilities, but also depends on the sizes of g_{12} and g_{21}, which can be interpreted as how quickly/slowly the packet loss process is varying in time. For example, when both g_{12} and g_{21} are small, transitions from one state to the other are rare, so that we can regard the Markov chain as being slow. Essentially, Theorem 2.3.3 says that for a given p_0, slower dynamics are worse

for state estimation in that we will get a higher probability of error (except perhaps when the signal is i.i.d. as noted in example (ii) of Section 2.3.2).

Theorem 2.3.3. *(i) For fixed A and p_0, the probability of error is non-increasing in g_{12} (equivalently in g_{21}).*
(ii) The probability of error converges, for fixed p_0 and as $g_{12} \to 0$, to $p_0 E_0$, where $E_0 \equiv \min(\frac{a_{12}}{a_{12}+a_{21}}, \frac{a_{21}}{a_{12}+a_{21}})$ is the probability of error in the complete absence of observations.

By Theorem 2.3.3 (ii), $p_0 E_0$ is therefore an upper bound on the error probability. A lower bound can also be derived, by noting that for $p_0 \leq \frac{1}{2}$, the largest possible values for g_{12} and g_{21} where one can obtain p_0 are $g_{12} = 1$ and $g_{21} = \frac{p_0}{1-p_0}$, and for $p_0 > \frac{1}{2}$, the largest possible values are $g_{21} = 1$ and $g_{12} = \frac{1-p_0}{p_0}$. Substituting these values into the formula for $P(\text{Error})$ and using Theorem 2.3.3 (i), we then obtain the lower bound

$$P(\text{Error}) \geq \begin{cases} p_0 t(1) & , \quad p_0 \leq 0.5 \\ \sum_{n=1}^{\infty} (1-p_0) \left(\frac{2p_0-1}{p_0}\right)^{n-1} t(n) & , \quad p_0 > 0.5. \end{cases}$$

where $t(n)$ is defined as (also see Section 2.6):

$$t(n) \equiv \begin{cases} \frac{a_{21}(a_{12}+a_{21})}{(a_{12}+a_{21})^2} & , \quad q_n^1 \leq 0, q_n^2 \leq 0 \\ \frac{a_{21}^2+2a_{21}a_{12}(1-a_{21}-a_{12})^n+a_{12}^2}{(a_{12}+a_{21})^2} & , \quad q_n^1 \leq 0, q_n^2 > 0 \\ \frac{2a_{21}a_{12}-2a_{21}a_{12}(1-a_{21}-a_{12})^n}{(a_{12}+a_{21})^2} & , \quad q_n^1 > 0, q_n^2 \leq 0 \\ \frac{a_{12}(a_{12}+a_{21})}{(a_{12}+a_{21})^2} & , \quad q_n^1 > 0, q_n^2 > 0. \end{cases}$$

As a special case, when $a_{12} = a_{21} < 0.5$ we have the more explicit expression

$$P(\text{Error}) \geq \begin{cases} a_{21} p_0 & , \quad p_0 \leq 0.5 \\ \frac{a_{21} p_0^2}{1-2a_{21}+(4a_{21}-1)p_0} & , \quad p_0 > 0.5. \end{cases}$$

For the i.i.d. signal case, it can also be shown that the upper and lower bounds coincide.

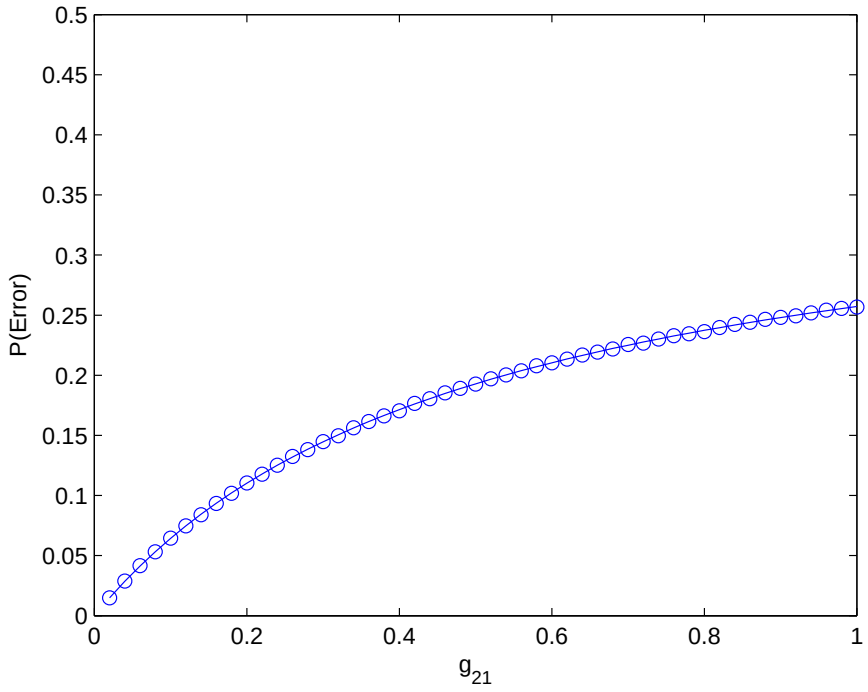

Figure 2.1: Noiseless probability of error for various g_{21}

2.3.4 Numerical studies

For the simulations in the first three graphs, the length of each run is one million. The solid lines represent the theoretical error probability. In Figure 2.1 we plot the simulated probability of error for 50 values of g_{21}, with $a_{12} = 0.3$, $a_{21} = 0.4$, and $g_{12} = 0.5$. In Figure 2.2 we plot the simulated probability of error for 50 values of g_{12}, with $a_{12} = 0.3$, $a_{21} = 0.4$, and $g_{21} = 0.5$. In Figure 2.3 we plot the simulated probability of error for 50 values of g_{21}, with $g_{12} = g_{21}$ (i.e. $p_0 = 0.5$ is fixed), $a_{12} = 0.1$ and $a_{21} = 0.1$. We can see that the results are in agreement with Theorems 2.3.1 - 2.3.3 respectively.

For the next two graphs, we randomly generate both g_{21} and g_{12}, and then form $p_0 = \frac{g_{21}}{g_{12}+g_{21}}$. The length of each simulation run is 100000. In Figure 2.4 we plot the simulated probability of error for 500 values of p_0, with $a_{12} = a_{21} = 0.1$, i.e. A is symmetric. The solid lines are plots of the upper and lower bounds on the error probability mentioned after the statement of Theorem 2.3.3. The simulation results can be seen to lie within the bounds. In Figure 2.5 we plot the simulated probability of error for 500 values of p_0, with $a_{12} = 0.3$ and $a_{21} = 0.7$, i.e. signal is i.i.d. The

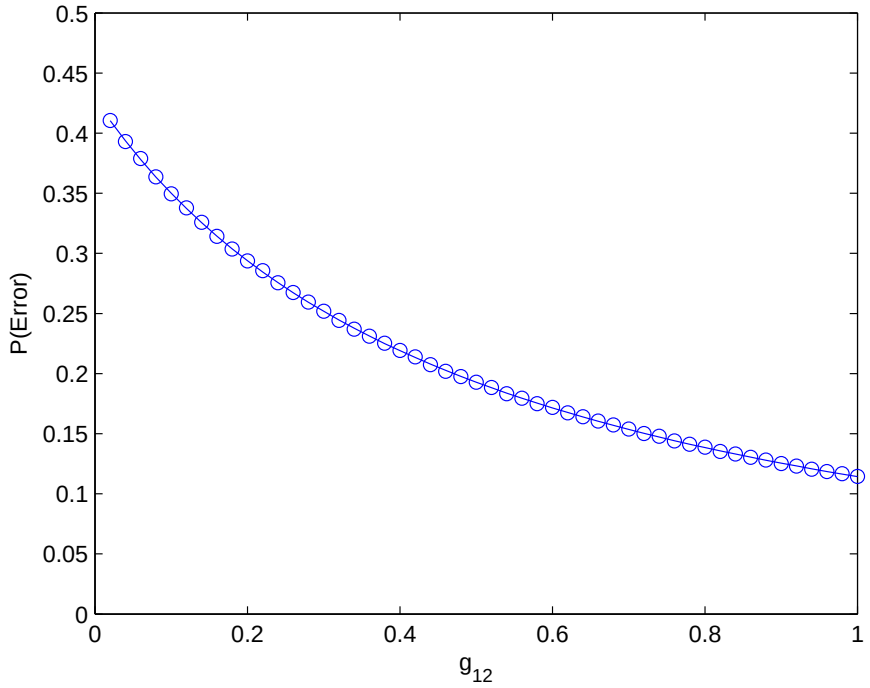

Figure 2.2: Noiseless probability of error for various g_{12}

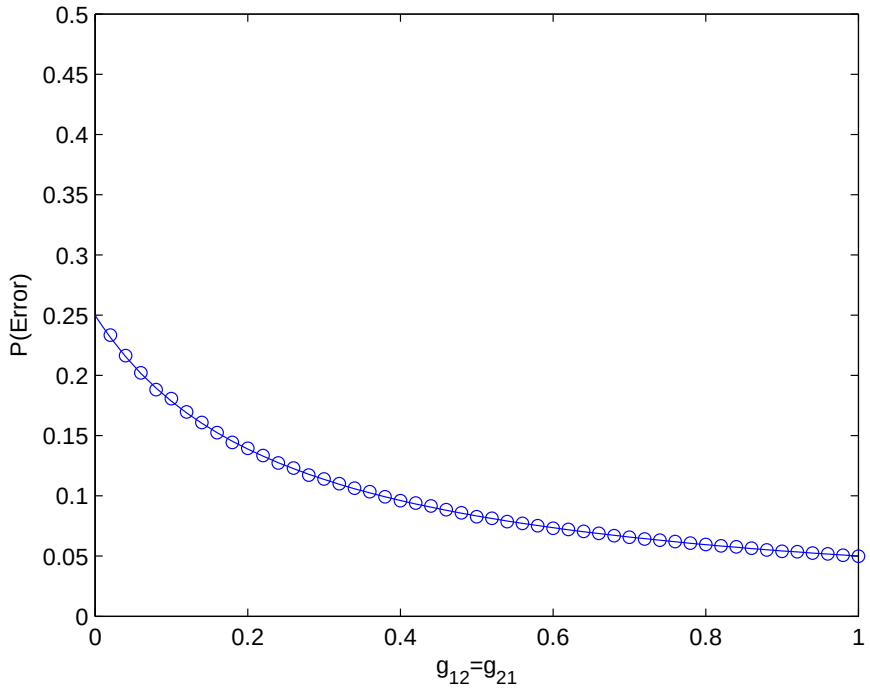

Figure 2.3: Noiseless probability of error for various g_{21}, with p_0 fixed

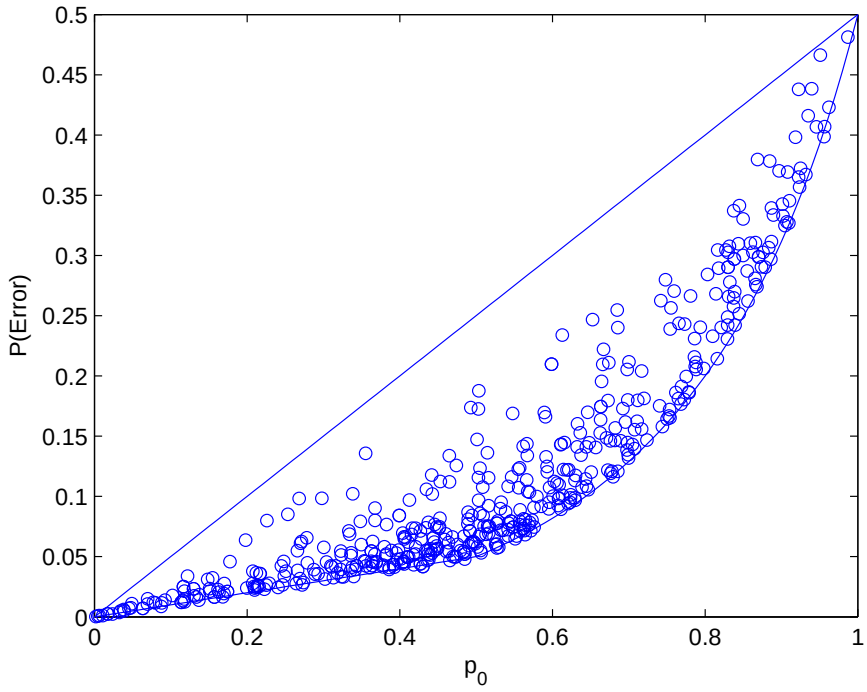

Figure 2.4: Noiseless probability of error and bounds for various p_0, symmetric A

linear dependence on the probability of not receiving a packet, agrees with example (ii) of Section 2.3.2.

2.3.5 Multiple states

We now consider $\{X_k\}$ as a Markov chain having $M > 2$ states, with the assumption that it has a unique stationary distribution. The packet loss process $\{Z_k\}$ is still the same two-state Markov chain.

The probability vectors are updated as:

$$\Pi_{k+1|k+1} = \begin{cases} A^T \Pi_{k|k} & , \quad \gamma(Z_{k+1}) = 0 \\ \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} = h(1) \\ \begin{bmatrix} 0 & 1 & \dots & 0 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} = h(2) \\ \vdots & \vdots \\ \begin{bmatrix} 0 & 0 & \dots & 1 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} = h(M) \end{cases}$$

and MAP estimates of the states are obtained as $\hat{X}_k = \arg\max_{j=1,\dots,M} \Pi^j_{k|k}$.

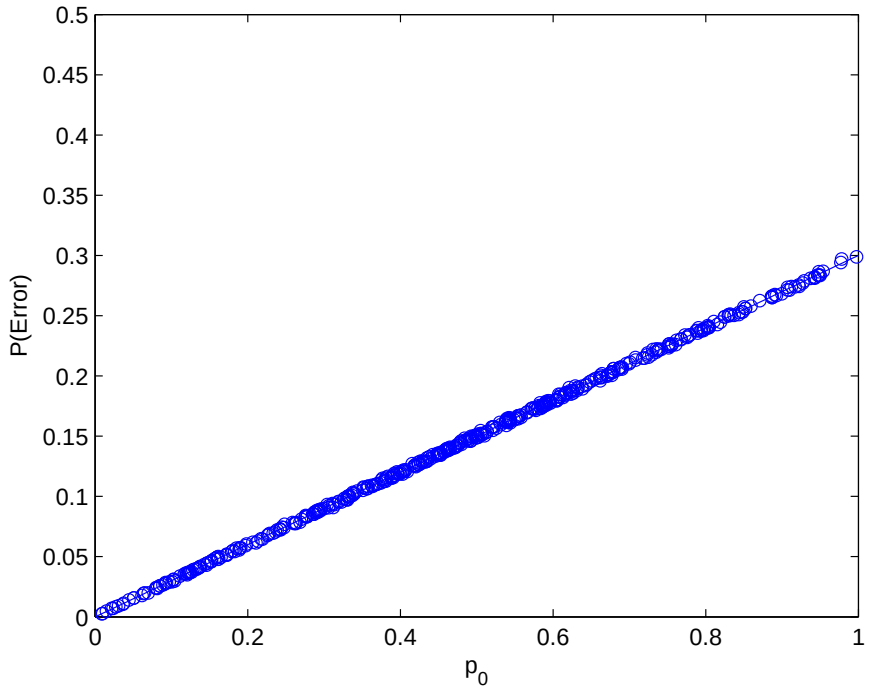

Figure 2.5: Noiseless probability of error for various p_0, signal i.i.d.

Define $\Pi_{i,n} \equiv (A^n)^T \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix}^T$ where the 1 is in the i-th position. Then in a straightforward extension of the two-state case (2.2), we can derive

$$p(n) = \frac{g_{12}}{g_{12}+g_{21}} g_{21}(1-g_{12})^{n-1} \sum_{r,s=1}^{M} P(X_{k-n}=r) a_{rs}^{(n)} I_{\text{argmax}_j \Pi_{r,n}^j \neq s} \qquad (2.6)$$

with the probability of error given by $P(\text{Error}) = \sum_{n=1}^{\infty} p(n)$. More explicit expressions (in terms of the Markov chain parameters) for the stationary probabilities $P(X_{k-n}=r)$ and the elements $a_{rs}^{(n)}$ would be very complicated to write down without additional structure in the transition matrix A, since for M states, the A matrix would have $M(M-1)$ free parameters in general. However, given a set of parameters these quantities can be evaluated on a computer quite easily.

As an example, let $g_{12} = 0.48$, $g_{21} = 0.26$ and

$$A = \begin{bmatrix} 0.18 & 0.76 & 0.06 \\ 0.56 & 0.36 & 0.08 \\ 0.13 & 0.27 & 0.60 \end{bmatrix}.$$

Computation of the error probability using (2.6) is 0.1527. Simulation results averaged over 10 runs, each of length 100000, gives an error probability of 0.1523, which is very close to the analytical expression.

As another example, where an explicit expression is available, assume that the signal is governed by an M-state Markov chain, with $\epsilon \leq \frac{1}{M}$ and transition matrix

$$A = \begin{bmatrix} 1-(M-1)\epsilon & \epsilon & \dots & \epsilon \\ \epsilon & 1-(M-1)\epsilon & \dots & \epsilon \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon & \epsilon & \dots & 1-(M-1)\epsilon \end{bmatrix}.$$

In (Khasminskii & Zeitouni, 1996), a similar model is considered where the off-diagonal entries are not necessarily identical (the diagonal entries are also not necessarily identical), but need to be close to zero for their asymptotic results to be valid. Here however, ϵ does not need to be asymptotically small, only that $\epsilon \leq \frac{1}{M}$ is satisfied. It can be verified that the (i,j)-th entries of A^n are

$$a_{ij}^{(n)} = \begin{cases} \frac{1}{M}\left(1+(M-1)(1-M\epsilon)^n\right) & , \quad i=j \\ \frac{1}{M}\left(1-(1-M\epsilon)^n\right) & , \quad i \neq j. \end{cases}$$

Hence

$$p(n) \;=\; \frac{g_{12}g_{21}}{g_{12}+g_{21}}(1-g_{12})^{n-1}\frac{M-1}{M}(1-(1-M\epsilon)^n)$$

and

$$P(\text{Error}) = \sum_{n=1}^{\infty} p(n) = \frac{(M-1)g_{21}\epsilon}{(g_{12}+g_{21})(M\epsilon+g_{12}-M\epsilon g_{12})}.$$

2.4 Noisy case

2.4.1 Derivation of probability of error

In this section we derive expressions for the probability of error in the noisy case. Unfortunately, the methods used for the noiseless case don't seem to extend to the

situation here. We will use a different method, whose analysis is based in part on (Shue *et al.*, 2001). Recall the model (2.1) from Section 2.2,

$$Y_k = \gamma(Z_k)h(X_k) + v_k.$$

Given X_k and $\gamma(Z_k) = 1$, the observations Y_k are conditionally distributed as

$$\begin{aligned} P(Y_k \in dy | X_k = i, Z_k = 2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_k - h(i))^2}{2\sigma^2}\right) dy \\ &\equiv b_i(y_k)dy. \end{aligned}$$

A straightforward modification of (1.3) to incorporate the packet loss shows that the probability vectors can be updated recursively as follows:

$$\Pi_{k+1|k+1} = \begin{cases} A^T\Pi_{k|k} & , \ \gamma(Z_{k+1}) = 0 \\ \dfrac{B_{y_{k+1}}A^T\Pi_{k|k}}{\begin{bmatrix} 1 & 1 \end{bmatrix} B_{y_{k+1}}A^T\Pi_{k|k}} & , \ \gamma(Z_{k+1}) = 1 \end{cases} \tag{2.7}$$

where

$$B_{y_{k+1}} = \begin{bmatrix} b_1(y_{k+1}) & 0 \\ 0 & b_2(y_{k+1}) \end{bmatrix}.$$

Using the definition $q_k \equiv \Pi^1_{k|k} - \Pi^2_{k|k}$ (noting that $-1 \leq q_k \leq 1$), the probability of filtering error can then be written as

$$\begin{aligned} P(\text{Error}) &= P(X_k = 1, \hat{X}_k = 2) + P(X_k = 2, \hat{X}_k = 1) \\ &= P(X_k = 1, Z_k = 1, q_k \leq 0) + P(X_k = 1, Z_k = 2, q_k \leq 0) \\ &\quad + P(X_k = 2, Z_k = 1, q_k > 0) + P(X_k = 2, Z_k = 2, q_k > 0) \\ &= \int_{-1}^{0} f_k^{1,1}(q)dq + \int_{-1}^{0} f_k^{1,2}(q)dq + \int_{0}^{1} f_k^{2,1}(q)dq + \int_{0}^{1} f_k^{2,2}(q)dq \end{aligned} \tag{2.8}$$

where $f_k^{i,I}(q)dq \equiv P(X_k = i, Z_k = I, q_k \in (q, q + dq))$.

To find a recursive relation which will allow us to characterize the densities

$f_k^{i,I}(q)$, first consider

$$\begin{aligned}
&P(X_{k+1}=i, Z_{k+1}=I, q_{k+1}\in(q,q+dq), X_k=j, Z_k=J, q_k\in(\tilde{q},\tilde{q}+d\tilde{q}))\\
&\quad= P(q_{k+1}\in(q,q+dq)|X_{k+1}=i, Z_{k+1}=I, q_k=\tilde{q})\\
&\qquad\times P(X_{k+1}=i, Z_{k+1}=I|X_k=j, Z_k=J)\times P(X_k=j, Z_k=J, q_k\in(\tilde{q},\tilde{q}+d\tilde{q}))\\
&\quad= S_{i,I}(q,\tilde{q})dq\times a_{ij}g_{IJ}\times f_k^{j,J}(\tilde{q})d\tilde{q}
\end{aligned}$$

where we have used the Markov property and the independence of $\{X_k\}$ and $\{Z_k\}$, and defined

$$S_{i,I}(q,\tilde{q})dq \equiv P(q_{k+1}\in(q,q+dq)|X_{k+1}=i, Z_{k+1}=I, q_k=\tilde{q}).$$

For $\gamma(Z_{k+1})=1$, i.e. $Z_{k+1}=2$, we can derive in a similar manner to (Shue *et al.*, 2001) (which only contained expressions for symmetric A) the recursion

$$\begin{aligned}
&q_{k+1}\\
&=\frac{\exp(-\frac{2y_{k+1}}{\sigma^2})[1+a_{21}-a_{12}+(1-a_{21}-a_{12})q_k]-1+[a_{21}-a_{12}+(1-a_{21}-a_{12})q_k]}{\exp(-\frac{2y_{k+1}}{\sigma^2})[1+a_{21}-a_{12}+(1-a_{21}-a_{12})q_k]+1-[a_{21}-a_{12}+(1-a_{21}-a_{12})q_k]}
\end{aligned}\tag{2.9}$$

and so

$$S_{i,2}(q,\tilde{q})dq=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(g(q,\tilde{q})-h(i))^2}{2\sigma^2}\right]\frac{\sigma^2}{1-q^2}dq$$

with

$$g(q_{k+1},q_k)=-\frac{\sigma^2}{2}\ln\frac{(1+q_{k+1})[1-(a_{21}-a_{12})-(1-a_{21}-a_{12})q_k]}{(1-q_{k+1})[1+(a_{21}-a_{12})+(1-a_{21}-a_{12})q_k]}.$$

For $\gamma(Z_{k+1})=0$, i.e. $Z_{k+1}=1$, it is straightforward to show that the recursion for q is now

$$q_{k+1}=a_{21}-a_{12}+(1-a_{21}-a_{12})q_k.\tag{2.10}$$

Thus

$$\begin{aligned} S_{i,1}(q,\tilde{q})dq &= P(q_{k+1} \in (q, q+dq) | X_{k+1} = i, Z_{k+1} = 1, q_k = \tilde{q}) \\ &= P(a_{21} - a_{12} + (1 - a_{21} - a_{12})\tilde{q} \in (q, q+dq)) \\ &= I_B(q, \tilde{q}, dq) \end{aligned}$$

where I is the indicator function and

$$B \equiv \{(q, \tilde{q}, dq) : a_{21} - a_{12} + (1 - a_{21} - a_{12})\tilde{q} \in (q, q+dq)\}.$$

Hence in the steady state we have the following relations for the densities, which is a system of four Fredholm integral equations

$$f^{i,I}(q)dq = \sum_{j=1}^{2}\sum_{J=1}^{2} a_{ji} g_{JI} \int_{-1}^{1} S_{i,I}(q,\tilde{q}) f^{j,J}(\tilde{q}) d\tilde{q} dq, \qquad i = 1,2, I = 1,2. \quad (2.11)$$

2.4.2 Numerical method

To compute the error probability we will need to numerically solve the system of integral equations (2.11). We will present an existing method which is slightly different from that of (Shue *et al.*, 2001), where convergence analysis is perhaps more readily obtained.[5]

By a result from (Tricomi, 1957, p.151), we can transform (2.11) into a single integral equation as follows. Define

$$\Phi(q) = \begin{cases} f^{1,1}(q) & , \quad -1 < q < 1 \\ f^{1,2}(q-2) & , \quad 1 < q < 3 \\ f^{2,1}(q-4) & , \quad 3 < q < 5 \\ f^{2,2}(q-6) & , \quad 5 < q < 7 \end{cases} \quad (2.12)$$

[5]The method presented here and its analysis can also be applied with slight modifications to the original problem in (Shue *et al.*, 2001)

$$K(q,\tilde{q}) = \begin{cases} a_{11}g_{11}S_{1,1}(q,\tilde{q}) & , \ -1<q<1, -1<\tilde{q}<1 \\ a_{11}g_{21}S_{1,1}(q,\tilde{q}-2) & , \ -1<q<1, \ \ 1<\tilde{q}<3 \\ a_{21}g_{11}S_{1,1}(q,\tilde{q}-4) & , \ -1<q<1, \ \ 3<\tilde{q}<5 \\ a_{21}g_{21}S_{1,1}(q,\tilde{q}-6) & , \ -1<q<1, \ \ 5<\tilde{q}<7 \\ a_{11}g_{12}S_{1,2}(q-2,\tilde{q}) & , \ \ 1<q<3, -1<\tilde{q}<1 \\ a_{11}g_{22}S_{1,2}(q-2,\tilde{q}-2) & , \ \ 1<q<3, \ \ 1<\tilde{q}<3 \\ a_{21}g_{12}S_{1,2}(q-2,\tilde{q}-4) & , \ \ 1<q<3, \ \ 3<\tilde{q}<5 \\ a_{21}g_{22}S_{1,2}(q-2,\tilde{q}-6) & , \ \ 1<q<3, \ \ 5<\tilde{q}<7 \\ a_{12}g_{11}S_{2,1}(q-4,\tilde{q}) & , \ \ 3<q<5, -1<\tilde{q}<1 \\ a_{12}g_{21}S_{2,1}(q-4,\tilde{q}-2) & , \ \ 3<q<5, \ \ 1<\tilde{q}<3 \\ a_{22}g_{11}S_{2,1}(q-4,\tilde{q}-4) & , \ \ 3<q<5, \ \ 3<\tilde{q}<5 \\ a_{22}g_{21}S_{2,1}(q-4,\tilde{q}-6) & , \ \ 3<q<5, \ \ 5<\tilde{q}<7 \\ a_{12}g_{12}S_{2,2}(q-6,\tilde{q}) & , \ \ 5<q<7, -1<\tilde{q}<1 \\ a_{12}g_{22}S_{2,2}(q-6,\tilde{q}-2) & , \ \ 5<q<7, \ \ 1<\tilde{q}<3 \\ a_{22}g_{12}S_{2,2}(q-6,\tilde{q}-4) & , \ \ 5<q<7, \ \ 3<\tilde{q}<5 \\ a_{22}g_{22}S_{2,2}(q-6,\tilde{q}-6) & , \ \ 5<q<7, \ \ 5<\tilde{q}<7. \end{cases}$$

Then it can be seen that (2.11) is equivalent to the homogeneous Fredholm equation

$$\Phi(q) - \int_{-1}^{7} K(q,\tilde{q})\Phi(\tilde{q})d\tilde{q} = 0 \tag{2.13}$$

with $\Phi(q)$ also satisfying the normalising condition $\int_{-1}^{7}\Phi(q)dq = 1$. For the numerical solution of (2.13), consider the related eigenvalue problem (Hackbusch, 1995)

$$\gamma\Phi(q) = \int_{-1}^{7} K(q,\tilde{q})\Phi(\tilde{q})d\tilde{q} \tag{2.14}$$

which corresponds to (2.13) when $\gamma = 1$. We will solve (2.14) using the Nyström method (Hackbusch, 1995; Atkinson, 1997). Replacing the integral by a $4N$-point

quadrature rule[6], and defining

$$\mathbf{K} \equiv \begin{bmatrix} w_1 K(t_1, t_1) & \dots & w_{4N} K(t_1, t_{4N}) \\ \vdots & \ddots & \vdots \\ w_1 K(t_{4N}, t_1) & \dots & w_{4N} K(t_{4N}, t_{4N}) \end{bmatrix}$$

we obtain

$$\mathbf{K} \begin{bmatrix} \Phi(t_1) \\ \vdots \\ \Phi(t_{4N}) \end{bmatrix} = \gamma \begin{bmatrix} \Phi(t_1) \\ \vdots \\ \Phi(t_{4N}) \end{bmatrix}$$

where w_j represent the weights and t_j the quadrature points of the quadrature rule. In the results presented in this chapter the midpoint rule is used, though other alternatives such as composite Gauss-Legendre quadrature are possible (Atkinson, 1997, p.110).[7]

To obtain an approximation for $\Phi(q)$, we then take the eigenvector that corresponds to the largest real eigenvalue of $\mathbf{K}$, and normalise it so that $\int_{-1}^{7} \Phi(q)dq = 1$ is satisfied. Using a "weak" version of the Perron-Frobenius theorem (Seneta, 1981, p.28) on $\mathbf{K}$ shows that this eigenvector will have non-negative entries, which is required if it is to approximate a probability density. The probability of error can then be calculated from (2.8) and (2.12).

We would like the largest eigenvalue to be close to one. As $N \to \infty$, $\mathbf{K}$ will tend to a column stochastic matrix, in the sense that the sum of each column will converge to one. For example, for the first column

$$\begin{aligned} \sum_{i=1}^{4N} w_1 K(t_i, t_1) \to \int_{-1}^{1} \Big(& a_{11} g_{11} S_{1,1}(q, t_1) + a_{11} g_{12} S_{1,2}(q, t_1) \\ & + a_{12} g_{11} S_{2,1}(q, t_1) + a_{12} g_{12} S_{2,2}(q, t_1) \Big) dq \\ = a_{11} g_{11} + a_{11} g_{12} + a_{12} g_{11} + a_{12} g_{12} = 1 & \end{aligned}$$

where convergence is achieved for any reasonable composite quadrature scheme, such

[6] We call it a $4N$-point rather an N-point quadrature rule for convenience, since $\Phi(q)$ is a combination of 4 densities

[7] The Gauss-Legendre quadrature rule will be used in Chapter 3 of this thesis.

as the midpoint rule (Ralston & Rabinowitz, 1978, p.116). Similarly this holds for the other columns of $\mathbf{K}$. Since a stochastic matrix has largest eigenvalue one, and by the continuity of eigenvalues (Stewart & Sun, 1990), it then follows that the largest eigenvalue of $\mathbf{K}$ can be made arbitrarily close to one for N sufficiently large. We can also show that the eigenvector corresponding to the largest eigenvalue is unique. Note that we can partition $\mathbf{K}$ as

$$\mathbf{K} = \begin{bmatrix} a_{11}g_{11}S_{11} & a_{11}g_{21}S_{11} & a_{21}g_{11}S_{11} & a_{21}g_{21}S_{11} \\ a_{11}g_{12}S_{12} & a_{11}g_{22}S_{12} & a_{21}g_{12}S_{12} & a_{21}g_{22}S_{12} \\ a_{12}g_{11}S_{21} & a_{12}g_{21}S_{21} & a_{22}g_{11}S_{21} & a_{22}g_{21}S_{21} \\ a_{12}g_{12}S_{22} & a_{12}g_{22}S_{22} & a_{22}g_{12}S_{22} & a_{22}g_{22}S_{22} \end{bmatrix} \tag{2.15}$$

with each element representing an $N \times N$ matrix. Referring to (2.15) and the definition of $\mathbf{K}$, it may be easily shown that the blocks $a_{ij}g_{IJ}S_{12}$ and $a_{ij}g_{IJ}S_{22}$ contain strictly positive entries. The blocks $a_{ij}g_{IJ}S_{11}$ and $a_{ij}g_{IJ}S_{21}$ can each be further divided into the form

$$\left[\begin{array}{c} 1 \\ \hline 2 \\ \hline 3 \end{array}\right]$$

where blocks 1 and 3 contain all zeros, while block 2 has at least one positive entry in each row and column, with the sizes of these blocks being the same for all $a_{ij}g_{IJ}S_{11}$ and $a_{ij}g_{IJ}S_{21}$. The matrix $\mathbf{K}$ is thus reducible since it will have a number of rows which consist entirely of zeros, and the Perron-Frobenius theorem (Seneta, 1981) is not directly applicable. We can however form a submatrix by deleting the all-zero rows and the associated columns, without changing the largest eigenvalue. Due to the structure in the blocks $a_{ij}g_{IJ}S_{11}$ and $a_{ij}g_{IJ}S_{21}$, it can be seen that this submatrix is primitive. The Perron-Frobenius theorem may then be applied to conclude that there is an eigenvalue with magnitude greater than any other, with a unique eigenvector (up to constant multiples). By the comments above, this will also hold for the original matrix $\mathbf{K}$.

In Table 2.2 we compare the numerical method with simulation results for a selection of different parameter values. Here we use $N = 1000$, and fix $\sigma^2 = 1$.

Table 2.2: Simulation and analytical comparison of noisy error probabilities, with $\sigma^2 = 1$

a_{12}	a_{21}	g_{12}	g_{21}	simulation	analytical
0.89	0.30	0.26	0.93	0.2241	0.2248
0.86	0.97	0.06	0.78	0.3882	0.3887
0.14	0.68	0.07	0.10	0.1415	0.1411
0.07	0.88	0.71	0.87	0.0658	0.0654
0.55	0.50	0.22	0.50	0.3793	0.3789

Simulations results were again averaged over 10 runs, each run of length 100000. It may be seen that there is very close agreement at the values considered here. However, for smaller values of σ, it has been observed (see Table 2.3, also (Shue *et al.*, 2001)) that the accuracy of the numerical method is not so good when using $N = 1000$. By the previous statements, the accuracy should increase with N, but due to memory limitations increasing N substantially is currently not feasible. In Section 2.4.5 we will derive an approximation for the error probability which is more computationally tractible and provides close agreement with simulations for small σ.

2.4.3 Numerical studies

We now show some plots which are analogues of Figures 2.1 - 2.5, with an additional noise term of variance $\sigma^2 = 1$. The simulation runs are of length one million for the first three graphs. In Figure 2.6 we plot the simulated probability of error for 50 values of g_{21}, with $a_{12} = 0.3$, $a_{21} = 0.4$, and $g_{12} = 0.5$. In Figure 2.7 we plot the simulated probability of error for 50 values of g_{12}, with $a_{12} = 0.3$, $a_{21} = 0.4$, and $g_{21} = 0.5$. In Figure 2.8 we plot the simulated probability of error for 50 values of g_{21}, with $g_{21} = g_{12}$, $a_{12} = 0.1$ and $a_{21} = 0.1$. The solid lines represents the analytical calculation using $N = 200$. In Figure 2.8 there is a slight but noticeable discrepancy between the simulation and analytical results at small values of g_{21}. This is due to the use of $N = 200$ in the numerical calculuation, as it is found that increasing N to say 500 would give much closer agreement. For Figures 2.9 and 2.10 the simulations runs are of length 100000. In Figure 2.9 we plot the simulated probability of error for 500 random values of p_0, with $a_{12} = a_{21} = 0.1$, i.e. A is symmetric. In Figure

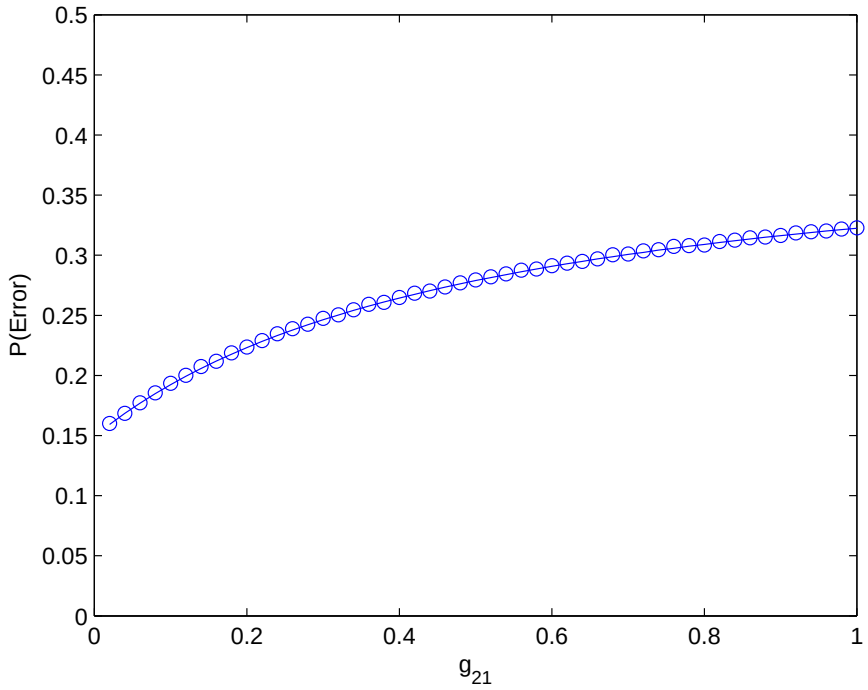

Figure 2.6: Noisy probability of error for various g_{21}

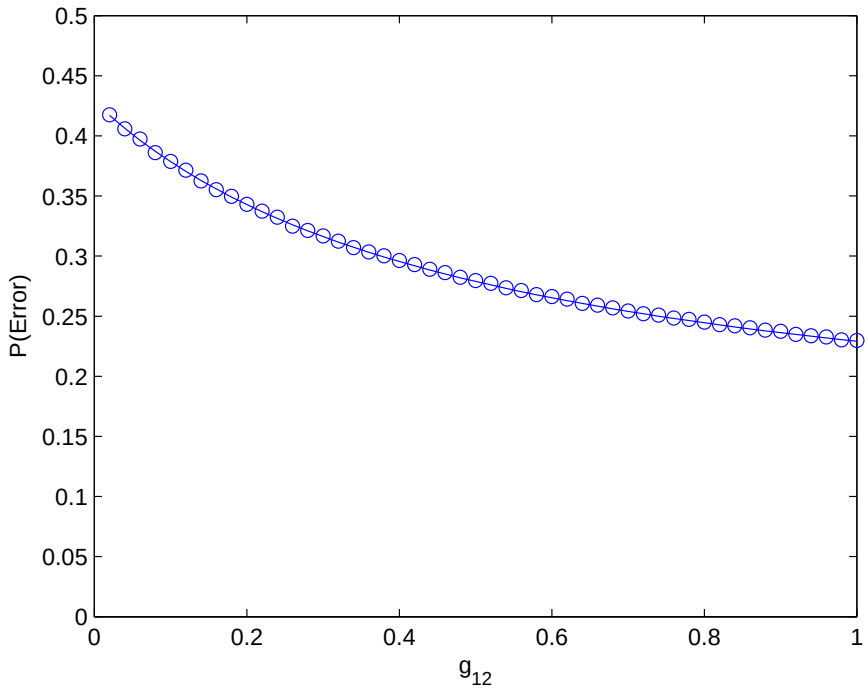

Figure 2.7: Noisy probability of error for various g_{12}

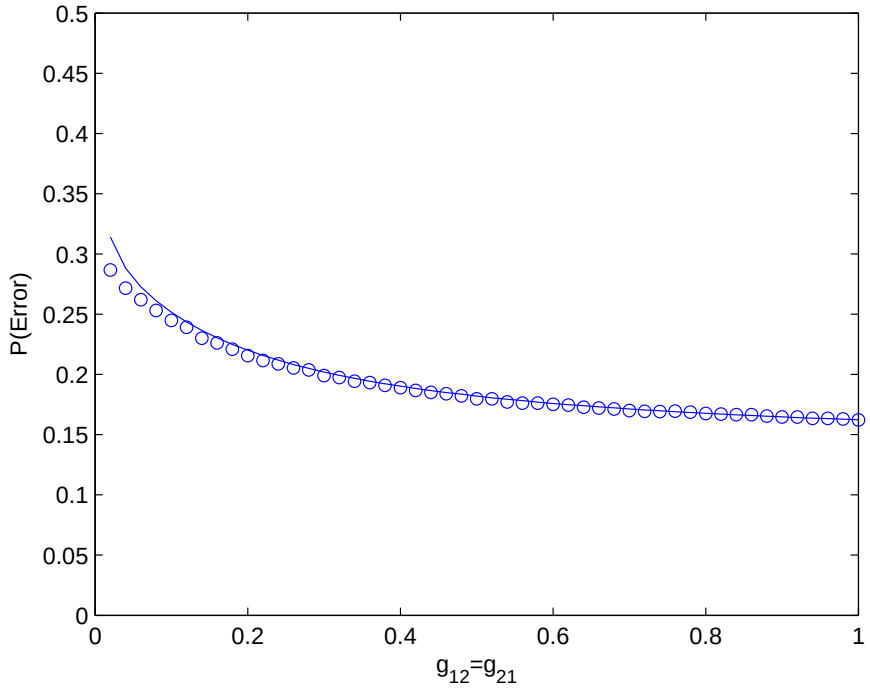

Figure 2.8: Noisy probability of error for various g_{21}, with p_0 fixed

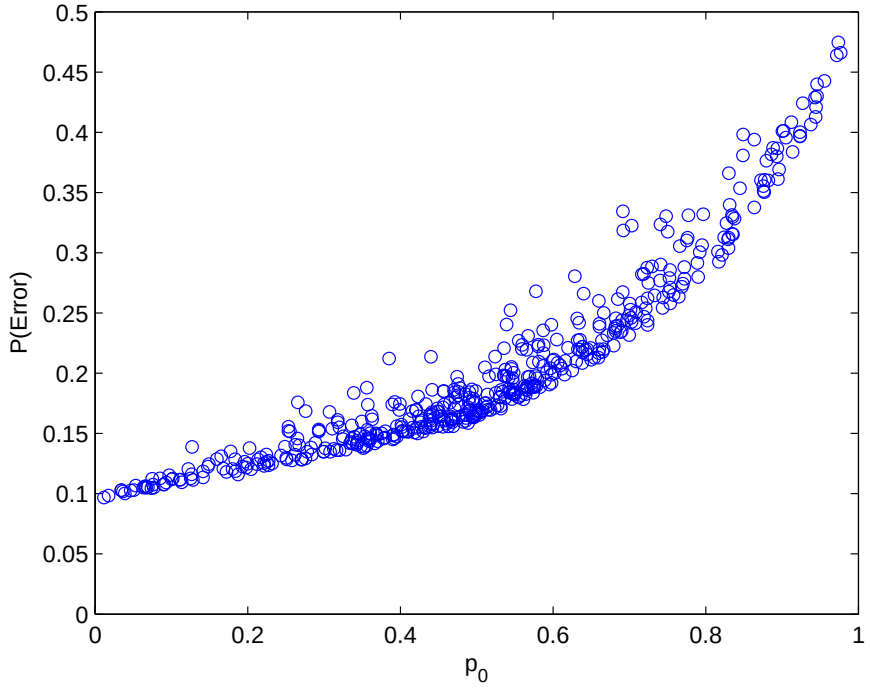

Figure 2.9: Noisy probability of error for various p_0, symmetric A

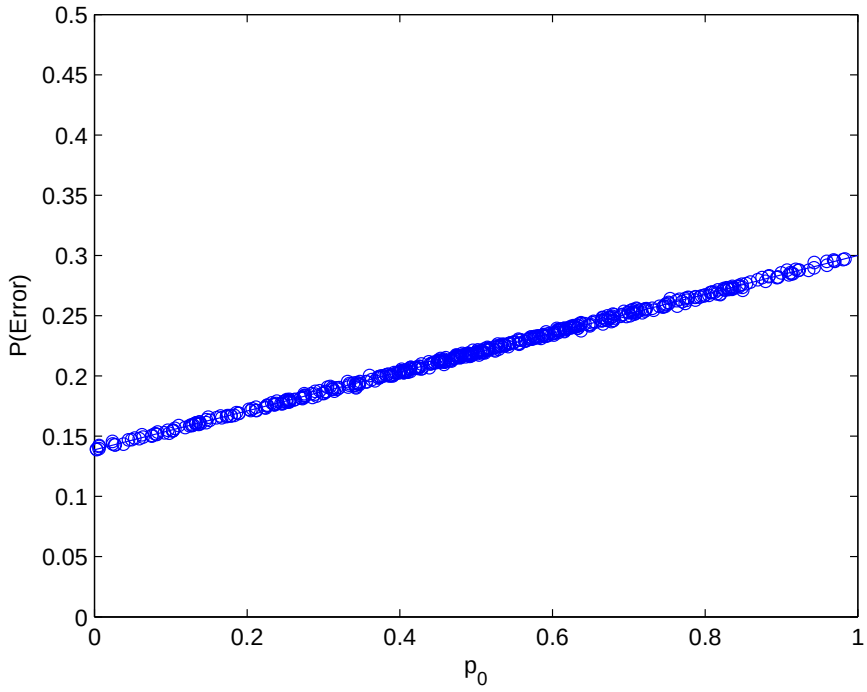

Figure 2.10: Noisy probability of error for various p_0, signal i.i.d.

2.10 we plot the simulated probability of error for 500 random values of p_0, with $a_{12} = 0.3$ and $a_{21} = 0.7$, i.e. signal i.i.d. Comparing these graphs with Figures 2.1 - 2.5, we can see that there is a noise floor introduced which tends to shift the graphs upwards. The noiseless Figures 2.1 - 2.5 also seem to cover a larger range of values as the parameters vary, one reason could be that there is a greater sensitivity to packet loss in the noiseless case, since losing packets is the only way in which one can make an error there.

Proving results similar to Theorems 2.3.1 - 2.3.3 in the noisy case seems to be very difficult. We will state these as conjectures.

Conjectures

1. For fixed A, σ and g_{12}, the probability of error is monotonically increasing in g_{21}.
2. For fixed A, σ and g_{21}, the probability of error is monotonically decreasing in g_{12}.
3. Let p_0 be the stationary probability that a measurement is not received, i.e. $p_0 = P(Z_k = 1) = \frac{g_{21}}{g_{12}+g_{21}}$. Then for fixed A, σ and p_0, the probability of error is non-increasing with g_{21}. Furthermore, this probability of error converges, for fixed p_0 and as $g_{21} \to 0$, to $p_0 E_0 + (1 - p_0)E_1$, where E_1 is the probability of error

obtained when measurements are always received, and $E_0 \equiv \min(\frac{a_{12}}{a_{12}+a_{21}}, \frac{a_{21}}{a_{12}+a_{21}})$ is the probability of error in the complete absence of observations.

The intuition behind conjecture 3 can be explained as follows: In (Le Gland & Mevel, 2000; Shue *et al.*, 1998), the exponential forgetting property of HMM filters is demonstrated. For slowly varying dynamics in the loss process, during the (usually) long periods where we always receive measurements, the probability of error there would be close to the error probability when we assume that measurements are always received. Similarly, for the periods when we don't obtain any observations, the error probability would be close to the error in the complete absence of observations. The overall error probability should then be able to be averaged over these two situations, giving $p_0E_0 + (1-p_0)E_1$ as the conjectured limit.

2.4.4 Signal is i.i.d.

Signals which are i.i.d. (or close to i.i.d.) are commonly encountered in digital communications. In Figure 2.10, we saw that when the signal $\{X_k\}$ is i.i.d., there seems to be a linear relationship between p_0 and the probability of error even when the loss process is Markovian. In this section we will show that this is indeed the case. If the signal is i.i.d., the A matrix has the form

$$A = \begin{bmatrix} a_{21} & 1-a_{21} \\ a_{21} & 1-a_{21} \end{bmatrix}$$

and the probability vector updates (2.7) become

$$\Pi_{k+1|k+1} = \begin{cases} \begin{bmatrix} a_{21} \\ 1-a_{21} \end{bmatrix} & , \gamma(Z_{k+1}) = 0 \\ \begin{bmatrix} \frac{a_{21}b_1(y_{k+1})}{a_{21}b_1(y_{k+1})+(1-a_{21})b_2(y_{k+1})} \\ \frac{(1-a_{21})b_2(y_{k+1})}{a_{21}b_1(y_{k+1})+(1-a_{21})b_2(y_{k+1})} \end{bmatrix} & , \gamma(Z_{k+1}) = 1 \end{cases}$$

which depends on the values Y_{k+1} and Z_{k+1} but which importantly does not depend on values at previous times. Using this fact, an explicit expression for the error probability can then be derived quite easily.

Given $Z_k = 1$, the recursion (2.10) for q is just $q_k = 2a_{21} - 1$. If $a_{21} > 0.5$, then $q_k > 0$, and

$$P(\text{Error}|Z_k = 1) = P(X_k = 2, q_k > 0|Z_k = 1) = P(X_k = 2) = 1 - a_{21}.$$

If $a_{21} \leq 0.5$, then $q_k \leq 0$, and so

$$P(\text{Error}|Z_k = 1) = P(X_k = 1, q_k \leq 0|Z_k = 1) = P(X_k = 1) = a_{21}.$$

This can be written more compactly as

$$P(\text{Error}|Z_k = 1) = \min(a_{21}, 1 - a_{21}).$$

Next, we have

$$\begin{aligned}
&P(\text{Error}|X_k = 1, Z_k = 2) \\
&\quad = P(a_{21}b_1(Y_k) \leq (1 - a_{21})b_2(Y_k)|X_k = 1) \\
&\quad = P\left(a_{21}\exp\left(\frac{-(Y_k + 1)^2}{2\sigma^2}\right) \leq (1 - a_{21})\exp\left(\frac{-(Y_k - 1)^2}{2\sigma^2}\right)|X_k = 1\right) \\
&\quad = P\left(Y_k \geq -\frac{\sigma^2}{2}\ln\left(\frac{1 - a_{21}}{a_{21}}\right)|X_k = 1\right) \\
&\quad = P\left(v_k \geq 1 - \frac{\sigma^2}{2}\ln\left(\frac{1 - a_{21}}{a_{21}}\right)\right) \\
&\quad = P\left(v'_k \leq -\frac{1}{\sigma}\left(1 - \frac{\sigma^2}{2}\ln\left(\frac{1 - a_{21}}{a_{21}}\right)\right)\right)
\end{aligned}$$

where v'_k is $N(0, 1)$. Similarly,

$$P(\text{Error}|X_k = 2, Z_k = 2) = P\left(v'_k \leq -\frac{1}{\sigma}\left(1 + \frac{\sigma^2}{2}\ln\left(\frac{1 - a_{21}}{a_{21}}\right)\right)\right),$$

and so

$$\begin{aligned} P(\text{Error}|Z_k=2) &= P(\text{Error}|X_k=1,Z_k=2)P(X_k=1) \\ &\quad + P(\text{Error}|X_k=2,Z_k=2)P(X_k=2) \\ &= a_{21}P\left(v'_k \le -\frac{1}{\sigma}\left(1-\frac{\sigma^2}{2}\ln\left(\frac{1-a_{21}}{a_{21}}\right)\right)\right) \\ &\quad + (1-a_{21})P\left(v'_k \le -\frac{1}{\sigma}\left(1+\frac{\sigma^2}{2}\ln\left(\frac{1-a_{21}}{a_{21}}\right)\right)\right). \end{aligned} \tag{2.16}$$

The probability of error in the case of i.i.d. data is therefore

$$P(\text{Error}) \;=\; P(\text{Error}|Z_k=1)P(Z_k=1) + P(\text{Error}|Z_k=2)P(Z_k=2)$$

where we use the expressions for $P(\text{Error}|Z_k=1)$ and (2.16), together with the stationary probabilities for the loss process $P(Z_k=1)=\frac{g_{21}}{g_{12}+g_{21}}$ and $P(Z_k=2)=\frac{g_{12}}{g_{12}+g_{21}}$. Since $P(Z_k=2)=1-P(Z_k=1)$, this probability of error is linear in $P(Z_k=1)$, for fixed A and σ. The solid line in Figure 2.10 is a plot of this linear expression, which can be seen to coincide with the simulation results.

2.4.5 High SNR approximation

In this section we will derive an approximate expression for the error in the high SNR, or small σ regime. When σ is small, we can see from (2.9) that for $Y_{k+1}>0$, $q_{k+1}\approx -1$, and for $Y_{k+1}<0$, $q_{k+1}\approx 1$. So at high SNR, the probability updates can be approximated by the simpler sub-optimal scheme

$$\Pi_{k+1|k+1} = \begin{cases} A^T\Pi_{k|k} & , \quad \gamma(Z_{k+1})=0 \\ \begin{bmatrix} 1 & 0 \end{bmatrix}^T & , \quad \gamma(Z_{k+1})=1 \quad , \quad Y_{k+1}\in D_1 \\ \begin{bmatrix} 0 & 1 \end{bmatrix}^T & , \quad \gamma(Z_{k+1})=1 \quad , \quad Y_{k+1}\in D_2. \end{cases}$$

where we have defined the sets $D_1 \equiv (-\infty, 0]$ and $D_2 \equiv (0,\infty)$. To derive the probability of error using such a scheme, first note that

$$P(Y_k\in D_2|X_k=1,Z_k=2) = P\left(v'_k > \frac{1}{\sigma}\right)$$

and

$$P(Y_k \in D_1 | X_k = 2, Z_k = 2) = P\left(v'_k > \frac{1}{\sigma}\right),$$

where v'_k is $N(0,1)$. Using D_m^c to denote the complement of D_m,

$$\begin{aligned} P(\text{Error}, Z_k = 2) &= P(Y_k \in D_1^c, X_k = 1, Z_k = 2) + P(Y_k \in D_2^c, X_k = 2, Z_k = 2) \\ &= P(Y_k \in D_1^c | X_k = 1, Z_k = 2) P(X_k = 1) P(Z_k = 2) \\ &\quad + P(Y_k \in D_2^c | X_k = 2, Z_k = 2) P(X_k = 2) P(Z_k = 2) \\ &= P\left(v'_k > \frac{1}{\sigma}\right) P(Z_k = 2) \end{aligned}$$

where we have again used the independence of $\{X_k\}$ and $\{Z_k\}$.

The derivation of the term $P(\text{Error}, Z_k = 1)$ is similar to the noiseless error probability derived in Section 2.3.1, we will use the same notation and point out the main differences. We can still write

$$P(\text{Error}, Z_k = 1) = \sum_{n=1}^{\infty} p(n)$$

where $p(n) \equiv P(\text{Error}, Z_{k-n} = 2, Z_{k-n+1} = 1, Z_{k-n+2} = 1, \ldots, Z_k = 1) \equiv P(\text{Error}, Z)$, but now

$$\begin{aligned} p(n) &= P(Z, X_k = 1, \arg\max_j \Pi_{1,n}^j \neq 1) + P(Z, X_k = 2, \arg\max_j \Pi_{1,n}^j \neq 2) \\ &\quad + P(Z, X_k = 1, \arg\max_j \Pi_{2,n}^j \neq 1) + P(Z, X_k = 2, \arg\max_j \Pi_{2,n}^j \neq 2) \\ &= P(Z, X_k = 1, \arg\max_j \Pi_{1,n}^j \neq 1, Y_{k-n} \in D_1, X_{k-n} = 1) \\ &\quad + P(Z, X_k = 1, \arg\max_j \Pi_{1,n}^j \neq 1, Y_{k-n} \in D_1, X_{k-n} = 2) \\ &\quad + \cdots + P(Z, X_k = 2, \arg\max_j \Pi_{2,n}^j \neq 2, Y_{k-n} \in D_2, X_{k-n} = 1) \\ &\quad + P(Z, X_k = 2, \arg\max_j \Pi_{2,n}^j \neq 2, Y_{k-n} \in D_2, X_{k-n} = 2) \\ &= \frac{g_{12} g_{21}}{g_{12} + g_{21}} (1 - g_{12})^{n-1} \\ &\quad \times \sum_{r,s,t=1}^{2} P(X_{k-n} = r) a_{rs}^{(n)} I_{\text{argmax}_j \Pi_{t,n}^j \neq s} P(Y_{k-n} \in D_t | X_{k-n} = r, Z_{k-n} = 2) \end{aligned}$$

As in the noiseless case, we can write this in a more explicit form

$$p(n) = \frac{g_{12}}{g_{12}+g_{21}} g_{21}(1-g_{12})^{n-1} \times \begin{cases} \left[\frac{a_{21}}{a_{12}+a_{21}} a_{11}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{21}^{(n)}\right] & , \quad q_n^1 \le 0, q_n^2 \le 0 \\ P\left(v_k' < \frac{1}{\sigma}\right)\left[\frac{a_{21}}{a_{12}+a_{21}} a_{11}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{22}^{(n)}\right] & \\ + P\left(v_k' > \frac{1}{\sigma}\right)\left[\frac{a_{21}}{a_{12}+a_{21}} a_{12}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{21}^{(n)}\right] & , \quad q_n^1 \le 0, q_n^2 > 0 \\ P\left(v_k' < \frac{1}{\sigma}\right)\left[\frac{a_{21}}{a_{12}+a_{21}} a_{12}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{21}^{(n)}\right] & \\ + P\left(v_k' > \frac{1}{\sigma}\right)\left[\frac{a_{21}}{a_{12}+a_{21}} a_{11}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{22}^{(n)}\right] & , \quad q_n^1 > 0, q_n^2 \le 0 \\ \left[\frac{a_{21}}{a_{12}+a_{21}} a_{12}^{(n)} + \frac{a_{12}}{a_{12}+a_{21}} a_{22}^{(n)}\right] & , \quad q_n^1 > 0, q_n^2 > 0. \end{cases} \tag{2.17}$$

The probability of error of this scheme is therefore

$$\begin{aligned} P(\text{Error}) &= P(\text{Error}, Z_k = 1) + P(\text{Error}, Z_k = 2) \\ &= \sum_{n=1}^{\infty} p(n) + P\left(v_k' > \frac{1}{\sigma}\right) \frac{g_{12}}{g_{12}+g_{21}}. \end{aligned}$$

with $p(n)$ given by (2.17) and v_k' begin $N(0,1)$.

In Table 2.3 we compare simulation results of the optimal filter together with the "exact" analytical calculation of Section 2.4.2, and the sub-optimal approximation just derived. We use $a_{12} = 0.3, a_{21} = 0.2, g_{12} = 0.8, g_{21} = 0.5$ and various values of σ. The computation using the method of Section 2.4.2 was done with $N = 1000$. For a further comparison, in the final column we also include simulation results when there is no packet loss. The simulation runs are of length 10 million.

Firstly, we can see that for values of σ smaller than approximately 0.8, the numerical method of Section 2.4.2 does not give accurate results when using $N = 1000$, in fact the accuracy worsens the smaller σ is. Improving the accuracy would involve increasing N substantially, which in turn increases the computation time and memory requirements dramatically. We can also see that the sub-optimal expression gives very good agreement with simulations for small values of σ, moreover it can be computed very easily with current computer algebra software. Indeed, in this example the noiseless probability of error can be computed to be 0.1022, so that

Table 2.3: Comparison of error probabilities for various values of σ

σ	simulation	approximation	Method of Section 2.4.2	no packet loss
1.0	0.2126	0.2279	0.2123	0.1351
0.8	0.1754	0.1859	0.1763	0.0889
0.6	0.1350	0.1400	0.1540	0.0395
0.4	0.1064	0.1071	0.2193	0.0052
0.3	0.1024	0.1025	0.3520	0.00037

even for $\sigma = 0.3$, the difference between the noisy and noiseless error probabilities are almost negligible. Since our approximation (2.17) converges to the noiseless expression (2.5) as $\sigma \to 0$, this is one reason why the approximation performs very well at high SNR.

Comparing the noiseless error probability of 0.1022, the error probabilities with no packet loss and the error probabilities with both packet loss and noise, it appears that for σ smaller than around 0.4-0.5, the packet loss starts to dominate for this example. In general, the σ value where the packet loss term starts to dominate will depend on other parameters such as g_{12} and g_{21}, and is an issue that requires further study. However through our numerical investigations, we have found that σ around 0.4-0.5 seems to be a reasonable figure for most (randomly generated) sets of parameter values.

2.4.6 Multiple states - High SNR

For the noisy case with multiple states (and no packet loss), asymptotic results for the error performance of slow Markov chains exist in the literature e.g. (Khasminskii & Zeitouni, 1996), (Golubev, 2000), but general expressions for arbitrary Markov chains are not known. In this section we will treat arbitrary Markov chains with packet loss at high SNR. We will choose the signal levels $h(X_k)$ to be of M-ary Pulse-amplitude-modulation (PAM) type. Without loss of generality, we let these levels be situated at $1, 3, \ldots, 2M-1$, i.e.

$$h(m) = 2m - 1 \quad , \quad m = 1, 2, \ldots, M.$$

Define the sets

$$D_m = \begin{cases} (-\infty, 2] & , \quad m = 1 \\ (2m-2, 2m] & , \quad m = 2, 3, \ldots, M-1 \\ (2m-2, \infty) & , \quad m = M. \end{cases}$$

The optimal way to update the probabilities is the obvious generalization of (2.7), but which appears to be very difficult to analyze. However, motivated by the high SNR approximation in the 2-state case, consider the following sub-optimal scheme:

$$\Pi_{k+1|k+1} = \begin{cases} A^T \Pi_{k|k} & , \quad \gamma(Z_{k+1}) = 0 \\ \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} \in D_1 \\ \begin{bmatrix} 0 & 1 & \ldots & 0 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} \in D_2 \\ \vdots & & & \vdots \\ \begin{bmatrix} 0 & 0 & \ldots & 1 \end{bmatrix}^T & , \quad \gamma(Z_{k+1}) = 1, Y_{k+1} \in D_M. \end{cases}$$

Similar to the two-state case in Section 2.4.5, we can derive

$$\begin{aligned} P(\text{Error}, Z_k = 2) &= P(Y_k \in D_1^c | X_k = 1, Z_k = 2) P(X_k = 1) P(Z_k = 2) + \ldots \\ &\quad + P(Y_k \in D_M^c | X_k = M, Z_k = 2) P(X_k = M) P(Z_k = 2) \\ &= P\left(v_k' > \frac{1}{\sigma}\right) P(Z_k = 2)[2 - P(X_k = 1) - P(X_k = M)] \end{aligned}$$

and

$$\begin{aligned} p(n) = &\frac{g_{12} g_{21}}{g_{12} + g_{21}} (1 - g_{12})^{n-1} \\ &\times \sum_{r,s,t=1}^{M} P(X_{k-n} = r) a_{rs}^{(n)} I_{\text{argmax}_j \Pi_{t,n}^j \neq s} P(Y_{k-n} \in D_t | X_{k-n} = r, Z_{k-n} = 2). \end{aligned}$$

Therefore

$$\begin{aligned} P(\text{Error}) &= P(\text{Error}, Z_k = 1) + P(\text{Error}, Z_k = 2) \\ &= P\left(v_k' > \frac{1}{\sigma}\right) \frac{g_{12}}{g_{12} + g_{21}} [2 - P(X_k = 1) - P(X_k = M)] + \sum_{n=1}^{\infty} p(n). \end{aligned}$$

Table 2.4: Comparison of noisy error probabilities for various values of σ, 3-state example

σ	simulation	approximation
1.0	0.2905	0.3276
0.8	0.2444	0.2691
0.6	0.1948	0.2054
0.4	0.1580	0.1595
0.3	0.1524	0.1531

with v_k' being $N(0,1)$. As in the noiseless multi-state case, more explicit expressions would be very complicated to write down in general.

In Table 2.4 we compare this approximation with simulation results of the optimal filter, using the 3-state example of Section 2.3.5 with $g_{12} = 0.48$, $g_{21} = 0.26$ and

$$A = \begin{bmatrix} 0.18 & 0.76 & 0.06 \\ 0.56 & 0.36 & 0.08 \\ 0.13 & 0.27 & 0.60 \end{bmatrix}$$

together with an additional noise term. We can again see that the sub-optimal expression gives good agreement with simulations for small values of σ.

2.5 Conclusion

In this chapter we have derived analytical expressions for the error probability of HMM filters in the presence of Markovian packet losses, with emphasis on two-state Markov chains. A number of relationships between the error probabilities and the parameters of the loss process have been shown via numerical studies, and theoretical justification has also been obtained in some cases. Chapter 3 will turn to the problem of detecting whether a signal modelled as a two-state Markov chain is actually present. While the two problems (in this chapter and Chapter 3) appear to be quite different, it turns out that some of the techniques used in this chapter can also be applied there.

2.6 Appendix

For convenience in our proofs, let us define

$$s(n) \equiv \frac{g_{12}g_{21}}{g_{12}+g_{21}}(1-g_{12})^{n-1}$$

$$t(n) \equiv \begin{cases} \frac{a_{21}(a_{12}+a_{21})}{(a_{12}+a_{21})^2} & , \quad q_n^1 \leq 0, q_n^2 \leq 0 \\ \frac{a_{21}^2+2a_{21}a_{12}(1-a_{21}-a_{12})^n+a_{12}^2}{(a_{12}+a_{21})^2} & , \quad q_n^1 \leq 0, q_n^2 > 0 \\ \frac{2a_{21}a_{12}-2a_{21}a_{12}(1-a_{21}-a_{12})^n}{(a_{12}+a_{21})^2} & , \quad q_n^1 > 0, q_n^2 \leq 0 \\ \frac{a_{12}(a_{12}+a_{21})}{(a_{12}+a_{21})^2} & , \quad q_n^1 > 0, q_n^2 > 0 \end{cases} \tag{2.18}$$

so that $p(n) = s(n)t(n)$.

2.6.1 Proof of Theorem 2.3.1

Proof.

$$\frac{dP(\text{Error})}{dg_{21}} = \sum_{n=1}^{\infty} \frac{ds(n)}{dg_{21}} t(n) = \sum_{n=1}^{\infty} \frac{g_{12}^2(1-g_{12})^{n-1}}{(g_{12}+g_{21})^2} t(n).$$

Term-by-term differentiation of the infinite series can be justified by using the Weierstrass M-test, see e.g. (Rudin, 1976). It is easy to see from (2.18) or (2.5) that the $t(n)$ terms are all positive. As each term in the sum is greater than zero, $\frac{dP(\text{Error})}{dg_{21}} > 0$. □

2.6.2 Proof of Theorem 2.3.2

While the statement of Theorem 2.3.2 looks similar to that of Theorem 2.3.1, the proof is not as straightforward. Before we prove this, we need the following technical result.

Lemma 2.6.1. *The terms $t(n)$ given by (2.18) form a non-decreasing sequence.*

Proof. We will prove this lemma assuming throughout that $a_{21} \leq a_{12}$. The arguments when $a_{21} > a_{12}$ are almost identical and will be omitted.

The proof will be divided up into the situations $a_{21}+a_{12}<1$, $a_{21}+a_{12}>1$ and $a_{21}+a_{12}=1$.

We first consider the situation $a_{21}+a_{12}<1$. We need to look at the following four cases.

1. $q_n^1 \le 0$ and $q_n^2 \le 0$

By definition (2.4), this implies $2a_{12}(1-a_{21}-a_{12})^n \le a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n \le a_{12}-a_{21}$. Then

$$a_{21}-a_{12}+2a_{12}(1-a_{21}-a_{12})^{n+1} \le a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12}) \le 0$$
$$a_{21}-a_{12}-2a_{21}(1-a_{21}-a_{12})^{n+1} \le a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12}) \le 0$$

so that $q_{n+1}^1 \le 0$ and $q_{n+1}^2 \le 0$, and hence $t(n+1)=t(n)=\frac{a_{21}}{a_{21}+a_{12}}$ from (2.18).

2. $q_n^1 \le 0$ and $q_n^2 > 0$

This would imply $2a_{12}(1-a_{21}-a_{12})^n \le a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$. However the second assertion is a contradiction, so this case cannot occur.

3. $q_n^1 > 0$ and $q_n^2 \le 0$

This would imply $2a_{12}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n \le a_{12}-a_{21}$. Then

$$a_{21}-a_{12}+2a_{12}(1-a_{21}-a_{12})^{n+1} > a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12})$$
$$a_{21}-a_{12}-2a_{21}(1-a_{21}-a_{12})^{n+1} \le a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12}) \le 0.$$

So either $q_{n+1}^1 > 0$ and $q_{n+1}^2 \le 0$, or $q_{n+1}^1 \le 0$ and $q_{n+1}^2 \le 0$. If $q_{n+1}^1 > 0$ and $q_{n+1}^2 \le 0$, then

$$\begin{aligned} t(n+1) &= \frac{2a_{21}a_{12}-2a_{21}a_{12}(1-a_{21}-a_{12})^{n+1}}{(a_{21}+a_{12})^2} \\ &> \frac{2a_{21}a_{12}-2a_{21}a_{12}(1-a_{21}-a_{12})^{n}}{(a_{21}+a_{12})^2} = t(n) \end{aligned}$$

since $(1-a_{21}-a_{12})^{n+1} < (1-a_{21}-a_{12})^n$.

If $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 \leq 0$, then

$$\begin{aligned}(a_{21}+a_{12})^2[t(n+1)-t(n)] &= a_{21}^2 + a_{21}a_{12} - 2a_{21}a_{12} + 2a_{21}a_{12}(1-a_{21}-a_{12})^n \\ &= a_{21}^2 - a_{21}a_{12} + 2a_{21}a_{12}(1-a_{21}-a_{12})^n \\ &> a_{21}^2 - a_{21}a_{12} + a_{21}(a_{12}-a_{21}) = 0.\end{aligned}$$

4. $q_n^1 > 0$ and $q_n^2 > 0$

This would imply $2a_{12}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$. The second assertion is a contradiction. We thus see that for $a_{21}+a_{12} < 1$, $t(n)$ is non-decreasing with n in each case.

Now consider the situation $a_{21}+a_{12} > 1$. We have the following four cases:

1. $q_n^1 \leq 0$ and $q_n^2 \leq 0$

This implies $2a_{12}(1-a_{21}-a_{12})^n \leq a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n \leq a_{12}-a_{21}$. We have $2a_{12}(1-a_{21}-a_{12})^{n+1} \geq (a_{12}-a_{21})(1-a_{21}-a_{12})$ and $-2a_{21}(1-a_{21}-a_{12})^{n+1} \geq (a_{12}-a_{21})(1-a_{21}-a_{12})$, or

$$-2a_{21}(1-a_{21}-a_{12})^{n+1} \leq -\frac{a_{21}}{a_{12}}(a_{12}-a_{21})(1-a_{21}-a_{12})$$

and

$$2a_{12}(1-a_{21}-a_{12})^{n+1} \leq -\frac{a_{12}}{a_{21}}(a_{12}-a_{21})(1-a_{21}-a_{12}).$$

Then

$$\begin{aligned}a_{21}-a_{12}+2a_{12}(1-a_{21}-a_{12})^{n+1} &\leq a_{21}-a_{12}+\frac{a_{12}}{a_{21}}(a_{21}-a_{12})(1-a_{21}-a_{12}) \\ &= (a_{21}-a_{12})\frac{(a_{21}+a_{12})(1-a_{12})}{a_{21}} \leq 0\end{aligned}$$

and

$$\begin{aligned}a_{21}-a_{12}-2a_{21}(1-a_{21}-a_{12})^{n+1} &\leq a_{21}-a_{12}+\frac{a_{21}}{a_{12}}(a_{21}-a_{12})(1-a_{21}-a_{12}) \\ &= (a_{21}-a_{12})\frac{(a_{21}+a_{12})(1-a_{21})}{a_{12}} \leq 0,\end{aligned}$$

so that $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 \leq 0$, and hence $t(n+1) = t(n)$.

2. $q_n^1 \leq 0$ and $q_n^2 > 0$

This implies $2a_{12}(1-a_{21}-a_{12})^n \leq a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$, which in turns implies that $(1-a_{21}-a_{12})^n < 0$. Then

$$a_{21}-a_{12}+2a_{12}(1-a_{21}-a_{12})^{n+1} \geq a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12})$$
$$a_{21}-a_{12}-2a_{21}(1-a_{21}-a_{12})^{n+1} < a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12}) \leq 0.$$

So either $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 \leq 0$, or $q_{n+1}^1 > 0$ and $q_{n+1}^2 \leq 0$. If $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 \leq 0$,

$$\begin{aligned}(a_{21}+a_{12})^2[t(n+1)-t(n)] &= a_{21}^2+a_{21}a_{12}-a_{21}^2-2a_{21}a_{12}(1-a_{21}-a_{12})^n-a_{12}^2\\ &= a_{21}a_{12}-2a_{21}a_{12}(1-a_{21}-a_{12})^n-a_{12}^2\\ &> a_{21}a_{12}+a_{12}(a_{12}-a_{21})-a_{12}^2=0.\end{aligned}$$

If $q_{n+1}^1 > 0$ and $q_{n+1}^2 \leq 0$,

$$\begin{aligned}&(a_{21}+a_{12})^2[t(n+1)-t(n)]\\ &\quad= 2a_{21}a_{12}-2a_{21}a_{12}(1-a_{21}-a_{12})^{n+1}-a_{21}^2-2a_{21}a_{12}(1-a_{21}-a_{12})^n-a_{12}^2\\ &\quad= -2a_{21}a_{12}(1-a_{21}-a_{12})^n(2-a_{21}-a_{12})-(a_{21}-a_{12})^2\\ &\quad> a_{12}(a_{12}-a_{21})(2-a_{21}-a_{12})-(a_{21}-a_{12})^2\\ &\quad= (1-a_{12})(a_{12}^2-a_{21}^2) \geq 0.\end{aligned}$$

3. $q_n^1 > 0$ and $q_n^2 \leq 0$

This implies $2a_{12}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n \leq a_{12}-a_{21}$, which in turns implies that $(1-a_{21}-a_{12})^n > 0$. Then

$$a_{21}-a_{12}+2a_{12}(1-a_{21}-a_{12})^{n+1} < a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12}) \leq 0$$
$$a_{21}-a_{12}-2a_{21}(1-a_{21}-a_{12})^{n+1} \geq a_{21}-a_{12}+(a_{12}-a_{21})(1-a_{21}-a_{12}).$$

So either $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 \leq 0$, or $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 > 0$. If $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 \leq 0$,

$$\begin{aligned}(a_{21}+a_{12})^2[t(n+1)-t(n)] &= a_{21}^2 + a_{21}a_{12} - 2a_{21}a_{12} + 2a_{21}a_{12}(1-a_{21}-a_{12})^n \\ &= a_{21}^2 - a_{21}a_{12} + 2a_{21}a_{12}(1-a_{21}-a_{12})^n \\ &> a_{21}^2 - a_{21}a_{12} + a_{21}(a_{12}-a_{21}) = 0.\end{aligned}$$

If $q_{n+1}^1 \leq 0$ and $q_{n+1}^2 > 0$,

$$\begin{aligned}&(a_{21}+a_{12})^2[t(n+1)-t(n)] \\ &\quad = a_{21}^2 + 2a_{21}a_{12}(1-a_{21}-a_{12})^{n+1} + a_{12}^2 - 2a_{21}a_{12} + 2a_{21}a_{12}(1-a_{21}-a_{12})^n \\ &\quad = (a_{21}-a_{12})^2 + 2a_{21}a_{12}[(1-a_{21}-a_{12})^{n+1} + (1-a_{21}-a_{12})^n] > 0\end{aligned}$$

since $(1-a_{21}-a_{12})^n > 0$ and $|(1-a_{21}-a_{12})^n| > |(1-a_{21}-a_{12})^{n+1}|$.

4. $q_n^1 > 0$ and $q_n^2 > 0$

This implies $2a_{12}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$ and $-2a_{21}(1-a_{21}-a_{12})^n > a_{12}-a_{21}$. This is not possible since of the terms $2a_{12}(1-a_{21}-a_{12})^n, -2a_{21}(1-a_{21}-a_{12})^n$, one must be positive and one negative, but $a_{12}-a_{21} \geq 0$. Hence for $a_{21}+a_{12} > 1$, $t(n)$ is also non-decreasing with n in all cases.

The final situation with $a_{21}+a_{12} = 1$ is straightforward, only the case $q_n^1 \leq 0$ and $q_n^2 \leq 0$ can occur, and

$$t(n) = \frac{a_{21}}{a_{21}+a_{12}}, \forall n.$$

□

Now we can prove Theorem 2.3.2.

Proof. First,

$$\begin{aligned}\frac{dP(\text{Error})}{dg_{12}} &= \sum_{n=1}^{\infty} \frac{ds(n)}{dg_{12}} t(n) \\ &= \sum_{n=1}^{\infty} -\frac{g_{21}(1-g_{12})^{n-2}[g_{12}^2(n-1) + g_{21}(ng_{12}-1)]}{(g_{12}+g_{21})^2} t(n).\end{aligned} \tag{2.19}$$

Showing that this quantity is negative is equivalent to showing that

$$\sum_{n=1}^{\infty} -(1-g_{12})^{n-2}[g_{12}^2(n-1)+g_{21}(ng_{12}-1)]t(n) < 0.$$

Unlike the proof of Theorem 2.3.1, not every term in the summation here is negative. However, it is not difficult to show that for $m \equiv \left\lfloor \frac{g_{12}^2+g_{21}}{g_{12}^2+g_{21}g_{12}} \right\rfloor$, the first m terms will be positive, while the rest will be negative. We may then use Lemma 2.6.1 to obtain the bound

$$\sum_{n=1}^{\infty} -(1-g_{12})^{n-2}[g_{12}^2(n-1)+g_{21}(ng_{12}-1)]t(n) \\ \leq t(m)\sum_{n=1}^{\infty} -(1-g_{12})^{n-2}[g_{12}^2(n-1)+g_{21}(ng_{12}-1)].$$

To complete the proof, we note the following closed form expression, which can be verified using induction:

$$\sum_{n=1}^{k} -(1-g_{12})^{n-2}[g_{12}^2(n-1)+g_{21}(ng_{12}-1)] = -1+(1-g_{12})^{k-1}[1+g_{12}(k-1)+g_{21}k].$$

We note that this expression also allows us to apply the Weierstrass M-test to justify the term-by-term differentiation (2.19). Hence

$$\sum_{n=1}^{\infty} -(1-g_{12})^{n-2}[g_{12}^2(n-1)+g_{21}(ng_{12}-1)] = -1$$

and therefore

$$\sum_{n=1}^{\infty} -(1-g_{12})^{n-2}[g_{12}^2(n-1)+g_{21}(ng_{12}-1)]t(n) \leq -t(m) < 0.$$

□

2.6.3 Proof of Theorem 2.3.3

Proof. That E_0 actually is the error probability in the complete absence of observations is not difficult to show. For example, one can use the fact that A^n converges to a rank 1 matrix (with the stationary probabilities in the rows) as $n \to \infty$, so that without observations, one would choose the state estimate which on average is more likely to occur.

(i) The proof of this part is similar to that of Theorem 2.3.2.

Since $P(\text{Error}) = \sum_{n=1}^{\infty} p_0 g_{12}(1-g_{12})^{n-1} t(n)$, we have

$$\frac{dP(\text{Error})}{dg_{12}} = \sum_{n=1}^{\infty} -p_0(1-g_{12})^{n-2}(ng_{21}-1)t(n).$$

It can be easily seen that for $m \equiv \left\lfloor \frac{1}{g_{12}} \right\rfloor$, the first m terms in the series will be positive, while the rest will be negative. Using Lemma 2.6.1, we obtain the bound

$$\sum_{n=1}^{\infty} -p_0(1-g_{12})^{n-2}(ng_{21}-1)t(n) \leq t(m) \sum_{n=1}^{\infty} -p_0(1-g_{12})^{n-2}(ng_{21}-1).$$

We have the following closed form expression:

$$\sum_{n=1}^{k} -p_0(1-g_{12})^{n-2}(ng_{21}-1) = (1-g_{12})^{k-1} p_0,$$

so

$$\sum_{n=1}^{\infty} -p_0(1-g_{12})^{n-2}(ng_{21}-1) = 0,$$

and therefore

$$\sum_{n=1}^{\infty} -p_0(1-g_{12})^{n-2}(ng_{21}-1)t(n) \leq 0.$$

(ii) We consider the cases $a_{21} \neq a_{12}$ and $a_{21} = a_{12}$ separately. First assume that $a_{21} \neq a_{12}$. From (2.4) it can be seen that there exists an N such that either $q_n^1 > 0, q_n^2 > 0, \forall n > N$ (when $a_{21} > a_{12}$), or $q_n^1 < 0, q_n^2 < 0, \forall n > N$ (when $a_{21} < a_{12}$).

Hence $t(n) = E_0, \forall n > N$ and

$$P(\text{Error}) = \sum_{n=1}^{N} p_0 g_{12}(1-g_{12})^{n-1} t(n) + E_0 \sum_{n=N+1}^{\infty} p_0 g_{12}(1-g_{12})^{n-1}.$$

Applying Lemma 2.6.1, we can obtain the bounds

$$t(1)\sum_{n=1}^{N} p_0 g_{12}(1-g_{12})^{n-1} + E_0 \sum_{n=N+1}^{\infty} p_0 g_{12}(1-g_{12})^{n-1}$$
$$\leq P(\text{Error}) \leq E_0 \sum_{n=1}^{\infty} p_0 g_{12}(1-g_{12})^{n-1}$$

or

$$t(1)[p_0 - (1-g_{12})^N p_0] + E_0(1-g_{12})^N p_0 \leq P(\text{Error}) \leq E_0 p_0.$$

Taking the limit as $g_{12} \to 0$ then gives the result for $a_{21} \neq a_{12}$.

Now assume that $a_{21} = a_{12}$. We further divide into 3 cases.

1) For $a_{21} = 0.5$, we have $q_n^1 \leq 0, q_n^2 \leq 0, \forall n$, and

$$\begin{aligned} P(\text{Error}) &= p_0 \times \frac{1}{2}\sum_{n=1}^{\infty} g_{12}(1-g_{12})^{n-1} \\ &= \frac{p_0}{2} = p_0 E_0 \end{aligned}$$

irrespective of g_{12}.

2) For $a_{21} < 0.5$, we have $q_n^1 > 0, q_n^2 < 0, \forall n$, so

$$\begin{aligned} P(\text{Error}) &= p_0 \sum_{n=1}^{\infty} g_{12}(1-g_{12})^{n-1} \times \frac{1}{2}[1-(1-2a_{21})^n] \\ &= \frac{a_{21}p_0}{2a_{21} - 2a_{21}g_{12} + g_{12}} \end{aligned}$$

which converges to $\frac{p_0}{2}$ as $g_{12} \to 0$.

3) For $a_{21} > 0.5$, we have $q_n^1 < 0, q_n^2 > 0$ for n odd, and $q_n^1 > 0, q_n^2 < 0$ for n even, so

$$\begin{aligned} P(\text{Error}) &= p_0 \sum_{m=1}^{\infty} g_{12}(1-g_{12})^{2m-2} \times \frac{1}{2}[1+(1-2a_{21})^{2m-1}] \\ &\quad + p_0 \sum_{m=1}^{\infty} g_{12}(1-g_{12})^{2m-1} \times \frac{1}{2}[1-(1-2a_{21})^{2m}] \\ &= p_0 \sum_{n=1}^{\infty} g_{12}(1-g_{12})^{n-1} \times \frac{1}{2}[1-(2a_{21}-1)^n] \\ &= \frac{(1-a_{21})p_0}{2(1-a_{21})-2(1-a_{21})g_{12}+g_{12}} \end{aligned}$$

which also converges to $\frac{p_0}{2}$ as $g_{12} \to 0$. □

Chapter 3

Neyman-Pearson Detection of Markov Chains in Noise

3.1 Introduction

The detection of random signals in noise is an important problem in engineering and signal processing. In general, performance analysis of detectors via closed form expressions for the probability of error or related quantities is intractible except for very simple test statistics. One alternative is to study the rate of decrease of the probability of error as the number of samples increases, or in the Neyman-Pearson formulation the probability of missed detection with a fixed false alarm probability constraint.

Let P_M represent the probability of missed detection and P_{FA} the probability of false alarm. The error exponent for Neyman-Pearson detection is defined as the limit (log in this chapter will refer to the natural logarithm unless stated otherwise)

$$K = \lim_{n\to\infty} -\frac{1}{n} \log P_M, \tag{3.1}$$

and represents the rate at which P_M decays for a fixed constraint on P_{FA} as the number of samples $n \to \infty$. Another interpretation of the error exponent is that P_M becomes proportional to $\exp(-nK)$ as $n \to \infty$. The error exponent K for the Neyman-Pearson detection problem can be shown (see (Sung *et al.*, 2006) and the references therein) to be the following almost sure limit (provided it exists) under

H_0[1]

$$K = \lim_{n\to\infty} \frac{1}{n} \log \frac{p_{0,n}}{p_{1,n}}(y^n) = \lim_{n\to\infty} -\frac{1}{n} \log L_n(y^n) \tag{3.2}$$

where $p_{0,n}$ and $p_{1,n}$ are the null and alternative joint densities respectively of the measurements

$$y^n \equiv (y_1, \ldots, y_n),$$

and

$$L_n(y^n) = \frac{p_{1,n}}{p_{0,n}}(y^n)$$

is the likelihood ratio.[2] Unfortunately closed form expressions for the error exponent are often still not possible except in rare cases. The subject of this chapter is the numerical computation of the error exponent for Neyman-Pearson detection of a two-state Markov chain in noise. Such a detection problem can arise when the signal is modelled as binary data.

Some previous work, e.g. (Benitz & Bucklew, 1990; Bahr & Bucklew, 1990; Chamberland & Veeravalli, 2003; Mergen *et al.*, 2007), have characterised the large sample behaviour of detectors using results from large deviations theory such as the Gärtner-Ellis theorem to determine the rate functions, though evaluating these numerically usually required further optimization techniques. Moreover, often the minimum probability of error criterion or detection with fixed thresholds is used instead of the Neyman-Pearson criterion, so that $\lim_{n\to\infty} -\frac{1}{n} \log P_e$ is analysed instead of (3.1). Neyman-Pearson detection of Gauss-Markov processes is however considered in (Sung *et al.*, 2006). By studying (3.2) and using properties of the innovations process and the state space structure, the authors managed to obtain a closed form expression for the error exponent, and derived relationships between the error exponent and parameters of the system.

On the other hand, in detection problems involving Markov chains, the problem

[1] In the i.i.d. case this is Stein's lemma, and (3.2) is equivalent to computing the Kullback-Leibler distance between the two densities

[2] One could also define error exponents for Bayesian detection. Here we would have two exponents representing $\lim_{n\to\infty} -\frac{1}{n} \log P_M$ and $\lim_{n\to\infty} -\frac{1}{n} \log P_{FA}$ which could be possibly different. However relationships such as (3.2) no longer hold, and one must resort to large deviations techniques to analyse these exponents, such as in (Mergen *et al.*, 2007).

of trying to decide between two Markov chains with different transition parameters (but no noise) is treated in standard textbooks such as (Dembo & Zeitouni, 1998). Likelihood ratios for detecting Markov chains in noise are derived in (Scharf & Nolte, 1977). Classification of HMMs with empirically observed statistics is studied in (Merhav, 1991). Error exponents in HMM order estimation are considered in e.g. (Liu & Narayan, 1994; Gassiat & Boucheron, 2003). However, the problem of determining the error exponent associated with detecting a Markov chain in noise and how this error exponent behaves as system parameters vary, has not been previously treated in the literature.

Moreover, apart from (Mergen *et al.*, 2007) the impact of fading has not been addressed in these previous works, though likelihood ratios for various decentralized detection schemes over fading channels have been derived recently, see e.g. (Chen *et al.*, 2004; Niu *et al.*, 2006). Intuitively fading reduces detection performance, and it is of interest to know what the error exponents are in fading environments, for example in wireless sensor networks where there are limited resources and delay constraints such that waiting to collect too many data samples for a hypothesis test is not desirable.

In this chapter we will adopt a similar philosophy to (Sung *et al.*, 2006) in studying the error exponent via (3.2) and properties of the likelihood ratio. While a closed form expression cannot be obtained, numerical methods for computing the error exponent when the Markov chain has two states will be presented; this situation can occur when binary data (e.g. the presence or absence of a target) is being detected over a communication channel with noise. We first treat the case where the channel is time-invariant and known in Section 3.3, and then extend our methods to channels which are randomly time-varying (e.g. a wireless link) in Section 3.4, for a Rayleigh fading channel with no instantaneous channel state information. We will numerically study the behaviour of the error exponent as the transition parameters of the Markov chain are varied, and also as the noise variance is varied. In particular, for Gaussian noise we shall see that at high SNR there is a marked contrast between the time-invariant and fading scenarios, with the scaling of the error exponent with SNR being linear and logarithmic respectively.

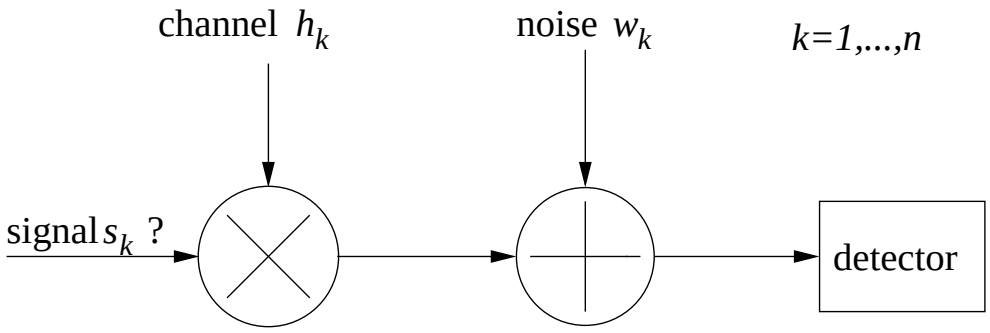

Figure 3.1: System model

3.2 System model

Let the null and alternative hypotheses be

$$\begin{aligned} H_0 &: y_k = w_k \\ H_1 &: y_k = s_k + w_k \end{aligned} \tag{3.3}$$

for $k = 1, 2, \ldots, n$, see Figure 3.1. The noise process $\{w_k\}$ is assumed to be i.i.d. and $N(0, \sigma^2)$.[3] The process $\{s_k\}$ which we wish to detect is a homogeneous two-state Markov chain, with state space $\{\psi_1, \psi_2\}$, and $\psi_2 = -\psi_1$. We use the convention that

$$a_{ij} = P(s_{k+1} = \psi_j | s_k = \psi_i)$$

for the transition probabilities, with the assumption that $0 < a_{ij} < 1, \forall i, j$. In Section 3.3 the process $\{h_k\}$ will be taken to be constant for all k, while in Section 3.4 $\{h_k\}$ will be a time-varying fading process.

3.3 Error exponent computation for time-invariant channels

We first consider time-invariant channels, where without loss of generality we take $h_k = 1, \forall k$ in (3.3). We will calculate the error exponent using the form (3.2), where

[3]The method presented here does not necessarily require the noise to be Gaussian (Shue *et al.*, 2001), though some of the expressions will change for different noise distributions

$L_n(y^n) = \frac{p_{1,n}}{p_{0,n}}(y^n)$ is the likelihood ratio. The likelihood ratio for the problem of detecting Markov chains in noise (3.3) has been previously derived in e.g. (Scharf & Nolte, 1977). One can write

$$L_n(y^n) = \prod_{k=1}^{n} l(y_k|y^{k-1})$$

where

$$\begin{aligned} l(y_k|y^{k-1}) &= \frac{p_1(y_k|s_k=\psi_j, y^{k-1})}{p_0(y_k|y^{k-1})} \frac{P(s_k=\psi_j|y^{k-1})}{P(s_k=\psi_j|y^k)} \\ &= \exp\left(\frac{\psi_j y_k}{\sigma^2} - \frac{1}{2}\frac{\psi_j^2}{\sigma^2}\right) \frac{P(s_k=\psi_j|y^{k-1})}{P(s_k=\psi_j|y^k)} \end{aligned}$$

for any $j \in \{1,2\}$. For definiteness we choose $j=1$. We have the recursive relations

$$P(s_k=\psi_j|y^{k-1}) = \sum_{i=1}^{2} a_{ij} P(s_{k-1}=\psi_i|y^{k-1}) \tag{3.4}$$

$$P(s_k=\psi_j|y^k) = \frac{p_1(y_k|s_k=\psi_j)P(s_k=\psi_j|y^{k-1})}{\sum_{i=1}^{2} p_1(y_k|s_k=\psi_i)P(s_k=\psi_i|y^{k-1})} \tag{3.5}$$

where $p_1(y_k|s_k=\psi_j)$ is the density of y_k under the alternative hypothesis, i.e.

$$p_1(y_k|s_k=\psi_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_k-\psi_j)^2}{2\sigma^2}\right).$$

Thus the error exponent can be written as

$$\begin{aligned} K &= \lim_{n\to\infty} -\frac{1}{n}\log\prod_{k=1}^{n} l(y_k|y^{k-1}) = \lim_{n\to\infty} -\frac{1}{n}\sum_{k=1}^{n}\log l(y_k|y^{k-1}) \\ &= \lim_{n\to\infty} -\frac{1}{n}\sum_{k=1}^{n}\left[\frac{\psi_1 y_k}{\sigma^2} - \frac{1}{2}\frac{\psi_1^2}{\sigma^2} + \log P(s_k=\psi_1|y^{k-1}) - \log P(s_k=\psi_1|y^k)\right] \end{aligned}$$

provided that the limit exists. Now under H_0, $\{y_k\}$ is i.i.d. zero mean Gaussian. Thus

$$-\frac{1}{n}\sum_{k=1}^{n}\frac{\psi_1 y_k}{\sigma^2} \to 0 \text{ a.s.}$$

by the strong law of large numbers. Since $0 < a_{ij} < 1, \forall i, j$ by assumption, the transition matrix for $\{s_k\}$ will be primitive. From (Le Gland & Mevel, 2000) it then follows that $P(s_k = \psi_1|y^{k-1})$ and $P(s_k = \psi_1|y^k)$ have invariant limiting distributions under H_0 as $k \to \infty$, by taking $h_\bullet$ as the zero mapping in Example 1.1 of (Le Gland & Mevel, 2000) and considering our situation as a case of a misspecified hidden Markov model (HMM). Hence by the arguments above and the ergodic theorem:

$$K = \frac{1}{2}\frac{\psi_1^2}{\sigma^2} - \mathbb{E}[\log P(s_k = \psi_1|y^{k-1}) - \log P(s_k = \psi_1|y^k)] \text{ a.s.} \tag{3.6}$$

We also note that an alternative way of showing the existence of the almost sure limit is by using Proposition 3.2 of (Le Gland & Mevel, 1997).

To calculate the above, it suffices to find the limiting distribution of $P(s_k = \psi_1|y^k)$, since $P(s_k = \psi_1|y^{k-1})$ is related by (3.4). We apply a method from (Shue *et al.*, 2001) (also Chapter 2), which was originally used for the problem of computing the probability of error in HMM filtering for two-state Markov chains.[4] Let $q_k \equiv P(s_k = \psi_1|y^k)$ and $f_k(q)dq \equiv P(q_k \in (q, q + dq))$. Then

$$\begin{aligned} &P(q_k \in (q, q+dq), q_{k-1} \in (\tilde{q}, \tilde{q}+d\tilde{q})) \\ &\qquad = P(q_k \in (q, q+dq)|q_{k-1} = \tilde{q})P(q_{k-1} \in (\tilde{q}, \tilde{q}+d\tilde{q})) \\ &\qquad = P(q_k \in (q, q+dq)|q_{k-1} = \tilde{q})f_{k-1}(\tilde{q})d\tilde{q}. \end{aligned} \tag{3.7}$$

Using (3.5) and recalling that $\psi_2 = -\psi_1$, we have

$$\begin{aligned} q_k &= \frac{\exp(\frac{-(y_k-\psi_1)^2}{2\sigma^2})[a_{11}q_{k-1} + a_{21}(1-q_{k-1})]}{\exp(\frac{-(y_k-\psi_1)^2}{2\sigma^2})[a_{11}q_{k-1} + a_{21}(1-q_{k-1})] + \exp(\frac{-(y_k+\psi_1)^2}{2\sigma^2})[a_{12}q_{k-1} + a_{22}(1-q_{k-1})]} \\ &= \frac{a_{11}q_{k-1} + a_{21}(1-q_{k-1})}{a_{11}q_{k-1} + a_{21}(1-q_{k-1}) + \exp(\frac{-2\psi_1 y_k}{\sigma^2})[a_{12}q_{k-1} + a_{22}(1-q_{k-1})]}. \end{aligned}$$

Thus

$$y_k = -\frac{\sigma^2}{2\psi_1}\ln\frac{[a_{11}q_{k-1} + a_{21}(1-q_{k-1})](1-q_k)}{[a_{12}q_{k-1} + a_{22}(1-q_{k-1})]q_k} \equiv g(q_k, q_{k-1})$$

[4] As was the case in Chapter 2, this method unfortunately doesn't appear to be extendable to Markov chains with more than two states.

and

$$\left|\frac{dy_k}{dq_k}\right| = \left|\frac{\sigma^2}{2\psi_1 q_k(1-q_k)}\right| = \frac{\sigma^2}{2|\psi_1| q_k(1-q_k)}.$$

Then by a change of variable, again noting that y_k is $N(0,\sigma^2)$ under H_0, we obtain

$$\begin{aligned} P(q_k \in (q, q+dq)|q_{k-1} = \tilde{q}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{g(q,\tilde{q})^2}{2\sigma^2}\right) \frac{\sigma^2}{2|\psi_1|q(1-q)} dq \\ &\equiv S(q,\tilde{q})dq. \end{aligned}$$

Hence from (3.7) we can see that in the limit as $k \to \infty$, $f_k(q)$ satisfies the following integral equation:

$$f(q)dq = \int_0^1 S(q,\tilde{q}) f(\tilde{q}) d\tilde{q} dq. \tag{3.8}$$

This integral equation may be solved numerically using a procedure described in (Leong *et al.*, 2007) and Chapter 2 (with slight modifications). Consider the eigenvalue problem

$$\lambda f(q) = \int_0^1 S(q,\tilde{q}) f(\tilde{q}) d\tilde{q},$$

which corresponds to (3.8) when $\lambda = 1$. We can solve this eigenvalue problem using the Nyström method (Atkinson, 1997). Replacing the integral by an N-point quadrature rule, and defining

$$\mathbf{K} \equiv \begin{bmatrix} w_1 S(t_1,t_1) & \dots & w_N S(t_1,t_N) \\ \vdots & \ddots & \vdots \\ w_1 S(t_N,t_1) & \dots & w_N S(t_N,t_N) \end{bmatrix}$$

we obtain

$$\mathbf{K} \begin{bmatrix} f(t_1) \\ \vdots \\ f(t_N) \end{bmatrix} = \lambda \begin{bmatrix} f(t_1) \\ \vdots \\ f(t_N) \end{bmatrix}$$

where w_j represent the weights and t_j the quadrature points of the quadrature rule. The Gauss-Legendre quadrature rule is commonly used for the Nyström method (Atkinson, 1997, p.110). In this chapter we will use the Gauss-Legendre rule in a composite scheme (Ralston & Rabinowitz, 1978, p.113); this allows a simple way to vary the precision by changing the number of sub-intervals. In the following sections

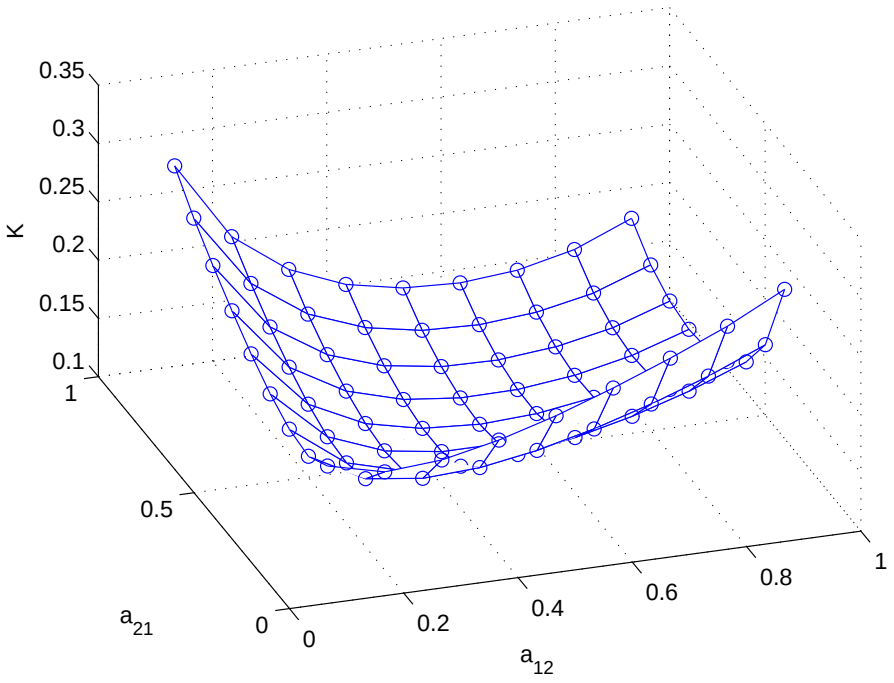

Figure 3.2: Error exponent K for various a_{12} and a_{21}

we use 200 sub-intervals, each sub-interval using a 4-th order Gauss-Legendre rule. Then in order to obtain an approximation for $f(q)$, we take the eigenvector that corresponds to the largest real eigenvalue of $\mathbf{K}$, and normalise so that $\int_0^1 f(q)dq = 1$ is satisfied. We may then calculate $\mathbb{E}[\log P(s_k = \psi_1|y^k)]$ and hence the error exponent via (3.6). Some numerical properties of this method are described in (Leong *et al.*, 2007) and Chapter 2.

3.3.1 Numerical studies

First we study how the error exponent varies when the transition parameters vary, while all other parameters are fixed. Figure 3.2 shows a 3D plot of the error exponent for different values of a_{12} and a_{21}. We fix $\psi_1 = -1, \psi_2 = 1, \sigma^2 = 1$. In Figure 3.3 we show plots for four different fixed values of a_{21}, and a_{12} is varied.[5] For a_{12} small, the error exponent K is large. As a_{12} increases, K decreases until a certain point at which it starts to increase again. An intuitive explanation as to why K is large

[5]By the symmetry of the situation, we will get the same behaviour when a_{21} is varied and the other parameters are fixed.

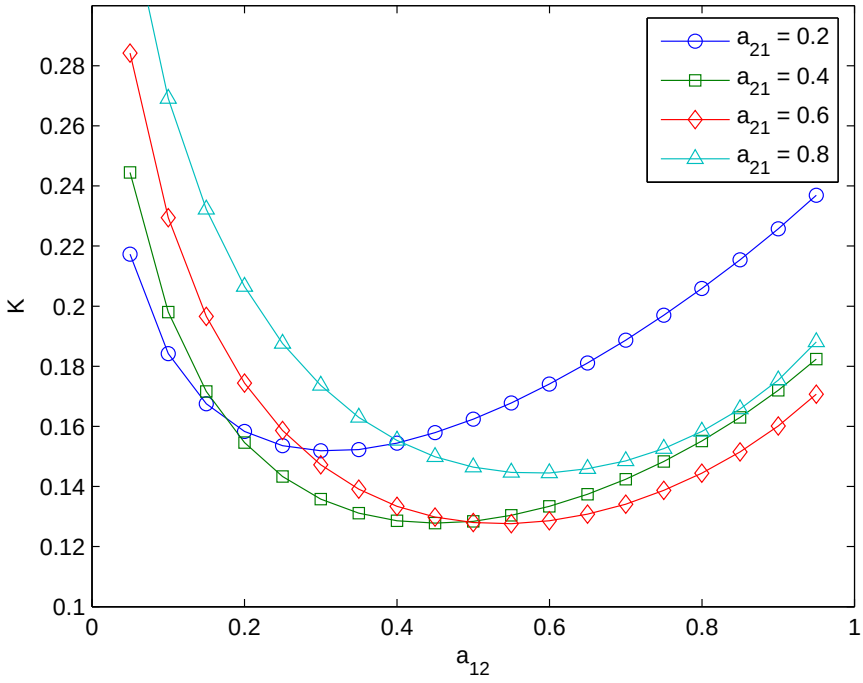

Figure 3.3: Error exponent K for various a_{12}, with a_{21} fixed

when a_{12} is close to zero or one is that in either of these situations the signal will have a certain amount of "correlation" which makes the signal easier to detect. This is because if a_{12} is close to zero then when it is in state 1, it tends to stay in the same state, whereas if a_{12} is close to one then when it is in state 1, the state at the next time instant is very likely to switch to state 2.

In Figure 3.4 we plot Monte Carlo simulations of the miss probability P_M versus the number of samples n, with a P_{FA} constraint of 0.01 and $\sigma^2 = 1$, for three different sets of transition parameters. We can see that the rate of decrease of P_M is greatest when $a_{12} = 0.9, a_{21} = 0.2$, followed by $a_{12} = 0.1, a_{21} = 0.2$ with the smallest for $a_{12} = 0.3, a_{21} = 0.2$. This is consistent with the results in Figure 3.3, where $a_{12} = 0.9, a_{21} = 0.2$ has the largest error exponent and $a_{12} = 0.3, a_{21} = 0.2$ the smallest error exponent for the three sets of values considered.

Next we look at how the error exponent behaves at different noise levels for fixed Markov chain transition parameters. In Figure 3.5 we plot the error exponent as the noise variance σ^2 varies, for four different sets of parameter values. We again set $\psi_1 = -1, \psi_2 = 1$. Observe that the error exponent appears to scale linearly with

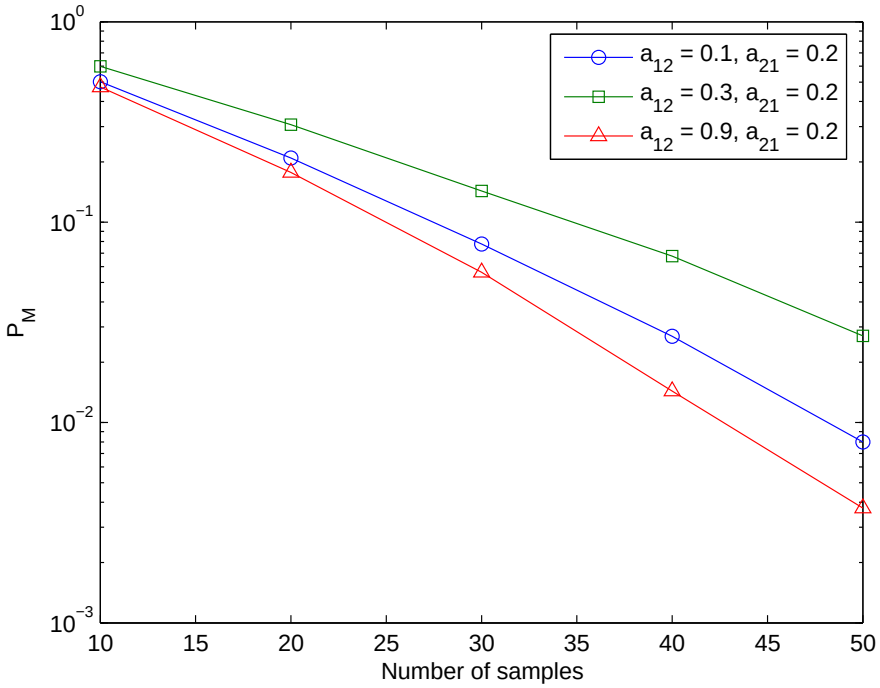

Figure 3.4: Miss probability versus number of samples, with $P_{FA} = 0.01$ and $\sigma^2 = 1$

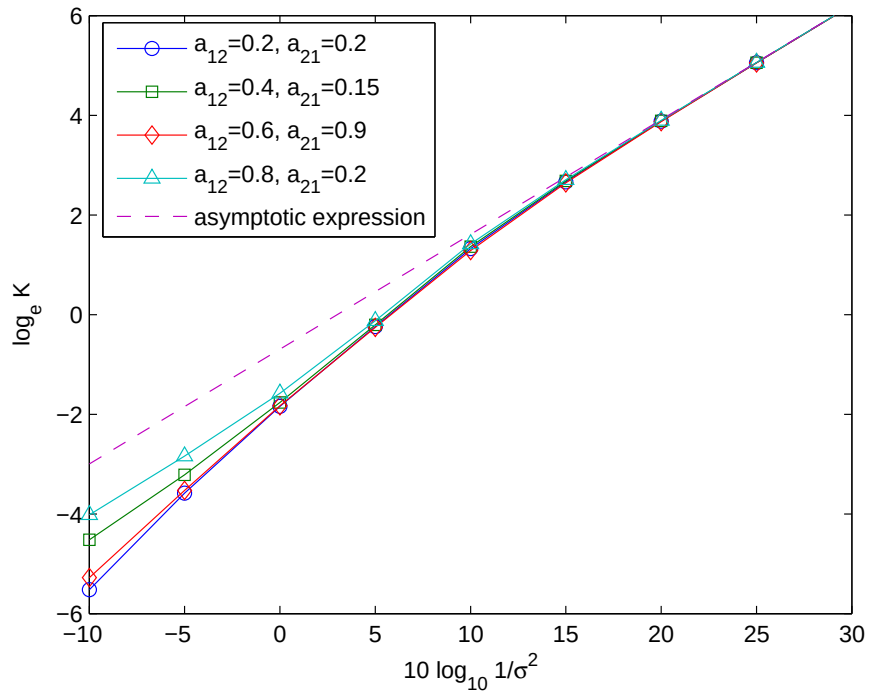

Figure 3.5: Error exponent K for various σ^2

the "signal-to-noise ratio" $\frac{\psi_1^2}{\sigma^2}$. At high SNR the error exponents also become very close to each other in a relative sense, for all four sets of parameter values used. We have the following result, which is proved in the appendix, Section 3.6.

Lemma 3.3.1. *With Gaussian noise and as* $\sigma^2 \to 0$*,*

$$K \sim \frac{\psi_1^2}{2\sigma^2} \tag{3.9}$$

where $f \sim g$ *means that* $\lim_{\sigma \to 0} \frac{f(\sigma)}{g(\sigma)} = 1$.

The dashed line in Figure 3.5 is the asymptotic expression (3.9). By Lemma 3.3.1 the scaling at high SNR is indeed linear. Moreover the asymptotic expression for K at high SNR does not depend on the parameters a_{12} and a_{21}, which completely specifies the two-state Markov chain. It is known that when detecting a constant signal in Gaussian noise, the error exponent scales linearly with SNR. In fact, it can be easily shown that for a constant signal of amplitude ψ_1, the error exponent is exactly $\frac{\psi_1^2}{2\sigma^2}$. So from Lemma 3.3.1, the effect of Markov chain state transitions on the performance of detectors become less important at high SNR. Another potential use of Lemma 3.3.1 is to show that the high SNR performance of a detector does not degrade too much even when the transition parameters may not be known exactly but only estimates are available, though this will require further investigation.

3.4 Error exponent computation for fading channels

We now consider a randomly time-varying Rayleigh[6] fading channel in the detection problem (3.3). Without loss of generality we will let the process $\{h_k\}$ have unit power, with density

$$p(h_k) = 2h_k \exp(-h_k^2), h_k \geq 0$$

[6]Our method should also work for other fading distributions, provided that after "averaging" over h_k (Niu *et al.*, 2006), $p_1(y_k|s_k = \psi_j)$ still has a tractible closed form such as (3.10)

as in (Niu *et al.*, 2006). We assume that h_k and h_l are independent for $k \neq l$. This is a reasonable assumption if the sampling time is greater than the channel coherence time, as might occur in low data rate systems such as wireless sensor networks, see also the block fading models commonly used in the communications theory literature (Ozarow *et al.*, 1994). We assume that the receiver knows the distribution of the fading process, but the instantaneous values of h_k are unknown, i.e. has no instantaneous channel state information (CSI). This is because acquiring channel knowledge might be too expensive when resources are limited, and also because the assumption of full CSI may not be appropriate in the context of a signal detection problem, depending on the nature of the source of the signal.

An inspection of the derivations in (Scharf & Nolte, 1977) shows that one can still write the likelihood ratio for problem (3.3) in the form $L_n(y^n) = \prod_{k=1}^n l(y_k|y^{k-1})$ where

$$l(y_k|y^{k-1}) = \frac{p_1(y_k|s_k = \psi_j, y^{k-1})}{p_0(y_k|y^{k-1})} \frac{P(s_k = \psi_j|y^{k-1})}{P(s_k = \psi_j|y^k)}$$

for any $j \in \{1, 2\}$ (the explicit expressions for the terms will now be different though). We have $p_1(y_k|s_k = \psi_j, y^{k-1}) = p_1(y_k|s_k = \psi_j)$ if the fades are independent. The recursive relations (3.4) and (3.5) will continue to hold. One also has, in a slight generalization of Lemma 1 in (Niu *et al.*, 2006),

$$p_1(y_k|s_k = \psi_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{2\sigma^2}{\psi_j^2 + 2\sigma^2} e^{-y_k^2/2\sigma^2} \left[1 + \sqrt{2\pi} a_j y_k e^{(a_j y_k)^2/2} Q(-a_j y_k)\right] \quad (3.10)$$

where

$$a_j \equiv \frac{\psi_j}{\sigma\sqrt{\psi_j^2 + 2\sigma^2}}$$

and

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

is the Q-function. For brevity, call

$$a \equiv a_1 = \frac{\psi_1}{\sigma\sqrt{\psi_1^2 + 2\sigma^2}}.$$

The error exponent can be written as

$$K = \lim_{n\to\infty} -\frac{1}{n}\sum_{k=1}^{n}\left[\log\frac{p_1(y_k|s_k=\psi_1)}{p_0(y_k|y^{k-1})} + \log P(s_k=\psi_1|y^{k-1}) - \log P(s_k=\psi_1|y^k)\right].$$

provided the limit exists. By similar arguments as in Section 3.3 and again using results from (Le Gland & Mevel, 2000), we can show that the almost sure limit does exist and is equal to

$$\begin{aligned} K = &-\log\frac{2\sigma^2}{\psi_1^2+2\sigma^2} - \int_{-\infty}^{\infty}\log\left[1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)\right]\frac{1}{\sqrt{2\pi\sigma^2}}e^{-y_k^2/2\sigma^2}dy_k \\ &-\mathbb{E}\left[\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}\right] \end{aligned} \tag{3.11}$$

where the second term (the integral) can be evaluated numerically. For calculation of the other terms, again let $q_k = P(s_k=\psi_1|y^k)$ and $f_k(q)dq = P(q_k\in(q,q+dq))$. The relation (3.7) still holds, but we now have from (3.5) that

$$\begin{aligned} q_k = &[1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)][a_{11}q_{k-1}+a_{21}(1-q_{k-1})]\Big/ \\ &\Big([1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)][a_{11}q_{k-1}+a_{21}(1-q_{k-1})] \\ &+[1-\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(ay_k)][a_{12}q_{k-1}+a_{22}(1-q_{k-1})]\Big) \end{aligned}$$

or

$$\frac{1-\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)} = \frac{a_{11}q_{k-1}+a_{21}(1-q_{k-1})}{a_{12}q_{k-1}+a_{22}(1-q_{k-1})}\times\frac{1-q_k}{q_k}. \tag{3.12}$$

Given q_k and q_{k-1}, (3.12) can be solved for y_k numerically. The solution obtained can be shown to be unique by the following argument. Since the right hand side of (3.12) is positive, for $a>0$,

$$\frac{1-\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)} \to \begin{cases} \infty & , \ y\to-\infty \\ 0 & , \ y\to+\infty. \end{cases}$$

Moreover, we obtain after some calculations that

$$\frac{d}{dy_k}\left(\frac{1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)}\right)=\frac{-\sqrt{2\pi}ae^{(ay_k)^2/2}}{\left[1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)\right]^2}<0$$

so that the left hand side of (3.12) is strictly decreasing in y_k, and hence a unique solution will be obtained. A similar argument holds if $a<0$.

Next, using implicit differentation on (3.12) we obtain

$$\frac{-\sqrt{2\pi}ae^{(ay_k)^2/2}}{\left[1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)\right]^2}\frac{dy_k}{dq_k}=-\frac{1}{q_k^2}\times\frac{a_{11}q_{k-1}+a_{21}(1-q_{k-1})}{a_{12}q_{k-1}+a_{22}(1-q_{k-1})}$$

or

$$\left|\frac{dy_k}{dq_k}\right|=\frac{\left[1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)\right]^2}{\sqrt{2\pi}\,|a|\,e^{(ay_k)^2/2}q_k^2}\times\frac{a_{11}q_{k-1}+a_{21}(1-q_{k-1})}{a_{12}q_{k-1}+a_{22}(1-q_{k-1})}. \tag{3.13}$$

By a change of variable

$$\begin{aligned}S(q,\tilde{q})dq &\equiv P(q_k\in(q,q+dq)|q_{k-1}=\tilde{q})\\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-y_k^2/2\sigma^2}\left|\frac{dy_k}{dq_k}\right|_{|q_k=q,q_{k-1}=\tilde{q}}dq\end{aligned}$$

where y_k is the numerical solution to (3.12) and $\left|\frac{dy_k}{dq_k}\right|$ is given by (3.13). We will then need to solve the integral equation

$$f(q)dq=\int_0^1 S(q,\tilde{q})f(\tilde{q})d\tilde{q}dq$$

using the same numerical procedure as in Section 3.3.

3.4.1 Numerical studies

In Figure 3.2 we show a 3D plot of the error exponent for different values of a_{12} and a_{21}. We fix $\psi_1=-1,\psi_2=1,\sigma^2=1$. In Figure 3.7 we show plots for four different fixed values of a_{21}, as a_{12} is varied. Similar behaviour to the case without fading in Figures 3.2 and 3.3 can be observed, though the error exponents are smaller

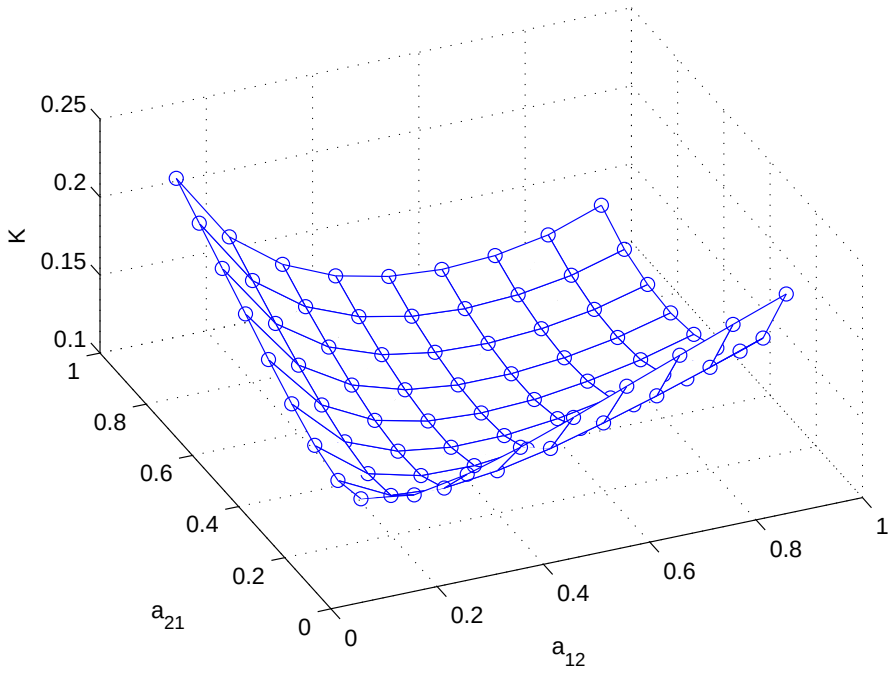

Figure 3.6: Error exponent K with Rayleigh fading for various a_{12} and a_{21}

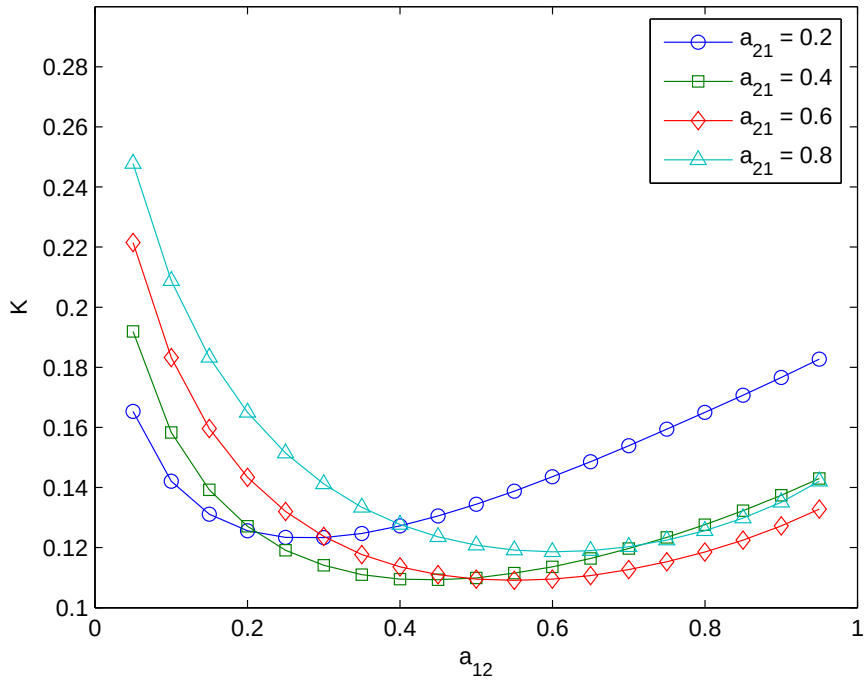

Figure 3.7: Error exponent K with Rayleigh fading for various a_{12}, with a_{21} fixed

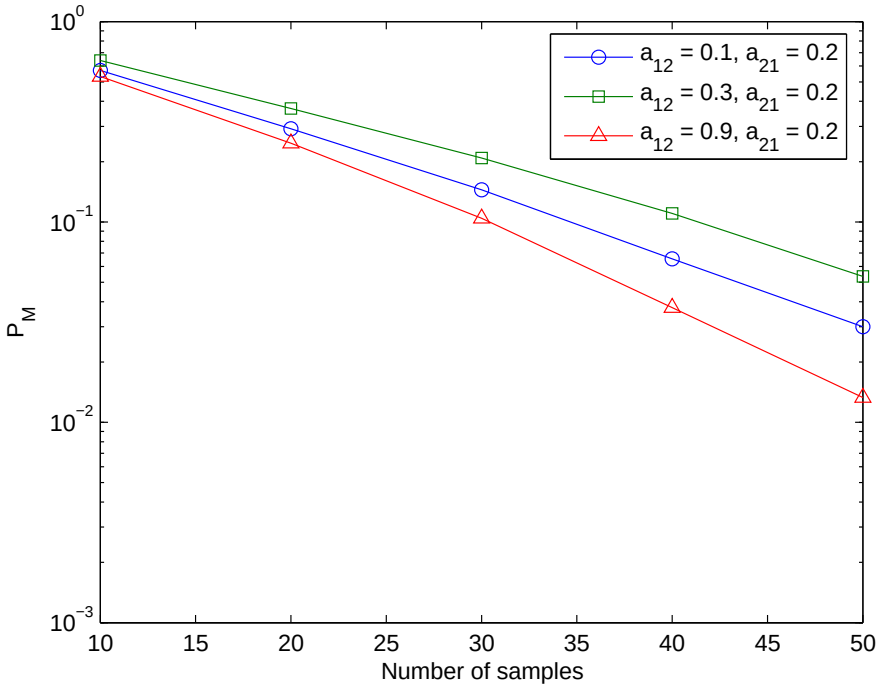

Figure 3.8: Miss probability with Rayleigh fading versus number of samples, with $P_{FA} = 0.01$ and $\sigma^2 = 1$

due to the presence of fading. Figure 3.8 plots Monte Carlo simulations of P_M versus the number of samples, which also demonstrates the deterioration in detection performance when there is fading when compared with Figure 3.4.

In Figure 3.9 we plot the error exponent with fading as the noise variance σ^2 varies, for four different sets of parameter values. We again set $\psi_1 = -1, \psi_2 = 1$. The error exponent now appears to scale logarithmically with the signal-to-noise ratio. As in Figure 3.5, at high SNR the error exponents approach each other (in a relative sense) for all four sets of parameter values used. We have the following:

Lemma 3.4.1. *With Gaussian noise and Rayleigh fading, as* $\sigma^2 \to 0$,

$$K \sim \log \frac{\psi_1^2}{2\sigma^2}. \tag{3.14}$$

The proof is in Section 3.6. The dashed line in Figure 3.9 is a plot of the expression (3.14). Note that we don't plot for smaller values of $1/\sigma^2$ as then the asymptotic expression takes on negative values. We see that the asymptotic expression is log-

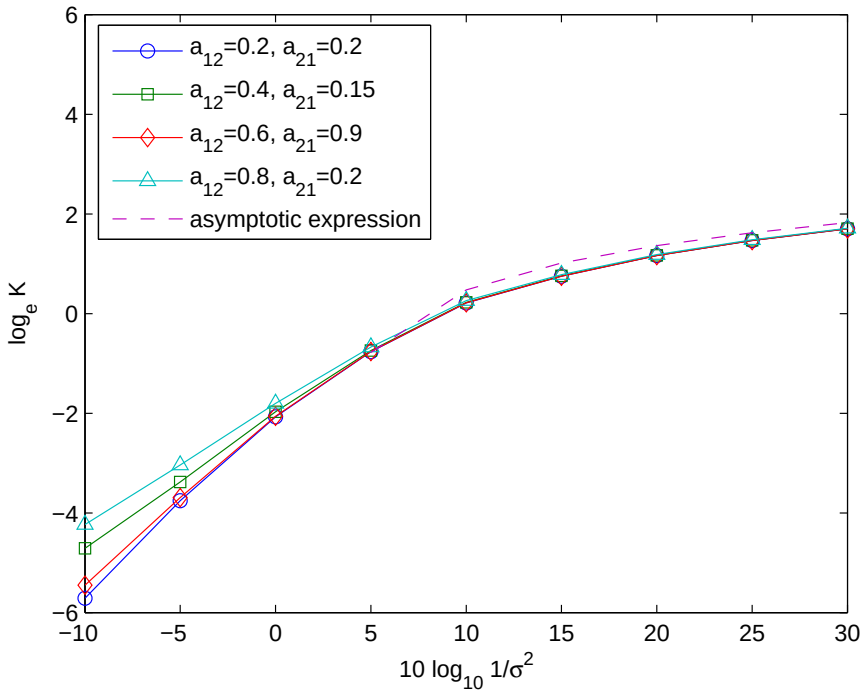

Figure 3.9: Error exponent K with Rayleigh fading for various σ^2

arithmic, where again there is no dependence on the parameters a_{12} and a_{21}. The error exponent for detecting a constant signal of amplitude ψ_1 with the Rayleigh fading model used here can be shown to be asymptotic to $\log \frac{\psi_1^2}{2\sigma^2}$, so that as in the time-invariant case, state transitions of the Markov chain have a less significant effect on performance at high SNR.

The difference in behaviour in the scaling of the error exponent with SNR in Figures 3.5 and 3.9 can possibly be attributed to the discrete nature of the states in the signal $\{s_k\}$ in the case without fading, whereas with fading the received signal $\{h_k s_k\}$ (even without noise) can take on a continuous range. Other situations where logarithmic scaling in the error exponent has also been observed include (see e.g. (Sung *et al.*, 2006)) the detection of Gaussian signals in Gaussian noise, and the detection of Gauss-Markov systems in noise, where in both cases the signals to be detected have a continuous range.

3.5 Conclusion

We have presented numerical methods to compute the error exponent associated with Neyman-Pearson detection of a two-state Markov chain in noise, both with and without fading. Numerical studies relating the error exponent to the parameters of the system have been presented. In particular, with Gaussian noise, at high SNR the error exponents scale linearly without fading but logarithmically with fading.

While the results presented in this chapter have dealt with Gaussian noise, the methods to compute the error exponent will in principle still work with other noise distributions. However, in general the high SNR behaviour of the error exponent in non-Gaussian noise is likely to be highly dependent on the nature of the distributions. For instance, consider the generalized Gaussian distribution (Miller & Thomas, 1972) with density

$$f(x) = \frac{c\eta(\sigma, c)}{2\Gamma(1/c)} \exp\{-[\eta(\sigma, c)|x|]^c\},$$

where $\eta(\sigma, c) \equiv \sigma^{-1}[\Gamma(3/c)/\Gamma(1/c)]^{1/2}$, σ^2 is the noise variance, and $c > 0$ is a parameter that controls the rate of decay of the tails. For $c = 2$, this reduces to the Gaussian density and for $c = 1$ the Laplacian density. Using this density for the noise, preliminary results for the time-invariant case suggest a $O(\frac{1}{\sigma^c})$ scaling of the error exponent at high SNR, so that for larger c (faster decaying tail probabilities) the rate of increase of the error exponent is greater than for distributions with heavier tails. Future work will include a more thorough investigation of the generalized Gaussian distribution and other types of noise and fading processes.

3.6 Appendix

3.6.1 Proof of Lemma 3.3.1

Note that $\frac{\psi_1^2}{2\sigma^2}$ is just the first term on the right hand side of (3.6), so it suffices to show that

$$\mathbb{E}\left[\log \frac{P(s_k = \psi_1|y^{k-1})}{P(s_k = \psi_1|y^k)}\right] = o\left(\frac{1}{\sigma^2}\right)$$

for small σ. Denote

$$p_o(y_k) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} e^{-y_k^2/2\sigma^2}.$$

Suppose $P(s_k = \psi_1 | y^{k-1}) = p$. Then from (3.5)

$$P(s_k = \psi_1 | y^k) = \frac{\exp\left(\frac{-(y_k-\psi_1)^2}{2\sigma^2}\right) p}{\exp\left(\frac{-(y_k-\psi_1)^2}{2\sigma^2}\right) p + \exp\left(\frac{-(y_k+\psi_1)^2}{2\sigma^2}\right)(1-p)}$$

and so

$$\log \frac{P(s_k = \psi_1 | y^{k-1})}{P(s_k = \psi_1 | y^k)} = \log\left(p + (1-p)\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right)\right).$$

Suppose $\psi_1 > 0$ (the calculations for $\psi_1 < 0$ are similar). For $y_k > 0$, one can easily show that

$$\log p \leq \log\left(p + (1-p)\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right)\right) \leq 0.$$

So

$$\int_0^\infty \log\left(p + (1-p)\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right)\right) p_0(y_k) dy_k = O(1).$$

For $y_k < 0$,

$$\begin{aligned}
&\int_{-\infty}^0 \log\left(p + (1-p)\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right)\right) p_0(y_k) dy_k \\
&\quad = \int_0^\infty \log\left(p + (1-p)\exp\left(\frac{2\psi_1 y_k}{\sigma^2}\right)\right) p_0(y_k) dy_k \\
&\quad = \int_0^\infty \log\left[\exp\left(\frac{2\psi_1 y_k}{\sigma^2}\right)\left(p\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right) + (1-p)\right)\right] p_0(y_k) dy_k \\
&\quad = \int_0^\infty \left(\frac{2\psi_1 y_k}{\sigma^2} + \log\left(p\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right) + (1-p)\right)\right) p_0(y_k) dy_k \\
&\quad = \sqrt{\frac{2}{\pi}}\frac{\psi_1}{\sigma} + \int_0^\infty \log\left(p\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right) + 1 - p\right) p_0(y_k) dy_k
\end{aligned}$$

and the term

$$\int_0^\infty \log\left(p\exp\left(-\frac{2\psi_1 y_k}{\sigma^2}\right) + 1 - p\right) p_0(y_k) dy_k$$

is $O(1)$ by a similar argument to the case when $y_k > 0$. So

$$\begin{aligned}\mathbb{E}\left[\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}\Big|p\right] &= O(1)+O\left(\frac{1}{\sigma}\right)+O(1)\\ &= O\left(\frac{1}{\sigma}\right)=o\left(\frac{1}{\sigma^2}\right)\end{aligned}$$

for small σ.

Hence

$$\mathbb{E}\left[\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}\right]=\mathbb{E}\left[\mathbb{E}\left[\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}\Big|p\right]\right]$$

is also $o\left(\frac{1}{\sigma^2}\right)$.

3.6.2 Proof of Lemma 3.4.1

Referring to (3.11), we have

$$-\log\frac{2\sigma^2}{\psi_1^2+2\sigma^2}\sim\log\frac{\psi_1^2}{2\sigma^2}.$$

Denote

$$p_o(y_k)\equiv\frac{1}{\sqrt{2\pi\sigma^2}}e^{-y_k^2/2\sigma^2}.$$

We first show that

$$\int_{-\infty}^{\infty}\log\left[1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)\right]p_0(y_k)dy_k=o\left(\log\frac{1}{\sigma^2}\right).$$

Assume $\psi_1>0$ (again the calculations for $\psi_1<0$ are similar), which implies a is positive. We can write

$$\begin{aligned}&\int_{-\infty}^{\infty}\log\left[1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)\right]p_0(y_k)dy_k\\ &\quad=\int_0^{\infty}\log\left[1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)\right]p_0(y_k)dy_k\\ &\qquad+\int_0^{\infty}\log\left[1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)\right]p_0(y_k)dy_k\end{aligned}$$

We have

$$\begin{aligned}
&\int_0^\infty \log\left[1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)\right]p_0(y_k)dy_k \\
&\quad\le \int_0^\infty \log\left[1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}\right]p_0(y_k)dy_k \\
&\quad= \int_0^\infty \log\left[e^{(ay_k)^2/2}\left(e^{-(ay_k)^2/2}+\sqrt{2\pi}ay_k\right)\right]p_0(y_k)dy_k \\
&\quad\le \int_0^\infty \left(\frac{(ay_k)^2}{2}+\log\left(1+\sqrt{2\pi}ay_k\right)\right)p_0(y_k)dy_k \\
&\quad\le \int_0^\infty \left(\frac{(ay_k)^2}{2}+\sqrt{2\pi}ay_k\right)p_0(y_k)dy_k \\
&\quad= \frac{a^2\sigma^2}{4}+a\sigma = O(1)
\end{aligned}$$

since $a \equiv \frac{\psi_1}{\sigma\sqrt{\psi_1^2+2\sigma^2}}$. So

$$\int_0^\infty \log\left[1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)\right]p_0(y_k)dy_k = O(1).$$

Next note that

$$\left|\log\left[1-\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(ay_k)\right]\right| \le \log\left[1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)\right]$$

using the result previously shown in Section 3.4 that

$$\frac{1-\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)}$$

is decreasing with y_k, and is equal to 1 when $y_k = 0$. So

$$\int_0^\infty \log\left[1-\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(ay_k)\right]p_0(y_k)dy_k$$

is also $O(1)$. Hence

$$\int_{-\infty}^\infty \log\left[1+\sqrt{2\pi}ay_k e^{(ay_k)^2/2}Q(-ay_k)\right]p_0(y_k)dy_k = O(1)+O(1) = o\left(\log\frac{1}{\sigma^2}\right).$$

Using the bounds just derived, together with the same arguments as for the proof

of Lemma 3.3.1, we can show that the remaining term in (3.11) satisfies

$$\mathbb{E}\left[\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}\right]=O(1)=o\left(\log\frac{1}{\sigma^2}\right),$$

as follows. Suppose $P(s_k=\psi_1|y^{k-1})=p$. Then

$$\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}=\log\left(p+(1-p)\frac{1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)}\right).$$

For $y_k>0$, one can show easily that

$$\log p\le\log\left(p+(1-p)\frac{1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)}\right)\le 0,$$

so

$$\int_0^\infty\log\left(p+(1-p)\frac{1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)}\right)p_0(y_k)dy_k=O(1).$$

For $y_k<0$,

$$\begin{aligned}
&\int_{-\infty}^0\log\left(p+(1-p)\frac{1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)}{1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)}\right)p_0(y_k)dy_k\\
&\quad=\int_0^\infty\log\left(p+(1-p)\frac{1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)}{1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)}\right)p_0(y_k)dy_k\\
&\quad\le\int_0^\infty\log\left(1+\frac{1+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)}{1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)}\right)p_0(y_k)dy_k\\
&\quad=\int_0^\infty\log\left(2+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)\right)p_0(y_k)dy_k\\
&\qquad\quad-\int_0^\infty\log\left(1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)\right)p_0(y_k)dy_k\\
&\quad\le\int_0^\infty\log\left(2+\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(-ay_k)\right)p_0(y_k)dy_k\\
&\qquad\quad-\int_0^\infty\log\left(1-\sqrt{2\pi}ay_ke^{(ay_k)^2/2}Q(ay_k)\right)p_0(y_k)dy_k
\end{aligned}$$

which by previous arguments is $O(1)$. Hence

$$\mathbb{E}\left[\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}\right]=\mathbb{E}\left[\mathbb{E}\left[\log\frac{P(s_k=\psi_1|y^{k-1})}{P(s_k=\psi_1|y^k)}\Big|p\right]\right]=O(1).$$

Chapter 4

Kalman Smoothing with Random Packet Loss

4.1 Introduction

Problems involving estimation over lossy communication networks have received considerable attention in recent years, due to their relevance in areas such as wireless sensor networks and networked control systems. When measurements from sensors are located at separate locations and have to be transmitted for processing through unreliable (e.g. wireless) channels losses can occur, and how these packet losses affect the performance of the estimator is of significant interest.

Early work on state estimation with measurements losses include (Nahi, 1969), who derived the optimal linear estimator for linear systems with i.i.d. Bernoulli losses, where the parameters of the loss process is known, but which of the individual measurements are lost/received is not explicitly known. This was later extended to the optimal linear smoother in (Monzingo, 1975). More recently, in the case where we know which measurements are lost/received, it was shown in (Sinopoli *et al.*, 2004) that for an unstable system with i.i.d. Bernoulli losses there exists a critical threshold such that the expected value of the error covariance (which is randomly time-varying due to the random losses) will be bounded if the packet arrival rate exceeds this threshold, but will diverge otherwise. Further avenues of research suggested in (Sinopoli *et al.*, 2004) include studying multiple sensors (Liu & Goldsmith, 2004), correlated loss processes such as Markov (Huang & Dey, 2007), consideration of delays (Schenato, 2006), and smoothing, which is the subject of this chapter. A different notion of estimator performance using probabilistic constraints was considered in (Shi *et al.*, 2005). A survey of these and other related results can be found in (Hespanha *et al.*, 2007).

If the sensor has local computation ability, which is sometimes referred to as a "smart sensor", then an alternative scheme is to transmit the state estimate instead of the raw measurements (Xu & Hespanha, 2005), which allows estimator stability to be achieved with lower packet arrival rates. With computation available at the sensors, distributed Kalman filtering with quantized measurements (Ribeiro *et al.*, 2006) has also been considered. Estimation problems with random measurement losses for other systems include: jump linear systems (Fletcher *et al.*, 2004), hidden Markov models (Leong *et al.*, 2007), robust finite horizon filtering (Wang *et al.*, 2005) and robust H_∞ filtering (Wang *et al.*, 2006). The related problem of control over packet dropping links has also received considerable attention, see (Schenato *et al.*, 2007; Hespanha *et al.*, 2007) and the references therein.

This chapter considers the situation where we allow for some additional delay and computational complexity so that Kalman smoothing can be done on the measurements, and whether this provides any advantages over filtering. We assume that the sensors transmit the raw measurements directly as they do not have enough computation ability to be a smart sensor. We first derive in Section 4.2 the equations for the Kalman fixed lag smoother with random packet loss and use these equations to analyse its performance. While intuitively we might expect smoothing to perform better than filtering, in Section 4.3 we show that with the stability notion via expected error covariance in (Sinopoli *et al.*, 2004), the use of Kalman smoothing does not actually provide any improvement over the Kalman filter. However, using instead the probabilistic notion of performance in (Shi *et al.*, 2005), we will see in Section 4.4 that the Kalman smoother can provide significant performance gains over the Kalman filter. In Section 4.5 we analyze Kalman filtering using two simple retransmission strategies, which we find provides the same performance as a Kalman filter without retransmission, and hence Kalman smoothing outperforms these retransmission strategies with the probabilistic performance measure of (Shi *et al.*, 2005).

4.2 Derivation of Kalman fixed lag smoother

Let the discrete time linear system be:

$$\begin{aligned} x_{k+1} &= Ax_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{4.1}$$

with w_k and v_k being Gaussian with zero mean and covariance matrices $Q \geq 0$ and $R > 0$ respectively. Let $\{\gamma_k\}$ be the random packet loss process which is equal to 1 if the measurement y_k is received at time k, and 0 otherwise. Define the state estimates and corresponding error covariances as

$$\begin{aligned} \hat{x}_{k|m} &= \mathbb{E}\left[x_k | \{Y_0, \ldots, Y_m\}, \{\gamma_0, \ldots, \gamma_m\}\right] \\ P_{k|m} &= \mathbb{E}\left[(x_k - \hat{x}_{k|m})(x_k - \hat{x}_{k|m})^T | \{Y_0, \ldots, Y_m\}, \{\gamma_0, \ldots, \gamma_m\}\right]. \end{aligned}$$

In (Sinopoli *et al.*, 2004), it is shown that the Kalman filter equations for the system (4.1) when there is packet loss can be written as

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + \gamma_k K_k (y_k - C\hat{x}_{k|k-1}) \\ P_{k+1|k} &= AP_{k|k-1}A^T - \gamma_k K_k C P_{k|k-1} A^T + Q \end{aligned} \tag{4.2}$$

where $K_k = AP_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}$ (note that K_k here is defined slightly differently from (Sinopoli *et al.*, 2004)).

Now from (Anderson & Moore, 1979, p.176-179), it is known that one way to derive a fixed lag Kalman smoother for the system (4.1), with smoothing lag N, is to consider the "augmented" model

$$\begin{bmatrix} x_{k+1} \\ x_{k+1}^{(1)} \\ x_{k+1}^{(2)} \\ \vdots \\ x_{k+1}^{(N+1)} \end{bmatrix} = \begin{bmatrix} A & 0 & \ldots & 0 & 0 \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_k^{(1)} \\ x_k^{(2)} \\ \vdots \\ x_k^{(N+1)} \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} w_k$$

$$y_k = \begin{bmatrix} C & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_k^{(1)} \\ x_k^{(2)} \\ \vdots \\ x_k^{(N+1)} \end{bmatrix} + v_k$$

so that

$$x_{k+1}^{(1)} = x_k, x_{k+1}^{(2)} = x_{k-1}, \dots, x_{k+1}^{(N+1)} = x_{k-N}.$$

Define

$$\hat{x}_{k+1|k}^{(i)} \equiv \mathbb{E}[x_{k+1}^{(i)} | \{Y_0, \dots, Y_k\}, \{\gamma_0, \dots, \gamma_k\}] = \hat{x}_{k-(i-1)|k}, i = 1, \dots, N+1$$

$$\hat{\mathbf{x}}_{k+1|k} = \begin{bmatrix} \hat{x}_{k+1|k} \\ \hat{x}_{k+1|k}^{(1)} \\ \vdots \\ \hat{x}_{k+1|k}^{(N+1)} \end{bmatrix}, \mathbf{K}_k = \begin{bmatrix} K_k \\ K_k^{(1)} \\ \vdots \\ K_k^{(N+1)} \end{bmatrix}$$

$$\mathbf{P}_{k+1|k} = \begin{bmatrix} P_{k+1|k}^{(0,0)} & P_{k+1|k}^{(0,1)} & \dots & P_{k+1|k}^{(0,N+1)} \\ P_{k+1|k}^{(1,0)} & P_{k+1|k}^{(1,1)} & \dots & P_{k+1|k}^{(1,N+1)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{k+1|k}^{(N+1,0)} & P_{k+1|k}^{(N+1,1)} & \dots & P_{k+1|k}^{(N+1,N+1)} \end{bmatrix}$$

where we make the identifications

$$P_{k+1|k}^{(0,0)} = P_{k+1|k}, P_{k+1|k}^{(0,i)} = P_{k+1|k}^{(i)}, P_{k+1|k}^{(i,i)} = P_{k-(i-1)|k}$$

Then applying the result of (Sinopoli *et al.*, 2004) (equation (4.2)) to the augmented model with $\hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}, \mathbf{K}_k$ in place of $\hat{x}_{k+1|k}, P_{k+1|k}, K_k$, the Kalman smoother

equations with packet loss can be extracted after some computation as follows:

$$\begin{aligned}
\hat{x}_{k+1|k} &= A\hat{x}_{k|k-1} + \gamma_k K_k(y_k - C\hat{x}_{k|k-1}) \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + \gamma_k K_k^{(1)}(y_k - C\hat{x}_{k|k-1}) \\
\hat{x}_{k-i|k} &= \hat{x}_{k-i|k-1} + \gamma_k K_k^{(i+1)}(y_k - C\hat{x}_{k|k-1}), i = 1, \ldots, N \\
K_{k+1} &= AP_{k+1|k}C^T(CP_{k+1|k}C^T + R)^{-1} \\
K_{k+1}^{(i)} &= P_{k+1|k}^{(i-1)}C^T(CP_{k+1|k}C^T + R)^{-1}, i = 1, \ldots, N+1 \\
P_{k+1|k} &= AP_{k|k-1}A^T - \gamma_k K_k C P_{k|k-1}A^T + Q \\
P_{k+1|k}^{(i)} &= AP_{k|k-1}^{(i-1)} - \gamma_k K_k C P_{k|k-1}^{(i-1)}, i = 1, \ldots, N+1
\end{aligned} \tag{4.3}$$

with the error covariances of the filtered and smoothed estimates given by

$$\begin{aligned}
P_{k|k} &= P_{k|k-1} - \gamma_k K_k^{(1)} C P_{k|k-1} \\
P_{k-i|k} &= P_{k-i|k-1} - \gamma_k K_k^{(i+1)} C P_{k|k-1}^{(i)}, i = 1, \ldots, N
\end{aligned} \tag{4.4}$$

and with initial conditions as stated in (Anderson & Moore, 1979).

When there are no packet losses at time k we have $\gamma_k = 1$, and we see that the equations reduce to the standard fixed lag smoother equations. However when a measurement at time k is lost, i.e. $\gamma_k = 0$, we have $\hat{x}_{k-i|k} = \hat{x}_{k-i|k-1}$ and $P_{k-i|k} = P_{k-i|k-1}$ and the previous estimates and covariance matrices will get propagated.

In Figure 4.1 we show an example comparing the error covariances of the predicted, filtered and smoothed (with lag 1) estimates, for a scalar system with $A = 1.3, C = 1, R = 1, Q = 0.5$. The sequence $\{\gamma_k\}$ here is 111101001110000111.... Propagation of the error covariances when there are measurement losses can be readily observed.

4.3 Stability of Kalman smoothing

In (Sinopoli *et al.*, 2004), the authors showed that for an unstable system (i.e. the matrix A has an eigenvalue with magnitude ≥ 1) with Bernoulli packet losses, there exists a critical threshold such that the expected value of the error covariance $P_{k|k-1}$ will be bounded if the packet arrival rate exceeds this threshold, but becomes

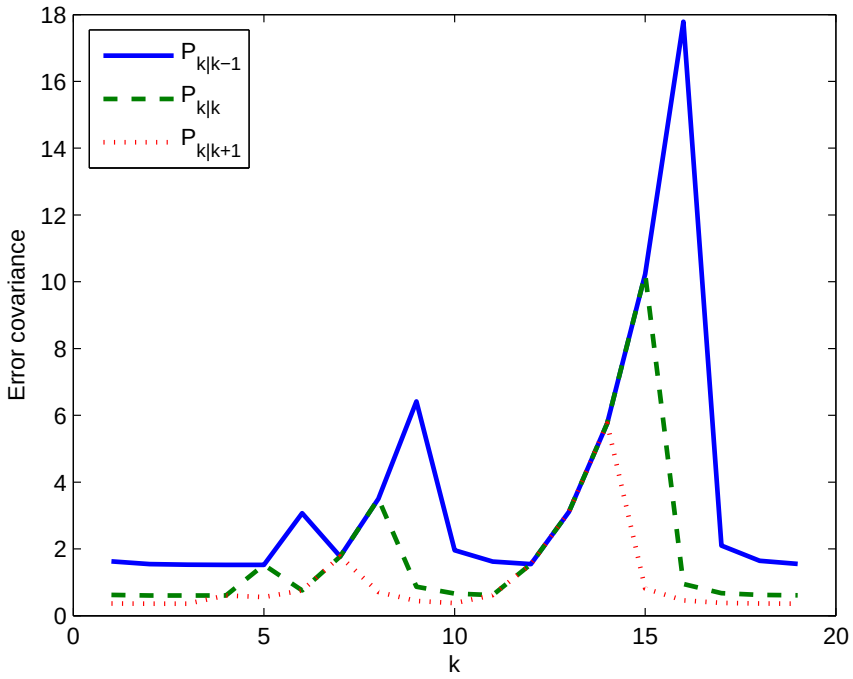

Figure 4.1: Comparison of the error covariances of the predicted, filtered and smoothed (with lag 1) estimates.

unbounded otherwise. Now from Figure 4.1 it might seem reasonable to suggest that smoothing gives an improvement over filtering. However in this section we will show the somewhat surprising result that for an unstable system with Bernoulli losses, smoothing does not improve the stability of the estimator, in the sense that the critical threshold for keeping the expected error covariances bounded in the smoothing case is the same as in the filtering case.

We first introduce some definitions. For $\{\gamma_k\}$ a Bernoulli process, let $\lambda = \mathbb{P}(\gamma_k = 1)$, then λ will also be the arrival rate of the measurements. For matrices P and Q, $P \leq Q$ will mean that $Q - P$ is positive semi-definite. Matrices P_k are said to be bounded if there exists a matrix $M < \infty$ such that $P_k \leq M, \forall k$, and unbounded if such an M does not exist. Then we have the following:

Lemma 4.3.1. *For unstable systems, as $k \to \infty$, $\mathbb{E}[P_{k|k+N}]$ is unbounded if and only if $\mathbb{E}[P_{k|k-1}]$ is unbounded.*

Proof. First, from the smoother equations (4.4) and the definitions of $K_k^{(i)}$ in (4.3), it is not hard to see that $P_{k|k+N} \leq P_{k|k-1}, \forall k$. So the implication that $\mathbb{E}[P_{k|k-1}]$

bounded $\Rightarrow \mathbb{E}[P_{k|k+N}]$ bounded, or the equivalent statement $\mathbb{E}[P_{k|k+N}]$ unbounded $\Rightarrow \mathbb{E}[P_{k|k-1}]$ unbounded, immediately follows.

We now show that $\mathbb{E}[P_{k|k-1}]$ unbounded $\Rightarrow \mathbb{E}[P_{k|k+N}]$ unbounded. Let $\lambda < \lambda_c$, where λ_c is the critical arrival rate such that $\mathbb{E}[P_{k|k-1}]$ will be unbounded as $k \to \infty$ if and only if $\lambda < \lambda_c$, whose existence for unstable systems is shown in (Sinopoli *et al.*, 2004).[1] Let $\mathcal{A}$ be the event that measurements at time $k, k+1, \ldots, k+N$ are lost. Using some elementary properties of positive semi-definite matrices and the smoother equations (4.4), we have

$$\begin{aligned} \mathbb{E}[P_{k|k+N}] &= \mathbb{E}[P_{k|k+N}|\mathcal{A}]\mathbb{P}(\mathcal{A}) + \mathbb{E}[P_{k|k+N}|\mathcal{A}^c]\mathbb{P}(\mathcal{A}^c) \\ &\geq \mathbb{E}[P_{k|k+N}|\mathcal{A}]\mathbb{P}(\mathcal{A}) \\ &= \mathbb{E}[P_{k|k-1}]\mathbb{P}(\mathcal{A}) \\ &= \mathbb{E}[P_{k|k-1}](1-\lambda)^{N+1} \end{aligned} \tag{4.5}$$

and $\mathbb{E}[P_{k|k-1}]$ by hypothesis is unbounded as $k \to \infty$ for $\lambda < \lambda_c$. Hence $\mathbb{E}[P_{k|k+N}]$ is also unbounded as $k \to \infty$. □

Since $\mathbb{E}[P_{k|k+N}]$ is unbounded if and only if $\mathbb{E}[P_{k|k-1}]$ is unbounded, the critical threshold on the arrival rate of packets for the Kalman smoother must therefore be the same as the critical threshold λ_c for the Kalman filter derived in (Sinopoli *et al.*, 2004). Thus from the stability point of view of keeping the expected error covariances bounded, there is no advantage to be gained in smoothing. One can compare this result with the work in (Schenato, 2006) where packets can be both lost or delayed, which showed that stability of the Kalman filter (using constant gains) does not depend on packet delay but only on the probability that the packet eventually arrives.

The above result can also be extended to the case of Markovian packet losses (Huang & Dey, 2007), at least for scalar systems. Theorem 6 of (Huang & Dey, 2007) gives the condition that the Kalman filter for a scalar system is stable if and

[1] In the general vector case λ_c cannot be determined even numerically, though upper and lower bounds can be derived (Sinopoli *et al.*, 2004), however exact knowledge of the critical rate will not be required in the proof.

only if q, the probability of receiving the next measurement given that the current measurement is lost, satisfies $A^2(1-q) < 1$. The statement and proof of Lemma 4.3.1 can be easily modified to deal with this situation. To show that $\mathbb{E}[P_{k|k-1}]$ unbounded $\Rightarrow \mathbb{E}[P_{k|k+N}]$ unbounded, we choose $q < 1 - \frac{1}{A^2}$ instead of $\lambda < \lambda_c$. Then the calculation (4.5) holds except that now we replace $\mathbb{P}(\mathcal{A}) = (1-\lambda)^{N+1}$ with $\mathbb{P}(\mathcal{A}) = \mathbb{P}(\gamma_k = 0)(1-q)^N$.

4.4 Kalman smoothing with probabilistic constraints

Rather than studying the expected error covariance, an alternative notion of performance for Kalman filtering which has been considered is putting probabilistic constraints on the error covariance (Shi *et al.*, 2005; Adlakha, 2005). The motivation for this is that low probability events, such as a long sequence of measurement losses, can cause the expected error covariance to become unbounded, even when "typical" behaviour is such that the error covariance will lie below a certain value with high probability.

Given an upper bound M and an $\epsilon \in (0,1)$, one can ask the question as to what packet arrival rate λ is required in order to satisfy the constraint

$$\mathbb{P}(P_{k|k-1} \leq M) > 1 - \epsilon \tag{4.6}$$

where ϵ is usually chosen to be small so that the error covariance satisfies $P_{k|k-1} \leq M$ with probability close to one. This can be extended naturally to Kalman smoothing as the constraint

$$\mathbb{P}(P_{k|k+N} \leq M) > 1 - \epsilon \tag{4.7}$$

While we showed in Section 4.3 that estimator stability in the sense of keeping expected error covariance bounded cannot be improved by smoothing, the situation is different when we consider probabilistic constraints. For instance, we shall see that given M and ϵ the arrival rate which is sufficient to satisfy (4.7) will be smaller than what is required to satisfy (4.6).

Define

$$k_1 \equiv \min\{k \in Z^+ : h^k(\bar{M}) \not\leq M\}$$

where $h^k(X)$ means that the operator $h(X) \equiv AXA^T + Q$ is applied k times to X. For C invertible (e.g. a scalar system), $\bar{M}$ is defined as $\bar{M} \equiv AC^{-1}R(C^T)^{-1}A^T + Q$. When C is not invertible the expression for $\bar{M}$ is more complicated and may be found in (Epstein *et al.*, 2007).

Then it is shown in (Shi *et al.*, 2005, Corollary 4) that if the losses are Bernoulli and the packet arrival rate λ satisifes the condition

$$\lambda \geq 1 - \epsilon^{\frac{1}{k_1}}, \tag{4.8}$$

then (4.6) will also be satisfied. The quantity k_1 specifies the number of successive packet losses that can be tolerated before the error covariance updates for $P_{k|k-1}$ applied to the matrix $\bar{M}$ can no longer be bounded by the threshold M, which is then used to derive the condition (4.8). Using the property of the smoother equations (4.4) that previous estimates will get propagated when no measurements are received, it can be seen that an extra $N + 1$ losses can be tolerated before the smoothing error covariances (with lag N) will exceed the threshold M. Hence replacing k_1 with $k_1 + N + 1$ allows us to translate the results of (Shi *et al.*, 2005) to the smoothing case, so that if

$$\lambda \geq 1 - \epsilon^{\frac{1}{k_1+N+1}} \tag{4.9}$$

then the condition (4.7) will be satisfied. From the condition (4.9), it may be seen that as the smoothing lag N increases, the packet arrival rate λ which is sufficient to guarantee (4.7) will be smaller, at the expense of additional delay and computational complexity.

As an illustration consider the scalar system $A = 1.3, C = 1, Q = 0.5, R = 1$, where it can be determined that $k_1 = 2$. Choose $M = 6.25$ and $\epsilon = 0.05$, so that we want the error covariances to lie below 6.25 for at least 95% of the time. The second row of Table 4.1 shows the arrival rates required for several different smoothing lags

Table 4.1: Arrival rates λ sufficient to satisfy probabilistic constraints (4.6)-(4.7) and simulated probabilities, with $A = 1.3, C = 1, Q = 0.5, R = 1, M = 6.25, \epsilon = 0.05$

$P_{k\|k+N}$	$P_{k\|k-1}$	$P_{k\|k}$	$P_{k\|k+1}$	$P_{k\|k+2}$	$P_{k\|k+3}$	$P_{k\|k+4}$
λ	0.776	0.632	0.527	0.451	0.393	0.348
$\mathbb{P}(P_{k\|k+N} \leq M)$	0.975	0.970	0.963	0.959	0.958	0.959

N obtained using the condition (4.9). The third row contains simulated probabilities $\mathbb{P}(P_{k|k+N} \leq M)$ using the corresponding λ values, where we simulate the system over 100000 time steps. We can see that in each case we have $\mathbb{P}(P_{k|k+N} \leq M) > 0.95$.

4.5 Kalman filtering with retransmissions

In this section we analyze the performance of Kalman filtering using some simple retransmission strategies. The packet losses will again be assumed to be Bernoulli with packet arrival rate λ.

4.5.1 Deterministic retransmission strategy

Consider the following retransmission strategy: If a measurement is lost at time k, ask for retransmission of this measurement up to N times (if the packet is still not received), while the measurements at times $k+1, k+2, \ldots$ are discarded when retransmission is occurring. Note that with this scheme only one packet is transmitted at any time instance and there is no queueing of packets, other than the single packet waiting to be transmitted. Assume that the probability of receiving a retransmitted packet is still λ. We are interested in the Kalman filtering performance (which could be delayed by up to N) of this retransmission scheme.

Recall the sequence $\{\gamma_k\}$ of 0s and 1s that specifies which measurements are lost and received. With retransmission we will look at the sequences of 0s and 1s together with whether the retransmitted packets are successful. To introduce notation, such a sequence for $N = 2$ might look like

$$1 \quad \not{0}^{01} \quad \_ \quad \_ \quad \not{0}^{00} \quad \_ \quad \_ \quad 1 \quad 1 \quad \not{0}^{1} \quad \_ \quad 1 \quad \ldots \tag{4.10}$$

where "$\not{0}^{01}$" represents that the retransmission is successful on the second attempt, "$\not{0}^{00}$" that both retransmissions are not successful, "$\not{0}^{1}$" that the retransmission succeeded on the first attempt and "_" that the measurement is not sent (so will be assumed to be 0), so that (4.10) is equivalent to the sequence $110000011101\ldots$. It is clear that the number of 0s and 1s in both sequences are the same, hence the probabilities of each occurring will be the same.

The key idea in analyzing the performance of this scheme is the following:

Lemma 4.5.1. *Ignoring the first entry,*[2] *there is a bijection between sequences* $\{\gamma_k\}$ *which can be obtained without retransmission and sequences that can be obtained with retransmission. Moreover, they have the same probabilities of occurring.*

Proof. Firstly, for any valid sequence of retransmissions, we will clearly obtain a corresponding $\{\gamma_k\}$ which has the same probability of occurring.

Now given a sequence $\{\gamma_k\}$, the following procedure will allow us to obtain a retransmission sequence which will match up with $\{\gamma_k\}$.

If $\{\gamma_k\}$ starts off with 0, go to step 1), otherwise go to step 2).

Step 1) Let m count the number of 0s before the first 1.

- If $m = 1$ then the first entry for the retransmission sequence is 1. Go to step 2)
- If $2 \leq m \leq N$, the sequence starts with 0, followed by retransmissions with success at the $(m-1)$-th retransmission. Go to step 2)
- If $m > N$, write this as $m = a(N+1) + b$, with a and b being non-negative integers and $b < N + 1$.
  - If $b = 0$, the sequence consists of 0s with all retransmissions failing repeated a times. Go to step 2)
  - If $b = 1$, the sequence consists of a 1, followed by 0s with all retransmissions failing repeated a times. Go to step 2)

[2] *The reason we ignore the first entry is that cases like* $01\ldots$ *cannot be obtained using the retransmission strategy considered here since the second measurement will always be discarded when we ask for retransmission of the first measurement.*

  - Else the sequence is 0 followed by retransmissions with success at the $(b-1)$-th retransmission, followed by 0s with all retransmissions failing repeated a times. Go to step 2)

Step 2) Let n count the number of 0s between two successive 1s.

- If $n = 0$, the next entry in the retransmission sequence is 1. Return to step 2)
- If $0 < n \leq N$, the next entries are a 0, followed by retransmissions with success at the n-th retransmission. Return to step 2)
- If $n > N$, write this as $n = c(N+1) + d$, with c and d being non-negative integers and $d < N + 1$.
  - If $d = 0$, the next entries are a 1, followed by 0s with all retransmissions failing repeated c times. Return to step 2)
  - Else the next entries are a 0 followed by retransmissions with success at the d-th retransmission, followed by 0s with all retransmissions failing repeated c times. Return to step 2)

Following this procedure we can find a retransmission sequence which will match up with any given$\{\gamma_k\}$, and the probabilities of obtaining both sequences are the same. Uniqueness comes from the fact that lost measurements are always retransmitted in this scheme. □

Step 1) in the proof of Lemma 4.5.1 takes care of the situation where there is a possible mismatch in the first entry due to $\{\gamma_k\}$ starting off with a 0. Step 2) of the proof is then repeatedly applied, and here we can always finding a matching set of retransmissions. Table 4.2 shows some simple examples of $\{\gamma_k\}$ and the corresponding retransmission sequences, with $N = 3$. The first three columns involve applications of step 1), and we see that the only possible mismatch in the sequences in the first and third columns is in the first entry. The last three columns will involve step 2), and we can see that the sequences in the fourth and sixth column are matched up. As an example of the procedure in full, consider $\{\gamma_k\} = 0000001000010000001$, and we want to find a retransmission sequence equivalent to this, using $N = 3$. Then

Table 4.2: Some simple $\{\gamma_k\}$ and retransmission sequences, with $N = 3$

$\{\gamma_k\}$	m	retransmission sequence	$\{\gamma_k\}$	n	retransmission sequence
01	1	11	11	0	11
001	2	$\not{0}^{1}$ _ 1	101	1	$\not{0}^{1}$ _ 1
0001	3	$\not{0}^{01}$ _ _ 1	1001	2	$\not{0}^{01}$ _ _ 1
00001	4	$\not{0}^{000}$ _ _ _ 1	10001	3	$\not{0}^{001}$ _ _ _ 1
000001	5	1 $\not{0}^{000}$ _ _ _ 1	100001	4	1 $\not{0}^{000}$ _ _ _ 1
0000001	6	$\not{0}^{1}$ _ $\not{0}^{000}$ _ _ _ 1	1000001	5	$\not{0}^{1}$ _ $\not{0}^{000}$ _ _ _ 1

in step 1) $m = 7 = 1 \times 4 + 3$, so $a = 1$ and $b = 3$. The first time we run step 2), we have $n = 4 = 1 \times 4 + 0$, so $c = 1$ and $d = 0$. The second time we run step 2), we have $n = 5 = 1 \times 4 + 1$, so $c = 1$ and $d = 1$. Following the procedure the retransmission sequence is constructed as

$$\not{0}^{01} \_ \_ \not{0}^{000} \_ \_ \_ 1 \not{0}^{000} \_ \_ \_ \not{0}^{1} \_ \not{0}^{000} \_ \_ \_ 1$$

which agrees with the original $\{\gamma_k\}$ apart from the first entry.

Thus apart from a possible mismatch in the first entry we know that there is always a sequence of retransmissions which will reproduce the same behaviour as a sequence $\{\gamma_k\}$ for the Kalman filter without retransmission, with the same probability of occurring. So asymptotically, the probability distributions when doing Kalman filtering with retransmissions will be the same as that for Kalman filtering with no retransmissions. Hence stability properties are the same as that of (Sinopoli *et al.*, 2004), i.e. the critical thresholds λ_c don't change. The probabilistic behaviour will also be the same as for Kalman filtering with no retransmissions so the bounds in (Shi *et al.*, 2005) will still apply. Therefore this retransmission scheme provides no advantages over filtering using both of the performance measures considered in this correspondence, while at the same time possibly introducing a delay up to N. This agrees somewhat with the intuition that using new measurements for estimation is better than retransmitting old measurements, though here we actually showed that their distributions are essentially the same. Consequently, comparing smoothing with retransmissions we find that Kalman smoothing will outperform this retransmission strategy using the probabilistic notion of performance as discussed in

Section 4.4. We remark that this retransmission scheme and Lemma 4.5.1 can also be applied to other estimators such as the hidden Markov model filter (Chapter 2), so long as the measurement losses are restricted to be an i.i.d. Bernoulli process.

4.5.2 Random retransmission strategy

An extension of the deterministic retransmission strategy is for retransmission requests to be random. Here if a packet is not received, then with probability p it asks for a retransmission independently up to a maximum of N times (if retransmissions were unsuccessful), otherwise it will wait for the next measurement. It turns out that the distribution is again asymptotically the same as that for Kalman filtering without retransmissions, hence the same conclusions on its performance and comparison with Kalman smoothing applies.

To analyze this scheme, let us look at the case of a 1 followed by n successive 0s. Consider for instance $n = 3$ (i.e. the sequence 1000) and $N = 2$. In contrast to the deterministic strategy, there are now seven possible ways in which we can obtain this via retransmissions: $1\ 0^*\ 0^*\ 0^*$, $1\ 0^*\ \not{0}^{0^*}\ \_$, $1\ \not{0}^{0^*}\ \_\ 0^*$, $1\ \not{0}^{\not{0}0}\ \_\ \_$, $\not{0}^1\ \_\ 0^*\ 0^*$, $\not{0}^1\ \_\ \not{0}^{0^*}\ \_$, $\not{0}^{\not{0}1}\ \_\ \_\ 0^*$, where the "$*$" here indicates that we did not ask for a retransmission. The corresponding probabilities of these events (here we're ignoring the $\lambda(1-\lambda)^3$ term which results from receiving a 1 and three 0s) are $(1-p)^3, p(1-p)^2, p(1-p)^2, p^2, p(1-p)^2, p^2(1-p), p^2(1-p)$. However one can check that the sum of these probabilities is equal to one, so that considering all these events as a whole gives a correspondence to the original sequence 1000, similar to Lemma 4.5.1. If we can show that this is true in general, then by independence and similar arguments as in the proof of Lemma 4.5.1 the probability distributions for this scheme will asymptotically be the same as for Kalman filtering without retransmissions.

To do this we first look at the block of 0s by themselves. Let $f(n)$ be the sum of the probabilities of the different ways in which we can get n successive zeroes. Then

$f(0) = 1, f(-1) = f(-2) = \cdots = 0$ and the following recursion holds:

$$\begin{aligned} f(n) = (1-p)f(n-1) + p(1-p)f(n-2) + p^2(1-p)f(n-3) + \dots \\ + p^{N-1}(1-p)f(n-N) + p^N f(n-N-1) \end{aligned} \tag{4.11}$$

This is because the term $(1-p)f(n-1)$ comes from not asking for a retransmission after the first 0, the term $p(1-p)f(n-2)$ comes for asking for retransmission once unsuccessfully, and then deciding not to ask again, and so on with the last term $p^N f(n-(N+1))$ the case where we asked for N retransmissions but all were unsuccessful.

Returning to the case of a 1 followed by n successive 0s, and let $g(n)$ be the sum of the probabilities of the different ways in which we can do this. Then $g(n)$ satisfies the relation

$$g(n) = f(n) + pf(n-1) + p^2 f(n-2) + \cdots + p^N f(n-N) \tag{4.12}$$

The term $f(n)$ represents the situation where we successfully received the first measurement, the other terms represent cases where there is retransmission of the first measurement until a successs, similar to how (4.11) is derived. Substituting (4.11) into (4.12) we find that

$$g(n) = f(n-1) + pf(n-2) + p^2 f(n-3) + \cdots + p^N f(n-N-1)$$

Comparing this with (4.12) we see that $g(n)$ must be constant, and hence $g(n) = \cdots = f(0) = 1$.

4.6 Conclusion

In this chapter we have derived the Kalman smoother in the presence of random packet losses, and compared it against the Kalman filter with packet losses as well as a simple retransmission scheme. We found that the Kalman smoother provided gains over the Kalman filter in the probabilistic sense of (Shi *et al.*, 2005), but not

in terms of the stability notion of (Sinopoli *et al.*, 2004). We also found that the simple retransmission schemes considered here will statistically perform the same as the Kalman filter without retransmission, while also introducing an additional delay. Future work will involve analysis of retransmission strategies where more than one packet can be sent at a time.

Chapter 5

Power Efficient State Estimation Using Multiple Sensors

5.1 Introduction

Previous chapters have analyzed the performance of existing algorithms which have been modified to deal with phenomena such as measurement losses and fading. The focus of this chapter is on estimation algorithms more specifically designed for use in wireless sensor networks, where issues such as power efficiency and medium access are also addressed. Since the sensors will have limited energy and computational ability, which imposes severe constraints on system design, signal processing algorithms which can efficiently utilise these resources are of great interest.

In recent years there has been a considerable literature on estimation and detection schemes for use in wireless sensor networks. Work on detection include (Chamberland & Veeravalli, 2003; Chamberland & Veeravalli, 2004) which studies the asymptotic optimality of using identical sensors in the presence of energy constraints, and (Chen *et al.*, 2004; Jiang & Chen, 2005) which derives fusion rules for distributed detection in the presence of fading. Parameter estimation or estimation of constant signals is studied in e.g. (Ribeiro & Giannakis, 2006; Xiao *et al.*, 2006) where issues of quantization and optimization of power usage are addressed. A hierarchical approach to estimation of fields is considered in (Nowak *et al.*, 2004). Type based methods for detection and estimation of discrete sources are proposed and analyzed in (Mergen & Tong, 2006; Mergen *et al.*, 2007; Liu & Sayeed, 2007).

A promising scheme for distributed estimation in sensor networks is analog forwarding, where measurements from the sensors are transmitted directly (possibly scaled) to the fusion center without any coding, which is motivated by optimal-

ity results on uncoded transmissions in point-to-point links (Goblick, 1965; Gastpar *et al.*, 2003). Other related information theoretic results include (Berger *et al.*, 1996; Viswanathan & Berger, 1997; Ishwar *et al.*, 2005). Analog forwarding schemes are attractive due to their simplicity as well as the possibility of real-time processing since there is no coding delay. In (Gastpar & Vetterli, 2003) the asymptotic (large number of sensors) optimality of analog forwarding for estimating an i.i.d. scalar Gaussian process is shown. Analog forwarding with optimal power allocation is studied in (Xiao *et al.*, 2005) and (Cui *et al.*, 2007) for multi-access and orthogonal schemes respectively. Lower bounds and asymptotic optimality results for estimating independent vector processes, is addressed in (Gastpar & Vetterli, 2005). In (Bahceci & Khandani, 2006) the vector data to be estimated is allowed to be correlated between sensors, and optimal power allocation problems are formulated and solved, though the processes are still *i.i.d. over time*.

Rather than the i.i.d. processes previously considered, in this chapter we address estimation of dynamical systems using analog forwarding of measurements. In particular, we will consider the problem of state estimation of discrete-time scalar linear systems using multiple sensors. As is well-known, optimal state estimation of a linear system can be achieved using a Kalman filter. Other work on Kalman filtering in sensor networks include (Zhang & Li, 2004) which studied optimal sensor data quantization, and (Ribeiro *et al.*, 2006), where Kalman filtering using one bit quantized observations is considered and performance is shown to lie within a constant factor of the standard Kalman filter. Another related area with a long history is that of distributed Kalman filtering, where the main objectives include doing local processing at the individual sensor level to reduce the computations required at the fusion centre (Willner *et al.*, 1976; Hashemipour *et al.*, 1988), or to form estimates at each of the individual sensors in a completely decentralized fashion without any fusion centre (Rao & Durrant-Whyte, 1991). However here we assume that computational resources available at the sensors are limited so that they will only take measurements and then transmit them to the fusion center for further processing, using uncoded analog forwarding.

In this chapter we will consider estimation of scalar linear systems using multiple

sensors communicating to the fusion centre via multi-access or orthogonal medium access schemes. In particular, our focus will be on deriving asymptotic behaviour of the error covariance with respect to the number of sensors for these schemes as well as optimal transmission power allocation to the sensors under a constraint on the error covariance at the fusion center or a sum power constraint at the sensor transmitters. We consider both static and fading channels and in the context of fading channels, we consider various levels of availability of channel state information (CSI) at the transmitters and the fusion center. The chapter is organized as follows. Section 5.2 specifies our model and preliminaries. Section 5.3 gives a number of examples between multi-access and orthogonal access schemes, which show that in general one scheme does not always perform better than the other. We investigate the asymptotic behaviour for a large number of sensors M in Section 5.4, where it is shown that the error covariance decays at the rate $1/M$, and in the multi-access case even where the total power is bounded. Optimal power allocation for static channels is considered in Section 5.5, where we formulate and solve optimization problems for 1) an error covariance constraint and 2) a sum power constraint. Section 5.6 will look at fading channels. In the case where we have channel state information (CSI) we can run the optimization at each time step. When we don't have CSI, we will derive a sub-optimal linear estimator similar to (Nahi, 1969; Rajasekaran *et al.*, 1971; Hadidi & Schwartz, 1979), which can be used for non-zero mean fading. The $1/M$ scaling behaviour is also observed here, and we will also perform optimization using this estimator. Some numerical studies are presented in Section 5.7.

5.2 Models and preliminaries

Throughout this chapter, i represents the sensor index and k represents the time index. Let the linear system be

$$x_{k+1} = Ax_k + w_k$$

with the M sensors each observing

$$y_{i,k} = C_i x_k + v_{i,k}, i = 1, \ldots, M$$

with w_k and $v_{i,k}$ being zero-mean Gaussians having covariances Q and R_i respectively, with the $v_{i,k}$'s being independent between sensors. Note that the sensors can have different observation matrices C_i and measurement noise variances R_i. In this chapter we will restrict ourselves to scalar systems, so that A, C_i, Q, R_i are scalar quantities. For the sake of generality, we allow A and C_i to take on both positive and negative values.[1] It is assumed that the parameters A, C_i, Q and R_i are known. Moreover, we assume that the system is stable, i.e. $|A| < 1$.

5.2.1 Multi-access scheme

In the (non-orthogonal) multi-access scheme the fusion center receives the sum

$$\tilde{z}_k = \sum_{i=1}^{M} \tilde{\alpha}_{i,k} \tilde{h}_{i,k} y_{i,k} + \tilde{n}_k \tag{5.1}$$

where $\tilde{n}_k$ is zero-mean complex Gaussian with variance $2N$, $\tilde{h}_{i,k}$ are the complex-valued channel gains, and $\tilde{\alpha}_{i,k}$ are the complex-valued multiplicative amplification factors in an amplify and forward scheme. See Figure 5.1. We assume that all transmitters have access to their complex channel state information (CSI), and use

$$\tilde{\alpha}_{i,k} = \alpha_{i,k} \frac{\tilde{h}^*_{i,k}}{|\tilde{h}_{i,k}|}$$

where $\alpha_{i,k}$ is real-valued, i.e. we do distributed transmitter beamforming. Calling $h_{i,k} \equiv |\tilde{h}_{i,k}|$, $z_k \equiv \Re[\tilde{z}_k]$, $n_k \equiv \Re[\tilde{n}_k]$, we then have

$$z_k = \sum_{i=1}^{M} \alpha_{i,k} h_{i,k} y_{i,k} + n_k \tag{5.2}$$

[1] This is with the aim of eventually extending our results to vector models, where e.g. A matrices would often contain both positive and negative entries.

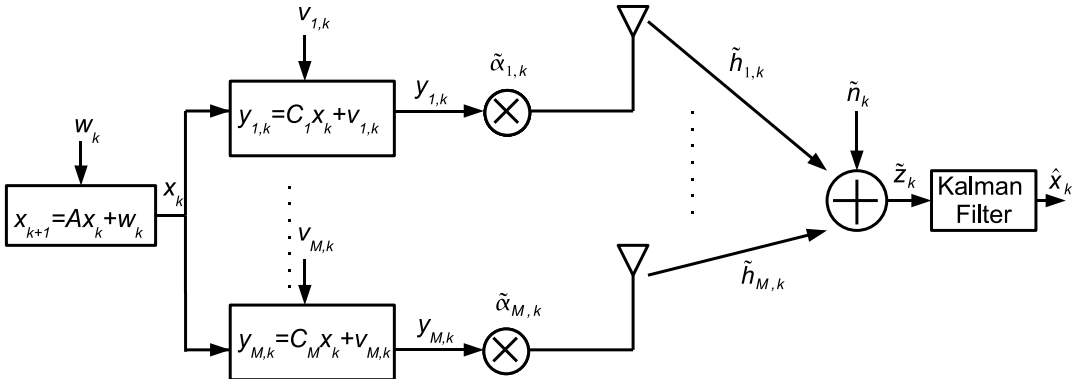

Figure 5.1: Analog Forwarding - Multi-Access

The assumption of CSI at the transmitters is crucial in order for the signals to add up coherently in (5.2), and may not be easy to achieve in large sensor networks.[2]

Continuing further, we may write

$$\begin{aligned} z_k &= \sum_{i=1}^{M} \alpha_{i,k} h_{i,k} C_i x_k + \sum_{i=1}^{M} \alpha_{i,k} h_{i,k} v_{i,k} + n_k \\ &= \bar{C}_k x_k + \bar{v}_k \end{aligned}$$

where $\bar{C}_k \equiv \sum_{i=1}^{M} \alpha_{i,k} h_{i,k} C_i$ and $\bar{v}_k \equiv \sum_{i=1}^{M} \alpha_{i,k} h_{i,k} v_{i,k} + n_k$. Hence, we have the following linear system

$$\begin{aligned} x_{k+1} &= Ax_k + w_k \\ z_k &= \bar{C}_k x_k + \bar{v}_k \end{aligned} \tag{5.3}$$

with $\bar{v}_k$ having variance $\bar{R}_k \equiv \sum_{i=1}^{M} \alpha_{i,k}^2 h_{i,k}^2 R_i + N$. Define the state estimate and error covariance as

$$\begin{aligned} \hat{x}_{k+1|k} &= \mathbb{E}\left[x_{k+1} | \{z_0, \ldots, z_k\}\right] \\ P_{k+1|k} &= \mathbb{E}\left[(x_{k+1} - \hat{x}_{k+1|k})^2 | \{z_0, \ldots, z_k\}\right] \end{aligned}$$

Then it is well known that optimal estimation of the state x_k in the minimum mean

[2]The case where the channel gains are unknown but channel statistics are available is addressed in Section 5.6.2. This can also be used to model the situation where perfect phase synchronization cannot be achieved (Gastpar & Vetterli, 2005).

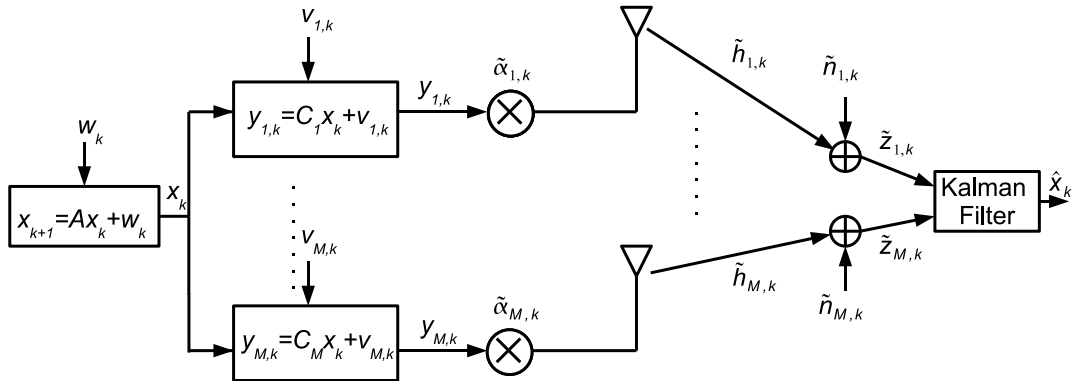

Figure 5.2: Analog Forwarding - Orthogonal Access

squared error sense can be achieved using a (time-varying) Kalman filter (Anderson & Moore, 1979). We recall that even if the noises are non-Gaussian, the Kalman filter is still the best *linear* estimator.

5.2.2 Orthogonal access scheme

In the orthogonal scheme each sensor transmits its measurement to the fusion center via orthogonal channels (e.g. using FDMA or CDMA), so the fusion center receives

$$\tilde{z}_{i,k} = \tilde{\alpha}_{i,k}\tilde{h}_{i,k}y_{i,k} + \tilde{n}_{i,k}, i = 1, \ldots, M$$

with $\tilde{n}_{i,k}$'s independent, zero mean complex Gaussian with variance $2N, \forall i$. See Figure 5.2. We will again assume CSI at the transmitters and use $\tilde{\alpha}_{i,k} = \alpha_{i,k}\frac{\tilde{h}^*_{i,k}}{|\tilde{h}_{i,k}|}, \alpha_{i,k} \in \mathbb{R}$. Let $h_{i,k} \equiv |\tilde{h}_{i,k}|$, $z_{i,k} \equiv \Re[\tilde{z}_{i,k}]$, $n_{i,k} \equiv \Re[\tilde{n}_{i,k}]$. We can then write the situation (using the superscript "o" to distinguish some quantities in the orthogonal scheme from the multi-access scheme) as the linear system

$$x_{k+1} = Ax_k + w_k$$
$$z^o_k = \bar{C}^o_k x_k + \bar{v}^o_k$$

where

$$z_k^o \equiv \begin{bmatrix} z_{1,k} \\ \vdots \\ z_{M,k} \end{bmatrix}, \bar{C}_k^o \equiv \begin{bmatrix} \alpha_{1,k} h_{1,k} C_1 \\ \vdots \\ \alpha_{M,k} h_{M,k} C_M \end{bmatrix}, \bar{v}_k^o \equiv \begin{bmatrix} \alpha_{1,k} h_{1,k} v_{1,k} + n_{1,k} \\ \vdots \\ \alpha_{M,k} h_{M,k} v_{M,k} + n_{M,k} \end{bmatrix}$$

with the covariance of $\bar{v}_k^o$ being

$$\bar{R}_k^o \equiv \begin{bmatrix} \alpha_{1,k}^2 h_{1,k}^2 R_1 + N & 0 & \dots & 0 \\ 0 & \alpha_{2,k}^2 h_{2,k}^2 R_2 + N & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_{M,k}^2 h_{M,k}^2 R_M + N \end{bmatrix}$$

The state estimate and error covariance are now defined as

$$\begin{aligned} \hat{x}_{k+1|k}^o &= \mathbb{E}\left[x_{k+1} | \{z_0^o, \dots, z_k^o\}\right] \\ P_{k+1|k}^o &= \mathbb{E}\left[(x_{k+1} - \hat{x}_{k+1|k}^o)^2 | \{z_0^o, \dots, z_k^o\}\right] \end{aligned}$$

Optimal estimation of x_k in the orthogonal scheme can also be achieved using a Kalman filter. The advantage of the orthogonal scheme is that we do not need carrier-level synchronization among all sensors, but only require synchronization between each individual sensor and the fusion center (Cui *et al.*, 2007).

5.2.3 Transmit powers

For stable scalar systems, it is well known that if X_k is stationary we have

$$\mathbb{E}[X_k^2] = \frac{Q}{1 - A^2}, \forall k.$$

In both the multi-access and orthogonal schemes, the power used at time k by the ith sensor in transmitting its measurement to the fusion center is then

$$p_{i,k} = \alpha_{i,k}^2 \left(C_i^2 \frac{Q}{1 - A^2} + R_i \right)$$

5.2.4 Steady state error covariance

For the moment we will let $\tilde{h}_{i,k} = \tilde{h}_i, \forall k$, and hence $h_{i,k} = h_i, \forall k$, be time-invariant, deferring the discussion of time-varying channels until Section 5.6. For stable systems and time-invariant channels we will also assume that $\alpha_{i,k} = \alpha_i, \forall k$, i.e. the amplification factors don't vary with time, and will drop the subscript k from quantities such as $\bar{C}_k$ and $\bar{R}_k$.

From Kalman filtering theory, we know that the steady state (as $k \to \infty$) error covariance P_∞ (provided it exists) in the multi-access scheme satisfies

$$P_\infty = A[P_\infty - P_\infty \bar{C}^T(\bar{C}P_\infty\bar{C}^T + \bar{R})^{-1}\bar{C}P_\infty]A^T + Q = \frac{A^2 P_\infty \bar{R}}{\bar{C}^2 P_\infty + \bar{R}} + Q$$

where $\bar{R}$ and $\bar{C}$ are the time-invariant versions of $\bar{R}_k$ and $\bar{C}_k$.[3] For stable systems, it is known that the steady state error covariance always exists (Anderson & Moore, 1979, p.77). For $\bar{C} \neq 0$, the solution to this can be easily shown to be

$$P_\infty = \frac{(A^2-1)\bar{R} + \bar{C}^2 Q + \sqrt{((A^2-1)\bar{R} + \bar{C}^2 Q)^2 + 4\bar{C}^2 Q\bar{R}}}{2\bar{C}^2} \tag{5.4}$$

We note that (5.4) can also be written as

$$P_\infty = \frac{(A^2-1) + QS + \sqrt{(A^2 - 1 + QS)^2 + 4QS}}{2S} \tag{5.5}$$

with $S \equiv \bar{C}^2/\bar{R}$ regarded as a signal-to-noise ratio (SNR). We have the following property, which is proved in Appendix 5.9.1. For a similar SNR improvement property of the Kalman filter, see (Anderson & Moore, 1979, p.118-120).

Lemma 5.2.1. *P_∞ is a decreasing function of S*

In the orthogonal access scheme, the steady state error covariance P^o_∞ satisfies

$$P^o_\infty = A[P^o_\infty - P^o_\infty \bar{C}^{oT}(\bar{C}^o P^o_\infty \bar{C}^{oT} + \bar{R}^o)^{-1}\bar{C}^o P^o_\infty]A^T + Q \tag{5.6}$$

[3]The assumption of time-invariance is important. For time-varying $\bar{R}_k$ and $\bar{C}_k$, the error covariance usually will not converge to a steady state value

where $\bar{R}^o$ and $\bar{C}^o$ are the time-invariant versions of $\bar{R}^o_k$ and $\bar{C}^o_k$. We have that P^o_∞ is a scalar, but $\bar{C}^o$ is now a vector and $\bar{R}^o$ a matrix. To simplify the expressions, first apply the matrix inversion lemma to get

$$\begin{aligned}\bar{C}^{o^T}(\bar{C}^o P^o_\infty \bar{C}^{o^T} + \bar{R}^o)^{-1}\bar{C}^o &= \bar{C}^{o^T}\left[\bar{R}^{o^{-1}} - \bar{R}^{o^{-1}}\bar{C}^o(P^{o^{-1}}_\infty + \bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o)^{-1}\bar{C}^{o^T}\bar{R}^{o^{-1}}\right]\bar{C}^o \\ &= \bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o - \frac{(\bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o)^2}{1/P^o_\infty + \bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o} \\ &= \frac{\bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o}{1 + P^o_\infty \bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o}\end{aligned} \tag{5.7}$$

and where we can easily compute $\bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o$ as

$$S^o \equiv \bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o = \sum_{i=1}^{M}\frac{\alpha_i^2 h_i^2 C_i^2}{\alpha_i^2 h_i^2 R_i + N}$$

with S^o regarded as a signal-to-noise ratio. Then the solution to (5.6) can be computed as

$$P^o_\infty = \frac{A^2 - 1 + QS^o + \sqrt{(A^2 - 1 + QS^o)^2 + 4QS^o}}{2S^o} \tag{5.8}$$

The error covariance P^o_∞ is a decreasing function of S^o using the same proof as in Appendix 5.9.1.

Comparing (5.5) and (5.8) we see that the functions for P_∞ and P^o_∞ are of the same form, except that in the multi-access case we have

$$S \equiv \frac{\bar{C}^2}{\bar{R}} = \frac{\left(\sum_{i=1}^{M}\alpha_i h_i C_i\right)^2}{\sum_{i=1}^{M}\alpha_i^2 h_i^2 R_i + N}$$

and in the orthogonal case we have

$$S^o \equiv \bar{C}^{o^T}\bar{R}^{o^{-1}}\bar{C}^o = \sum_{i=1}^{M}\frac{\alpha_i^2 h_i^2 C_i^2}{\alpha_i^2 h_i^2 R_i + N}$$

5.3 Examples of multi-access vs orthogonal access

Now one might ask the question as to whether one scheme is always better the other, e.g. whether $S \geq S^o$ given the same values for $\alpha_i, h_i, C_i, R_i, N$ are used in both expressions. However in general this is not true. We present below a number of examples showing different behaviour in different situations. Assume for simplicity that the α_i's are chosen such $\alpha_i C_i$ are positive for all $i = 1, \ldots, M$.

1) Consider first the case when $N = 0$. Then we have the inequality

$$\sum_{i=1}^{M} \frac{\alpha_i^2 h_i^2 C_i^2}{\alpha_i^2 h_i^2 R_i} \geq \frac{\left(\sum_{i=1}^{M} \alpha_i h_i C_i\right)^2}{\sum_{i=1}^{M} \alpha_i^2 h_i^2 R_i}$$

which can be shown by applying Theorem 65 of (Hardy *et al.*, 1952). So when $N = 0$, $S^o \geq S$ and consequently P_∞^o will be smaller than P_∞. The intuitive explanation for this is that if there is no noise introduced at the fusion center, then receiving the individual measurements from the sensors is better than receiving a linear combination of the measurements.

2) Next we consider the case when the noise variance N is large. We write

$$\begin{aligned}
&S - S^o \\
&= \frac{\left(\sum_{i=1}^{M} \alpha_i h_i C_i\right)^2}{\sum_{i=1}^{M} \alpha_i^2 h_i^2 R_i + N} - \sum_{i=1}^{M} \frac{\alpha_i^2 h_i^2 C_i^2}{\alpha_i^2 h_i^2 R_i + N} \\
&= \frac{1}{\left(\sum_{i=1}^{M} \alpha_i^2 h_i^2 R_i + N\right) \prod_{i=1}^{M} (\alpha_i^2 h_i^2 R_i + N)} \left(\left(\sum_{i=1}^{M} \alpha_i h_i C_i \right)^2 \prod_{i=1}^{M} (\alpha_i^2 h_i^2 R_i + N) \right. \\
&\quad - \alpha_1^2 h_1^2 C_1^2 \left(\sum_{i=1}^{M} \alpha_i^2 h_i^2 R_i + N \right) \prod_{i:i\neq 1} (\alpha_i^2 h_i^2 R_i + N) - \ldots \\
&\quad \left. - \alpha_M^2 h_M^2 C_M^2 \left(\sum_{i=1}^{M} \alpha_i^2 h_i^2 R_i + N \right) \prod_{i:i\neq M} (\alpha_i^2 h_i^2 R_i + N) \right).
\end{aligned}$$

The coefficient of the N^M term in the numerator is

$$\left(\sum_{i=1}^{M} \alpha_i h_i C_i \right)^2 - \alpha_1^2 h_1^2 C_1^2 - \cdots - \alpha_M^2 h_M^2 C_M^2 > 0.$$

For N sufficiently large, this term will dominate, hence $S > S^o$ and the multi-access case will now have smaller error covariance than the orthogonal case.

3) Now look at the "symmetric" situation where $\alpha_i = \alpha, C_i = C, R_i = R, h_i = h, \forall i$. Then we have

$$S = \frac{M^2\alpha^2h^2C^2}{M\alpha^2h^2R + N} = \frac{M\alpha^2h^2C^2}{\alpha^2h^2R + N/M}$$
$$S^o = \frac{M\alpha^2h^2C^2}{\alpha^2h^2R + N}$$

so $S \geq S^o$ with equality only when $N = 0$ (or $M = 1$). Thus in the symmetric case, multi-access outperforms orthogonal access.

4) Suppose $N \neq 0$. In the multi-access case we have a sum of measurements plus noise introduced at the fusion center, whereas in the orthogonal case the fusion center adds noise onto each of the individual measurements. Thus a reasonable question to ask is whether one can always achieve $S > S^o$ for M sufficiently large. However the answer is no as the following counterexample illustrates. Let $\alpha_i = 1, h_i = 1, R_i = 1, \forall i$. Let $M/2$ of the sensors have $C_i = 1$, and the other $M/2$ sensors have $C_i = 2$. We find that

$$S = \frac{(M/2 + M)^2}{M + N} = \frac{9}{4}\frac{M}{1 + N/M}$$
$$S^o = \frac{M}{2}\frac{1 + 4}{1 + N} = \frac{5}{2}\frac{M}{1 + N}$$

If say $N = 1/8$, then it may be verified that $S^o > S$ for $M < 10$, $S^o = S$ for $M = 10$, and $S > S^o$ for $M > 10$, so eventually the multi-access scheme outperforms the orthogonal scheme. On the other hand, if

$$\frac{5}{2(1 + N)} > \frac{9}{4}$$

or $N < 1/9$, we will have $S^o > S$ no matter how large M is.

Similar to example 1), the possible intuitive explanation for this is that even if the individual measurements have a very small amount of noise added at the fusion center, receiving these individual measurements from the sensors may still be better

than receiving a linear combination of the measurements.

5.4 Asymptotic behaviour

Since P_∞ is a decreasing function of S (similar comments apply for the orthogonal case), increasing S will provide an improvement in performance. As $S \to \infty$, we can see from (5.5) that $P_\infty \to Q$, the process noise variance. Note that unlike e.g. (Gastpar & Vetterli, 2003; Cui *et al.*, 2007) where the mean squared error (MSE) can be driven to zero in some situations such as when there is a large number of sensors, here the lower bound Q on performance is always strictly greater than zero. Now if the number of sensors is fixed, then it is not too difficult to show that S will be bounded no matter how large (or small) one makes the α_i's, so getting arbitrarily close to Q is not possible. On the other hand, if instead the number of sensors M is allowed to increase, then $P_\infty \to Q$ as $M \to \infty$ can be achieved. Moreover we will be interested in the rate at which this convergence occurs.

In this section we will investigate two simple strategies, 1) $\alpha_i = 1, \forall i$, and 2) $\alpha_i = 1/\sqrt{M}, \forall i$.[4] We do this first for the "symmetric" case (i.e. the parameters are the same for each sensor) where we can obtain explicit asymptotic expressions. We then use these results to bound the performance in the general asymmetric case.

We remark that the results in this section assume that large M is possible, e.g. ability to synchronize a large number of sensors in the multi-access scheme, or the availability of a large number of non-overlapping frequency bands in FDMA or a large number of orthogonal spreading sequences in CDMA in the orthogonal scheme, which may not always be possible in practice. However, in numerical investigations we have found that the results derived in this section are quite accurate even for 20-30 sensors, see Section 5.7

[4]These strategies are related to the case of "equal power constraint" and "total power constraint" in (Liu *et al.*, 2007), and various versions have also been considered in the work of (Gastpar & Vetterli, 2003; Gastpar & Vetterli, 2005; Xiao *et al.*, 2005; Cui *et al.*, 2007), in the context of estimation of i.i.d. processes.

5.4.1 No scaling: $\alpha_i = 1, \forall i$

Let $\alpha_i = 1, \forall i$, so measurements are forwarded to the fusion center without any scaling. Assume for simplicity the symmetric case, where $C_i = C, R_i = R, h_i = h, \forall i$.

In the multi-access scheme, $\bar{C} = MhC$ and $\bar{v}_k$ has variance $\bar{R} = Mh^2R + N$. Substituting into (5.4)

$$\begin{aligned} P_\infty =& \frac{(A^2-1)(Mh^2R+N) + M^2h^2C^2Q}{2M^2h^2C^2} \\ &+ \frac{\sqrt{((A^2-1)(Mh^2R+N) + M^2h^2C^2Q)^2 + 4M^2h^2C^2Q(Mh^2R+N)}}{2M^2h^2C^2}. \end{aligned}$$

We are interested in the behaviour of this as $M \to \infty$. Now

$$\begin{aligned} &\sqrt{((A^2-1)(Mh^2R+N) + M^2h^2C^2Q)^2 + 4M^2h^2C^2Q(Mh^2R+N)} \\ &\quad = \left(h^4C^4Q^2M^4 + 2(A^2-1)Rh^4C^2QM^3 + 4h^4C^2QRM^3 + O(M^2)\right)^{1/2} \\ &\quad = h^2C^2QM^2\left(1 + \frac{2(A^2+1)R}{C^2QM} + O\left(\frac{1}{M^2}\right)\right)^{1/2} \\ &\quad = h^2C^2QM^2\left(1 + \frac{1}{2}\frac{2(A^2+1)R}{C^2QM} + O\left(\frac{1}{M^2}\right)\right) \\ &\quad = h^2C^2QM^2 + (A^2+1)h^2RM + O(1) \end{aligned} \tag{5.9}$$

where we have used the expansion

$$(1+x)^{1/2} = 1 + \frac{x}{2} + O(x^2)$$

for $|x| < 1$ (Abramowitz & Stegun, 1965, p.15), which is valid when M is sufficiently large. Hence

$$P_\infty = Q + \frac{A^2R}{C^2}\frac{1}{M} + O\left(\frac{1}{M^2}\right)$$

So in this case the steady state error covariance for the multi-access scheme converges to the process noise variance Q, at a rate of $1/M$. This result matches the rate of $1/M$ achieved for estimation of i.i.d. processes using multi-access schemes, e.g. (Gastpar & Vetterli, 2003; Liu *et al.*, 2007).

In the orthogonal scheme we have

$$S^o = \frac{Mh^2C^2}{h^2R+N}$$

and

$$\begin{aligned} P_\infty^o =& \frac{(A^2-1)(h^2R+N)+Mh^2C^2Q}{2Mh^2C^2} \\ &+ \frac{\sqrt{((A^2-1)(h^2R+N)+Mh^2C^2Q)^2+4Mh^2C^2Q(h^2R+N)}}{2Mh^2C^2}. \end{aligned}$$

By similar calculations to (5.9) we find that

$$P_\infty^o = Q + \frac{A^2(h^2R+N)}{h^2C^2}\frac{1}{M} + O\left(\frac{1}{M^2}\right) = Q + \frac{A^2(R+N/h^2)}{C^2}\frac{1}{M} + O\left(\frac{1}{M^2}\right).$$

The steady state error covariance again converges to Q at a rate of $1/M$, but the constant $\frac{A^2(R+N/h^2)}{C^2}$ in front is larger. This agrees with example 3) of Section 5.3 that in the symmetric situation multi-access will perform better than orthogonal.

5.4.2 Scaling $\alpha_i = 1/\sqrt{M}, \forall i$

In the previous case with $\alpha_i = 1, \forall i$, the power received at the fusion center will grow unbounded as $M \to \infty$. Suppose instead we let $\alpha_i = 1/\sqrt{M}, \forall i$, which will keep the power received at the fusion center bounded (and is constant in the symmetric case), while the transmit power used by each sensor will tend to zero as $M \to \infty$. Again assume for simplicity that $C_i = C, R_i = R, h_i = h, \forall i$.

In the multi-access scheme we now have $\bar{C} = \sqrt{M}hC$ and $\bar{R} = h^2R + N$, and

$$S = \frac{Mh^2C^2}{h^2R+N},$$

so that

$$P_\infty = Q + \frac{A^2(R+N/h^2)}{C^2}\frac{1}{M} + O\left(\frac{1}{M^2}\right).$$

So we again have the steady state error covariance converging to the process noise variance Q at a rate of $1/M$. In fact, we see that this is the same expression

as in the orthogonal case but where we were using $\alpha_i = 1, \forall i$. The advantage here is that the transmit power used by each individual sensor can decrease to zero as the number of sensors increases, which is possibly more desirable in power constrained environments such as wireless sensor networks. For i.i.d. processes, this $1/M$ behaviour when the total received power is bounded has also been observed (Gastpar & Vetterli, 2005; Liu *et al.*, 2007).

In the orthogonal scheme we have

$$P_\infty^o = \frac{(A^2-1)(h^2R/M+N)+h^2C^2Q}{2h^2C^2} + \frac{\sqrt{((A^2-1)(h^2R/M+N)+h^2C^2Q)^2+4h^2C^2Q(h^2R/M+N)}}{2h^2C^2}.$$

One can show by similar computations to (5.9) that for large M,

$$P_\infty^o = \frac{(A^2-1)N+h^2C^2Q+\sqrt{(A^2-1)^2N^2+2(A^2+1)Nh^2C^2Q+h^4C^4Q^2}}{2h^2C^2} + \left[\frac{(A^2-1)R}{2C^2} + \frac{(A^2+1)h^4RC^2Q+(A^2-1)^2Nh^2R}{2h^2C^2\sqrt{(A^2-1)^2N^2+2(A^2+1)Nh^2C^2Q+h^4C^4Q^2}}\right]\frac{1}{M} + O\left(\frac{1}{M^2}\right).$$

Note that

$$\frac{(A^2-1)N+h^2C^2Q+\sqrt{(A^2-1)^2N^2+2(A^2+1)Nh^2C^2Q+h^4C^4Q^2}}{2h^2C^2} > Q$$

so that the steady state error covariance converges as $M \to \infty$ to a value strictly greater than Q, though the convergence is still at a rate $1/M$. An easier way to see that convergence of P_∞^o to Q cannot be achieved is to note that here

$$S^o = \frac{h^2C^2}{h^2R/M+N}$$

which is bounded even as $M \to \infty$. Another observation is that this is also the behaviour which can be achieved in the multi-access case but with a scaling $\alpha_i = 1/M$ (rather than $\alpha_i = 1/\sqrt{M}$), as the reader may verify. Analogously, for i.i.d. processes it has been shown that in the orthogonal case the MSE does not go to zero

as $M \to \infty$ when the total power used is bounded (Cui *et al.*, 2007).

5.4.3 General parameters

The behaviour shown in the two previous cases can still hold under more general conditions on C_i, R_i and h_i. Suppose for instance that they can be bounded from both above and below, i.e.

$$0 < C_{min} \leq |C_i| \leq C_{max} < \infty, \forall i$$
$$0 < R_{min} \leq R_i \leq R_{max} < \infty, \forall i$$
$$0 < h_{min} \leq h_i \leq h_{max} < \infty, \forall i$$

We first treat the multi-access scheme. We have

$$Mh_{min}C_{min} \leq \sum_{i=1}^{M} h_i C_i \leq Mh_{max}C_{max}$$

and

$$Mh_{min}^2 R_{min} \leq \sum_{i=1}^{M} h_i^2 R_i \leq Mh_{max}^2 R_{max}$$

Recall from Lemma 5.2.1 that P_∞ is a decreasing function of $S = \bar{C}^2/\bar{R}$. If we choose $\alpha_i \in \{+1, -1\}$ such that $\alpha_i C_i$ is positive for all i, we have

$$\frac{Mh_{min}^2 R_{min} + N}{M^2 h_{max}^2 C_{max}^2} \leq \frac{\bar{R}}{\bar{C}^2} \leq \frac{Mh_{max}^2 R_{max} + N}{M^2 h_{min}^2 C_{min}^2}$$

and by a similar calculation to (5.9) we can show that as $M \to \infty$

$$Q + \frac{A^2 h_{min}^2 R_{min}}{h_{max}^2 C_{max}^2}\frac{1}{M} + O\left(\frac{1}{M^2}\right) \leq P_\infty \leq Q + \frac{A^2 h_{max}^2 R_{max}}{h_{min}^2 C_{min}^2}\frac{1}{M} + O\left(\frac{1}{M^2}\right)$$

If instead we choose $\alpha_i \in \{1/\sqrt{M}, -1/\sqrt{M}\}$ such that $\alpha_i C_i$ is positive for all i, we have

$$\frac{h_{min}^2 R_{min} + N}{Mh_{max}^2 C_{max}^2} \leq \frac{\bar{R}}{\bar{C}^2} \leq \frac{h_{max}^2 R_{max} + N}{Mh_{min}^2 C_{min}^2}$$

and we can show that as $M \to \infty$

$$Q+\frac{A^2(h_{min}^2 R_{min}+N)}{h_{max}^2 C_{max}^2}\frac{1}{M}+O\left(\frac{1}{M^2}\right) \leq P_\infty \leq Q+\frac{A^2(h_{max}^2 R_{max}+N)}{h_{min}^2 C_{min}^2}\frac{1}{M}+O\left(\frac{1}{M^2}\right)$$

In either case, as the upper and lower bounds both converge to Q at a rate of $1/M$, P_∞ itself will also do so.

For the orthogonal scheme, a similar argument to the above shows that choosing $\alpha_i \in \{+1, -1\}$ gives convergence of P_∞^o to Q at the rate $1/M$ for general parameters. However, if we choose $\alpha_i \in \{1/\sqrt{M}, -1/\sqrt{M}\}$, then P_∞^o does not converge to a limit as $M \to \infty$ for general parameters. For instance, suppose there are two distinct sets of parameters, such that if all the sensors had the first set of parameters the error covariance would converge to $P_{\infty,1}^o$, and if all the sensors had the second set of parameters the error covariance would converge to $P_{\infty,2}^o$, with $P_{\infty,2}^o \neq P_{\infty,1}^o$. Suppose the first M_1 sensors have the first set of parameters, the next M_2 (with $M_2 >> M_1$) sensors the second set, the next M_3 (with $M_3 >> M_2$) sensors the first set, the next M_4 (with $M_4 >> M_3$) sensors the second set, etc... Then P_∞^o would alternate between approaching $P_{\infty,1}^o$ and $P_{\infty,2}^o$, and will not converge to a limit as $M \to \infty$.

5.4.4 Equal power allocation

When the parameters are asymmetric, the above rules in general will allocate different powers to the individual sensors. An alternative is to use equal power allocation. Recall that the transmit power used by each sensor is

$$p_i = \alpha_i^2\left(C_i^2\frac{Q}{1-A^2}+R_i\right)$$

If we allocate power p to each sensor, i.e. $p_i = p, \forall i$, then

$$\alpha_i = \sqrt{\frac{p(1-A^2)}{C_i^2 Q + R_i(1-A^2)}} \tag{5.10}$$

If instead the total power p_{total} is to be shared equally amongst sensors, then $p_i = p_{total}/M, \forall i$, and

$$\alpha_i = \sqrt{\frac{p_{total}(1-A^2)}{M\left(C_i^2 Q + R_i(1-A^2)\right)}} \tag{5.11}$$

Asymptotic behaviour of the different cases using equal power allocation will be similar to Section 5.4.3, and can be treated using similar arguments there.

5.4.5 Remarks

1) Most of the previous rules in this section give a convergence rate of $1/M$. We might wonder whether we can achieve an even better rate (e.g. $1/M^2$) using other choices for α_i. The answer is no. Following (Gastpar & Vetterli, 2003), consider the "ideal" case where sensor measurements are received perfectly at the fusion center, and which mathematically corresponds to the orthogonal scheme with $N = 0, \alpha_i = 1, h_i = 1, \forall i$. This idealized situation provides a lower bound on the achievable error covariance. We will have

$$S^o = \sum_{i=1}^{M} \frac{C_i^2}{R_i}$$

which can be used to show that P_∞^o converges to Q at the rate $1/M$ by a similar argument to Section 5.4.3. Hence $1/M$ is the best rate that can be achieved using any coded/uncoded scheme.

2) In the previous derivations we have not actually used the assumption that $|A| < 1$, so the results in Sections 5.4.1 - 5.4.3 will hold even when the system is unstable. However for unstable systems, $\mathbb{E}[X_k^2]$ becomes unbounded as $k \to \infty$, so if the $\alpha_{i,k}$'s are time invariant, then more and more power is used by the sensors as time passes. If the application is a wireless sensor network where power is limited, then the question is whether one can choose these $\alpha_{i,k}$'s such that *both* the power used by the sensors and the error covariances will be bounded for all times. Now if there is no noise at the fusion center, i.e. $n_k = 0$ then a simple scaling of the measurements at the individual sensors will work. But when $n_k \neq 0$, as will usually be the case in analog forwarding, we have not been able to find a scheme which can achieve this for unstable systems. However in practice, in most cases, we will

be interested in finite horizon results for unstable systems where the system states and measurements can take on large values but are still bounded. In such finite horizon situations, one can perform optimum power allocation either at each time step similar to Section 5.6 for a finite number of time steps, or use a finite horizon dynamic programming approach. These problems will be considered in future work.

5.5 Optimal power allocation

When there are a large number of sensors, one can use simple strategies such as $\alpha_i = 1/\sqrt{M}, \forall i$, or the equal power allocation (5.11), which will both give a convergence of the steady state error covariance to Q at a rate of $1/M$ (in the multi-access scheme), while bounding the total power used by all the sensors. But when the number of sensors is small, one may perhaps do better with different choices of the α_i's. In this section we will address two relevant optimization problems.

5.5.1 Multi-access

Minimizing sum power

One possible formulation is to minimize the sum of transmission powers used by the sensors subject to a bound on the steady state error covariance. More formally, the problem is

$$\min \sum_{i=1}^{M} p_i = \sum_{i=1}^{M} \alpha_i^2 \left(\frac{C_i^2 Q}{1 - A^2} + R_i \right)$$
$$\text{subject to } P_\infty \leq D$$

with P_∞ given by (5.4).

Some straightforward manipulations show that the constraint can be simplified to

$$\bar{R} \left(A^2 D + Q - D \right) + \bar{C}^2 D(Q - D) \leq 0 \tag{5.12}$$

i.e.

$$\left(\sum_{i=1}^{M}\alpha_i^2 h_i^2 R_i + N\right)\left(A^2 D + Q - D\right) + \left(\sum_{i=1}^{M}\alpha_i h_i C_i\right)^2 D(Q-D) \leq 0$$

Now define $s = h_1 C_1 \alpha_1 + \cdots + h_M C_M \alpha_M$. Then the optimization problem becomes

$$\begin{aligned} &\min_{\alpha_1,\ldots,\alpha_M,s} \sum_{i=1}^{M}\alpha_i^2\left(\frac{C_i^2 Q}{1-A^2} + R_i\right) \\ &\text{subject to } \left(\sum_{i=1}^{M}\alpha_i^2 h_i^2 R_i + N\right)\left(A^2 D + Q - D\right) \leq s^2 D(D-Q) \\ &\qquad s = \sum_{i=1}^{M} h_i C_i \alpha_i \end{aligned} \tag{5.13}$$

Before going further, we determine some upper and lower bounds on D. From Section 5.4, a lower bound is $D \geq Q$, where Q is the process noise variance. For an upper bound, suppose $\bar{C} = 0$ so we don't have any information about x_k. Since we are assuming the system is stable, one can still achieve an error covariance of $\frac{Q}{1-A^2}$ (just let $\hat{x}_k = 0, \forall k$), so $D \leq \frac{Q}{1-A^2}$. Hence in problem (5.13) both $D - Q \geq 0$ and $A^2 D + Q - D \geq 0$.

To reduce the amount of repetition in later sections, consider the slightly more general problem

$$\begin{aligned} &\min_{\alpha_1,\ldots,\alpha_M,s} \sum_{i=1}^{M}\alpha_i^2\gamma_i \\ &\text{subject to } \left(\sum_{i=1}^{M}\alpha_i^2 b_i + N\right) x \leq s^2 y \\ &\qquad s = \sum_{i=1}^{M}\alpha_i a_i \end{aligned} \tag{5.14}$$

where $x > 0, y > 0, \gamma_i > 0, a_i \in \mathbb{R}, b_i > 0, i = 1, \ldots, M$ are constants. In the context of (5.13), $x = A^2 D + Q - D$, $y = D(D-Q)$, $a_i = h_i C_i$, $b_i = h_i^2 R_i$ and $\gamma_i = \frac{C_i^2 Q}{1-A^2} + R_i$ for $i = 1, \ldots, M$.

The objective function of problem (5.14) is clearly convex. We can divide the feasible region into two regions corresponding to $s > 0$ and $s < 0$. Then in each of

the two regions, the function

$$\frac{\sum_{i=1}^{M} \alpha_i^2 b_i + N}{s^2}$$

is convex, by noting that each of the functions α_i^2/s^2 is convex. Hence the two regions corresponding to $s > 0$ and $s < 0$ are both convex and the global solution can be easily found numerically. Moreover, following similar steps to (Xiao *et al.*, 2005), a solution in (mostly) closed form can be obtained. Some details on the derivations are in Appendix 5.9.2. Below we shall summarise what is required. One first solves numerically for λ the equation

$$\sum_{i=1}^{M} \frac{\lambda a_i^2}{\gamma_i + \lambda b_i x} = \frac{1}{y}$$

Since the left hand side is increasing with λ solutions to this equation will be unique provided it exists. Taking limits as $\lambda \to \infty$, we see that a solution exists if and only if

$$\sum_{i=1}^{M} \frac{a_i^2}{b_i} > \frac{x}{y} \tag{5.15}$$

Equation (5.15) thus also provides a feasibility check for the optimization problem (5.14). In the context of (5.13), one can easily derive that (5.15) implies

$$\sum_{i=1}^{M} \frac{C_i^2}{R_i} > \frac{A^2 D + Q - D}{D(D-Q)}$$

which indicates that the sum of the sensor signal to noise ratios must be greater than a threshold (dependent on the error covariance threshold D) for the optimization problem (5.13) to be feasible.

Next, we compute μ from

$$\mu^2 = Nx \left(\sum_{i=1}^{M} \frac{a_i^2 \gamma_i}{4\lambda(\gamma_i + \lambda b_i x)^2} \right)^{-1}$$

Finally we obtain the optimal α_i's (denoted by α_i^*)

$$\alpha_i^* = \frac{\mu a_i}{2(\gamma_i + \lambda b_i x)}, i = 1, \ldots, M.$$

with the resulting powers

$$p_i = \alpha_i^{*2}\left(C_i^2\frac{Q}{1-A^2}+R_i\right), i = 1,\ldots,M$$

Note that depending on whether we choose μ to be positive or negative, two different sets of α_i^*'s will be obtained, one of which is the negative of the other, though the p_i's and hence the optimal value of the objective function remains the same.

Minimizing error covariance

A related problem is to minimize the steady state error covariance subject to a sum power constraint. Formally, this is

$$\begin{aligned}&\min P_\infty\\ &\text{subject to } \sum_{i=1}^{M}\alpha_i^2\left(\frac{C_i^2Q}{1-A^2}+R_i\right)\le p_{total}\end{aligned}$$

with P_∞ again given by (5.4). For this problem, the feasible region is clearly convex, but the objective function is complicated. To simplify the objective, recall from Lemma 5.2.1 that P_∞ is a decreasing function of $S = \bar{C}^2/\bar{R}$. Thus maximizing $\bar{C}^2/\bar{R}$ (or minimizing $\bar{R}/\bar{C}^2$) is equivalent to minimizing P_∞. If we regard $\bar{C}^2/\bar{R}$ as a signal-to-noise ratio, then this result has the interpretation that maximizing the SNR minimizes P_∞. Hence the problem is equivalent to

$$\begin{aligned}&\min_{\alpha_1,\ldots,\alpha_M,s}\frac{\sum_{i=1}^{M}\alpha_i^2h_i^2R_i+N}{s^2}\\ &\text{subject to } \sum_{i=1}^{M}\alpha_i^2\left(\frac{C_i^2Q}{1-A^2}+R_i\right)\le p_{total}\\ &s=\sum_{i=1}^{M}h_iC_i\alpha_i\end{aligned}$$

We again introduce a more general problem

$$\begin{aligned} \min_{\alpha_1,\ldots,\alpha_M,s} \quad & \frac{\sum_{i=1}^{M} \alpha_i^2 b_i + N}{s^2} \\ \text{subject to} \quad & \sum_{i=1}^{M} \alpha_i^2 \gamma_i \leq p_{total} \\ & s = \sum_{i=1}^{M} \alpha_i a_i \end{aligned} \tag{5.16}$$

with $x > 0, y > 0, \gamma_i > 0, a_i \in \mathbb{R}, b_i > 0, i = 1, \ldots, M$ being constants. Arguing as in the paragraph after (5.14), the objective function will be convex in each of the two regions $s > 0$ and $s < 0$, so we will again obtain two sets of solutions. Unlike problem (5.14), we have not been able to obtain an analytical solution here, though numerical solutions can still be found easily since the problem is convex.

5.5.2 Orthogonal access

Minimizing sum power

The corresponding problem in the orthogonal scheme is

$$\begin{aligned} \min \sum_{i=1}^{M} p_i &= \sum_{i=1}^{M} \alpha_i^2 \left(\frac{C_i^2 Q}{1 - A^2} + R_i \right) \\ \text{subject to } & P_\infty^o \leq D \end{aligned}$$

with P_∞^o now given by (5.8). By a rearrangement of the constraint, this can be shown to be equivalent to

$$\begin{aligned} \min_{\alpha_1,\ldots,\alpha_M} & \sum_{i=1}^{M} \alpha_i^2 \left(\frac{C_i^2 Q}{1 - A^2} + R_i \right) \\ \text{subject to } & \sum_{i=1}^{M} \frac{\alpha_i^2 h_i^2 C_i^2}{\alpha_i^2 h_i^2 R_i + N} \geq \frac{A^2 D + Q - D}{D(D - Q)} \end{aligned} \tag{5.17}$$

Since each of the functions

$$\frac{-\alpha_i^2 h_i^2 C_i^2}{\alpha_i^2 h_i^2 R_i + N}$$

is convex for $\alpha_i > 0$ (also for $\alpha_i < 0$), this problem is convex. Note that without further restrictions on α_i we will get 2^M solutions with the same values of the objective function, corresponding to the different choices of positive and negative signs on the α_i's. This is in contrast to the multi-access case where there were two sets of solutions. For simplicity we can take the solution corresponding to all $\alpha_i \geq 0$.[5]

An analytical solution can also be obtained. To reduce repetition in later sections, consider the more general problem

$$\begin{aligned} &\min_{\alpha_1,\ldots,\alpha_M} \sum_{i=1}^{M} \alpha_i^2 \gamma_i \\ &\text{subject to } \sum_{i=1}^{M} \frac{\alpha_i^2 a_i^2}{\alpha_i^2 b_i + N} \geq \frac{x}{y} \end{aligned} \tag{5.18}$$

where $x > 0, y > 0, \gamma_i > 0, a_i \in \mathbb{R}, b_i > 0, i = 1, \ldots, M$ are constants and γ_i, x, y, a_i, b_i have identical interpretations as in Section 5.5.1. Since the derivation of the analytical solution is similar to that found in (Cui *et al.*, 2007) (though what they regard as α_k is α_i^2 here), it will be omitted and we will only present the solution.

Firstly, the problem will be feasible if and only if

$$\sum_{i=1}^{M} \frac{a_i^2}{b_i} > \frac{x}{y}$$

Interestingly, this is the same condition for feasibility (5.15) in the multi-access case, problem (5.14), indicating that the total SNR for the sensor measurements must be greater than a certain threshold (dependent on D). Now assume that the sensors are ordered so that

$$\frac{a_1^2}{\gamma_1} \geq \cdots \geq \frac{a_M^2}{\gamma_M}$$

[5] In general this is not possible in the multi-access case. For instance if we have two sensors with C_1 being positive and C_2 negative, the optimal solution will involve α_1 being positive and α_2 negative, or vice versa. Restricting both α_i's to be positive in the multi-access case will result in a sub-optimal solution.

Note that in the context of problem (5.17),

$$\frac{a_i^2}{\gamma_i} = \frac{h_i^2}{Q/(1-A^2) + R_i/C_i^2}.$$

Clearly, this ordering favours the sensors with better channels and higher measurement quality. Then the optimal values of α_i^2 (and hence α_i^*) is given by

$$\alpha_i^{*2} = \begin{cases} \frac{1}{b_i}(\sqrt{\frac{\lambda a_i^2 N}{\gamma_i}} - N) & , \quad i \leq M_1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

where

$$\sqrt{\lambda} = \frac{\sum_{i=1}^{M_1} \frac{|a_i|}{b_i}\sqrt{\gamma_i N}}{\sum_{i=1}^{M_1} \frac{a_i^2}{b_i} - \frac{x}{y}}$$

and the number of sensors which are active, M_1 (which can be shown to be unique (Xiao *et al.*, 2006)), satisfies $\sum_{i=1}^{M_1} \frac{a_i^2}{b_i} - \frac{x}{y} > 0$,

$$\frac{\sum_{i=1}^{M_1} \frac{|a_i|}{b_i}\sqrt{\gamma_i N}}{\sum_{i=1}^{M_1} \frac{a_i^2}{b_i} - \frac{x}{y}} \sqrt{\frac{a_{M_1}^2 N}{\gamma_{M_1}}} - N > 0 \text{ and } \frac{\sum_{i=1}^{M_1+1} \frac{|a_i|}{b_i}\sqrt{\gamma_i N}}{\sum_{i=1}^{M_1+1} \frac{a_i^2}{b_i} - \frac{x}{y}} \sqrt{\frac{a_{M_1+1}^2 N}{\gamma_{M_1+1}}} - N \leq 0.$$

Minimizing error covariance

The corresponding problem in the orthogonal scheme is equivalent to

$$\begin{aligned} &\min_{\alpha_1,\ldots,\alpha_M} -\sum_{i=1}^{M} \frac{\alpha_i^2 h_i^2 C_i^2}{\alpha_i^2 h_i^2 R_i + N} \\ &\text{subject to } \sum_{i=1}^{M} \alpha_i^2 \left(\frac{C_i^2 Q}{1-A^2} + R_i \right) \leq p_{total} \end{aligned}$$

which is again a convex problem. For an analytical solution (Cui *et al.*, 2007), consider a more general problem

$$\begin{aligned} &\min_{\alpha_1,\ldots,\alpha_M} -\sum_{i=1}^{M} \frac{\alpha_i^2 a_i^2}{\alpha_i^2 b_i + N} \\ &\text{subject to } \sum_{i=1}^{M} \alpha_i^2 \gamma_i \leq p_{total} \end{aligned} \tag{5.19}$$

where $x > 0, y > 0, \gamma_i > 0, a_i \in \mathbb{R}, b_i > 0, i = 1, \ldots, M$ are constants and interpreted similarly as before. Assuming that the sensors are ordered so that

$$\frac{a_1^2}{\gamma_1} \geq \cdots \geq \frac{a_M^2}{\gamma_M}$$

the optimal α_i^2 to problem (5.19) is

$$\alpha_i^{*2} = \begin{cases} \frac{1}{b_i}(\sqrt{\frac{a_i^2 N}{\lambda \gamma_i}} - N) & , \quad i \leq M_1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

where

$$\frac{1}{\sqrt{\lambda}} = \frac{p_{total} + \sum_{i=1}^{M_1} \frac{\gamma_i}{b_i} N}{\sum_{i=1}^{M_1} \frac{|a_i|}{b_i} \sqrt{\gamma_i N}}$$

and the number of sensors which are active, M_1 (which is again unique (Xiao *et al.*, 2006)), satisfies

$$\frac{p_{total} + \sum_{i=1}^{M_1} \frac{\gamma_i}{b_i} N}{\sum_{i=1}^{M_1} \frac{|a_i|}{b_i} \sqrt{\gamma_i N}} \sqrt{\frac{a_{M_1}^2 N}{\gamma_{M_1}}} - N > 0 \text{ and } \frac{p_{total} + \sum_{i=1}^{M_1+1} \frac{\gamma_i}{b_i} N}{\sum_{i=1}^{M_1+1} \frac{|a_i|}{b_i} \sqrt{\gamma_i N}} \sqrt{\frac{a_{M_1+1}^2 N}{\gamma_{M_1+1}}} - N \leq 0$$

In the orthogonal case, the solutions of the optimization problems (5.18) and (5.19) will involve some sensors with low quality measurements not transmitting, whereas in the multi-access case all sensors will still get allocated some power when we perform the optimizations. Therefore issues to do with sensor lifetime due to some sensors transmitting more than others may be more serious in the orthogonal case.

5.5.3 Distributed implementations

The optimization problems (5.14), (5.18) and (5.19) have analytical solutions, and can admit distributed implementations, which may be important in large sensor networks. For problem (5.14) the fusion center can calculate the values λ and μ and broadcast them to all sensors, which can then be used by the sensors along with their local information to determine the optimal α_i's, see (Xiao *et al.*, 2005). For problems (5.18) and (5.19), the fusion center can compute and broadcast λ to all

sensors, which then determine their optimal α_i's using λ and their local information, see (Cui *et al.*, 2007).

5.6 Fading channels

The results presented so far have assumed that the channels are time-invariant. But in applications such as sensor networks, measurements may be transmitted over randomly time-varying (or fading) wireless channels. In this section we let the channel gains $\tilde{h}_{i,k}$ be randomly time varying. We consider both the case where the channel gains are known at the sensors and the fusion center, i.e. both transmitter and receiver have channel state information (CSI), and where the channel gains are not known at either the transmitters (sensors) or receiver (fusion center), in which case we derive a suboptimal linear estimator.

5.6.1 With CSI

First we let both the sensors and fusion center to have channel state information (CSI), so that the $h_{i,k}$'s are known. We now also allow the amplification factors $\alpha_{i,k}$ to be time-varying. As a shorthand let $P_k = P_{k|k-1}$.

Multi-access

The Kalman filter recursion for the error covariances is (see (Anderson & Moore, 1979))

$$\begin{aligned} P_{k+1} &= A^2 P_k - \frac{A^2 P_k^2 \bar{C}_k^2}{\bar{C}_k^2 P_k + \bar{R}_k} + Q \\ &= \frac{A^2 P_k \bar{R}_k}{\bar{C}_k^2 P_k + \bar{R}_k} + Q \end{aligned}$$

where $\bar{C}_k \equiv \sum_{i=1}^{M} \alpha_{i,k} h_{i,k} C_i$ and $\bar{R}_k \equiv \sum_{i=1}^{M} \alpha_{i,k}^2 h_{i,k}^2 R_i + N$.

One way in which we can formulate an optimization problem is to minimize the sum of powers used at each time instant, subject to $P_{k+1|k}$ being $\leq D$ at all time

instances k. That is, for all k, we want to solve

$$\begin{aligned} \min \sum_{i=1}^{M} p_{i,k} &= \sum_{i=1}^{M} \alpha_{i,k}^2 \left(\frac{C_i^2 Q}{1-A^2} + R_i \right) \\ \text{subject to } P_{k+1} &= \frac{A^2 P_k \bar{R}_k}{\bar{C}_k^2 P_k + \bar{R}_k} + Q \leq D \end{aligned} \tag{5.20}$$

The constraint can be rearranged to be equivalent to

$$\bar{R}_k \left(A^2 P_k + Q - D \right) + \bar{C}_k^2 P_k (Q - D) \leq 0$$

which looks rather similar to (5.12). In fact, once we've solved the problem (5.20) at an initial time instance, e.g. $k = 1$, then $P_2 = D$ is true, so that the problem becomes essentially identical to what was solved in Section 5.5.1. Therefore, the only slight difference is in the initial optimization problem, which is also covered by problem (5.14).

Another possible optimization problem is to minimize $P_{k+1|k}$ at each time instant subject to a sum power constraint at each time k, i.e.

$$\begin{aligned} \min P_{k+1} &= \frac{A^2 P_k \bar{R}_k}{\bar{C}_k^2 P_k + \bar{R}_k} + Q \\ \text{subject to } &\sum_{i=1}^{M} \alpha_{i,k}^2 \left(\frac{C_i^2 Q}{1-A^2} + R_i \right) \leq p_{total} \end{aligned} \tag{5.21}$$

As we can rewrite the objective as

$$\frac{A^2 P_k \bar{R}_k / \bar{C}_k^2}{P_k + \bar{R}_k / \bar{C}_k^2} + Q$$

it is clear that minimizing the objective function is equivalent to minimizing $\bar{R}_k / \bar{C}_k^2$. So at each time step we essentially solve the same problem considered in Section 5.5.1, while updating the value of P_{k+1} every time.

Orthogonal access

For time varying channels with CSI, the Kalman filter recursion for the error covariances can be shown by a similar argument to (5.7) to be

$$P^o_{k+1} = \frac{A^2 P^o_k}{1 + P^o_k \bar{C}^{o^T}_k \bar{R}^{o^{-1}}_k \bar{C}^o_k} + Q$$

If we wish to minimize the sum power while keeping $P^o_{k+1} \leq D$ at all time instances, the constraint becomes

$$\bar{C}^{o^T}_k \bar{R}^{o^{-1}}_k \bar{C}^o_k = \sum_{i=1}^{M} \frac{\alpha^2_{i,k} h^2_{i,k} C^2_i}{\alpha^2_{i,k} h^2_{i,k} R_i + N} \geq \frac{A^2 P^o_k + Q - D}{P^o_k (D - Q)}$$

If we wish to minimize P^o_{k+1} at each time instance subject to a sum power constraint at each time k, then this is the same as maximizing

$$\bar{C}^{o^T}_k \bar{R}^{o^{-1}}_k \bar{C}^o_k = \sum_{i=1}^{M} \frac{\alpha^2_{i,k} h^2_{i,k} C^2_i}{\alpha^2_{i,k} h^2_{i,k} R_i + N}$$

In both cases, the resulting optimization problems which are to be solved at each time instant are variants of problems (5.18) and (5.19), and can be handled using the same techniques.

Dynamic programming formulation

The optimization problems formulated in this section have constraints that must be satisfied at each time step. An alternative formulation is to consider constraints on the long term average. For instance, instead of problem (5.21), one might consider instead the problem

$$\min \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}\left[P_{k+1}\right]$$

$$\text{subject to } \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} \mathbb{E}[\sum_{i=1}^{M} p_{i,k}] \leq p_{total}$$

where we try to determine policies that will minimize the expected error covariance subject to the average sum power being less than a threshold p_{total}. Solving such problems will require dynamic programming techniques, and will be studied in future work.

5.6.2 No CSI

Multi-access

Suppose now that CSI is not available at either the sensors or fusion center, though channel statistics are known. We remark that this can also be used to model the situation where the sensors are not perfectly synchronized (Gastpar & Vetterli, 2005). The optimal filter in this case will be nonlinear and highly complex, see e.g. (Jaffer & Gupta, 1971). An alternative is to consider the best *linear* estimator in the minimum mean squared error (MMSE) sense, based on (Rajasekaran *et al.*, 1971). In our notation, the situation considered in (Rajasekaran *et al.*, 1971) would be applicable to the model

$$\begin{aligned} x_{k+1} &= Ax_k + w_k \\ z_k &= \alpha_k h_k C x_k + v_k \end{aligned}$$

While this is not quite the same as the situation that we are considering, their techniques can be suitably extended. We return to the full complex model (5.1). Since we don't have CSI, we will restrict $\tilde{\alpha}_{i,k} = \tilde{\alpha}_i, \forall k$ to be time invariant. The main difference from (Rajasekaran *et al.*, 1971) is that the innovations is now defined as

$$\begin{bmatrix} \Re[\tilde{z}_k] \\ \Im[\tilde{z}_k] \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{M} \mathbb{E}[\Re[\tilde{\alpha}_i \tilde{H}_i]] C_i \\ \sum_{i=1}^{M} \mathbb{E}[\Im[\tilde{\alpha}_i \tilde{H}_i]] C_i \end{bmatrix} \hat{x}_{k|k-1}$$

Assuming that the processes $\{\tilde{h}_{i,k}\}, i = 1, \ldots, M$ are i.i.d. over time, with real and imaginary compenents independent of each other, and $\{\tilde{h}_{i,k}\}$ independent of $\{w_k\}$ and $\{v_{i,k}\}, i = 1, \ldots, M$, the linear MMSE estimator for scalar systems can be

derived following the methods of (Rajasekaran *et al.*, 1971) as follows:

$$\begin{aligned}
\hat{x}_{k+1|k} &= A\hat{x}_{k|k} \\
P_{k+1|k} &= A^2 P_{k|k} \\
\hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + P_{k+1|k}\bar{\bar{C}}^T \left(\bar{\bar{C}} P_{k+1|k} \bar{\bar{C}}^T + \bar{\bar{R}}\right)^{-1} (\tilde{z}_{k+1} - \bar{\bar{C}}\hat{x}_{k+1|k}) \\
P_{k+1|k+1} &= P_{k+1|k} - P_{k+1|k}^2 \bar{\bar{C}}^T \left(\bar{\bar{C}} P_{k+1|k} \bar{\bar{C}}^T + \bar{\bar{R}}\right)^{-1}
\end{aligned} \tag{5.22}$$

where

$$\bar{\bar{C}} \equiv \begin{bmatrix} \sum_{i=1}^{M} \mathbb{E}[\Re[\tilde{\alpha}_i \tilde{H}_i]] C_i \\ \sum_{i=1}^{M} \mathbb{E}[\Im[\tilde{\alpha}_i \tilde{H}_i]] C_i \end{bmatrix}$$

$$\bar{\bar{R}} \equiv \begin{bmatrix} \bar{\bar{R}}_{11} & \bar{\bar{R}}_{12} \\ \bar{\bar{R}}_{21} & \bar{\bar{R}}_{22} \end{bmatrix}$$

with

$$\begin{aligned}
\bar{\bar{R}}_{11} &= \sum_{i=1}^{M} \left(\mathrm{Var}[\Re[\tilde{\alpha}_i \tilde{H}_i]] \frac{C_i^2 Q}{1 - A^2} + \mathbb{E}[\Re^2[\tilde{\alpha}_i \tilde{H}_i]] R_i \right) + N \\
\bar{\bar{R}}_{12} &= \bar{\bar{R}}_{21} = \sum_{i=1}^{M} \mathbb{E}[\Re[\tilde{\alpha}_i \tilde{H}_i]] \mathbb{E}[\Im[\tilde{\alpha}_i \tilde{H}_i]] R_i \\
\bar{\bar{R}}_{22} &= \sum_{i=1}^{M} \left(\mathrm{Var}[\Im[\tilde{\alpha}_i \tilde{H}_i]] \frac{C_i^2 Q}{1 - A^2} + \mathbb{E}[\Im^2[\tilde{\alpha}_i \tilde{H}_i]] R_i \right) + N
\end{aligned}$$

These look like the Kalman filter equations but with different C and R matrices, so much of the previous analysis in this chapter will apply.[6] For instance, there *will* be a steady state error covariance given by

$$P_\infty = \frac{(A^2 - 1) + QS + \sqrt{(A^2 - 1 + QS)^2 + 4QS}}{2S}$$

with $S = \bar{\bar{C}}^T \bar{\bar{R}}^{-1} \bar{\bar{C}}$. Note that for circularly symmetric fading channels e.g. Rayleigh, we have $\bar{\bar{C}} = [\ 0 \quad 0\ ]$ and estimates obtained using this estimator will not be useful. Other work where there are difficulties with circularly symmetric fading include (Mergen & Tong, 2006; Gastpar & Vetterli, 2005; Liu *et al.*, 2007).

[6] In fact one can regard it as an "equivalent" linear system along the lines of (Tugnait, 1981).

Thus we will now restrict ourselves to non-zero mean fading. Motivated by transmitter beamforming in the case with CSI, let us choose

$$\tilde{\alpha}_i = \alpha_i \frac{(\mathbb{E}[\tilde{H}_i])^*}{|\mathbb{E}[\tilde{H}_i]|}$$

Then S simplifies to

$$S = \frac{\left(\sum_{i=1}^{M} \mathbb{E}[\Re[\tilde{\alpha}_i \tilde{H}_i]] C_i\right)^2}{\sum_{i=1}^{M} \left(\mathrm{Var}[\Re[\tilde{\alpha}_i \tilde{H}_i]] C_i^2 \frac{Q}{1-A^2} + \mathbb{E}[\Re^2[\tilde{\alpha}_i \tilde{H}_i]] R_i\right) + N}$$

where we can find

$$\begin{aligned}
\mathbb{E}[\Re[\tilde{\alpha}_i \tilde{H}_i]] &= \alpha_i |\mathbb{E}[\tilde{H}_i]| \\
\mathrm{Var}[\Re[\tilde{\alpha}_i \tilde{H}_i]] &= \frac{\alpha_i^2}{|\mathbb{E}[\tilde{H}_i]|^2} \left(\mathbb{E}^2[\Re \tilde{H}_i] \mathrm{Var}[\Re \tilde{H}_i] + \mathbb{E}^2[\Im \tilde{H}_i] \mathrm{Var}[\Im \tilde{H}_i]\right) \\
\mathbb{E}[\Re^2[\tilde{\alpha}_i \tilde{H}_i]] &= \frac{\alpha_i^2}{|\mathbb{E}[\tilde{H}_i]|^2} \left(\mathbb{E}^2[\Re \tilde{H}_i] \mathbb{E}[\Re^2 \tilde{H}_i] + 2\mathbb{E}^2[\Re \tilde{H}_i] \mathbb{E}^2[\Im \tilde{H}_i] + \mathbb{E}^2[\Im \tilde{H}_i] \mathbb{E}[\Im^2 \tilde{H}_i]\right)
\end{aligned} \tag{5.23}$$

using the shorthand $\mathbb{E}^2[X] = (\mathbb{E}[X])^2, \Re^2[X] = (\Re[X])^2$ and $\Im^2[X] = (\Im[X])^2$. If the real and imaginary parts are identically distributed, we have the further simplifications $\mathrm{Var}[\Re[\tilde{\alpha}_i \tilde{H}_i]] = \alpha_i^2 \mathrm{Var}[\Re \tilde{H}_i]$ and $\mathbb{E}[\Re^2[\tilde{\alpha}_i \tilde{H}_i]] = \alpha_i^2 \left(\mathbb{E}[\Re^2 \tilde{H}_i] + \mathbb{E}^2[\Re \tilde{H}_i]\right)$.

Asymptotic behaviour using this estimator can then be analyzed using the techniques in Sections 5.4. In particular, one can achieve $P_\infty = Q + O(1/M)$ behaviour, even when the total power is bounded. So even though we don't have CSI, and are using a suboptimal linear filter, we can still achieve a decay to Q at a rate of $1/M$ when we have a large number of sensors and the fading is non-zero mean. The intuition behind this result is that adding up a lot of random $\tilde{h}_{i,k}$'s will tend to average things out, so that we can replace the unknown $\tilde{h}_{i,k}$'s with their statistics, e.g. means and variances. A similar scaling result in the context of i.i.d. processes and no CSI but non-zero mean fading was shown in (Liu *et al.*, 2007).

Optimal power allocation using this estimator can be done, and the resulting problems will be variants of problems (5.14) and (5.16).

Orthogonal access

In the case of orthogonal access and no CSI, the equations for the linear MMSE can also be similarly derived and will be of the form

$$\begin{aligned}
\hat{x}^o_{k+1|k} &= A\hat{x}^o_{k|k} \\
P^o_{k+1|k} &= A^2 P^o_{k|k} \\
\hat{x}^o_{k+1|k+1} &= \hat{x}^o_{k+1|k} + P^o_{k+1|k}\bar{\bar{C}}^{oT}\left(\bar{\bar{C}}^o P^o_{k+1|k}\bar{\bar{C}}^{oT} + \bar{\bar{R}}^o\right)^{-1}(\tilde{z}^o_{k+1} - \bar{\bar{C}}^o\hat{x}^o_{k+1|k}) \\
P^o_{k+1|k+1} &= P^o_{k+1|k} - (P^o_{k+1|k})^2\bar{\bar{C}}^{oT}\left(\bar{\bar{C}}^o P^o_{k+1|k}\bar{\bar{C}}^{oT} + \bar{\bar{R}}^o\right)^{-1}
\end{aligned}$$

We have

$$\bar{\bar{C}}^o \equiv \left[\begin{array}{cccc} \mathbb{E}[\Re[\tilde{\alpha}_1\tilde{H}_1]]C_1 & \mathbb{E}[\Im[\tilde{\alpha}_1\tilde{H}_1]]C_1 & \ldots & \mathbb{E}[\Re[\tilde{\alpha}_M\tilde{H}_M]]C_M \quad \mathbb{E}[\Im[\tilde{\alpha}_M\tilde{H}_M]]C_M \end{array}\right]^T$$

$$\bar{\bar{R}}^o \equiv \left[\begin{array}{ccc} \bar{\bar{R}}^o_{11} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \bar{\bar{R}}^o_{MM} \end{array}\right]$$

Each $\bar{\bar{R}}^o_{ii}$ is a block matrix of the form

$$\bar{\bar{R}}^o_{ii} = \left[\begin{array}{cc} \bar{\bar{R}}^o_{ii,11} & \bar{\bar{R}}^o_{ii,12} \\ \bar{\bar{R}}^o_{ii,21} & \bar{\bar{R}}^o_{ii,22} \end{array}\right]$$

with

$$\begin{aligned}
\bar{\bar{R}}^o_{ii,11} &= \mathrm{Var}[\Re[\tilde{\alpha}_i\tilde{H}_i]]C_i^2\frac{Q}{1-A^2} + \mathbb{E}[\Re^2[\tilde{\alpha}_i\tilde{H}_i]]R_i + N \\
\bar{\bar{R}}^o_{ii,12} &= \bar{\bar{R}}^o_{ii,21} = \mathbb{E}[\Re[\tilde{\alpha}_i\tilde{H}_i]]\mathbb{E}[\Im[\tilde{\alpha}_i\tilde{H}_i]]R_i \\
\bar{\bar{R}}^o_{ii,22} &= \mathrm{Var}[\Im[\tilde{\alpha}_i\tilde{H}_i]]C_i^2\frac{Q}{1-A^2} + \mathbb{E}[\Im^2[\tilde{\alpha}_i\tilde{H}_i]]R_i + N
\end{aligned}$$

There will be a steady state error covariance given by

$$P^o_\infty = \frac{(A^2-1) + QS^o + \sqrt{(A^2-1+QS^o)^2 + 4QS^o}}{2S^o}$$

with $S^o = \bar{\bar{C}}^{o^T} \bar{\bar{R}}^{o^{-1}} \bar{\bar{C}}^o$. If we choose

$$\tilde{\alpha}_i = \alpha_i \frac{(\mathbb{E}[\tilde{H}_i])^*}{|\mathbb{E}[\tilde{H}_i]|}$$

then S^o simplifies to

$$S^o = \sum_{i=1}^{M} \frac{\left(\mathbb{E}[\Re[\tilde{\alpha}_i \tilde{H}_i]] C_i\right)^2}{\left(\mathrm{Var}[\Re[\tilde{\alpha}_i \tilde{H}_i]] C_i^2 \frac{Q}{1-A^2} + \mathbb{E}[\Re^2[\tilde{\alpha}_i \tilde{H}_i]] R_i\right) + N}$$

where we also refer to (5.23) for further simplification of these quantities.

Asymptotic behaviour and optimal power allocation can also be analyzed using the techniques in Sections 5.4 and 5.5 respectively.

5.7 Numerical studies

5.7.1 Static channels

First we show some plots for the asymptotic results of Section 5.4. In Fig. 5.3 (a) we plot P_∞ vs M in the multi-access scheme for the symmetric situation with $\alpha_i = 1/\sqrt{M}$ and $A = 0.8, Q = 1.5, N = 1, C = 1, R = 1, h = 0.8$. We compare this with the asymptotic expression

$$Q + \frac{A^2(R + N/h^2)}{C^2} \frac{1}{M}$$

Fig. 5.3 (b) plots the difference between P_∞ and Q. We can see that P_∞ is well approximated by the asymptotic expression even for 20-30 sensors.

In Fig. 5.4 we plot P_∞ vs M in the multi-access scheme with $\alpha_i = 1/\sqrt{M}, A = 0.9, Q = 1, N = 1$ and values for C_i, R_i, h_i chosen from the range $0.5 \leq C_i \leq 1, 0.5 \leq R_i \leq 1, 0.5 \leq h_i \leq 1$. We also plot the (asymptotic) lower bound

$$Q + \frac{A^2(h_{min}^2 R_{min} + N)}{h_{max}^2 C_{max}^2} \frac{1}{M}$$

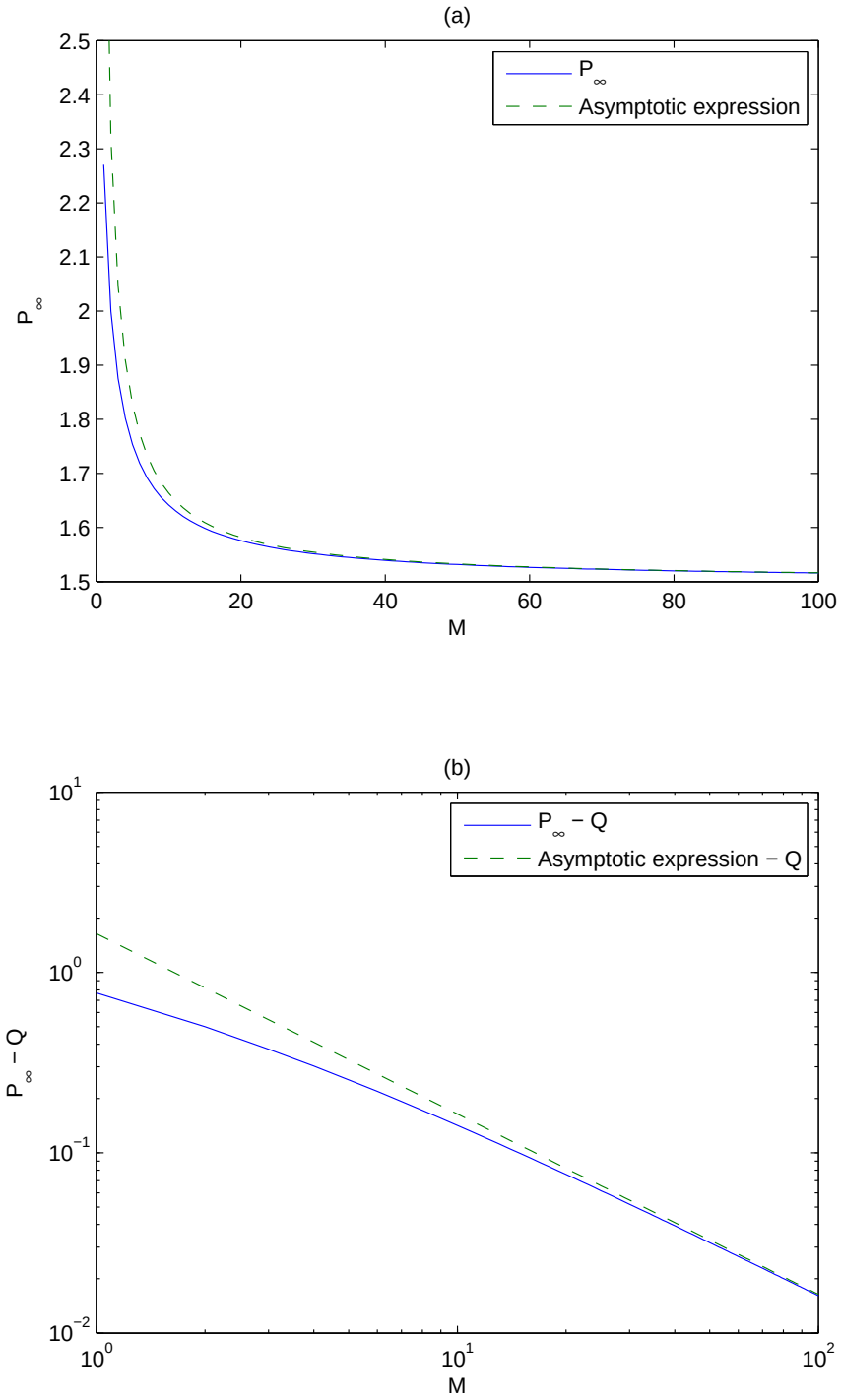

Figure 5.3: Comparison between P_∞ and asymptotic expression: Multi-access scheme with $\alpha_i = 1/\sqrt{M}$

and (asymptotic) upper bound

$$Q + \frac{A^2(h_{max}^2 R_{max} + N)}{h_{min}^2 C_{min}^2} \frac{1}{M}$$

It can be seen that P_∞ does lie between the two bounds, both of which converge to Q at the rate $1/M$.

Next we look at numerical results for optimal power allocation in Section 5.5. In Fig. 5.5 we compare between using optimal power allocation, equal power allocation and using the same value of α for all sensors, for the multi-access scheme. We use $A = 0.9, N = 1, Q = 1$ with various values for C_i, R_i drawn from a uniform distribution $U(0, 2)$, and values of h_i drawn from $U(0, 1)$. In (a) we keep $D = 2$, while in (b) we keep $p_{total} = 1$. In Fig. 5.6 we do the same thing for orthogonal access. What can be observed is that as the number of sensors M increases there is a general trend downwards for both graphs, though for equal power allocation and using the same α's for all sensors the behaviour is not necessarily monotonic. This is due to the fact that some sensors might have low quality measurements, e.g. sensor 5 in Fig. 5.5, so that extra resources are needed to compensate if this sensor is to be used. In the multi-access scheme, the performance gain in using optimal power allocation over simpler strategies is quite small when the number of sensors M is large. Also, there doesn't appear to be a clear performance advantage in using either equal power allocation or simply using the same α's for all sensors, though in terms of increasing the lifetime of a wireless sensor network, equal power allocation may be preferable. Finally, optimal power allocation in the orthogonal scheme will involve some sensors not transmitting, so even if we continually add in more sensors, they will not be utilised if their measurements are of low quality. Thus we sometimes see flat behaviour in Fig. 5.6 over a range of different numbers of sensors.

5.7.2 Fading channels

In Fig. 5.7 we compare between the full CSI and no CSI situations, using $A = 0.9, N = 1, Q = 1$ with various values for C_i, R_i drawn from a uniform distribution $U(0, 2)$. The $\tilde{h}_{i,k}$'s are chosen to be Rician distributed with real and imaginary parts

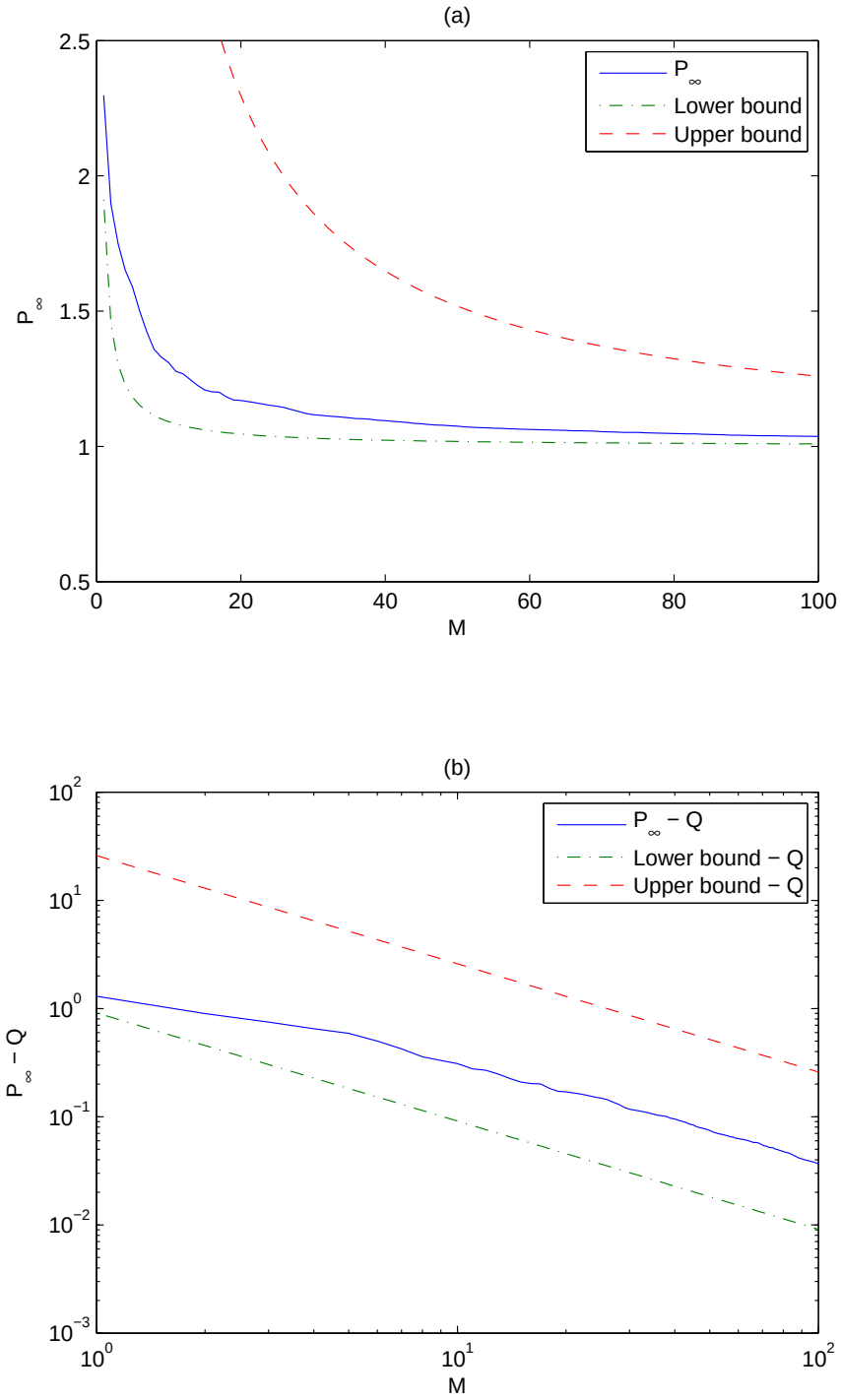

Figure 5.4: P_∞ with general parameters and bounds: Multi-access scheme with $\alpha_i = 1/\sqrt{M}$

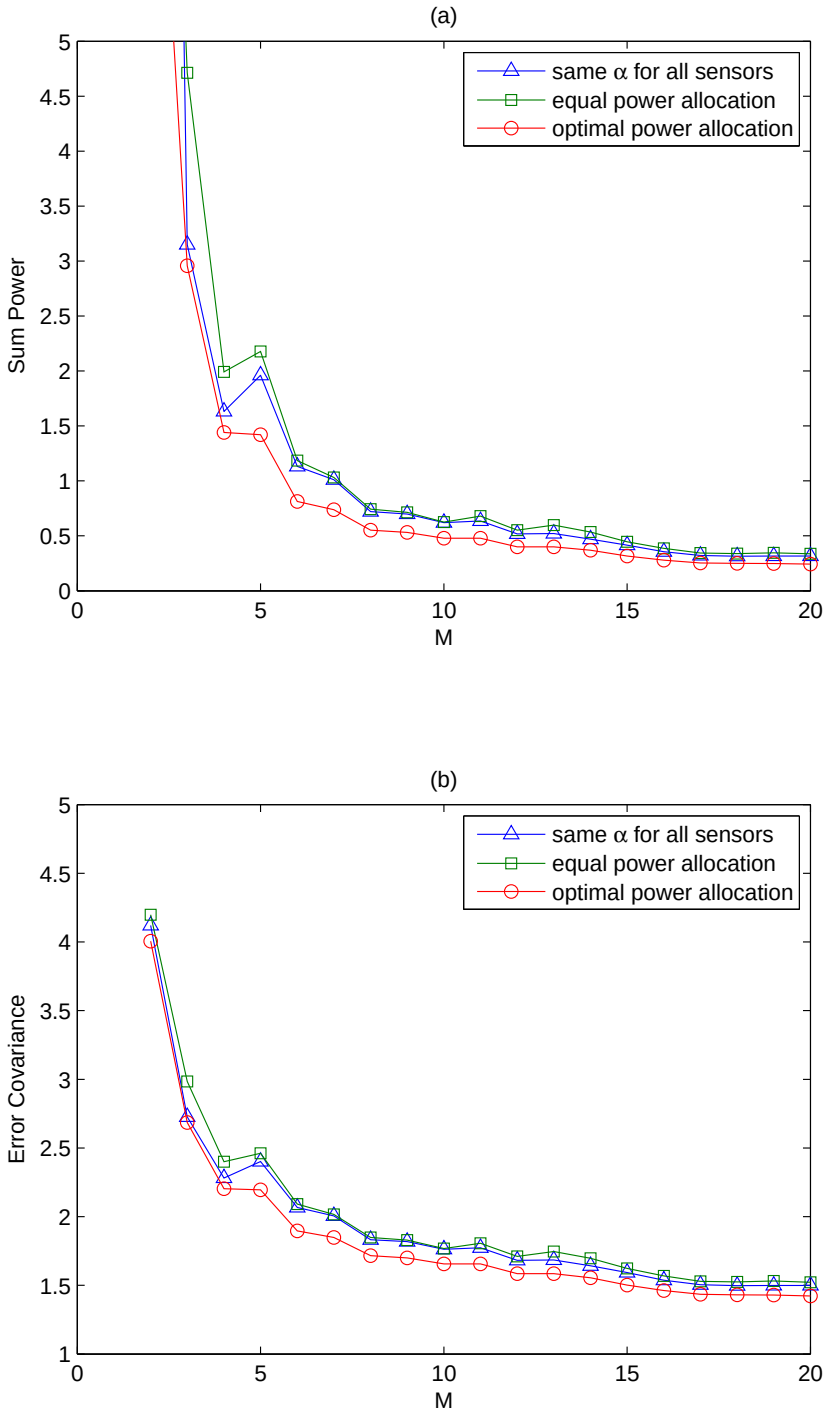

Figure 5.5: Multi-access. Comparison between three power allocation schemes, with (a) an error covariance constraint and (b) a sum power constraint

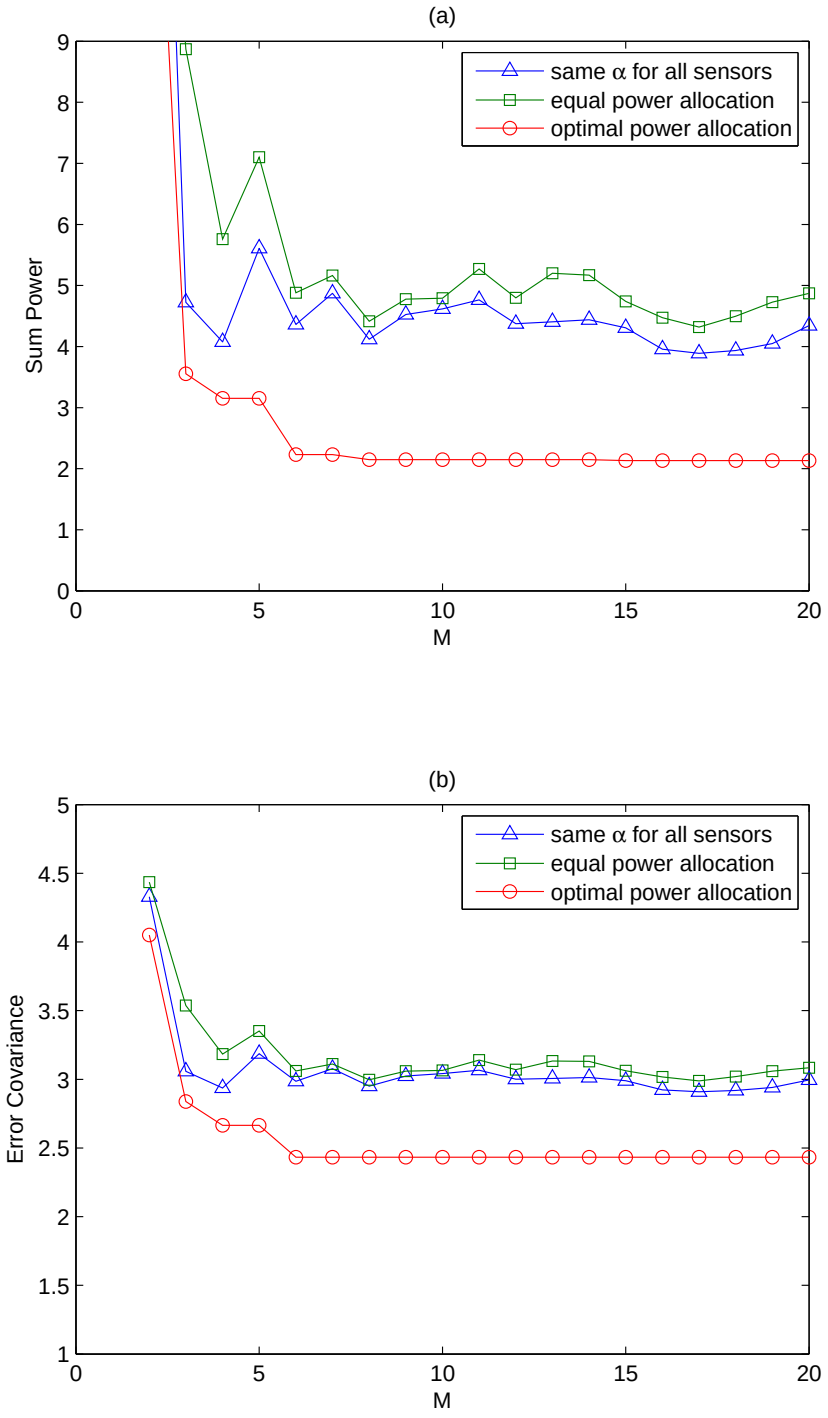

Figure 5.6: Orthogonal access. Comparison between three power allocation schemes, with (a) an error covariance constraint and (b) a sum power constraint

both being $N(\mu_i, \sigma_i^2)$, with μ_i and σ_i^2 drawn from $U(0,1)$. In (a) we keep $D = 2$, and in (b) we keep $p_{total} = 1$. In the full CSI case the values are averaged over 10000 time steps, and in the no CSI case they are the steady state values using the linear MMSE estimator (5.22). In Fig. 5.8 we do the same for orthogonal access. We can see that the performance loss in the no CSI case is not too great when compared to the case with full CSI. Thus even if one has full CSI, but doesn't want to perform power allocation at every time step, using the linear MMSE estimator (5.22) instead could be an attractive alternative. On the other hand, in Fig. 5.8 there is a significant performance loss in the situation with no CSI, when compared to the multi-access situation in Fig 5.7, possibly because we can't get the averaging benefits that are possible when we add measurements together in the multi-access case.

5.8 Conclusion

This chapter has investigated the use of analog forwarding in the distributed estimation of stable scalar linear systems. We have shown a $1/M$ scaling behaviour of the error covariance in a number of different situations, and have also formulated and solved some optimal power allocation problems. A possible extension of this work include extending our results to state estimation of vector linear systems. For vector linear systems, additional issues such as whether sensors will make observations of the entire state or whether sensors should only measure certain components of the state vector, observability of the resulting matrix when these measurements are added together in the multi-access scheme, and the appropriate criteria with which to formulate power allocation problems, will need to be addressed. Optimal power allocation problems in the case of unstable systems in a finite horizon setting, as well as the dynamic programming formulation approach to optimization, can also be studied. These problems will form the topics of future investigations.

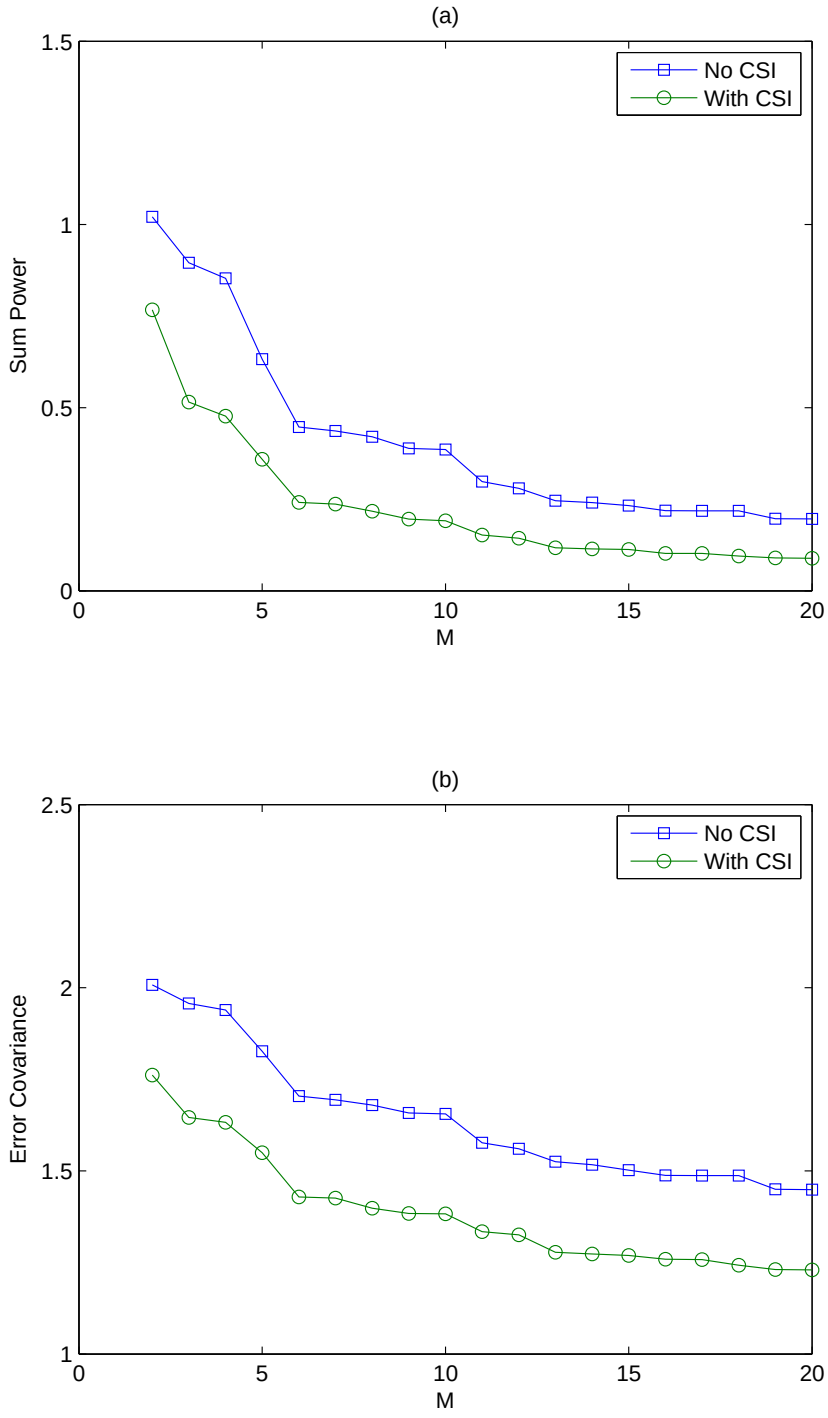

Figure 5.7: Multi-access. Comparison between the full CSI and no CSI situations, with (a) an error covariance constraint and (b) a sum power constraint

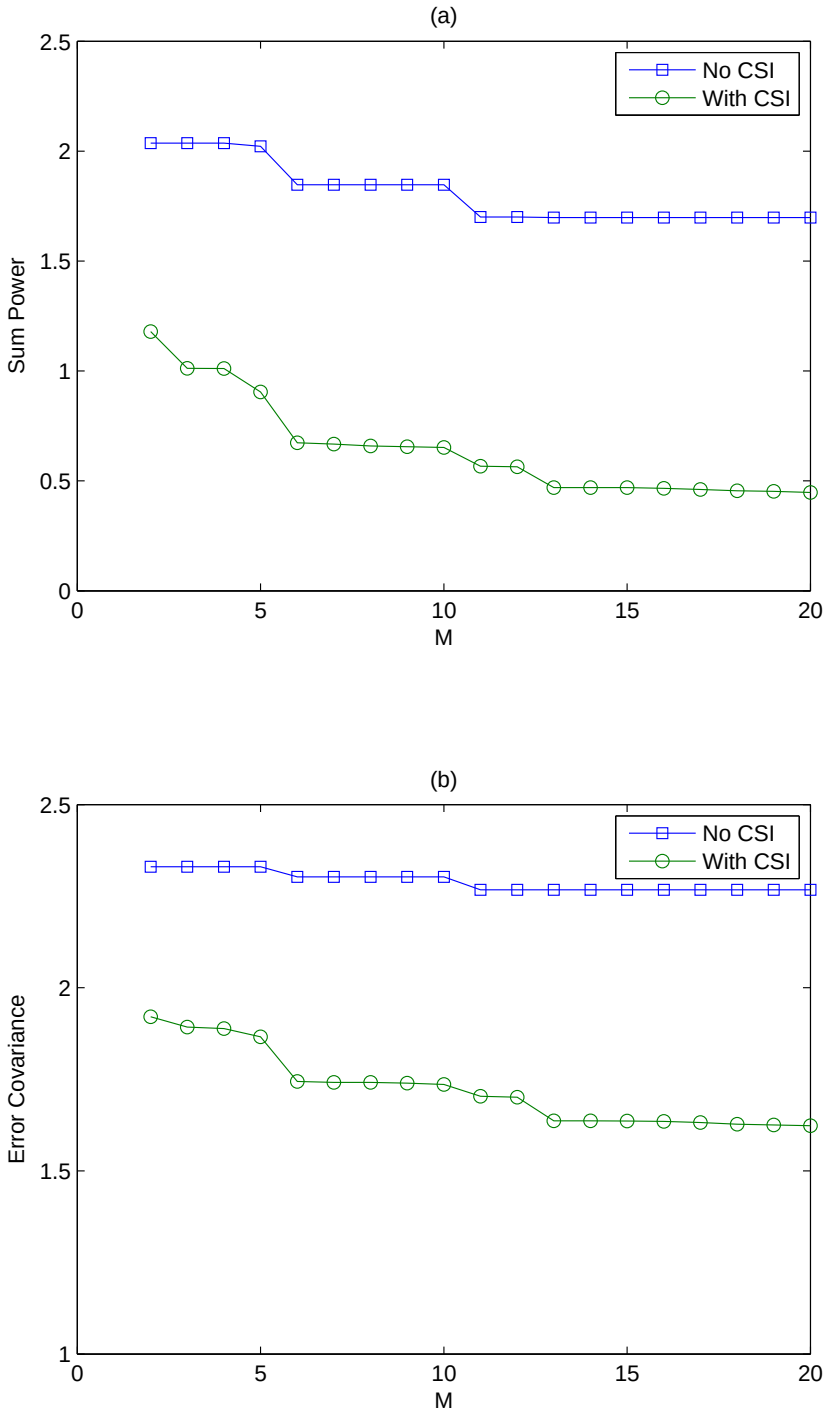

Figure 5.8: Orthogonal access. Comparison between the full CSI and no CSI situations, with (a) an error covariance constraint and (b) a sum power constraint

5.9 Appendix

5.9.1 Proof of Lemma 5.2.1

Rewrite (5.5) as

$$P_\infty = \frac{(A^2-1)}{2}\frac{1}{S} + \frac{Q}{2} + \sqrt{\frac{(A^2-1)^2}{4}\frac{1}{S^2} + \frac{(A^2+1)Q}{2}\frac{1}{S} + \frac{Q^2}{4}}.$$

Taking the derivative with respect to S we get

$$\frac{dP_\infty}{dS} = -\frac{A^2-1}{2}\frac{1}{S^2} - \frac{(A^2-1)^2\frac{1}{S^3} + (A^2+1)Q\frac{1}{S^2}}{4\sqrt{\frac{(A^2-1)^2}{4}\frac{1}{S^2} + \frac{(A^2+1)Q}{2}\frac{1}{S} + \frac{Q^2}{4}}}.$$

To show that $\frac{dP_\infty}{dS} \leq 0$, it is sufficient to show that

$$\left(\frac{(A^2-1)^2\frac{1}{S^3} + (A^2+1)Q\frac{1}{S^2}}{4\sqrt{\frac{(A^2-1)^2}{4}\frac{1}{S^2} + \frac{(A^2+1)Q}{2}\frac{1}{S} + \frac{Q^2}{4}}}\right)^2 \geq \left(\frac{A^2-1}{2}\frac{1}{S^2}\right)^2.$$

Expanding and rearranging, this is equivalent to

$$\begin{aligned}(A^2-1)^4\frac{1}{S^6} + 2(A^2-1)^2(A^2+1)Q\frac{1}{S^5} + (A^2+1)^2Q^2\frac{1}{S^4} \\ \geq (A^2-1)^4\frac{1}{S^6} + 2(A^2-1)^2(A^2+1)Q\frac{1}{S^5} + (A^2-1)^2Q^2\frac{1}{S^4}\end{aligned}$$

or

$$(A^2+1)^2Q^2 \geq (A^2-1)^2Q^2,$$

which is certainly true.

5.9.2 Derivation of analytical solution to problem (5.14)

Recall the problem (5.14)

$$\begin{aligned} &\min_{\alpha_1,\ldots,\alpha_M,s} \sum_{i=1}^{M} \alpha_i^2 \gamma_i \\ &\text{subject to } \left(\sum_{i=1}^{M} \alpha_i^2 b_i + N \right) x \leq s^2 y \\ &s = \sum_{i=1}^{M} \alpha_i a_i \end{aligned}$$

where $x > 0, y > 0, \gamma_i > 0, a_i \in \mathbb{R}, b_i > 0, i = 1, \ldots, M$ are constants.

Following (Xiao *et al.*, 2005) one can obtain a closed form solution as follows. Form the Lagrangian

$$L = \sum_{i=1}^{M} \alpha_i^2 \gamma_i + \lambda \left(\sum_{i=1}^{M} \alpha_i^2 b_i x + Nx - s^2 y \right) + \mu \left(s - \sum_{i=1}^{M} \alpha_i a_i \right).$$

The Karush-Kuhn-Tucker (KKT) conditions are:

$$\frac{\partial L}{\partial s} = -2\lambda s y + \mu = 0 \tag{5.24}$$

$$\frac{\partial L}{\partial \alpha_i} = 2\alpha_i \left(\gamma_i + \lambda b_i x \right) - \mu a_i = 0, i = 1, \ldots, M \tag{5.25}$$

$$\lambda \left(\sum_{i=1}^{M} \alpha_i^2 b_i x + Nx - s^2 y \right) = 0 \tag{5.26}$$

$$\lambda \geq 0 \tag{5.27}$$

$$\sum_{i=1}^{M} \alpha_i^2 b_i x + Nx - s^2 y \leq 0 \tag{5.28}$$

$$s - \sum_{i=1}^{M} \alpha_i a_i = 0 \tag{5.29}$$

Below we derive the solutions to the KKT equations (5.24)-(5.29). We have

$$\begin{aligned} \gamma_i + \lambda b_i x &= \frac{\mu a_i}{2\alpha_i} \\ &= \frac{\lambda s y a_i}{\alpha_i} \end{aligned}$$

$$= \frac{\lambda y a_i \sum_{i=1}^{M} \alpha_i a_i}{\alpha_i}$$

where the first, second and third lines used (5.25), (5.24) and (5.29) respectively. Then

$$\frac{\lambda a_i^2}{\gamma_i + \lambda b_i x} = \frac{\alpha_i a_i}{y \sum_{i=1}^{M} \alpha_i a_i}$$

and hence

$$\sum_{i=1}^{M} \frac{\lambda a_i^2}{\gamma_i + \lambda b_i x} = \frac{1}{y}, \tag{5.30}$$

from which λ may be found numerically. Since the left hand side is increasing with λ solutions to this equation will be unique provided it exists. Taking limits as $\lambda \to \infty$, we see that a solution exists if and only if

$$\sum_{i=1}^{M} \frac{a_i^2}{b_i} > \frac{x}{y}.$$

This provides a feasibility check for the optimization problem.

Next, we have from (5.26) that

$$\sum_{i=1}^{M} \alpha_i^2 b_i x + Nx - s^2 y = 0.$$

Substituting expressions for α_i from (5.25) and s from (5.24) we get

$$\sum_{i=1}^{M} \frac{\mu^2 a_i^2 b_i x}{4(\gamma_i + \lambda b_i x)^2} + Nx - \frac{\mu^2}{4\lambda^2 y} = 0$$

and hence μ can be computed since

$$\begin{aligned} \mu^2 &= Nx \left(\frac{1}{4\lambda^2 y} - \sum_{i=1}^{M} \frac{a_i^2 b_i x}{4(\gamma_i + \lambda b_i x)^2} \right)^{-1} \\ &= Nx \left(\sum_{i=1}^{M} \frac{a_i^2 \gamma_i}{4\lambda(\gamma_i + \lambda b_i x)^2} \right)^{-1}, \end{aligned} \tag{5.31}$$

where the second line can be obtained by using the relation (5.30) and some algebraic manipulations. Depending on whether we choose μ to be positive or negative, we

will obtain the solution in one of the regions $s > 0$ of $s < 0$.

From (5.25) we may then determine the optimal α_i's as

$$\alpha_i^* = \frac{\mu a_i}{2(\gamma_i + \lambda b_i x)}, i = 1, \ldots, M.$$

Finally, we also have the relation

$$\sum_{i=1}^{M} \alpha_i^2 \gamma_i = \sum_{i=1}^{M} \frac{\mu^2 a_i^2}{4(\gamma_i + \lambda b_i x)^2} \gamma_i = \lambda N x$$

using the expression (5.31) for μ^2.

Chapter 6

Conclusions

This thesis has looked at methods to analyze the performance of a selection of estimation and detection algorithms operating in wireless environments. Characterizing the achievable performance and performance limitations of these algorithms is of importance both for engineers who wish to apply these algorithms to wireless networks, as well as motivating the design of new algorithms which may be more suited to these environments. Below we shall summarise our work, and list some possible ideas for future research which are related to the topics in this thesis.

6.1 Summary

In Chapter 2 we studied the probability of error for maximum a posteriori estimation of hidden Markov models, where measurements could be either lost or received according to another Markov process. Analytical expressions for the error probabilities were derived for the noiseless and noisy cases. We demonstrated some relationships between the error probability and the parameters of the loss process via both analysis and numerical results. In the high SNR regime, approximate expressions which are more easily computed than the exact analytical form for the noisy case were also presented.

In Chapter 3 a numerical method for computing the error exponent for Neyman-Pearson detection of two-state Markov chains in noise was presented, for both time-invariant and fading channels. Numerical studies showing the behaviour of the error exponent as the transition parameters of the Markov chain and the SNR were varied were given. Comparisons between the high SNR asymptotics in Gaussian noise for the time-invariant and fading situations were made, with different scalings in the

error exponent with SNR observed and proved.

Chapter 4 studied the performance of the Kalman fixed lag smoother with random packet losses, and how it compares with the Kalman filter with packet loss. In terms of estimator stability via boundedness of the expectation of the error covariance, we showed that smoothing did not provide any benefit over filtering. On the other hand, it was demonstrated that using a probabilistic notion of performance, smoothing could provide significant gains when compared to Kalman filtering. An analysis of Kalman filtering using two simple retransmission schemes and its comparison with Kalman smoothing was also made.

In Chapter 5 we considered state estimation of scalar linear systems using analog forwarding with multiple sensors, for both multiple access and orthogonal access schemes. It was shown that optimal state estimation can be achieved at the fusion centre using a time-varying Kalman filter. We showed that in many situations, the error covariance decays at a rate of $1/M$ when the number of sensors M is large. Optimal allocation of transmission powers subject to constraints on the error covariance or sum power was also considered. In the case of fading channels without channel state information, a sub-optimal linear estimator was derived, and under optimal power allocation, numerical studies showed that its performance in the multi-access scheme is comparable to the case where full channel state information is available.

6.2 Future Research

For HMM filtering, incorporating fading with no CSI similar to Chapter 3 is one possible extension. Finding ways to compute the probability of error for multi-state Markov chains with noise will be highly desirable, though existing techniques which can be used for the two-state case don't appear to be sufficient to solve the general case. Similarly, methods for computing the error exponent for detection of multi-state Markov chains is another interesting open problem.

For the detection of Markov chains in noise, a more thorough analysis of non-Gaussian noise and its effect on the error exponent at high SNR can be done, as outlined in the conclusion of Chapter 3. It would be interesting to consider complex

fading channels, and whether there is a dramatic change in behaviour in the case of e.g. circularly symmetric channels (Mergen & Tong, 2006). Characterising the error exponents for Bayesian detection rather than Neyman-Pearson detection is another topic worth investigating. Analysis of detector behaviour in the case of uncertain knowledge of system parameters, and how it relates to Lemmas 3.3.1 and 3.4.1 will be of interest. The effect of uncertain parameters for the other algorithms in this thesis can also be investigated.

For problems of linear filtering with random packet losses, more thorough investigations of retransmission schemes would be interesting, to see whether performance can be improved if e.g. we allow more than one packet to be retransmitted at a time. Another intriguing problem is to look at using quantized measurements, possibly using the structural results of (Nair & Evans, 1999) or one bit quantization schemes (Ribeiro *et al.*, 2006), and see how stability of the estimator is affected.

For the work on power efficient state estimation in Chapter 5, we could investigate more situations on the availability of CSI, e.g. CSI available at receiver but not transmitter. In the case of no CSI, rather than the block fading model used, we could potentially allow correlations by using the results of (Chow & Birkemeier, 1989; Chow & Birkemeier, 1990). Extension of the analog forwarding scheme to vector linear systems is also important. For vector linear systems, additional issues such as whether sensors will make observations of the entire state or whether sensors should only measure certain components of the state vector, observability of the resulting matrix when these measurements are added together in the multi-access scheme, and the appropriate criteria with which to formulate power allocation problems, will need to be addressed. For instance, some sensors might only measure one component e.g. position, of a state, and other sensors another component e.g. velocity. Then one possibility is for all those sensors measuring the position to communicate in a multi-access scheme to the fusion center, while the sensors measuring the velocity communicates in a multi-access scheme but using a different orthogonal channel. Estimation of unstable systems (both scalar and vector) over a finite horizon, as well as solving the dynamic programming formulation of our optimization problems, could also be considered. Finally, one could envisage state estimation problems using

analog forwarding for other dynamical systems such as hidden Markov models, and analyzing its performance there.

References

Abramowitz, M., & Stegun, I. A. 1965. *Handbook of Mathematical Functions*. New York: Dover Publications.

Adlakha, S. 2005. *Kalman Filtering in Correlated Losses*. Tech. rept. Stanford University.

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. 2002. A Survey on Sensor Networks. *IEEE Commun. Mag.*, **40**(8), 102–114.

Anderson, B. D. O., & Moore, J. B. 1979. *Optimal Filtering*. New Jersey: Prentice Hall.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. 2002. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Trans. Signal Processing*, **50**(2), 174–188.

Atkinson, K. E. 1997. *The Numerical Solution of Integral Equations of the Second Kind*. Cambridge: Cambridge University Press.

Bahceci, I., & Khandani, A. K. 2006 (Mar.). Energy-Efficient Estimation of Correlated Data in Wireless Sensor Networks. *In: Proc. Conf. Information Sciences and Systems*.

Bahr, R. K., & Bucklew, J. A. 1990. Optimal Sampling Schemes for the Gaussian Hypothesis Testing Problem. *IEEE Trans. Acoust., Speech, Signal Processing*, **38**(10), 1677–1686.

Baum, L. E., & Petrie, T. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Annals of Mathematical Statistics*, **37**(6), 1554–1563.

Benitz, G. R., & Bucklew, J. A. 1990. Large Deviation Rate Calculations for Nonlinear Detectors in Gaussian Noise. *IEEE Trans. Inform. Theory*, **36**(2), 358–371.

Berger, T., Zhang, Z., & Viswanathan, H. 1996. The CEO Problem. *IEEE Trans. Inform. Theory*, **42**(3), 887–902.

Biglieri, E., Proakis, J., & Shamai (Shitz), S. 1998. Fading Channels: Information-Theoretic and Communications Aspects. *IEEE Trans. Inform. Theory*, **44**(6), 2619–2692.

Casella, G., & Berger, R. L. 2002. *Statistical Inference*. 2nd edn. California: Duxbury Press.

Chamberland, J.-F., & Veeravalli, V. V. 2003. Decentralized Detection in Sensor Networks. *IEEE Trans. Signal Processing*, **51**(2), 407–416.

Chamberland, J.-F., & Veeravalli, V. V. 2004. Asymptotic Results for Decentralized Detection in Power Constrained Wireless Sensor Networks. *IEEE J. Select. Areas Commun.*, **22**(6), 1007–1015.

Chen, B., Jiang, R., Kasetkasem, T., & Varshney, P. K. 2004. Channel Aware Decision Fusion in Wireless Sensor Networks. *IEEE Trans. Signal Processing*, **52**(12), 3454–3458.

Chong, C.-Y., & Kumar, S. P. 2003. Sensor Networks: Evolution, Opportunities and Challenges. *Proc. IEEE*, **91**(8), 1247–1256.

Chow, B. S., & Birkemeier, W. P. 1989. A New Structure of Recursive Estimator. *IEEE Trans. Automat. Contr.*, **34**(5), 568–572.

Chow, B. S., & Birkemeier, W. P. 1990. A New Recursive Filter for Systems With Multiplicative Noise. *IEEE Trans. Inform. Theory*, **36**(6), 1430–1435.

Cui, S., Xiao, J.-J., Goldsmith, A., Luo, Z.-Q., & Poor, H. V. 2007. Estimation Diversity and Energy Efficiency in Distributed Sensing. *IEEE Trans. Signal Processing*, **55**(9), 4683–4695.

Dembo, A., & Zeitouni, O. 1998. *Large Deviations Techniques and Applications*. 2nd edn. New York: Springer Verlag.

Elliott, R. J., Aggoun, L., & Moore, J. B. 1995. *Hidden Markov Models: Estimation and Control*. New York: Springer Verlag.

Ephraim, Y., & Merhav, N. 2002. Hidden Markov Processes. *IEEE Trans. Inform. Theory*, **48**(6), 1518–1569.

Epstein, M., Shi, L., Tiwari, A., & Murray, R. M. 2007. Probabilistic Performance of State Estimation Across a Lossy Network. *Automatica*. to appear.

Fletcher, A. K., Rangan, S., & Goyal, V. K. 2004 (Apr.). Estimation From Lossy Sensor Data: Jump Linear Modelling and Kalman Filtering. *Pages 251–258 of: Proc. IPSN*.

Fraser, D. C., & Potter, J. E. 1969. The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters. *IEEE Trans. Automat. Contr.*, **14**(4), 387–390.

Gassiat, E., & Boucheron, S. 2003. Optimal Error Exponents in Hidden Markov Models Order Estimation. *IEEE Trans. Inform. Theory*, **49**(4), 964–980.

Gastpar, M., & Vetterli, M. 2003. Source-Channel Communication in Sensor Networks. *Springer Lecture Notes in Computer Science*, **2634**(Apr.), 162–177.

Gastpar, M., & Vetterli, M. 2005. Power, Spatio-Temporal Bandwidth, and Distortion in Large Sensor Networks. *IEEE J. Select. Areas Commun.*, **23**(4), 745–754.

Gastpar, M., Rimoldi, B., & Vetterli, M. 2003. To Code, or Not to Code: Lossy Source-Channel Communication Revisited. *IEEE Trans. Inform. Theory*, **49**(5), 1147–1158.

Goblick, T. J. 1965. Theoretical Limitations on the Transmission of Data From Analog Sources. *IEEE Trans. Inform. Theory*, **11**(4), 558–567.

Golubev, G. K. 2000. On Filtering for a Hidden Markov Chain Under Square Performance Criterion. *Problems of Information Transmission*, **36**(3), 213–219.

Hackbusch, W. 1995. *Integral Equations: Theory and Numerical Treatment*. Basel, Switzerland: Birkhauser Verlag.

Hadidi, M. T., & Schwartz, S. C. 1979. Linear Recursive State Estimators Under Uncertain Observations. *IEEE Trans. Automat. Contr.*, **24**(6), 944–948.

Hardy, G. H., Littlewood, J. E., & Pólya, G. 1952. *Inequalities*. 2nd edn. Cambridge, Great Britain: Cambridge University Press.

Hashemipour, H. R., Roy, S., & Laub, A. J. 1988. Decentralized Structures for Parallel Kalman Filtering. *IEEE Trans. Automat. Contr.*, **33**(1), 88–94.

Hespanha, J. P., Naghshtabrizi, P., & Xu, Y. 2007. A Survey of Recent Results in Networked Control Systems. *Proc. IEEE*, **95**(1), 138–162.

Huang, M., & Dey, S. 2007. Stability of Kalman Filtering With Markovian Packet Losses. *Automatica*, **43**, 598–607.

Ishwar, P., Puri, R., Ramchandran, K., & Pradhan, S. S. 2005. On Rate-Constrained Distributed Estimation in Unreliable Sensor Networks. *IEEE J. Select. Areas Commun.*, **23**(4), 765–775.

Jaffer, A. G., & Gupta, S. C. 1971. Recursive Bayesian Estimation With Uncertain Observations. *IEEE Trans. Inform. Theory*, **17**(5), 614–616.

Jiang, R., & Chen, B. 2005. Fusion of Censored Decisions in Wireless Sensor Networks. *IEEE Trans. Wireless Commun.*, **4**(6), 2668–2673.

Julier, S. J., & Uhlmann, J. K. 2004. Unscented Filtering and Nonlinear Estimation. *Proc. IEEE*, **92**(3), 401–422.

Kailath, T., Sayed, A. H., & Hassibi, B. 2000. *Linear Estimation*. New Jersey: Prentice Hall.

Kalman, R. E. 1960. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, **82**(Series D), 35–45.

Kalman, R. E., & Bucy, R. S. 1961. New Results in Linear Filtering and Prediction Theory. *Transactions of the ASME–Journal of Basic Engineering*, **83**(Series D), 95–107.

Kay, S. M. 1998. *Fundamentals of Statisical Signal Processing, Volume II: Detection Theory*. New Jersey: Prentice Hall.

Khasminskii, R., & Zeitouni, O. 1996. Asymptotic Filtering for Finite State Markov Chains. *Stochastic Processes and their Applications*, **63**, 1–10.

Le Gland, F., & Mevel, L. 1997 (July). Asymptotic Properties of the MLE in Hidden Markov Models. *In: Proc. 4th European Control Conf.*

Le Gland, F., & Mevel, L. 2000. Exponential Forgetting and Geometric Ergodicity in Hidden Markov Models. *Mathematics of Control, Signals, and Systems*, **13**, 63–93.

Leong, A. S., Dey, S., & Evans, J. S. 2007. Probability of Error Analysis for Hidden Markov Model Filtering With Random Packet Loss. *IEEE Trans. Signal Processing*, **55**(3), 809–821.

Liu, C.-C., & Narayan, P. 1994. Order Estimation and Sequential Universal Data Compression of a Hidden Markov Source by the Method of Mixtures. *IEEE Trans. Inform. Theory*, **40**(4), 1167–1180.

Liu, K., & Sayeed, A. M. 2007. Type-based Decentralized Detection in Wireless Sensor Networks. *IEEE Trans. Signal Processing*, **55**(5), 1899–1910.

Liu, K., Gamal, H. El, & Sayeed, A. M. 2007. Decentralized Inference Over Multiple-Access Channels. *IEEE Trans. Signal Processing*, **5**(7), 3445–3455.

Liu, X., & Goldsmith, A. J. 2004 (Dec.). Kalman Filtering With Partial Observation Losses. *Pages 1413–1418 of: Proc. IEEE Conf. Decision and Control*.

Mergen, G., & Tong, L. 2006. Type Based Estimation Over Multiaccess Channels. *IEEE Trans. Signal Processing*, **54**(2), 613–626.

Mergen, G., Naware, V., & Tong, L. 2007. Asymptotic Detection Performance of Type-Based Multiple Access Over Multiaccess Fading Channels. *IEEE Trans. Signal Processing*, **55**(3), 1081–1092.

Merhav, N. 1991. Universal Classification for Hidden Markov Models. *IEEE Trans. Inform. Theory*, **37**(6), 1586–1594.

Miller, J. H., & Thomas, J. B. 1972. Detectors for Discrete-Time Signals in Non-Gaussian Noise. *IEEE Trans. Inform. Theory*, **18**(2), 241–250.

Monzingo, R. A. 1975. Discrete Optimal Linear Smoothing for Systems With Uncertain Observations. *IEEE Trans. Inform. Theory*, **21**(3), 271–275.

Mostofi, Y., & Murray, R. M. 2005 (Dec.). Receiver Design Principles for Estimation over Fading Channels. *Pages 464–469 of: Proc. IEEE Conf. Decision and Control*.

Nahi, N. E. 1969. Optimal Recursive Estimation With Uncertain Observation. *IEEE Trans. Inform. Theory*, **15**(4), 457–462.

Nair, G. N., & Evans, R. J. 1999 (Feb.). Structual Results for Finite Bit-Rate State Estimation. *Pages 47–52 of: Proc. Information, Decision and Control*.

Neyman, J., & Pearson, E. S. 1933. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Phil. Trans. Roy. Soc., Ser. A*, **231**, 289–337.

Niu, R., Chen, B., & Varshney, P. K. 2006. Fusion of Decisions Transmitted Over Rayleigh Fading Channels in Wireless Sensor Networks. *IEEE Trans. Signal Processing*, **54**(3), 1018–1027.

Nowak, R., Mitra, U., & Willett, R. 2004. Estimating Inhomogenous Fields Using Wireless Sensor Networks. *IEEE J. Select. Areas Commun.*, **22**(6), 999–1006.

Olver, F. W. J. 1974. *Asymptotics and Special Functions*. New York: Academic Press.

Ozarow, L. H., Shamai, S., & Wyner, A. D. 1994. Information Theoretic Considerations for Cellular Mobile Radio. *IEEE Trans. Veh. Technol.*, **43**(2), 359–378.

Poor, H. V. 1994. *An Introduction to Signal Detection and Estimation*. 2nd edn. New York: Springer.

Rabiner, L. R. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE*, **77**(2), 257–286.

Rajasekaran, P. K., Satyanarayana, N., & Srinath, M. D. 1971. Optimum Linear Estimation of Stochastic Signals in the Presence of Multiplicative Noise. *IEEE Trans. Aerosp. Electron. Syst.*, **7**(3), 462–468.

Ralston, A., & Rabinowitz, P. 1978. *A First Course in Numerical Analysis*. 2nd edn. New York: McGraw-Hill.

Rao, B. S., & Durrant-Whyte, H. F. 1991. Fully Decentralised Algorithm for Multisensor Kalman Filtering. *IEE Proceedings-D*, **138**(5), 413–420.

Rauch, H. E. 1963. Solutions to the Linear Smoothing Problem. *IEEE Trans. Automat. Contr.*, **8**(4), 371–372.

Ribeiro, A., & Giannakis, G. B. 2006. Bandwidth-Constrained Distributed Estimation for Wireless Sensor Networks, Part I: Gaussian Case. *IEEE Trans. Signal Processing*, **54**(3), 1131–1143.

Ribeiro, A., Giannakis, G. B., & Roumeliotis, S. I. 2006. SOI-KF: Distributed Kalman Filtering With Low-Cost Communications Using the Sign of Innovations. *IEEE Trans. Signal Processing*, **54**(12), 4782–4795.

Rudin, W. 1976. *Principles of Mathematical Analysis*. 3rd edn. New York: McGraw-Hill.

Scharf, L. L., & Nolte, L. W. 1977. Likelihood Ratios for Sequential Hypothesis Testing on Markov Sequences. *IEEE Trans. Inform. Theory*, **23**(1), 101–109.

Schenato, L. 2006 (Dec.). Optimal Estimation in Networked Control Systems Subject to Random Delay and Packet Drop. *In: Proc. IEEE Conf. Decision and Control*.

Schenato, L., Sinopoli, B., Franceschetti, M., Poolla, K., & Sastry, S. 2007. Foundations of Control and Estimation over Lossy Networks. *Proc. IEEE*, **95**(1), 163–187.

Seneta, E. 1981. *Non-negative Matrices and Markov Chains*. 2nd edn. New York: Springer-Verlag.

Shi, L., Epstein, M., Tiwari, A., & Murray, R. M. 2005 (Dec.). Estimation With Information Loss: Asymptotic Analysis and Error Bounds. *Pages 1215–1221 of: Proc. IEEE Conf. Decision and Control*.

Shiryaev, A. N. 1996. *Probability*. 2nd edn. New York: Springer-Verlag.

Shue, L., Anderson, B. D. O., & Dey, S. 1998. Exponential Stability of Filters and Smoothers for Hidden Markov Models. *IEEE Trans. Signal Processing*, **46**(8), 2180–2194.

Shue, L., Anderson, B. D. O., & De Bruyne, F. 2000. Asymptotic Smoothing Errors for Hidden Markov Models. *IEEE Trans. Signal Processing*, **48**(12), 3289–3300.

Shue, L., Dey, S., Anderson, B. D. O., & De Bruyne, F. 2001. On State-Estimation of a Two-State Hidden Markov Model With Quantization. *IEEE Trans. Signal Processing*, **49**(1), 202–208.

Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M. I., & Sastry, S. S. 2004. Kalman Filtering With Intermittent Observations. *IEEE Trans. Automat. Contr.*, **49**(9), 1453–1464.

Smith, S. C., & Seiler, P. 2003. Estimation With Lossy Measurements: Jump Estimators for Jump Systems. *IEEE Trans. Automat. Contr.*, **48**(12), 2163–2171.

Stewart, G. W., & Sun, J.-G. 1990. *Matrix Perturbation Theory*. San Diego, CA.: Academic Press.

Sung, Y., Tong, L., & Poor, H. V. 2006. Neyman-Pearson Detection of Gauss-Markov Signals in Noise: Closed-form Error Exponent and Properties. *IEEE Trans. Inform. Theory*, **52**(4), 1354–1365.

Tiwari, A., Jun, M., Jeffcoat, D. E., & Murray, R. M. 2005 (July). Analysis of Dynamic Sensor Coverage Problem Using Kalman Filters for Estimation. *In: Proc. IFAC World Congress.*

Tricomi, F. G. 1957. *Integral Equations.* New York: Interscience Publishers.

Tugnait, J. K. 1981. Stability of Optimum Linear Estimators of Stochastic Signals in White Multiplicative Noise. *IEEE Trans. Automat. Contr.*, **26**(3), 757–761.

Viswanathan, H., & Berger, T. 1997. The Quadratric Gaussian CEO Problem. *IEEE Trans. Inform. Theory*, **43**(5), 1549–1559.

Wang, Z., Yang, F., Ho, D. W. C., & Liu, X. 2005. Robust Finite-Horizon Filtering for Stochastic Systems With Missing Measurements. *IEEE Signal Processing Lett.*, **12**(6), 437–440.

Wang, Z., Yang, F., Ho, D. W. C., & Liu, X. 2006. Robust H_∞ Filtering for Stochastic Time-Delay Systems With Missing Measurements. *IEEE Trans. Signal Processing*, **54**(7), 2579–2587.

Willner, D., Chang, C. B., & Dunn, K. P. 1976 (Dec.). Kalman Filter Algorithms for a Multi-Sensor System. *Pages 570–574 of: Proc. IEEE Conf. Decision and Control.*

Wonham, W. M. 1965. Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering. *SIAM J. Control*, **2**(3), 347–369.

Xiao, J.-J., Luo, Z.-Q., Cui, S., & Goldsmith, A. J. 2005 (June). Power-Efficient Analog Forwarding Transmission in an Inhomogeneous Gaussian Sensor Network. *In: Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications.*

Xiao, J.-J., Cui, S., Luo, Z.-Q., & Goldsmith, A. J. 2006. Power Scheduling of Universal Decentralized Estimation in Sensor Networks. *IEEE Trans. Signal Processing*, **54**(2), 413–422.

Xu, Y., & Hespanha, J. P. 2005 (Dec.). Estimation Under Uncontrolled and Controlled Communications in Networked Control Systems. *Pages 842–847 of: Proc. IEEE Conf. Decision and Control.*

Zhang, K., & Li, X. R. 2004 (Dec.). Optimal Sensor Data Quantization for Best Linear Unbiased Estimation Fusion. *Pages 2656–2661 of: Proc. IEEE Conf. Decision and Control.*