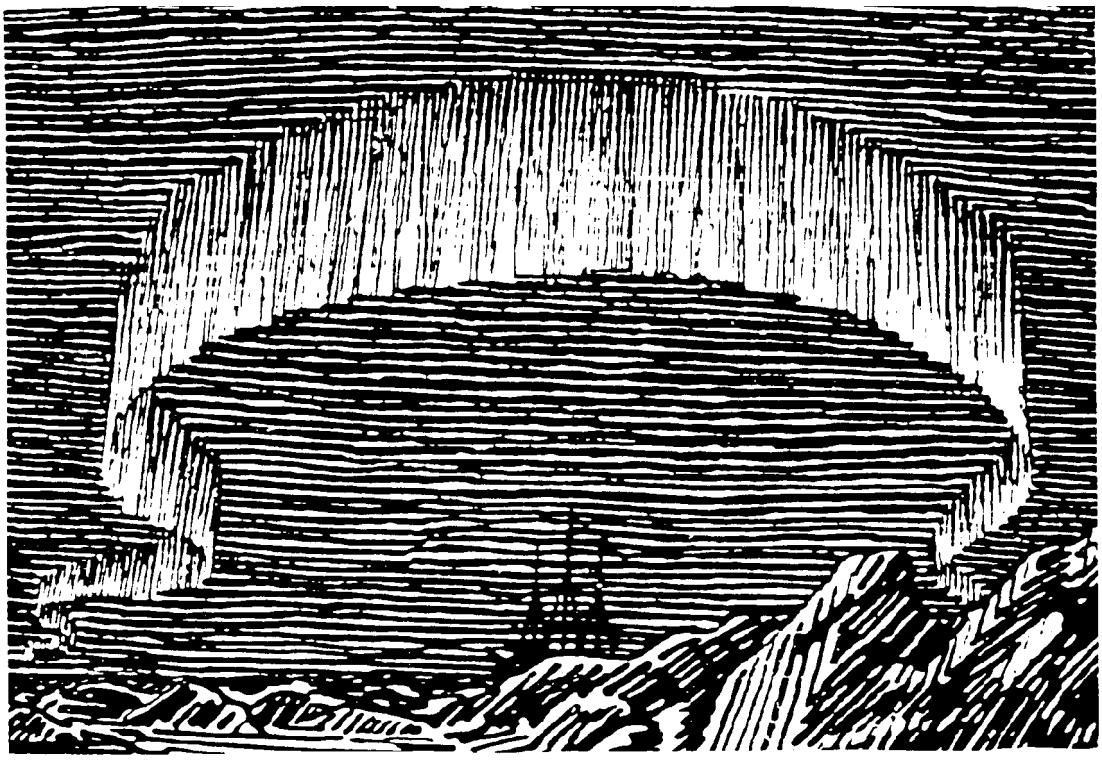
CAND. SCIENT. THESIS
UNIVERSITY OF TROMSØ

A Detection Theoretical Approach to Digital Communications Using Autoregressive Process Shift Keying

By
Stian Normann Anfinsen
February 2000



Original by F. Nansen

FACULTY OF SCIENCE
Physics Department
Auroral Observatory
University of Tromsø, 9037 Tromsø, Telephone: +47 77 64 51 50, Telefax: +47 77 64 55 80

Acknowledgements

I thank my supervisor, Associate Professor Alfred Hanssen, for being everything that one can expect from a supervisor, and a little more. I thank him for sharing of his enthusiasm and energy as well as his knowledge, and for motivating me at the times when I needed it the most.

I thank Ph.D. student Arnt-Børre Salberg, who has been working on the same project as me, for the time spent at rewarding discussions, for proposals and vital contributions to my understanding of the problems under study. His help has been greatly appreciated. I also thank the other students and members of staff that have contributed to the excellent learning environment and social environment at the Department.

Special thanks go to my family, who has supported me throughout my student career. Moreover I thank my friends and colleagues, within and outside the University, for giving me plenty of opportunities to forget the world of Physics and Mathematics. In that concern, my deepest thanks go to Kjersti.

Contents

Chapter 1

Introduction

In conventional digital communications, transmission of a bitstream over a channel is performed by modulating certain apects of a deterministic carrier wave. Familiar examples include amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK) [Gibson 1993, Proakis 1995]. The receiver estimates the parameters of the deterministic information-carrying signal and uses some detection rule to classify the received waveform as one of the possible parametric signals.

Conventional methods provide no protection against eavesdropping and unauthorised decoding of the signal. Recent methods promising some amount of protection against eavesdropping include so called spread spectrum techniques [Dixon 1994, Peterson et al. 1995, Viterbi 1995, Glisic and Vucetic 1997, Ojanpera and Prasad 1998] and chaotic digital encoders [Frey 1993, Brownhead et al. 1995, Aislam and Edwards 1996, Lee et al. 1997]. Such techniques demand precise synchronisation between transmitter and receiver. Even small synchronisation errors may cause high bit error rates (BER) at the receiver.

In this project a new concept of digital communication has been studied, which is based on realisations of stochastic processes as information-carrying signals. The concept has an inherent security against eavesdropping. At the same time, it is possible to device decoders that are simpler than those of spread spectrum and chaotic encoding.

The project aims to address some fundamental issues concerning the new technique: (i) How can the distance between the information carrying stochastic processes be measured in a statistical sense? (ii) What detector should be used to decode the information sequence modulated by stochastic processes and how does it perform? (iii) How should the stochastic transmission processes be chosen?

1.1 Secure Communications

The purpose of this project is to investigate and develop aspects of a new modulation technique with applications in secure digital communication. By secure communications we mean information transmission which is protected against attempts by unauthorised listeners to capture the information. Such hostile activity is also known as eavesdropping.

Most conventional modulation techniques offer no protection against eavesdropping. Examples of commercial modulation schemes are frequency shift keying (FSK), amplitude shift keying (ASK) and phase shift keying (PSK) [Gibson 1993, Proakis 1995]. All such signals are, as should they be, easily decoded by any receiver. Instead, security is normally provided by encryption [Welsh 1988, Golomb et al. 1994, Goldreich 1999]. Encryption is defined as the process of disguising data so that they become unintelligible to an unauthorised receiver.

In electronic computers, data is encrypted by applying mathematical operations on the information sequence, i.e. the bit stream that is produced at the transmitter. There are two kinds of basic operations: rearrangement of data without changing the symbols themselves (transposition), and substitution of data (single symbols or blocks of symbols) with other symbols or blocks of symbols without changing the sequence in which they occur. Modern encryption algorithms implement these operations through complex nonlinear schemes.

A personal encryption key, known only to the transmitter and intended receiver, controls the encryption algorithm. It ensures that the encrypted data can only be decrypted with the same key (symmetric encryption) or an associated key (asymmetric or public-key encryption). The best encryption algorithms are considered almost impregnable.

We will distinguish between two types of approaches to secure communications. First, we have the techniques that operate directly on the source symbols by altering the information sequence, as described above. Secondly, there are techniques that are concerned with the representation of source symbols when they are transmitted through a physical medium. The idea is to obscure the identity of the source symbols in the demodulation process. If this works, an eavesdropper will not be able to decode the information sequence from the physical waveform that is received through the medium.

Hence, approaches to secure communications are divided into two main categories:

- Code layer methods
- Physical layer methods

The most common approaches are found in the first group, which spans over the wide field of cryptography and coding theory. However, the technique which has been studied in this project belongs to the second category. This group also includes spread spectrum techniques, an approach to secure communications which has been investigated in interest of military applications for a long time. Another technique, which has been proposed more recently, is chaotic encoding. These methods will be explained in more detail subsequently.

We like to see the proposed technique as a supplement to, rather than a competitor to the methods in the first group. Physical layer security does not exclude the need for code layer security, and vice versa. In fact, one would often combine encryption with physical layer methods.

1.2 Spread Spectrum Techniques

The recent interest in spread spectrum communications (SSC) has been associated with applications like the global positioning system (GPS) and code division multiple access (CDMA), which is a multiuser system for personal mobile communications [Dixon 1994, Viterbi 1995, Ojanpera and Prasad 1998]. Nevertheless, the consept was first developed for secure communications in military applications [Glisic and Vucetic 1997]. The first approaches were undertaken more than half a century ago.

The original idea behind SSC is that a narrowband carrier signal will be more resistant to intensional interference from a hostile source if it is spread over a larger bandwidth. Let $x_{nb}(t)$ be a narrowband information signal with signal power P_x and bandwidth B_{nb}. Next, let $i_{nb}(t)$ be an intensive jamming signal with signal power $P_i > P_x$. The jamming signal is also relatively narrowband.

Define an invertible linear spreading operator $\mathcal{S}[\cdot]$ with the property that $\mathcal{S}[\cdot] = \mathcal{S}^{-1}[\cdot]$. The spreading operator transforms the narrowband information signal into a wideband signal $x_{wb}(t) = \mathcal{S}[x_{nb}(t)]$ with bandwidth $B_{wb} \gg B_{nb}$ before it is transmitted. Thus, if the jamming signal is present in the communications channel, the receiver receives the sum $x_{wb}(t) + i_{nb}(t)$ and applies the inverse spreading operator to obtain

$$\begin{aligned} \mathcal{S}^{-1}[x_{wb}(t) + i_{nb}(t)] &= \mathcal{S}[x_{wb}(t)] + \mathcal{S}[i_{nb}(t)] \\ &= x_{nb}(t) + i_{wb}(t) \end{aligned} \tag{1.1}$$

where $i_{wb}(t) = \mathcal{S}[i_{nb}(t)]$. The result can now be filtered by passband filter that matches

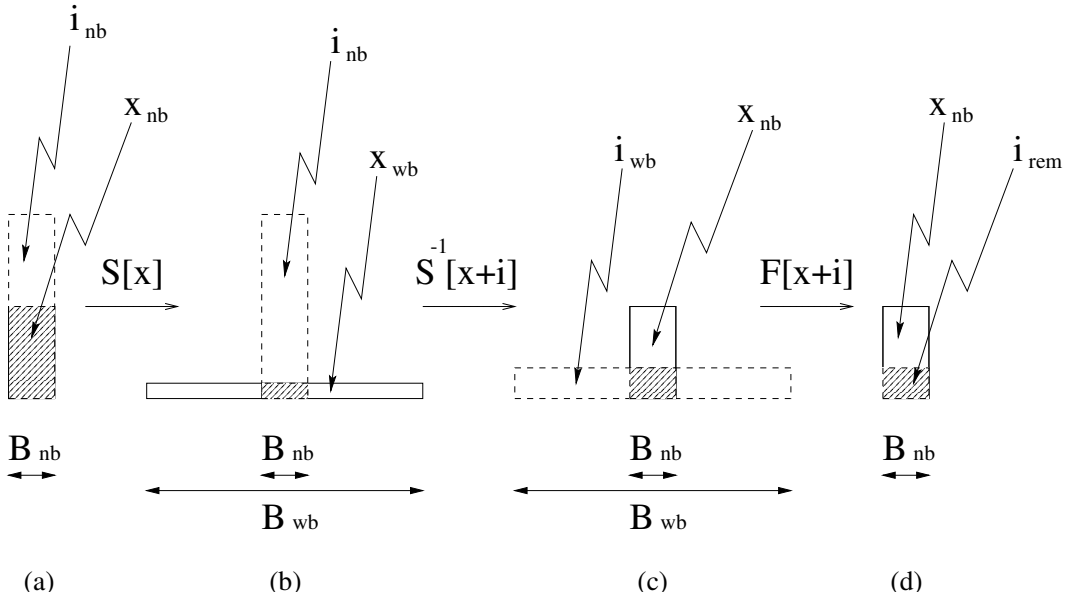


Figure 1.1: Antijamming principle in spread spectrum communications.

the bandwidth of $x_{nb}(t)$. Let the filter operation be denoted by $\mathcal{F}[\cdot]$. We then have

$$\mathcal{F}[x_{nb}(t) + i_{wb}(t)] = x_{nb}(t) + i_{rem}(t) \tag{1.2}$$

where $i_{rem}(t)$ is the remainder of the interference signal after bandpass-filtering. If $i_{wb}(t)$ is white, then only a fraction B_{nb}/B_{wb} of its signal energy will pass through the filter. Hence, the signal power of $i_{rem}(t)$ is $P_i(B_{nb}/B_{wb}) \ll P_x$, which explains that spreading of the signal bandwidth is an efficient tool to combat jamming.

The described antijamming procedure is illustrated by figure 1.1. The width of the rectangles represents the relative bandwidths of the assigned signals and the height represents the relative signal power. A solid rectangle denotes the information signal, while a dashed rectangle denotes the interference signal. The shaded areas represent the degree of interference or destructive jamming. The different stages are: (a) before spreading, (b) after spreading the information signal at the transmitter, (c) after despreading the received signal, (d) after bandpass-filtering the despread received signal.

We shall now explain how the spreading operation is performed, with reference to figure 1.2. If we assume that the information signal is a discrete bipolar sequence $x_{nb}(n)$ (e.g., it takes only the values $x_{nb}(n) = \pm 1$), then spreading is achieved by modulating the information sequence with a bipolar pseudo-random noise (PN) sequence $c(n)$. THe PN-sequence is also referred to as a chip sequence. For bipolar sequences, plain multiplication

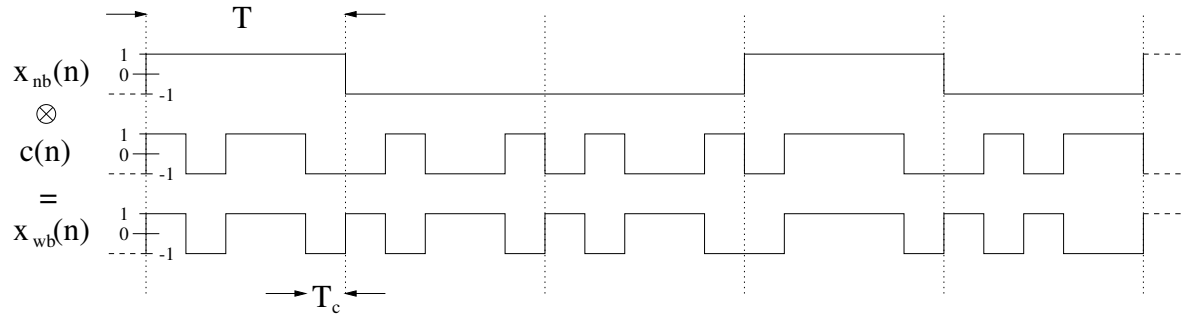


Figure 1.2: Modulation of discrete bipolar sequence x_{nb} with a discrete bipolar pseudorandom noise sequence $c(n)$ (chip sequence). The information rate is $1/T$, while the chip rate and the data rate of the PN-modulated sequence $x_{wb}(n)$ is $1/T_c$.

can be used as the spreading operator, as shown in the figure. For a unipolar signal (which takes only the value $x_{nb}(n) \in [0, 1]$), the requirement $\mathcal{S}[\cdot] = \mathcal{S}[\cdot]^{-1}$ is satisfied by the modulo 1 addition operator: $\mathcal{S}[x_{nb}(n)] = \{[x_{nb}(n) + c(n)] \mod 1\}$.

As illustrated by the figure, the chip rate $1/T_c$ should be mush higher than the information rate $1/T$, since the degree of spreading is proportional to T/T_c. The idea is that the transmitted wideband signal $x_{wb}(n)$ should be as uncorrelated and noise-like as possible. The PN-sequence is a deterministic and periodic sequence, and will never be truly random. Nevertheless, $c(n)$ can be chosen as a sequence which asympotically satisfies certain randomness criteria [Golomb 1967, Viterbi 1995] as the sequence period increases. Hence, the desired effect is obtained if the period of $c(n)$ is sufficiently large.

1.3 Chaotic Encoding

Chaos theory [Drazin 1992, Strogatz 1994] has been developed by physicists and mathematicians to describe apparently random or unpredictable behaviour generated by simple deterministic systems. Chaotic behaviour is observed in some nonlinear systems as a result of sensitivity to initial conditions. The interest in chaos in the fields of signal processing and communications has arised mainly because the signals produced by such deterministic systems may look like noise when displayed in either the time or frequency domain [Giannakis 1999, Lee et al. 1997].

Let the system state at a given time be a point in state space. The time development of a chaotic system can then be described by a trajectory in state space. Any slight change

in initial conditions creates a totally different state space trajectory. That is, two identical chaotic systems with nearly identical initial conditions will diverge. The trajectories are deterministic, but one cannot predict a future state without knowing the initial conditions exactly.

Despite the divergence property, Tang et al. [Tang et al. 1983] discovered that identical chaotic behaviour can be achieved by isolated systems. The theoretical framework was further developed by Pecora et al. [Pecora and Caroll 1990, Pecora and Caroll 1991, Ditto and Pecora 1993]. They proved that for certain stable systems, two separate systems driven by the same chaotic signal can be synchronised. Different curcuits that exhibit this synchronising property have been proposed [Chua et al. 1993, Cuomo et al. 1993]. They can be used to implement synchronised chaotic systems that suppress rather than enhance differences between them, thus enabling secure communications by means of chaotic encoding.

Lee [Lee et al. 1997] classifies existing secure communications schemes based on chaotic signals and systems into four categories. The first is chaotic modulation, where a wideband chaotic signal is used to modulate the information sequence. The chaotic signal is aperiodic and multivalued, which makes it suited as a spreading sequence. The drawback is that generation is critically sensitive to initial conditions.

Second, chaotic switching is a group of techniques where different source symbols are mapped to distinct chaotic signals. The schemes differ by the way signals and decision statistics are chosen. Again, sensitivity to initial conditions is the main practical hinder.

A third category is chaotic masking. The information signal is masked by adding a chaotic signal, and one of the described self-synchronising curcuits is used to extract the information at the receiver. Synchronisation is possible only when the power of the information signal is sufficiently smaller than the masking signal. Thus, synchronisation is sensitive to additive noise.

The fourth category is chaotic parameter modulation. Parameters of the carrier signal are perturbed at the transmitter by a chaotic signal. The information signal is recovered by use of a self-synchronising curcuit at the receiver. Also this technique suffers because the receiver requires high signal-to-noise ratio (SNR).

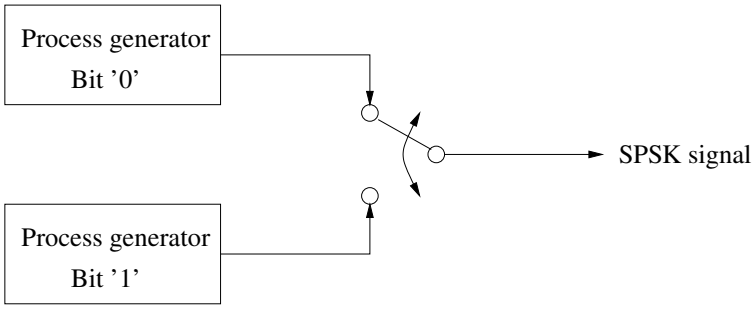


Figure 1.3: Generation of SPSK signal at transmitter.

1.4 Stochastic Process Shift Keying

The new technique coined Stochastic Process Shift Keying (SPSK) was first proposed by Hanssen in [Hanssen 1997]. The concept was developed further by Salberg and Hanssen in [Salberg and Hanssen 1999a, Salberg and Hanssen 1999b, Salberg and Hanssen 2000].

The idea behind SPSK is rather simple. Bit '0' of a binary signal is represented by the stochastic process $X_0(t)$. Bit '1' is represented by another stochastic process $X_1(t)$ with different parameters. The transmitter consists of two stochastic process generators and a switch between these, as shown in figure 1.3. Bit '0' is transmitted as a realisation of the process $X_0(t)$, $0 \leq t \leq T$ and bit '1' as a realisation of the process $X_1(t)$, $0 \leq t \leq T$, where T is the symbol period or Baud interval.

The continuous processes $X_0(t)$ and $X_1(t)$ can be made discrete by sampling them N times on the Baud interval $0 \leq t \leq T$. This produces the discrete stochastic processes $X_0(n)$ and $X_1(n)$, where n is the discrete time argument. The realisation $x(n)$ of any of the discrete stochastic processes is a sequence of N samples. Generation of a certain sequence $x(n)$, $n = 1, \ldots, N$ is associated with the probability

$$P\Big([X_i(1), \ldots, X_i(N)] = [x(1), \ldots, x(N)]\Big), \; i = 0, 1. \tag{1.3}$$

SPSK has two fundamental properties, due to the stochastic nature of the carrier signal. First, we note that two equal source bits will always be transmitted as different physical waveforms. Secondly, two different source bits will be transmitted as statistically similar, but not equal, waveforms. In addition, the stochastic signal is noise-like, which makes it difficult for unauthorised listeners to determine whether a meaningful message is sent at all.

Different processes can be used as carrier signals. A natural choice is linear Gaussian processes, or autoregressive/moving-average (ARMA) processes [Kay 1993, Box et al. 1994]. This class of processes has a simple structure, they have been extensively studied and have simple detectors. Other choices could be flicker noise ($1/f^{\gamma}$) processes [Mandelbrot 1999, Malakhov and Yakimov 1993, West and Schlesinger 1990] with different spectral exponents γ_0 and γ_1, bilinear and nonlinear processes [Priestley 1988]. Chaotic communications [Lee et al. 1997] can be viewed as a special case of SPSK with nonlinear processes. On the whole, there is a lot of freedom is the choice of processes.

In this thesis, we have restricted ourselves to a study of SPSK with autoregressive (AR) processes, a technique which will be referred to as autoregressive process shift keying (ARPSK). The AR-process is preferred to the moving-average (MA) process and the ARMA process because it is more resistant to additive white noise. The power spectral density (PSD) of a (higher-order) MA-process typically contains notches, wheras the PDS of a (higher-order) AR-process typically contains peaks. As a feature that contribute to detectability, a peak is more robust to white noise since notches can be "drowned" in the power of additive noise. The AR-process will thus be thoroughly presented and discussed in the following. At this point, it is sufficient to note that an AR-process has an order p, which specifies its number of characteristic parameters (disregarding the driving noise variance).

In order to capture the information in an ARPSK signal, an unauthorised listener will have to estimate the process order p and the parameters of the two AR-processes (p parameters each), as well as the pulse length N and synchronisation delay. Apart from the synchronisation, these figures are all known a priori to the authorised (intended) receiver.

ARPSK communications is an attempt to conceal information behind the variance in the estimates of the unknown parameters. The challenge is to specify processes whose AR-parameters are close enough to prevent eavesdropping, while at the same time enabling the decoder to meet the required bit error rate.

1.5 Overview of Thesis

Chapter one gives an overview of the problem. Secure communications is defined and some existing approaches are presented. The principles and properties of the novel technique coined stochastic process shift keying (SPSK) is described, before we restrict the choice of

stochastic transmission processes to autoregressive (AR) processes.

Chapter two gives the necessary theoretical background for a study of autoregressive process shift keying (ARPSK), with emphasis on probability and statistical signal theory. Chapter three is a review of statistical distance measures that can be used to quantify the distance between two autoregressive transmission processes.

In chapter four we propose two detectors for the ARPSK communications system and derive their respective detection error probability. We also assess the effect of additive white noise and synchronisation errors on the detectors. In chapter five we propose a set of criteria for selection of the transmission processes. The discussion of these criteria leads to a process selection procedure.

In chapter six we evaluate the theoretical expressions for the detection error probabilities of the deviced detectors. These results are compared with the results of numerical simulations. In chapter seven we give the conclusions of the thesis and propose topics of future research.

Chapter 2

Fundamentals on Autoregressive Processes

2.1 Stochastic Processes

A stochastic process is a waveform exhibiting some kind of random behaviour. In constrast to a deterministic signal, whose signal value is fully specified for all argument values, a stochastic process must be specified by the joint probability density function (PDF) of its possible outcomes [Papoulis 1991, Peebles 1993].

The stochastic process can be a deterministic waveform with a stochastic parameter, e.g. $X(t) = \sin(\omega_0 t + \Theta)$ where the phase Θ is a random variable taking on values $0 \leq \theta \leq 2\pi$. $X(t)$ is clearly deterministic after Θ is realised. The waveform can also be entirely random, like a noise signal. In this case, there exists no functional form of $X(t)$.

In engineering problems we encounter stochastic processes both as the signal of interest, and as noise that is contaminating our desired signal, whether it be stochastic or deterministic. In some cases, the nature of a process is truly stochastic. More commonly, the underlying physical model is so complex that stochastical modelling is the most practical approach.

A stochastic process is a generalisation of stochastic variables, to include one or more dimensions. Both the stochastic process and the independent variable can be continuous or discrete. We will be concerned only with stochastic processes as a continuous function of discrete time n. To specify a stochastic process $X(n)$ of length N, we thus need to know the PDF $f_X(x_1, \ldots, x_N)$, where x_i is the sample realisation of $X(n)$ at discrete time $n = i$.

2.2 Gaussian Probability Density Function

The Gaussian probability density function is without doubt the most important probability distribution in science and engineering. The joint PDF of N Gaussian random variables, denoted $\mathbf{X} = [X(1), \ldots, X(N)]^T$, is given by [Papoulis 1991, Peebles 1993]

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{2.1}$$

where $\boldsymbol{\mu} = E\{\mathbf{X}\}$ and $\mathbf{\Sigma} = E\{(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T\}$ are the mean vector and covariance matrix, respectively, and $|\mathbf{\Sigma}|$ denotes the determinant of $\mathbf{\Sigma}$. For $\boldsymbol{\mu} = \mathbf{0}$, $\mathbf{\Sigma}$ reduces to the correlation matrix $\mathbf{R} = E\{\mathbf{X}\mathbf{X}^T\}$, and the PDF becomes [Peebles 1993]

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}|\mathbf{R}|^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{R}^{-1}\mathbf{x}\right\}. \tag{2.2}$$

which is the form that will see all through the thesis.

A Gaussian distribution is completely specified by $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ [Peebles 1993]. Thus, the notation $\mathcal{N}[\boldsymbol{\mu}, \mathbf{\Sigma}]$ is a specification of a multivariate Gaussian PDF. When $\mathbf{x}$ is zero-mean, a necessary and sufficient description is $\mathcal{N}[\mathbf{0}, \mathbf{R}]$.

2.3 Central χ^2 Probability Density Function

Another important probability distribution is the χ^2 distribution. The χ^2 probability density function and other PDFs with similar functional form appear when we deal with quadratic forms in multivariate Gaussian random variables. The sum

$$Y = \sum_{i=1}^{N} X_i^2 \tag{2.3}$$

is centrally χ^2 distributed with N degrees of freedom when the X_i are statistically independent and identically distributed (i.i.d.) $\mathcal{N}[0,1]$ random variables (standardised Gaussian variables) [Scharf 1991]. This is denoted: $Y \sim \chi^2_N$. Equivalently, this is also true for the sum

$$Y = \sum_{i=1}^{N} (X_i - \mu)^2/\sigma^2 \tag{2.4}$$

when the $X_i \sim \mathcal{N}[\mu, \sigma^2]$. In any case, the PDF of Y is [Scharf 1991]

$$f_Y(y) = \frac{1}{\Gamma(N/2)\,2^{N/2}}\, y^{\frac{N-2}{2}} e^{-y/2} \tag{2.5}$$

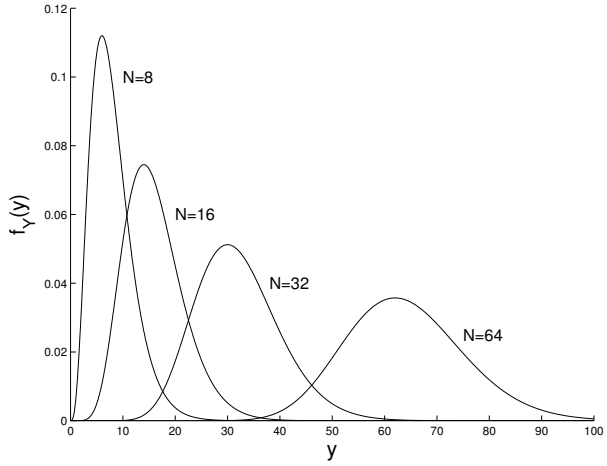


Figure 2.1: Probability density functions of central χ^2-distributed variables.

where the gamma function is defined as

$$\Gamma(N) = \int_0^\infty e^{-t}\, t^{N-1}\, dt, \quad \text{for } N > 0\ . \tag{2.6}$$

The mean value and the variance of a χ^2_N distributed variable Y is $E\{Y\} = N$ and $Var\{Y\} = 2N$, respectively. The maximum of the PDF occurs at $N-2$ and the skewness is $2\sqrt{2/N}$ [Scharf 1991].

A set of χ^2_N PDFs for different choices of N is shown in figure 2.1. From the figure we see that the the mean and the PDF maximum increases with increasing N in agreement with theory, and so does the variance. We also see that the skewness decreases as N increases. In the limiting case, when $N \to \infty$, the skewness vanishes and the χ^2_N PDF approaches a one-dimensional Gaussian PDF specified by $\mathcal{N}[N, 2N]$.

A more general result exists for the multivariate Gaussian random variable $\mathbf{X} \sim \mathcal{N}[\boldsymbol{\mu}, \boldsymbol{\Sigma}]$. The quadratic form

$$Q = (\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \tag{2.7}$$

is χ^2_N distributed. If the sequence $\{X(n)\}$, $n = 1, \ldots, N$ is white Gaussian noise, then Q reduces to Eq. (2.4), with PDF given by Eq. (2.5).

The characteristic function of a multivariate random variable $\mathbf{X}$ is defined as [Scharf 1991, Peebles 1993]

$$\begin{aligned} \Phi_{\mathbf{X}}(\boldsymbol{\omega}) &= \mathrm{E}\{e^{j\boldsymbol{\omega}^T\mathbf{X}}\} \\ &= \int f_{\mathbf{X}}(\mathbf{x}) e^{j\boldsymbol{\omega}^T\mathbf{X}}\, d\mathbf{X} \end{aligned} \tag{2.8}$$

where $j = \sqrt{-1}$, $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_N]^T$ and $\int(\cdot)\, d\mathbf{x}$ denotes a multi-dimensional integral. The close relationship between the characteristic function and multi-dimensional Fourier transform of $f_{\mathbf{X}}(\mathbf{x})$ is obvious. It is easily found that

$$\Phi_{\mathbf{X}}(-\boldsymbol{\omega}) = \mathfrak{F}\{f_{\mathbf{X}}(\mathbf{x})\} \tag{2.9}$$

where $\mathfrak{F}\{\cdot\}$ denotes the multi-dimensional Fourier transform. This result that will be used in subsequent chapters. The other major application of $\Phi_{\mathbf{X}}(\boldsymbol{\omega})$ is that it enables calculation of moments. The mth moment of $\mathbf{X}$ is given by [Papoulis 1991, Peebles 1993]

$$\mathrm{E}\{\mathbf{X}^m\} = (-j)^m \left.\frac{d^m \Phi_{\mathbf{X}}(\boldsymbol{\omega})}{d\boldsymbol{\omega}^m}\right|_{\boldsymbol{\omega}=\mathbf{0}} \quad . \tag{2.10}$$

The characteristic function of Q is readily found as [Scharf 1991]

$$\Phi_Q(\omega) = \frac{1}{(1-2j\omega)^{N/2}} \quad . \tag{2.11}$$

From $\Phi_Q(\omega)$, the mean and variance of Q is obtained as [Scharf 1991]

$$E\{Q\} = N \tag{2.12}$$

$$Var\{Q\} = 2N \; . \tag{2.13}$$

A more general result exist for the quadratic form

$$\tilde{Q} = (\mathbf{X}-\boldsymbol{\mu})^T \mathbf{P}\,(\mathbf{X}-\boldsymbol{\mu}) \tag{2.14}$$

in the symmetric matrix $\mathbf{P}$. The characteristic function of the modified $\tilde{Q}$ is found from a straight-forward derivation [Scharf 1991] as

$$\Phi_{\tilde{Q}}(\omega) = \frac{1}{|\mathbf{I}-2j\omega\mathbf{PR}|^{1/2}} \quad . \tag{2.15}$$

The mean and variance now become

$$E\{\tilde{Q}\} = \mathrm{tr}(\mathbf{PR}) \tag{2.16}$$

$$Var\{\tilde{Q}\} = 2\mathrm{tr}\big((\mathbf{PR})^2\big) \tag{2.17}$$

where the trace operator applied on a $N \times N$ matrix $\mathbf{A}$ is defined as the sum of all elements on the main diagonal:

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{N} [\mathbf{A}]_{ii} \; . \tag{2.18}$$

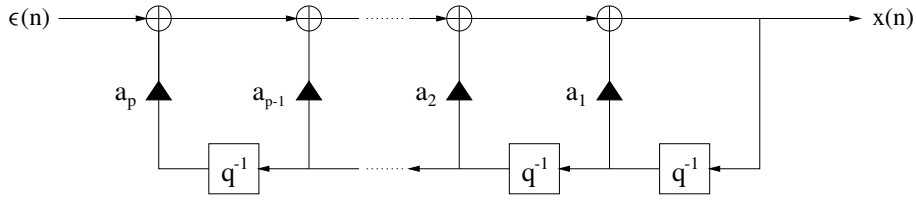


Figure 2.2: Block diagram of AR-process.

2.4 Definition of an AR-process

Many real-world signals can be described by stochastic processes assuming parametric models. One such model is the autoregressive (AR) process [Kay 1993, Box et al. 1994]. In an AR-model of order p, the present output $x(n)$ depends on a linear combination of the p previous outputs, driven by a random component $\epsilon(n)$, which is termed the driving noise of the process. This has the mathematical form

$$x(n) = -\sum_{i=1}^{p} a_i x(n-i) + \epsilon(n). \tag{2.19}$$

We assume that the driving noise is zero-mean, white and Gaussian, i.e. that $E\{\epsilon(n)\} = 0$ and $E\{\epsilon(n)\epsilon(n+k)\} = \sigma_\epsilon^2 \delta_{k,0}$ where $\delta_{k,0}$ is the Kronecker delta function. The output signal is then completely specified by the AR-parameters $a_i, i = 1, \ldots, p$ and the variance σ_ϵ^2 of the driving noise. A block diagram of an AR-process is shown in figure 2.2. The unit time delay operator is denoted q^{-1}.

The AR-process $x(n)$ can also be interpreted as a filtered version of the driving noise $\epsilon(n)$. In the time domain, the filtering operation is equal to the convolution

$$x(n) = h(n) * \epsilon(n) \tag{2.20}$$

where the filter has infinite impulse response (IIR) $h(n)$ [Oppenheim et al. 1983]. The filter interpretation is shown in figure 2.3. The filter coefficients cannot be written explicitly, but

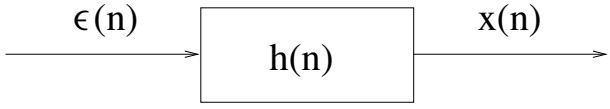


Figure 2.3: Filter interpretation of the AR-process.

they are given by the inverse discrete Fourier transform (IDFT) [Oppenheim et al. 1983]

$$h(n) = \frac{1}{2\pi} \int_{2\pi} H(\omega) e^{j\omega n} \, d\omega \tag{2.21}$$

where $H(\omega) = 1/A(\omega)$, as will be shown in the next section. Here, $H(\omega)$ and $A(\omega)$ is the discrete Fourier transform (DFT) of the filter impulse response and the AR-parameters, respectively.

2.5 Power Spectral Density

Define the AR-parameter of order zero as $a_0 = 1$. We may then rewrite the AR-model in Eq. (2.19) using the time domain operator q^{-d} representing a delay of d discrete time units (symbol periods), to obtain

$$\begin{aligned} \epsilon(n) &= x(n) \sum_{k=0}^{p} a_k q^{-k} \\ &= a(n) * x(n) \end{aligned} \tag{2.22}$$

where $a(n) = [a_0 \, , \ldots , a_p]^T$. From statistical signal theory [Scharf 1991, Kay 1993, Peebles 1993] we know that for a linear time invariant (LTI) system with input-output relation

$$y(n) = h(n) * x(n) \tag{2.23}$$

the power spectral density (PSD) of the output $y(n)$ is

$$S_{yy}(\omega) = |H(\omega)|^2 S_{xx}(\omega) \tag{2.24}$$

where $H(\omega)$ is the DFT of the system impulse response and $S_{xx}(\omega)$ is the PSD of the input signal $x(n)$. Hence, from Eqs. (2.20) and (2.22) we obtain the relations

$$S_{xx}(\omega) = |H(\omega)|^2 S_{\epsilon\epsilon}(\omega) \tag{2.25}$$

$$S_{\epsilon\epsilon}(\omega) = |A(\omega)|^2 S_{xx}(\omega) \tag{2.26}$$

which proves that

$$H(\omega) = \frac{1}{A(\omega)} \, . \tag{2.27}$$

Since the PSD of white noise equals the noise variance, the PSD of $x(n)$ is given by [Kay 1993, Box et al. 1994]

$$S_{xx}(\omega) = \sigma_\epsilon^2 / \left| 1 + \sum_{k=1}^{p} a_k e^{-j\omega k} \right|^2 . \tag{2.28}$$

We next evaluate the denominator as an explicit function of the AR-parameters.

$$|A(\omega)|^2 = \sum_{k=0}^{p} a_k^2 + 2 \sum_{k=1}^{p} \left(\sum_{l=0}^{p-k} a_l a_{k+l} \right) \cos(k\omega). \tag{2.29}$$

The denominator function can also be written as

$$|A(\omega)|^2 = A_0 + 2A_1 \cos(\omega) + \ldots + 2A_p \cos(p\omega) \tag{2.30}$$

where the A_k are Fourier series cosine terms coefficients of the inverse PSD $1/S_{xx}(\omega)$, weighted by the driving noise variance [Itakura and Saito 1970]. Hence, we identify the relation

$$A_k = \frac{\sigma_\epsilon^2}{2\pi} \int_{-\pi}^{\pi} \frac{\cos(k\omega)}{S_{xx}(\omega)} = \sum_{l=0}^{p-k} a_l a_{k+l}. \tag{2.31}$$

We achieve rather simple expressions for the power spectral densities of lower order AR-processes. For instance, we have [Box et al. 1994]

$$\text{AR(1)}: \quad Sxx(\omega) = \sigma_\epsilon^2 [(1 + a_1^2) + 2a_1 \cos(\omega)]^{-1}. \tag{2.32}$$

$$\text{AR(2)}: \quad Sxx(\omega) = \sigma_\epsilon^2 [(1 + a_1^2 + a_2^2) + 2a_1(1 + a_2)\cos(\omega) + 2a_2 \cos(2\omega)]^{-1}. \tag{2.33}$$

The power spectral densities of two AR-processes of order $p = 2$ are shown in figure 2.4. The first process has AR-parameters $a_1 = 0.4$ and $a_2 = -0.2$. It is clearly a high-frequency process, which should be expected from the sign of a_1. The second process has AR-parameters $a_1 = -0.4$ and $a_2 = 0.2$. As a consequence, this is a low-frequency process. With increasing order, the features of the power spectral density become more complex.

Inserting $z = e^{j\omega}$ into Eq. (2.28), we see that an AR-process is causal and stable if the roots of the characteristic denominator polynomial $A(z)$ (the Z transform of the AR-parameters a(n) [Oppenheim et al. 1983]) all have magnitude less than one. That is, for stability of a causal AR-process, we require [Box et al. 1994]

$$|z_i| < 1 \quad \text{for} \quad \Big\{ z_i \; : \; \sum_{k=0}^{p} a_k z^{-i} = 0 \Big\}, \; i = 1, \ldots, p. \tag{2.34}$$

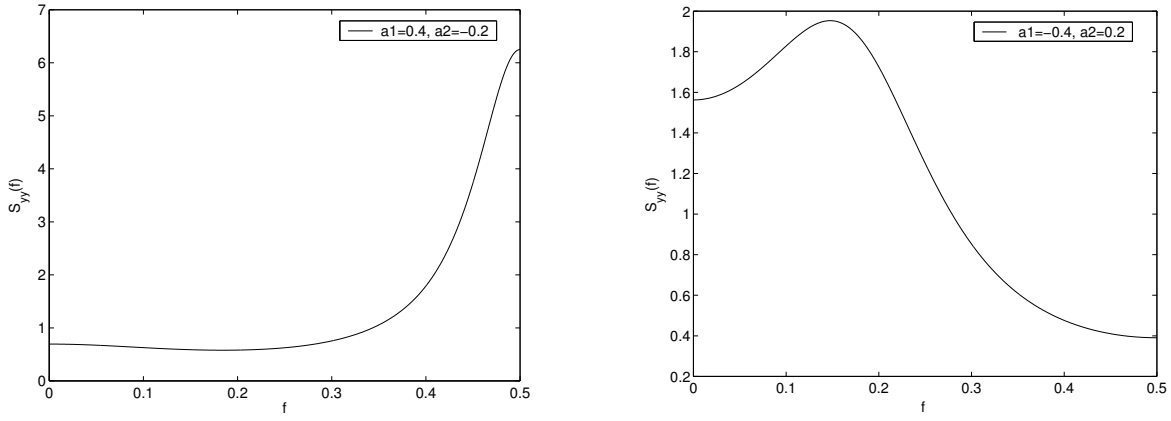


Figure 2.4: power spectral densities of two second-order AR-processes with parameters: $[a_1, a_2] = [0.4, -0.2]$ (left), $[a_1, a_2] = [-0.4, 0.2]$ (right) and $\sigma_\epsilon^2 = 1$ (both).

2.6 Autocorrelation Function

Since the AR-process regresses on previous values of itself, it has an infinite autocorrelation function (ACF). The ACF of an AR-process can be defined recursively, but the resulting expression become complex very rapidly with increasing order p.

To study the autocorrelation of an AR-process, we now derive the Yule-Walker equations [Box et al. 1994, Haykin 1996]. Starting from Eq. (2.19), we multiply both sides by $x(n-k)$ and take the expectation value,

$$E\left\{x(n)x(n-k) + \sum_{i=1}^{p} a_i x(n-i)x(n-k)\right\} = E\{\epsilon(n)x(n-k)\}. \tag{2.35}$$

The left hand side evaluates to a sum of scaled autocorrelations of varying time lag. The right hand side is non-zero only for zero lag ($k=0$), since the driving noise is uncorrelated. Hence, with the ACF of $x(n)$ defined as $r_{xx}(k) = E\{x(n)x(n+k)\}$ [Papoulis 1991, Peebles 1993], this becomes

$$r_{xx}(-k) + \sum_{i=1}^{p} a_i r_{xx}(i-k) = \sigma_\epsilon^2 \delta_{k,0}\ . \tag{2.36}$$

If we evaluate this equation for $k = 1, \ldots, p$, we obtain a set of equations in the AR-parameters, which can be rephrased as the well-known Yule-Walker equations [Box et al. 1994, Haykin 1996]

$$\sum_{i=1}^{p} a_i r_{xx}(-k+i) = -r_{xx}(-k),\ k = 1, \ldots, p. \tag{2.37}$$

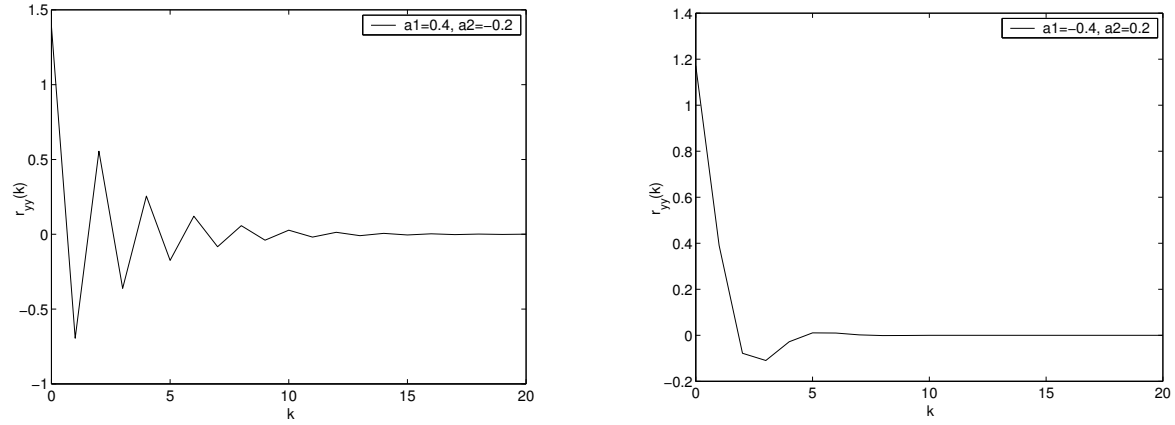


Figure 2.5: Autocorrelation functions of two second-order AR-processes with parameters: $a_1 = 0.4$, $a_2 = -0.2$ (left); $a_1 = -0.4$, $a_2 = 0.2$ (right) and $\sigma_\epsilon^2 = 1$ (both).

The expression gives an implicit solution for the ACF for $k > 0$. Eq. (2.36) evaluated for $k = 0$ provides an expression for the variance of an AR-process (referred to as process variance). Since $x(n)$ is zero-mean, we have $\sigma_x^2 = r_{xx}(0)$ and

$$\sigma_x^2 = \sigma_\epsilon^2 - \sum_{k=1}^{p} a_k r_{xx}(k). \tag{2.38}$$

Suppose that we write the Yule-Walker equations in Eq. (2.37) as $\sum_{i=0}^{p} a_i r_{xx}(k-i) = 0$ with $a_0 \triangleq 1$. We may then define the operator $A(q) = 1 + a_1 q^{-1} + \ldots + a_p q^{-p}$ with q^{-1} denoting a discrete unit time delay operator, such that $A(q) r_{xx}(k) = 0$. The operator $A(q)$ can also be written [Box et al. 1994]

$$A(q) = \prod_{i=1}^{p} (1 - q_i q^{-1}) \tag{2.39}$$

where the $\{q_i\}$ are roots of the characteristic equation $A(q) = 0$. The stability requirement again appears as the condition that $|q_k| < 1 \; \forall \, i$. The general solution of $A(q) r_{xx}(k) = 0$ is [Box et al. 1994]

$$r_{xx}(k) = \alpha_1 q_1^k + \ldots + \alpha_p q_p^k \tag{2.40}$$

for some constants $\{\alpha_i\}$. If a root q_i is real-valued, then the term $\alpha_i q_i^k$ is a damped exponential that decays to zero as k increases. If a pair of roots q_i, q_j are complex conjugates, their contribution to the ACF will be a damped sinusoid $|q_i|^k \sin(\omega k + \phi)$ with frequency [Box et al. 1994]

$$\omega = \cos^{-1}(|\mathrm{Re}\{q_i\}|/|q_i|) \; . \tag{2.41}$$

Figure 2.5 displays the autocorrelation function of two AR(2)-processes. The processes are the same there was used to generate the power spectral densities in figure 2.4. The effects of both damped exponentials and damped sinusoidals can be seen to appear in both functions.

2.7 Yule-Walker Estimation of AR-parameters

The system of equations in (2.37) can be expressed compactly on matrix form as

$$\mathbf{R}_x\mathbf{a} = -\mathbf{r}_x \tag{2.42}$$

where the correlation matrix is defined as $[\mathbf{R}_x]_{ij} = r_{xx}(i-j),\ i,j = 1,\ldots,p$ and the correlation vector as $\mathbf{r}_x = [r_{xx}(1),\ldots,r_{xx}(p)]^T$. Thus, we can solve for the parameter vector $\mathbf{a} = [a_1,\ldots,a_p]^T$ and insert estimates of $\mathbf{R}_x$ and $\mathbf{r}_x$ to obtain the Yule-Walker (YW) estimate [Box et al. 1994, Haykin 1996]

$$\hat{\mathbf{a}}_{YW} = -\hat{\mathbf{R}}_x^{-1}\hat{\mathbf{r}}_x\ . \tag{2.43}$$

Evaluation of Eq. (2.36) for $k = 0$ gives the variance of the driving noise as

$$\begin{aligned}\sigma_\epsilon^2 &= r_{xx}(0) + \mathbf{r}_x^T\mathbf{a} \\ &= r_{xx}(0) - \mathbf{r}_x^T\mathbf{R}_x^{-1}\mathbf{r}_x.\end{aligned} \tag{2.44}$$

The ACF can be estimated from a length N realisation of the AR-process, for instance using the biased estimator [Kay 1993]

$$\hat{r}_{xx}(k) = \frac{1}{N}\sum_{n=1}^{N-|k|} x(n)x(n+|k|),\ k = 0,\pm1,\ldots,\pm N. \tag{2.45}$$

Recalling that for real-valued data, $r_{xx}(k) = r_{xx}(-k)$, we need only estimate the ACF for non-negative lags. Inserting the estimated ACF-values for $k = 0,\ldots,p$ into Eqs. (2.43) and (2.44), the estimators $\hat{\mathbf{a}}$ and $\hat{\sigma}_\epsilon^2$ follow straight-forward.

2.8 Maximum Likelihood Estimation of AR-parameters

We have assumed that the driving noise of the AR-process is zero-mean, white and Gaussian. Hence, the PDF of the noise sequence $\boldsymbol{\epsilon} = [\epsilon(1), \ldots, \epsilon(N)]^T$ is

$$\begin{aligned} f_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) &= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{-\epsilon^2(n)/2\sigma^2} \\ &= (2\pi\sigma^2)^{-N/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{n=1}^{N} \epsilon^2(n) \right\}. \end{aligned} \tag{2.46}$$

From the definition of the AR-process, it follows that

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi\sigma^2)^{-N/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{n=1}^{N} \left[\sum_{i=0}^{p} a_i x(n-i) \right]^2 \right\} \tag{2.47}$$

which can be rewritten as

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi\sigma_\epsilon^2)^{-N/2} \exp\left\{ -\frac{N}{2\sigma_\epsilon^2} \boldsymbol{\alpha}^T \hat{\mathbf{R}}_x \boldsymbol{\alpha} \right\} \tag{2.48}$$

where we introduce $\boldsymbol{\alpha} = [a_0, \ldots, a_p]^T = [1, \mathbf{a}]^T$, and the $(p+1)\times(p+1)$ empirical correlation matrix, which is given by

$$\hat{\mathbf{R}}_x = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n^{p+1} \mathbf{x}_n^{p+1T}. \tag{2.49}$$

Here $\mathbf{x}_n^{p+1}$ denotes the sequence $[x(n-p), \ldots, x(n)]^T$ of $p+1$ samples of $x(n)$, up to and including discrete time n. Now consider the PDF as a likelihood function, by taking the parameter vector $\mathbf{a}$ to be the variable instead of $\mathbf{x}$, and denoting it $f_{\mathbf{X}}(\mathbf{x}|\mathbf{a})$. The maximum likelihood (ML) estimate [Scharf 1991, Box et al. 1994] of $\mathbf{a}$ is the parameter vector that maximises $f_{\mathbf{X}}(\mathbf{x}|\mathbf{a})$,

$$\hat{\mathbf{a}}_{ML} = \arg\left\{ \max_{\mathbf{a}} \{ f_\epsilon(\epsilon|\mathbf{a}) \} \right\}. \tag{2.50}$$

It is seen that a maximum is obtained when the term $\sum_{n=1}^{N} \epsilon^2(n)/N = \boldsymbol{\alpha}^T \hat{\mathbf{R}}_x \boldsymbol{\alpha}$ is minimum. Hence, the maximum likelihood estimate is identical to the least squares estimate, assuming a Gaussian model for the driving noise [Scharf 1991, Box et al. 1994]. We note that the least squares estimate is independent of the driving noise PDF.

Assume that the innovation variance is unknown. The maximum likelihood estimate of σ_ϵ^2 is obtained from the definition as $[\partial f_{\mathbf{X}}(\mathbf{x}|\mathbf{a})/\partial\sigma_\epsilon^2]|_{\hat{\sigma}_\epsilon^2} \triangleq 0$, which yields

$$\hat{\sigma}_\epsilon^2 = \boldsymbol{\alpha}^T \hat{\mathbf{R}}_x \boldsymbol{\alpha}. \tag{2.51}$$

Maximum likelihood estimates for the AR-parameters are obtained component-wise from $\partial f_{\mathbf{X}}(\mathbf{x}|\mathbf{a})/\partial a_k|_{\hat{a}_k} \triangleq 0$. Using Eq. (2.47), the result is the system of equations [Scharf 1991, Box et al. 1994]

$$\sum_{i=0}^{p} a_i \sum_{n=1}^{N} x(n-i)x(n-k) = 0, \quad k = 1, \ldots, p \tag{2.52}$$

which can be rewritten as

$$\sum_{i=1}^{p} a_i \hat{r}_{xx}(k-i) = -\hat{r}_{xx}(k), \quad k = 1, \ldots, p. \tag{2.53}$$

This is seen to be precisely the Yule-Walker estimate.

2.9 Approximate Likelihood Function

Since the AR-process is a linear combination of Gaussian variables, the PDF of $x(n)$ is also Gaussian. Due to the complex correlation matrix, it is difficult to derive an exact PDF or likelihood function for the AR-process in terms of the AR-parameters. In a search for an approximate likelihood function, the exact likelihood function can be factorised into [Itakura and Saito 1970, Box et al. 1994]

$$f(\mathbf{x}_n^N|\mathbf{a}, \sigma_\epsilon^2) = f(\mathbf{x}_n^{N-p}|\mathbf{x}_p^p, \mathbf{a}, \sigma_\epsilon^2) f(\mathbf{x}_p^p|\mathbf{a}, \sigma_\epsilon^2) \tag{2.54}$$

where the notation $\mathbf{x}_n^N$ should still be read as the length N sequence ending with the datum $x(n)$. In [Box et al. 1994] it is shown that this results in

$$\begin{aligned} f(\mathbf{x}_n^N|\mathbf{a}, \sigma_\epsilon^2) &= \frac{C}{(2\pi\sigma_\epsilon^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \left[\boldsymbol{\alpha}^T \mathbf{X} \boldsymbol{\alpha} + \sum_{i=p+1}^{N} \boldsymbol{\alpha}^T \mathbf{x}_i^{p+1} \mathbf{x}_i^{p+1\,T} \boldsymbol{\alpha} \right] \right\} \\ &= \frac{C}{(2\pi\sigma_\epsilon^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \boldsymbol{\alpha}^T \left[\mathbf{X} + (N-p)\hat{\mathbf{R}}_x \right] \boldsymbol{\alpha} \right\}. \end{aligned} \tag{2.55}$$

where we define the correlation matrix estimator as

$$\hat{\mathbf{R}}_x = \frac{1}{N-p} \sum_{i=p+1}^{N} \mathbf{x}_i^{p+1} \mathbf{x}_i^{p+1\,T}. \tag{2.56}$$

The constant $C = |\mathbf{R}_x/\sigma_\epsilon^2|^{1/2}$ and the elements of the $(p+1) \times (p+1)$ matrix $\mathbf{X}$ are

$$[\mathbf{X}]_{ij} = x_i x_j + x_{i+1}\, x_{j+1} + \ldots + x_{n+1-i}\, x_{n+1-j} \tag{2.57}$$

with summations consisting of $n-(i-1)-(j-1)$ terms. The last term of the exponent in Eq. (2.55) will dominate for $N \gg p$, and the likelihood function can be approximated by [Itakura and Saito 1970, Box et al. 1994]

$$f(\mathbf{x}_n^N|\mathbf{a},\sigma_\epsilon^2) = \frac{C}{(2\pi\sigma_\epsilon^2)^{N/2}} \exp\left\{-\frac{N}{2\sigma_\epsilon^2}\boldsymbol{\alpha}^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}\right\}. \tag{2.58}$$

The log-likelihood function is defined as

$$l(\mathbf{x}_n^N|\mathbf{a},\sigma_\epsilon^2) = \ln f(\mathbf{x}_n^N|\mathbf{a},\sigma_\epsilon^2) \tag{2.59}$$

An approximation for the log-likelihood function is thus found from Eq. (2.58) as

$$\ell(\mathbf{x}_n^N|\mathbf{a},\sigma_\epsilon^2) = \ln C - \frac{N}{2}\ln 2\pi\sigma_\epsilon^2 - \frac{N}{2\sigma_\epsilon^2}\boldsymbol{\alpha}^T\hat{\mathbf{R}}_x\boldsymbol{\alpha} \tag{2.60}$$

invoking the same assumption on the length of the data sequence. Under these approximations, we observe that Eqs. (2.48) and (2.58) have the same mathematical form. The maximum likelihood estimates for σ_ϵ^2 and $\mathbf{a}$ obtained from the likelihood function of the AR-model are therefore identical to those presented in section 2.8.

2.10 Approximate Log-Likelihood Ratio

The likelihood function $f(\mathbf{x}_n^N|\mathbf{a},\sigma_\epsilon^2)$ was introduced in section 2.8 as an equivalent to the PDF, when the statistical parameter vector of the probability model is regarded as the independent variable, after a data vector is observed. The log-likelihood function was defined in section 2.9, and denoted $\ell(\mathbf{x}_n^N|\mathbf{a},\sigma_\epsilon^2)$. We now define the log-likehood ratio as the logarithm of the ratio of two likelihood functions [Scharf 1991],

$$\begin{aligned} L(\mathbf{x}_n^N) &= \ln\left[f(\mathbf{x}_n^N|H_1)/f(\mathbf{x}_n^N|H_0)\right] \\ &= l(\mathbf{x}_n^N|H_1) - l(\mathbf{x}_n^N|H_0) \end{aligned} \tag{2.61}$$

where H_i denotes the hypotesis that $\mathbf{x}_n^N$ is a realisation of process $X_i,\ i \in [0,1]$. From now on, all process dependent function will be conditioned by the appropriate hypothesis, instead of respective parameter vector and innovation variance.

From the approximation of the log-likelihood function in Eq. (2.60), we now propose an approximate log-likelihood ratio (ALR)

$$\begin{aligned} \mathcal{L}(\mathbf{x}_n^N) &= \ell(\mathbf{x}_n^N|H_1) - \ell(\mathbf{x}_n^N|H_0) \\ &= \frac{N}{2}\left(\frac{1}{\sigma_{\epsilon_0}^2}\boldsymbol{\alpha}_0^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_0 - \frac{1}{\sigma_{\epsilon_1}^2}\boldsymbol{\alpha}_1^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_1\right) + \frac{N}{2}\ln\left(\frac{\sigma_{\epsilon_0}^2}{\sigma_{\epsilon_1}^2}\right). \end{aligned} \tag{2.62}$$

The interesting point about the ALR is that it can be used to derive approximations to Bayes detectors and Neyman-Pearson detectors [Scharf 1991]. Detection theory [Kazakos and Papantoni 1990, Scharf 1991] will be discussed in a later chapter, and the ALR will be used to design one of the alternative detectors in ARPSK communications. The simple mathematical form and low complexity makes it an attractive choice. In particular, we note that the ALR requires estimation of the first $p+1$ lags of the ACF only.

2.11 Orthogonal Decomposition

Although we cannot find an explicit expression in terms of the AR-process parameters, the exact log-likelihood ratio has a rather simple mathematical form. For a multivariate Gaussian process, whose PDF was defined in Eq. (2.2), the exact log-likelihood ratio is readily found as

$$L(\mathbf{x}_n^N) = \frac{1}{2}\mathbf{x}_n^T \left\{ \left[\mathbf{R}_x^{(0)}\right]^{-1} - \left[\mathbf{R}_x^{(1)}\right]^{-1} \right\} \mathbf{x}_n + \frac{1}{2} \ln \frac{\left|\mathbf{R}_x^{(0)}\right|}{\left|\mathbf{R}_x^{(1)}\right|} \,. \tag{2.63}$$

The true correlation matrices of process X_0 and X_1 are here denoted $\mathbf{R}_x^{(0)}$ and $\mathbf{R}_x^{(1)}$, respectively. The superscript of $\mathbf{x}_n^N$, denoting the segment length, was left out in the above equation and will be suppressed from now on, whenever it is convenient.

In the sequel, we will for different reasons need to write the log-likelihood ratio in an alternative form. Let $\mathbf{y}_n = \mathbf{T}^T\mathbf{x}_n$ be a linear transformation with the transformation matrix $\mathbf{T}$. Since $\mathbf{x}_n$ is zero-mean, we find that $\mathbf{y}_n \sim \mathcal{N}[\mathbf{0}, \mathbf{T}^T\mathbf{R}_x\mathbf{T}]$, provided that $\mathbf{T}^T\mathbf{R}_x\mathbf{T}$ is non-singular [Peebles 1993, Scharf 1991]. It can be shown [Peebles 1993] that

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|\mathbf{T}|} f_{\mathbf{X}}\Big(\mathbf{x} = (\mathbf{T}^T)^{-1}\mathbf{y}\Big) \,. \tag{2.64}$$

It follows that $L(\mathbf{y}_n^N) = L(\mathbf{x}_n^N)$ for any linear transformation. In particular, since the correlation matrix is positive semi-definite, we can use the N orthonormal eigenvectors $\{\mathbf{u}_k\}, k = 1, \ldots, N$ of the generalised eigenvector problem

$$\mathbf{R}_x^{(1)}\mathbf{u}_k = \lambda_k \mathbf{R}_x^{(0)}\mathbf{u}_k \tag{2.65}$$

to build a transformation matrix $\mathbf{U} = [\mathbf{u}_1\mathbf{u}_2\cdots\mathbf{u}_N]$ that defines an orthogonal decomposition [Fukunaga 1990, Scharf 1991]. Eigenvector $\mathbf{u}_k$ corresponds to eigenvalue λ_k and $\mathbf{R}_x^{(i)}$

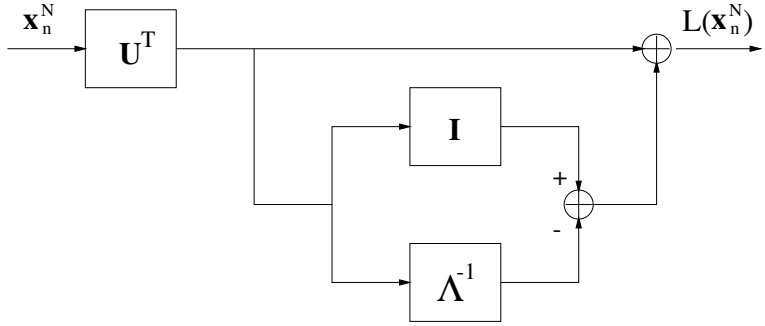


Figure 2.6: Computation of log-likelihood ratio by means of orthogonal transformation.

is the $N \times N$ correlation matrix of process $X_i,\ i \in [0,1]$. The orthogonal transformation $\mathbf{y}_n = \mathbf{U}^T\mathbf{x}_n$ gives diagonal correlation matrices

$$\mathbf{R}_y^{(0)} = \mathbf{I} \tag{2.66}$$

and

$$\mathbf{R}_y^{(1)} = \mathbf{\Lambda}\ . \tag{2.67}$$

The correlation matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ has eigenvalues on the diagonal. After transformation, the log-likelihood ratio is computed as

$$\begin{aligned} L(\mathbf{y}_n^N) &= L(\mathbf{y}_n^N|H_1) - L(\mathbf{y}_n^N|H_0) \\ &= \frac{1}{2}\mathbf{y}_n^{N\,T}\big(\mathbf{I} - \mathbf{\Lambda}^{-1}\big)\mathbf{y}_n^N - \frac{1}{2}\ln\big|\mathbf{\Lambda}\big| = L(\mathbf{x}_n^N)\ . \end{aligned} \tag{2.68}$$

Computation of the log-likelihood ratio using the orthogonal transformation matrix is illustrated by figure 2.6. In the block diagram, boxes denote left multiplication matrix operators. The plus sign denotes vector addition and the multiplication sign (diagonal cross) denotes the appropriate scalar product that produces $L(\mathbf{x}_n^N)$.

Chapter 3

Statistical Distance Measures

In the receiver of the binary ARPSK communications system, we need a detection rule that decides which one of the two possible AR-processes is transmitted. The detector should quantify the likelihood that the received sequence is produced by each one of the respective models of the transmitted signal, and base its decision on this information.

We might imagine that the detector somehow measures the similarity between the received process realisation and its parent models, in an implicit or explicit sense, by use of a statistical distance measure. Such distance measures have been proposed, both in the area of speech processing and communications [Gray and Markel 1976, Gray et al. 1980, Rabiner and Juang 1993], as well as in pattern recognition and statistical decision theory [Basseville 1989, Fukunaga 1990]. These distance measures will also be valuable tools when we attempt to choose the optimal processes for our communications system.

In communications and information theory, design of distance measures [Jeffreys 1948, Kullback 1959, Ali and Silvey 1966] has been motivated by the problem of selecting carrier signals that provide minimum detection error probability, denoted P_e. The analytical expression for the P_e of a given system may be too complex for analytical or numerical optimisation methods to be applied. Therefore, minimisation of P_e is replaced by weaker criteria that involve distance measures that are more mathematically tractable [Kailath 1967].

According to this problem formulation, an optimal distance measure $d(\mathbf{a}_0, \mathbf{a}_1)$ between the processes with parameter vectors $\mathbf{a}_0$ and $\mathbf{a}_1$ should have the property

$$P_e(\mathbf{a}_0, \mathbf{a}_1) > P_e(\mathbf{a}_0, \mathbf{a}_1') \quad \Longrightarrow \quad d(\mathbf{a}_0, \mathbf{a}_1) < d(\mathbf{a}_0, \mathbf{a}_1') \tag{3.1}$$

when $\mathbf{a}_1 \neq \mathbf{a}_1'$. I.e., the distance $d(\cdot)$ should be a monotone functional of the detection

probability $(1 - P_e)$. Apparently, we must seek distance measures that satisfy weaker criteria. But, as a general statement, a good distance measure would be one that mimics the behaviour of $(1-P_e)$. A weaker, but realistic constraint, is that $d(\cdot)$ should be a convex functional of the likelihood ratio [Kailath 1967].

The strict analogy between distance measure and P_e is important from the receiver's point of view, if the first is not used as a replacement for the other when the purpose is to assess or optimise the perfermance of the communications system. If we look at the problem from an eavesdropper's point of view, then an optimal detector is not available, and the theoretical P_e has no practical value. Hence, other statistical distance measures with less relation to the P_e may prove to be more intuitive tools.

We may expect that an unauthorised listener will try to decode the transmitted signal by means of a segmentation or change detection algorithm [Basseville 1988], or detection could be done in the domain of second-order statistics, attempting to distinguish between the power spectral densities of the transmission processes. Therefore, we should choose processes that have similar spectral characteristics. This suggests that we may employ distance measures that are designed for the frequency domain.

Spectral distance measures are explicit functions of second-order statistics, which would normally mean the power spectral densities of the processes. More general distance measures are derived, as will be seen, from the PDFs of the processes. However, since we deal with Gaussian processes that are complitely specified by their second-order statistics, spectral distance measures will not discard any inherent information.

Several spectral distance measures have been defined and studied in the area of speech processing. Speech is commonly modelled as an AR-process. Assuming that a segment of a speech signal can be described by one of a number of AR-models, a hypothesis test is carried out by measuring the distance between the estimated speech spectrum and the model spectra [Gray and Markel 1976].

We shall in this chapter examine a number of different statistical distance measures. For a distance measure $d(x, y)$ to be a true metric, it must satisfy three conditions:

$$\begin{array}{lllll} \text{(i)} & d(x,y) & = & d(y,x) & \text{(symmetry)} \\ \text{(ii)} & d(x,y) & \geq & 0, \quad \forall\, x, y & \\ & d(x,y) & = & 0, \quad \text{iff } x = y & \text{(positive definiteness)} \\ \text{(iii)} & d(x,y) & \leq & d(x,z) + d(y,z) & \text{(triangle inequality)} \end{array} \tag{3.2}$$

These requirements are not met for all the distance measures presented in this thesis. Still,

we find that some of the distance measures serves our purpose.

3.1 Euclidean Distance

The simplest possible distance measure for AR-processes would be the Euclidean distance between the AR-parameters,

$$\begin{aligned} d_E &= \|\mathbf{a}_1 - \mathbf{a}_0\| \\ &= \left[\sum_{k=1}^{p} \left(a_k^{(1)} - a_k^{(0)}\right)^2\right]^{1/2} \end{aligned} \tag{3.3}$$

where $\|\cdot\|$ denotes the Euclidean vector norm. This is not a good choice, because a large Euclidean distance does not always imply large distance in the feature space where process discrimination is performed [Rabiner and Juang 1993]. Moreover, d_E has no spectral theoretical interpretation [Basseville 1988].

3.2 Jeffreys Divergence

This divergence measure was first introduced by Jeffreys [Jeffreys 1946, Jeffreys 1948]. It measures the dispersion of the log-likelihood ratio expected values under the two hypotheses, and is defined by

$$d_J = \mathrm{E}\{L(\mathbf{x})|H_1\} - \mathrm{E}\{L(\mathbf{x})|H_0\}. \tag{3.4}$$

The constituent terms $d_{KL}(0,1) = \mathrm{E}\{L(\mathbf{x})|H_1\}$ and $d_{KL}(1,0) = -\mathrm{E}\{L(\mathbf{x})|H_0\}$ can also be used as distance measures. These are known as the Kullback-Leibler numbers or Kullback information [Kailath 1967, Basseville 1989]. In general, we have $d_{KL}(0,1) \neq d_{KL}(0,1)$. The sum, on the other hand, is symmetric. It is also known as the Kullback divergence [Kailath 1967, Basseville 1989].

For a zero mean multivariate Gaussian random variable $\mathbf{x}$ with covariance matrices $\mathbf{R}_x^{(0)}$ and $\mathbf{R}_x^{(1)}$ under the respective hypotheses, the Jeffreys divergence is [Scharf 1991]

$$d_J = \frac{1}{2}\mathrm{tr}\left(\mathbf{R}_x^{(1)}\mathbf{R}_x^{(0)^{-1}} + \mathbf{R}_x^{(0)}\mathbf{R}_x^{(1)^{-1}} - 2\mathbf{I}\right) . \tag{3.5}$$

Basseville [Basseville 1989] has classified the Jeffreys divergence as belonging to a class of likelihood distance measures [Rabiner and Juang 1993] related to the Csiszar I-divergence [Csiszar 1975]. Class members measure the distance between two probability

distributions $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$ with aid of the dispersion of the likelihood ratio with respect to $f_0(\mathbf{x})$. Hence, these distance measures can be written as

$$d\big(f_0(\mathbf{x}), f_1(\mathbf{x})\big) = h\left\{E_0\left[g\left(\frac{f_1(\mathbf{x})}{f_0(\mathbf{x})}\right)\right]\right\} \tag{3.6}$$

for some functions $g(\cdot)$ and $h(\cdot)$, where $E_0\{\cdot\}$ is the expectation with respect to $f_0(\mathbf{x})$. The Jeffreys divergence is written on this form with $g(x) = (x-1)\ln(x)$ and $h(x) = x$. From information theory, this measure is known as the relative entropy between the probability distributions.

3.3 Bhattacharyya Distance

The Bhattacharyya coefficient for two probability distributions $f_0(\mathbf{x})$ and $f_1(\mathbf{x})$ is given by [Bhattacharyya 1943]

$$\rho_B = \int \sqrt{f_0(\mathbf{x}) f_1(\mathbf{x})}\, d\mathbf{x}\,. \tag{3.7}$$

The Bhattacharyya distance is given by

$$d_B = -\ln \rho_B \tag{3.8}$$

where $0 \leq d_B < \infty$. Alternatively, we can define $\tilde{d}_B = \sqrt{1-\rho_B}$ with bounds $0 \leq \tilde{d}_B \leq 1$, which obeys the triangle inequality in Eq. (3.2). The Bhattacharyya distance for different statistical distributions is reported by Kailath [Kailath 1967]. For a multivariate Gaussian random variable $\mathbf{x}$ with zero mean vector, we have

$$d_B = \frac{1}{2}\ln\left(\frac{\left|\big(\mathbf{R}_x^{(0)} + \mathbf{R}_x^{(1)}\big)/2\right|}{\sqrt{\left|\mathbf{R}_x^{(0)}\right|\left|\mathbf{R}_x^{(1)}\right|}}\right). \tag{3.9}$$

An upper and lower bound for P_e is also given in terms of the Bhattacharyya distance. For symmetrical sources, we find that [Kailath 1967]

$$\frac{1}{2}\left(1 - \sqrt{1-\rho_B^2}\right) \leq P_e \leq \frac{1}{2}\rho_B \tag{3.10}$$

where $\rho_B = e^{-d_B}$.

In terms of Eq. (3.6), the Bhattacharyya distance is defined by functions $g(x) = -\sqrt{x}$ and $h(x) = -\ln(-x)$. It is a special case of the Chernoff distance [Kailath 1967], defined by $g(x) = -x^{1-r}$ for $0 \leq r \leq 1$ and $h(x) = -\ln(-x)$. Both the Jeffreys divergence and the Bhattacharyya distance are convex functionals of the likelihood ratio [Kailath 1967].

3.4 Log-Spectral Distance Measures

Assume two spectral models, $S(\omega)$ and $S'(\omega)$. The log-spectral difference per angular frequency between the models is defined as [Gray and Markel 1976, Rabiner and Juang 1993]

$$\begin{aligned} V(\omega) &= \ln\left(\frac{S(\omega)}{S'(\omega)}\right) \\ &= \ln(S(\omega)) - \ln(S'(\omega)) \end{aligned} \tag{3.11}$$

where ω is the normalised angular frequency ($\omega \in [0, 2\pi]$). The log-spectral distance measures d_{p}, also known as the L_{p} norms, are a set of true metrics, defined as [Gray and Markel 1976, Rabiner and Juang 1993]

$$[d_{\mathsf{p}}(S, S')]^{\mathsf{p}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} |V(\omega)|^{\mathsf{p}}\, d\omega. \tag{3.12}$$

where the argument in the power spectral densities are omitted for brevity. Different measures are obtained for different choices of p (which must not be confused with the order of the AR-model), e.g. the mean absolute log spectral measure ($\mathsf{p} = 1$), the root mean squared (RMS) log spectral measure ($\mathsf{p} = 2$) and the peak log spectral difference ($\mathsf{p} \to \infty$). Most commonly used is the RMS log spectral distance

$$d_2(S, S') = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} |V(\omega)|^2\, d\omega\right]^{1/2}. \tag{3.13}$$

The effect of large values of $V(\omega)$ is more heavily weighted as p is increased. In the limiting case, $d_\infty = \max|V(\omega)|$. As would be expected, measures for different choices of P are heavily correlated. A commonly used approximation to d_2 is found in the cepstral distance measure [Gray and Markel 1976, Rabiner and Juang 1993].

3.5 Itakura-Saito Distance Measure

Another spectral distance measure based on the log-spectral difference $V(\omega)$ defined in Eq. (3.11) was proposed in [Itakura and Saito 1970]. The Itakura-Saito distance is given by

$$\begin{aligned} d_{IS}(S, S') &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left[e^{V(\omega)} - V(\omega) - 1\right] d\omega \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{S(\omega)}{S'(\omega)} d\omega - \ln\frac{\sigma^2}{\sigma'^2} - 1 \end{aligned} \tag{3.14}$$

where σ^2 and σ'^2 are called the one-step prediction errors [Rabiner and Juang 1993] of the spectral models $S(\omega)$ and $S'(\omega)$, respectively. For general processes, σ^2 is calculated from

$$\sigma^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S(\omega)\, d\omega\right\} \tag{3.15}$$

and σ'^2 from the same expression replacing $S(\omega)$ with $S'(\omega)$. For AR-processes, the one-step prediction error is equal to the driving noise variance.

From the integrand of Eq. (3.14) we see that d_{IS} is asymmetric. In effect, positive values of $V(\omega)$ are weighted more than negative values. Therefore, d_{IS} itself is not useful for our purpose, but we shall show some relations to other distance measures in the following.

3.6 Cosh Distance Measure

We will now show that the Jeffreys divergence also has a spectral interpretation. Recall the orthogonal decomposition from section 2.11. It was shown that the log-likelihood ratio can be written as

$$L(\mathbf{x}_n^N) = (\mathbf{y}_n^N)^T(\mathbf{I} - \mathbf{\Lambda}^{-1})\mathbf{y}_n^N - \ln|\mathbf{\Lambda}| \,. \tag{3.16}$$

using the linear transformation $\mathbf{y} = \mathbf{U}^T\mathbf{x}$ where the transformation matrix $\mathbf{U} = [\mathbf{u}_1 \ldots \mathbf{u}_n]$ is constructed from eigenvectors of

$$\mathbf{R}_x^{(1)}\mathbf{u}_k = \lambda_k \mathbf{R}_x^{(0)}\mathbf{u}_k \,, \quad k = 1, \ldots, N \,. \tag{3.17}$$

From the definition of the Jeffreys Divergence we find that

$$\begin{aligned} d_J &= \mathrm{tr}(\mathbf{\Lambda} + \mathbf{\Lambda}^{-1} - 2\mathbf{I}) \\ &= \sum_{k=1}^{N}(\lambda_k + \lambda_k^{-1} - 2) \,. \end{aligned} \tag{3.18}$$

Scharf [Scharf 1987] conclude that the contribution of eigenvalue λ_k to the divergence and detectability is highest when $0 < \lambda_k \ll 1$, or $1 \ll \lambda_k$, such that the sum $(\lambda_k + \lambda_k^{-1})$ is large. He also provides an interpretation of λ_k, which relies on theory of circulant matrices.

An $N \times N$ Toeplitz matrix $\mathbf{M}$ is circulant if and only if its elements obey [Bellman 1970]

$$[\mathbf{M}]_{mn} = f\big((m-n) \mod N\big) \tag{3.19}$$

where $f(\cdot)$ is an arbitrary function. That is, each column of $\mathbf{M}$ must be equal to the previous column rotated downwards by one element. Every wide-sense stationary time

series $x(n)$ [Papoulis 1991, Peebles 1993] has a Toeplitz correlation matrix $\mathbf{R}_x$ which is asympotically circulant as $N \to \infty$ [Scharf 1987]. A circulant matrix $\mathbf{R}_x$ further has an orthogonal representation [Bellman 1970]

$$\mathbf{R}_x = \mathbf{F}\mathbf{S}\mathbf{F}^H \tag{3.20}$$

where $\mathbf{F}$ is a discrete Fourier Transform (DFT) matrix, $\mathbf{S}$ is a diagonal PSD matrix and $[\,\cdot\,]^H$ denotes the Hermitian transpose. The entries of the DFT and PSD matrices are

$$[\mathbf{F}]_{mn} = e^{j2\pi mn/N} \tag{3.21}$$

and

$$[\mathbf{S}]_{nn} = \sum_{k=0}^{N-1} r_{xx}(k) e^{j2\pi kn/N} \; . \tag{3.22}$$

After further derivations [Scharf 1987], we identify the eigenvalues as

$$\lambda_k = \frac{S_{xx}^{(1)}\big(\exp(j2\pi k/N)\big)}{S_{xx}^{(0)}\big(\exp(j2\pi k/N)\big)} \tag{3.23}$$

where $S_{xx}^{(i)}\big(\exp(j\Omega)\big)$ is the PSD of process X_i and $\Omega = 2\pi k/N$ is a discrete angular frequency. Hence, λ_k and its reciprocal value are the quotients of the process spectral densities at $\Omega = 2\pi k/N$. It is intuitive that large and small values of the quotients signify the strongest contribution to detectability, since this indicates a large difference between the spectra at that particular frequency.

The eigenvalues sample the function $S_{xx}^{(1)}(\omega)/S_{xx}^{(0)}(\omega)$ at N equally spaced frequencies. By comparison of the λ_k, we can identify the frequency windows that are important for detection. Moreover, as an asympotic result, we find the following spectral interpretation of the Jeffreys divergence

$$\begin{aligned} \lim_{N\to\infty} d_J &= \int_{-\pi}^{\pi} \left(\frac{S_{xx}^{(1)}(\omega)}{S_{xx}^{(0)}(\omega)} + \frac{S_{xx}^{(0)}(\omega)}{S_{xx}^{(1)}(\omega)} - 2 \right) \, d\omega \\ &= \int_{-\pi}^{\pi} \frac{\Big(S_{xx}^{(1)}(\omega) - S_{xx}^{(0)}(\omega)\Big)^2}{S_{xx}^{(0)}(\omega) S_{xx}^{(1)}(\omega)} \, d\omega \; . \end{aligned} \tag{3.24}$$

Define the dispersion spectral density as

$$D(\omega) = \frac{\Big(S_{xx}^{(1)}(\omega) - S_{xx}^{(0)}(\omega)\Big)^2}{S_{xx}^{(0)}(\omega) S_{xx}^{(1)}(\omega)} \; . \tag{3.25}$$

By integrating the dispersion spectral density over all frequencies and dividing by the frequency range 2π, we obtain a measure of the mean spectral dispersion. We have now arrived at the well known Cosh distance measure [Gray and Markel 1976, Rabiner and Juang 1993], which is defined as

$$d_{COSH} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{D(\omega)}{2} \, d\omega \; . \tag{3.26}$$

The Cosh distance can also be derived as a symmetrised version of the Itakura-Saito distance [Rabiner and Juang 1993]. We have $d_{IS}(S, S') \neq d_{IS}(S', S)$ for $S(\omega) \neq S'(\omega)$. Hence, a symmetric distance measure can be defined as $[d_{IS}(S, S') + d_{IS}(S, S')]/2$. It turns out that this is the Cosh distance, on the form

$$\begin{aligned} d_{COSH}(S, S') &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [e^{V(\omega)} + e^{-V(\omega)} - 2]/2 \, d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} [\cosh V(\omega) - 1] \, d\omega \end{aligned} \tag{3.27}$$

where $\cosh x = [\exp(x) + \exp(-x)]/2$. It is easily seen that this expression is equivalent to Eq. (3.26).

From Eq. (3.25), we note that a linear spectral difference of a given value is more significant to discrimination, the lower the process PSDs are at the frequency it occurs. This is also an implicit result of the d_P measures that operate on the logarithmic spectral difference. In particular, we experience in practice that the Cosh distance and the RMS log spectral distance have similar properties. This can be explained if we compare the integrands of d_2^2 and d_{COSH} in Eqs. (3.13) and (3.27), respectively. The serial expansion

$$\cosh[V(\omega)] = 1 + \frac{V(\omega)^2}{2!} + \frac{V(\omega)^4}{4!} + \cdots \tag{3.28}$$

proves that

$$d_{COSH}(S, S') \geq \frac{1}{2}[d_2(S, S')]^2 \; . \tag{3.29}$$

In figure 3.6, the curve $\cosh[V(\omega)] - 1$ is plotted as function of $V(\omega)$, together with $V(\omega)^2/2$, which is the basis of the RMS log spectral distance. We see that the weighting of spectral differences is relatively close at small spectral differences, while large spectral differences are weighted much more in d_{COSH}-measure.

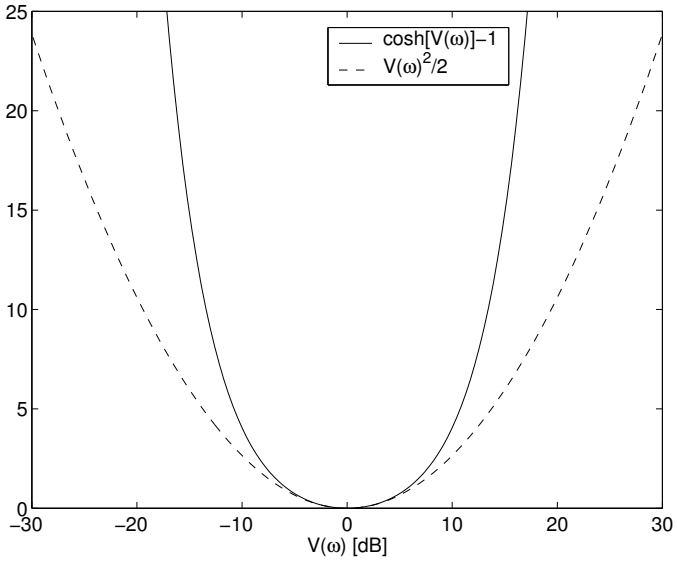


Figure 3.1: Comparison of the integrands $\cosh[V(\omega)] - 1$ and $V(\omega)^2/2$ in the spectral distance measures d_{COSH} and d_2, respectively.

3.7 Prediction Residual Power Ratio

If we have p observations $\mathbf{x}_n^p = [x(n-p), \ldots, x(n-1)]^T$ of a process that is assumed to be described by an AR-model with parameters $\mathbf{a} = [a_1, \ldots, a_p]^T$, then the next observation can be predicted from the linear predictor [Gray and Markel 1976, Scharf 1991]

$$\hat{x}(n) = -\sum_{i=1}^{p} a_i x(n-i) \; . \tag{3.30}$$

The prediction error is given by

$$\begin{aligned} \varepsilon(n) &= x(n) - \hat{x}(n) \\ &= \sum_{i=0}^{p} a_i x(n-i) \end{aligned} \tag{3.31}$$

where $a_0 \triangleq 1$. A minimum mean squared error (MMSE) estimate of the process model parameter vector is given as the $\boldsymbol{\alpha} = [a_0, \ldots, a_p]^T$ that produces the minimum prediction residual power (MPRP)

$$\begin{aligned} \alpha &= \min_{\mathbf{a}} \left\{ E\left[\varepsilon^2(n)\right] \right\} \\ &= \min_{\mathbf{a}} \left\{ E\left[(\boldsymbol{\alpha}^T \mathbf{x}_n^p)^2\right] \right\} \; . \end{aligned} \tag{3.32}$$

If $x(n)$ is truly described by an AR-model, then we have $\alpha = \sigma_\epsilon^2$. From Eq. (3.32) we see that $\sigma_\epsilon^2 = \boldsymbol{\alpha}^T \mathbf{R}_x \boldsymbol{\alpha}$, where $\mathbf{R}_x$ is the true correlation matrix of the AR-process. The prediction filter interpretation is an alternative way of looking at the AR-model.

Assume that a data sequence $\tilde{x}(n)$ is generated by another AR-process with parameter vector $\tilde{\boldsymbol{\alpha}} = [\tilde{a}_0, \ldots, \tilde{a}_p]^T$. If prediction vector $\boldsymbol{\alpha}$ is applied on $\tilde{x}(n)$, the linear predictor will produce the generally non-minimum prediction residual power (PRP)

$$\begin{aligned} \beta &= \boldsymbol{\alpha}^T \tilde{\mathbf{R}}_x \boldsymbol{\alpha} \\ &= \frac{\alpha}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\tilde{A}(\omega)}{A(\omega)} \right|^2 d\omega \geq \alpha \end{aligned} \tag{3.33}$$

where $A(\omega)$ and $\tilde{A}(\omega)$ are the DFTs of $\boldsymbol{\alpha}$ and $\tilde{\boldsymbol{\alpha}}$, respectively, while $\tilde{\mathbf{R}}_x$ is the true correlation matrix of the process with parameter vector $\tilde{\boldsymbol{\alpha}}$. Equality in $\beta \geq \alpha$ holds only if $\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}$. The prediction residual energy ratio β/α can be used as a distance measure. It is related to the Itakura-Saito distance of two AR-processes with unity driving noise [Rabiner and Juang 1993],

$$\frac{\beta}{\alpha} - 1 = d_{IS}\left(\frac{1}{|A(\omega)|^2}, \frac{1}{|\tilde{A}(\omega)|^2} \right) . \tag{3.34}$$

This distance measure puts emphasis on spectral shape, and totally disregards the driving noise variances of the processes. So does the gain-normalised Itakura distance [Itakura 1975, Rabiner and Juang 1993], which is defined for two general AR(p)-processes as

$$\begin{aligned} d_I\left(\frac{\sigma_\epsilon^2}{|A(\omega)|^2}, \frac{\tilde{\sigma}_\epsilon^2}{|\tilde{A}(\omega)|^2} \right) &= d_{IS}\left(\frac{\alpha}{|A(\omega)|^2}, \frac{\beta}{|\tilde{A}(\omega)|^2} \right) \\ &= \ln\left(\frac{\beta}{\alpha} \right) \end{aligned} \tag{3.35}$$

where $\alpha = \sigma_\epsilon^2$ and β is calculated according to their definition, assuming parameter model $\boldsymbol{\alpha}$ and disregarding the value of $\tilde{\sigma}_\epsilon^2$, which is generally different from σ_ϵ^2.

There is obviously a connection between β/α and the approximated log-likelihood ratio (ALR) from section 2.10. We can write the ALR as

$$\begin{aligned} \mathcal{L}(\mathbf{x}_n^N) &= \frac{N}{2}\left[\left(\frac{\boldsymbol{\alpha}_0^T \hat{\mathbf{R}}_x \boldsymbol{\alpha}_0}{\boldsymbol{\alpha}_0^T \mathbf{R}_x^{(0)} \boldsymbol{\alpha}_0} \right) - \left(\frac{\boldsymbol{\alpha}_1^T \hat{\mathbf{R}}_x \boldsymbol{\alpha}_1}{\boldsymbol{\alpha}_1^T \mathbf{R}_x^{(1)} \boldsymbol{\alpha}_1} \right) + \ln\left(\frac{\boldsymbol{\alpha}_0^T \mathbf{R}_x^{(0)} \boldsymbol{\alpha}_0}{\boldsymbol{\alpha}_1^T \mathbf{R}_x^{(1)} \boldsymbol{\alpha}_1} \right) \right] \\ &= \frac{N}{2}\left[\left(\frac{\hat{\beta}|\boldsymbol{\alpha}_0}{\alpha\,|H_0} \right) - \left(\frac{\hat{\beta}|\boldsymbol{\alpha}_1}{\alpha\,|H_1} \right) + \ln\left(\frac{\alpha\,|H_0}{\alpha\,|H_1} \right) \right] . \end{aligned} \tag{3.36}$$

The process-dependent MPRP values are here denoted $\alpha\,|H_0 = \sigma^2_{\epsilon_0}$ and $\alpha\,|H_1 = \sigma^2_{\epsilon_1}$. The data dependent terms $\hat{\beta}|\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_0^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_0$ and $\hat{\beta}|\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_1$ are interpreted as estimators of the PRP that is produced when the data sequence $\mathbf{x}_n^N$ is fitted to the AR-model with parameter vector $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_1$, respectively. The expectation values of PRP estimators under the two hypotheses are

$$E\{\hat{\beta}|\boldsymbol{\alpha}_0\} = \begin{cases} \boldsymbol{\alpha}_0^T\mathbf{R}_x^{(0)}\boldsymbol{\alpha}_0 = \sigma^2_{\epsilon_0} \;:\; H_0 \\ \boldsymbol{\alpha}_0^T\mathbf{R}_x^{(1)}\boldsymbol{\alpha}_0 > \sigma^2_{\epsilon_0} \;:\; H_1 \end{cases} \tag{3.37}$$

and

$$E\{\hat{\beta}|\boldsymbol{\alpha}_1\} = \begin{cases} \boldsymbol{\alpha}_1^T\mathbf{R}_x^{(0)}\boldsymbol{\alpha}_1 > \sigma^2_{\epsilon_1} \;:\; H_0 \\ \boldsymbol{\alpha}_1^T\mathbf{R}_x^{(1)}\boldsymbol{\alpha}_1 = \sigma^2_{\epsilon_0} \;:\; H_1 \end{cases} . \tag{3.38}$$

The first term in Eq. (3.36) is a measure of the distance or dissimilarity between the data sequence and process X_0. The second term measures the distance with respect to process X_1, and the relative magnitude of the terms tells something about how a source bit represented by $\mathbf{x}_n^N$ should be classified. The third term reflects the fact that the driving noise variances must be taken into account, since, in ARPSK communications, we must require that $\sigma^2_{\epsilon_0} \neq \sigma^2_{\epsilon_1}$ to obtain equal average process power. The application of the ALR to detection is examined in detail in the next chapter.

We find that the ALR can be expressed in terms of the Itakura-Saito distance as

$$\mathcal{L}(\mathbf{x}_n^N) = \frac{N}{2}\left[d_{IS}\left(\frac{\hat{\beta}_0}{|A_0|^2}, S_0\right) - d_{IS}\left(\frac{\hat{\beta}_1}{|A_1|^2}, S_1\right) - \ln\left(\frac{\hat{\beta}_1}{\hat{\beta}_0}\right)\right] \tag{3.39}$$

where $S_0 = \sigma^2_{\epsilon_0}/|A_0|^2$ and $S_1 = \sigma^2_{\epsilon_1}/|A_1|^2$ denotes the PSD of process X_0 and X_1, respectively, and frequency arguments are omitted. For brevity, the PRP estimators for model-fitted data are also subscripted, defining $\hat{\beta}_i = \hat{\beta}|\boldsymbol{\alpha}_i$, $i \in [0, 1]$.

Chapter 4

Detection

In this section we will discuss some possible detectors for the ARPSK-modulated signal, and their associated detection error probability P_e or bit error rate (BER).

The Neyman-Pearson detector [Scharf 1991] is optimum in a minimum detection error probability sense, and is one obvious candidate. It can be considered as a special case of the Bayes detector [Scharf 1991], which is another conventional detector of great importance. An alternative is to use the theory of statistical distance measures that have been reviewed in the previous section. We have already forecasted that the approximate likelihood ratio can be used as a detector. But before we look into the details of various detectors, we establish a constraint that must be satisfied for SPSK communications.

4.1 Process Power Equalisation

The average power of the two stochastic processes used in SPSK communication must evidently be equal. Otherwise, a variation in the transmitted power can be sufficient for an eavesdropper to distinguish between different source bits.

The average power of a discrete time stochastic process $X(n)$ is given by [Peebles 1993]

$$P_{xx} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{xx}(\omega)\, d\omega \; . \tag{4.1}$$

This integral can be interpreted as the inverse Fourier transform of the power density spectrum at time index $k = 0$. The Wiener-Khinchin relation [Peebles 1993] proves that

the following expression is equivalent to the average process power

$$r_{xx}(0) = \frac{1}{2\pi}\left[\int_{-\pi}^{\pi} \frac{\sigma_\epsilon^2}{|A(\omega)|^2}\, e^{j\omega k}\, d\omega\right]\Bigg|_{k=0}, \tag{4.2}$$

where the known power density spectrum of an AR-process is substituted into the expression. It should come as no surprise that this is identical to the process variance σ_x^2 [Peebles 1993].

The power equalisation constraint states that $P_0 = P_1$ for the two processes $X_0(n)$ and $X_1(n)$. This should be tranformed into a constraint on the driving noise variances, which are our equalisation tools. The constrained process variance obtained when $\sigma_\epsilon^2 = 1$ is given by

$$\rho_{xx}(0) = \int_{-\pi}^{\pi} \frac{1}{|A(\omega)|^2}\, d\omega\ . \tag{4.3}$$

It follows that the unconstrained process variance $r_{xx}(0) = \sigma_\epsilon^2 \rho_{xx}(0)$. Hence, from the power equalisation constraint, the innovation variances must obey

$$\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_0}^2 \frac{\rho_{xx}^{(0)}(0)}{\rho_{xx}^{(1)}(0)} \tag{4.4}$$

where process indices are introduced on both innovation variances and constrained process variances.

4.2 Neyman-Pearson Detection

The important Neyman-Pearson lemma [Scharf 1991] states that the hypothesis test which minimises the detection error probability is a log-likelihood ratio test on the form

$$L(\mathbf{x}_n) \underset{\Omega_0}{\overset{\Omega_1}{\gtrless}} \eta. \tag{4.5}$$

The likelihood ratio $l(\mathbf{x}_n)$ has previously been defined as the ratio of two likelihood functions, and the log-likelihood ratio of two zero-mean Gaussian processes was given in section 2.10 as

$$\begin{aligned} L(\mathbf{x}_n) &= \ln[l(\mathbf{x}_n)] \\ &= \frac{1}{2}\mathbf{x}_n^T \left\{ \left[\mathbf{R}_x^{(0)}\right]^{-1} - \left[\mathbf{R}_x^{(1)}\right]^{-1} \right\} \mathbf{x}_n + \frac{1}{2} \ln \frac{|\mathbf{R}_x^{(0)}|}{|\mathbf{R}_x^{(1)}|}\ . \end{aligned} \tag{4.6}$$

A complete test should also contain a decision for the case when $L(\mathbf{x}_n) = \eta$. In this marginal case, decision Ω_0 is made with probability $0 \leq \psi \leq 1$ for a chosen ψ.

The test is optimum for a chosen false alarm probability: $P_{\text{FA}} = P(\Omega_1|H_0)$. The targetted false alarm probability is used to determine the threshold η from

$$P_{\text{FA}} = \int_{\substack{\cdots \\ \forall \mathbf{x}_n \,:\, L(\mathbf{x}_n) > \eta}} \int f(\mathbf{x}_n|\mathbf{a}_0)\, d\mathbf{x}. \tag{4.7}$$

The terms "false alarm probability" and "miss probability" are widely used in detection theory. They stem from applications like radar and sonar, but make no sense in binary communications. In our problem, the erroneous decisions $\Omega_0|H_1$ and $\Omega_1|H_0$ are associated with the same cost. Hence, we want the class-specific detection error probabilities to be equal, $P(\Omega_1|H_0) = P(\Omega_0|H_1)$. This leads to a threshold value of $\eta = 0$, which provides the following decision rule for the Neyman-Pearson detector

$$Q(\mathbf{x}_n) \triangleq \mathbf{x}_n^T \left\{ \left[\mathbf{R}_x^{(0)}\right]^{-1} - \left[\mathbf{R}_x^{(1)}\right]^{-1} \right\} \mathbf{x}_n \underset{\Omega_0}{\overset{\Omega_1}{\gtrless}} \ln \frac{|\mathbf{R}_x^{(1)}|}{|\mathbf{R}_x^{(0)}|}. \tag{4.8}$$

The statistic $Q(\mathbf{x}_n)$ on the left-hand side of the inequalities is an inner product of data vectors weighted by a matrix difference. Such a quadratic form is known to be centrally χ^2-distributed with N degrees of freedom, if the weighting matrix is the inverse covariance matrix of $\mathbf{x}_n$ [Scharf 1991]. This is not true in our case, but the quadratic form must clearly have a χ^2-like distribution.

We assume equal a priori probabilities for the two processes, $p_0 = p_1 = 1/2$. The detection error probability is then given by

$$\begin{aligned} P_e &= p_0\, P(\Omega_1|H_0) + p_1\, P(\Omega_0|H_1) \\ &= \frac{1}{2}\left[\int_{\zeta}^{+\infty} f_Q^{(0)}(q)\, dq + \int_{-\infty}^{\zeta} f_Q^{(1)}(q)\, dq \right] \end{aligned} \tag{4.9}$$

where $f_Q^{(0)}(q)$ and $f_Q^{(1)}(q)$ is the PDF of $Q(\mathbf{x}_n)$ under H_0 and H_1, respectively. The threshold and integration limit is defined as $\zeta = \ln |\mathbf{R}_x^{(1)}|/|\mathbf{R}_x^{(0)}|$.

The characteristic function $\Phi_X(\omega)$ of a stochastic variable X was defined in Eq. (2.8), and we recall that $\Phi_X(-\omega) = \mathfrak{F}\{f_X(x)\}$. This relationship can now be exploited, since it is hard to find the PDF of X explicitly. We have

$$\begin{aligned} P_e &= \frac{1}{2}\left[\int_{\zeta}^{+\infty} \mathfrak{F}^{-1}\{\Phi_Q^{(0)}(-\omega)\}\, dq + \int_{-\infty}^{\zeta} \mathfrak{F}^{-1}\{\Phi_Q^{(1)}(-\omega)\}\, dq \right] \\ &= \frac{1}{4\pi}\left[\int_{\zeta}^{+\infty} \int_{-\infty}^{+\infty} \Phi_Q^{(0)}(-\omega) e^{j\omega q}\, d\omega\, dq + \int_{-\infty}^{\zeta} \int_{-\infty}^{+\infty} \Phi_Q^{(1)}(-\omega) e^{j\omega q}\, d\omega\, dq \right] \end{aligned} \tag{4.10}$$

where $\mathfrak{F}^{-1}\{\Phi_Q^{(i)}(q)\}$ denotes the inverse Fourier transform of the characteristic function of $Q(\mathbf{x}_n)$ under H_i, $i \in [0,1]$. This suggests that P_e can be computed by means of a discrete version of Eq. (4.10), using inverse FFTs and numerical integrations. This is feasible, provided we have an expression for $\Phi_Q(\omega)$, which is indeed the case.

The characteristic function of $Q(\mathbf{x}_n)$ is in this case given by [Scharf 1991]

$$\Phi_Q^{(i)}(\omega) = \left| \mathbf{I} + 2j\omega \left(\mathbf{R}_x^{(1)} - \mathbf{R}_x^{(0)} \right)^{-1} \mathbf{R}_x^{(i)} \right|^{-1/2} \quad : \quad H_i. \tag{4.11}$$

An improved method for evaluation of P_e will be demonstrated in the next section.

4.3 Bayes Detection

The decision rule of a Bayes detector says that a data vector should be classified as belonging to the class i whose joint PDF $f(\mathbf{x}_n, \mathbf{a}_i), i \in [0,1]$ is maximised for the given observation $\mathbf{x}_n$. From Bayes rule, the joint PDF can be expanded as $f(\mathbf{x}_n, \mathbf{a}_i) = p_i\, f(\mathbf{x}_n|\mathbf{a}_i)$. For our two-class detection problem with $p_0 = p_1 = 1/2$, the decision rule becomes

$$f(\mathbf{x}_n|\mathbf{a}_1) \underset{\Omega_0}{\overset{\Omega_1}{\gtrless}} f(\mathbf{x}_n|\mathbf{a}_0)\ . \tag{4.12}$$

This decision rule is equivalent to the log-likelihood ratio test in Eq. (4.5) with threshold $\eta = 0$, and as such, it represents nothing new. However, we shall in the following benefit from the studies of the detection error probability of a general Bayes detector, carried out in [Fukunaga and Krile 1969].

Fukunaga and Krile have derived the P_e for a general Bayes detection problem with two classes modelled by multivariate Gaussian distributions. Fukunaga assumes unequal a priori probabilities p_i, mean vectors $\boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{\Sigma}_i$ for the two classes. He proceeds by deriving the exact PDF of a generalised version of the log-likelihood ratio $L(\mathbf{x}_n)$ defined in Eq. (2.63), going the way through the characteristic function. The elegant derivation concludes with an expression for the detection error of a Bayes detector. Fukunagas and Kriles results are repeated here on a simplified form, after invoking the conditions specific to our problem.

The crux of their approach is to transform the data vector into a vector with statistically independent components, $\mathbf{y}_n = [y_1, \ldots, y_N]^T$, which once again leads to an application of the orthogonal decomposition in section 2.11. The transformation provides characteristic

functions and PDFs that depend only on the eigenvalues $\{\lambda_k\}$, $k = 1, \ldots, N$ obtained from the generalised eigenvalue problem

$$\mathbf{R}_x^{(1)}\mathbf{u}_k = \lambda_k \mathbf{R}_x^{(0)}\mathbf{u}_k \ . \tag{4.13}$$

From the transformation $\mathbf{y}_n = [\mathbf{u}_1 \cdots \mathbf{u}_N]^T \mathbf{x}_n$ and the alternative formulation of the log-likelihood ratio given by Eq. (2.68), it follows [Fukunaga and Krile 1969] that the characteristic function of $L(\mathbf{x}_n)$ under hypothesis H_i can be calculated as

$$\Phi_L^{(i)}(\omega) = \prod_{k=1}^{N} \frac{1}{(1 - 2j\omega\phi_{ik})^{1/2}} \exp(-j\omega \ln \lambda_k) \tag{4.14}$$

with hypothesis dependent parameters defined as

$$\phi_{ik} = \begin{cases} 1 - 1/\lambda_k & : \ i = 0 \\ \lambda_k - 1 & : \ i = 1 \ . \end{cases} \tag{4.15}$$

The total characteristic function is a product of N statistically independent characteristic functions, which have the same functional form and differ only in the parameter λ_k. Each independent function can be expressed as a product of magnitude and angle. This decomposition is helpful in the sequel, and is given by

$$\Phi_L^{(i)}(\omega) = \prod_{k=1}^{N} M_k^{(i)}(\omega) \exp\left[j \sum_{k=1}^{N} \Theta_k^{(i)}(\omega) \right] \tag{4.16}$$

where the magnitude component function is

$$M_k^{(i)}(\omega) = \left[1 + (2\phi_{ik}\omega)^2\right]^{-1/4} \tag{4.17}$$

and the phase component function is

$$\Theta_k^{(i)}(\omega) = \tan^{-1}(2\phi_{ik}\omega)/2 - \omega \ln \lambda_k \ . \tag{4.18}$$

We shall now follow the example of [Fukunaga and Krile 1969] and derive a general relationship between the detection error probability of the Bayes detector and the characteristic function of the log-likelihood ratio. The P_e for a Bayes detector can be expressed in terms of the cumulative distribution function (CDF) of the log-likelihood function, evaluated at $L = 0$ for both hypotheses.

The CDF of $L(\mathbf{x}_n)$ under hypothesis H_i is defined as

$$F_L^{(i)}(L) = \int_{-\infty}^{L} f_L^{(i)}(l)\, dl \quad , \; i = 0, 1 \; . \tag{4.19}$$

where $f_L^{(i)}(L)$ is the PDF of $L(\mathbf{x}_n)$. For the Bayes detector, we then have

$$\begin{aligned} P_e &= p_0 P(\omega_1|H_0) + p_1 P(\omega_0|H_1) \\ &= \frac{1}{2}\big[(1 - F_L^{(0)}(0)) + F_L^{(1)}(0)\big] \; . \end{aligned} \tag{4.20}$$

This is equal to the P_e of a Neyman-Pearson detector defined by Eq. (4.5) with decision threshold $\eta = 0$. Hence, we need to find an expression for $F_L^{(i)}(0)$, $i \in [0, 1]$.

The Fourier transform property for integrals [Oppenheim et al. 1983] is given by the transform pair

$$\int_{-\infty}^{t} x(\tau)\, d\tau = \frac{X(0)}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{X(\omega)}{j\omega} e^{j\omega t}\, d\omega \; . \tag{4.21}$$

If $x(t)$ is real-valued, the following symmetries hold [Oppenheim et al. 1983]: $\mathrm{Re}[X(\omega)] = \mathrm{Re}[X(-\omega)]$ and $\mathrm{Im}[X(\omega)] = -\mathrm{Im}[X(-\omega)]$, where $\mathrm{Re}[\,\cdot\,]$ and $\mathrm{Im}[\,\cdot\,]$ denote real and imaginary part, respectively. Odd parts of the integrand cancel under the doubly infinite integral on the right hand side. After separating the integrand into even and odd functions and inserting $t = 0$, Eq. (4.21) reduces to

$$\int_{-\infty}^{0} x(\tau)\, d\tau = \frac{X(0)}{2} - \frac{1}{\pi}\int_{0}^{\infty} \frac{\mathrm{Im}\{X(-\omega)\}}{\omega}\, d\omega \; . \tag{4.22}$$

From the Fourier transform relation between characteristic function and PDF and Eq. (4.21), we can now show that

$$F_L^{(i)}(L) = \frac{\Phi_L^{(i)}(0)}{2} + \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{\Phi_L^{(i)}(\omega)}{j\omega} \exp(j\omega L)\, d\omega \; . \tag{4.23}$$

and it follows from Eq. (4.22) that

$$F_L^{(i)}(0) = \frac{1}{2} + \frac{1}{\pi}\int_{0}^{\infty} -\frac{\mathrm{Im}[\Phi_L^{(i)}(-\omega)]}{\omega}\, d\omega \tag{4.24}$$

where we use the general result that $\Phi(\omega = 0) = 1$, for any characteristic function. Using the magnitude-phase decomposition of the characteristic function as defined in Eqs. (4.17) and (4.18), Fukunaga and Krile show that

$$F_L^{(i)}(0) = \frac{1}{2} + \frac{1}{\pi}\int_{0}^{\infty} \frac{\prod_{k=1}^{N} M_k^{(i)}(\omega)}{\omega} \sin\left[-\sum_{k=1}^{N} \Theta_k^{(i)}(\omega)\right] d\omega \; . \tag{4.25}$$

This is expression is inserted into Eq. (4.20) to obtain the detection error probability of the Neyman-Pearson detector.

We see that evaluation of P_e requires numerical integration, and unfortunately, the integration range is infinite. However, Fukunaga has shown that the magnitude component functions $M_k^{(i)}(\omega)$ are monotonically decreasing functions of ω. Moreover, the integrand contains the product of these N functions in the numerator, as well as the factor ω in the denominator. Hence, the range of integration and the number of samples needed for the integral to converge in numerical integration, is relatively small.

Note that the described Neyman-Pearson detector and likelihood ratio tests in general can be implemented recursively [Salberg and Hanssen 1999b, Salberg and Hanssen 2000, Basseville 1988]. These implementations yield suboptimal detectors, but have the advantage that synchronisation information about the ARPSK sequence is provided implicitly. In this thesis we have restricted ourselves to a study of detectors that operate on blocks of data that are assumed to be synchronised. I.e., until further notice we assume that the data vector $\mathbf{x}_n$ contains samples produced by a single process generator alone.

4.4 Approximate Log-Likelihood Ratio Detection

We shall now pursue a detector which is motivated by the statistical distance measures presented in section 3. Distance measures whose derivation is based on the likelihood function approximation for AR-processes were first presented in the work of Itakura and Saito [Itakura and Saito 1970, Itakura 1975]. These were designed within a framework of speech recognition, where speech signals are segmented and fitted to different AR-models describing different sounds.

In our problem setting, we have only two process models, and these are known exactly. Hence, we need not estimate the model parameters. We further assume that the innovation variances $\sigma_{\epsilon 0}^2$ and $\sigma_{\epsilon 1}^2$ have been chosen, such that they satisfy the power equalisation constraint. These must also be taken into consideration in the detection problem, unlike in speech recognition, where only the spectral shape is of general interest. As a consequence, there are other requirements to a distance measure that can be used for detection in ARPSK communications.

The detector that we seek is an approximation to the Neyman-Pearson detector defined

in Eq. (4.8). The approximated log-likelihood ratio (ALR) was derived in section 2.10 as

$$\mathcal{L}(\mathbf{x}_n^N) = \frac{N}{2}\left(\frac{1}{\sigma_{\epsilon_0}^2}\boldsymbol{\alpha}_0^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_0 - \frac{1}{\sigma_{\epsilon_1}^2}\boldsymbol{\alpha}_1^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_1\right) + \frac{N}{2}\ln\left(\frac{\sigma_{\epsilon_0}^2}{\sigma_{\epsilon_1}^2}\right). \tag{4.26}$$

In the optimal log-likelihood ratio test in Eq. (4.5), the exact log-likelihood ratio can be replaced with the ALR. This yields the decision rule

$$\frac{1}{\sigma_{\epsilon_0}^2}\boldsymbol{\alpha}_0^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_0 - \frac{1}{\sigma_{\epsilon_1}^2}\boldsymbol{\alpha}_1^T\hat{\mathbf{R}}_x\boldsymbol{\alpha}_1 + c \underset{\Omega_0}{\overset{\Omega_1}{\gtrless}} 0 \tag{4.27}$$

where the constant c is the threshold $\ln(\sigma_{\epsilon_0}^2/\sigma_{\epsilon_1}^2)$. This detector is also found in [Dickinson 1981]. We note that neither the ALR detector nor the Neyman-Pearson needs to know the exact values of the driving noise variances, but only the ratio $\sigma_{\epsilon_0}^2/\sigma_{\epsilon_1}^2$ which is determined by the average power equalisation constraint and the parameter vectors $\mathbf{a}_0$ and $\mathbf{a}_1$.

To calculate the P_e of the ALR detector, we must determine the PDF of $\mathcal{L}(\mathbf{x}_n^N)$. We will from now on omit the scaling factor $N/2$ in Eq. (4.26). The data dependent part of the ALR can then be written as

$$\begin{aligned} D &= \sum_{n=p+1}^{N}\left\{\mathbf{x}_n^{p+1T}\left[\frac{1}{N-p}\left(\frac{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0^T}{\sigma_{\epsilon_0}^2} - \frac{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T}{\sigma_{\epsilon_1}^2}\right)\right]\mathbf{x}_n^{p+1}\right\} \\ &= \frac{1}{N-p}\sum_{n=p+1}^{N}\left(\mathbf{x}_n^{p+1T}\mathbf{A}\mathbf{x}_n^{p+1}\right) \end{aligned} \tag{4.28}$$

where we define

$$\mathbf{A} = \mathbf{A}_0 - \mathbf{A}_1 = \frac{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0^T}{\sigma_{\epsilon_0}^2} - \frac{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T}{\sigma_{\epsilon_1}^2}. \tag{4.29}$$

The individual quadratic forms $Q_n = \mathbf{x}_n^{p+1T}\mathbf{A}\mathbf{x}_n^{p+1}$ in the summation follow a central χ^2-like distribution with $p+1$ degrees of freedom. However, the $\{Q_n\}$ are correlated. Therefore, on the above form, it is difficult to achieve a statistical description of D. This is why we seek the alternative form

$$D = (\mathbf{x}_n^N)^T\mathbf{P}\mathbf{x}_n^N\ . \tag{4.30}$$

This is a block formulation of the sum in Eq. (4.28). The matrix $\mathbf{P}$ is constructed by letting matrix $\mathbf{A}$ slide down along the diagonal, summing up the $N-p$ contributing block

matrices and dividing by $N-p$.

$$\begin{aligned}\mathbf{P} &= \frac{1}{N-p}\left(\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}+\begin{bmatrix}0 & \cdots & 0\\ \vdots & \mathbf{A} & \mathbf{0}\\ 0 & \mathbf{0} & \mathbf{0}\end{bmatrix}+\begin{bmatrix}0 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \mathbf{A} & \mathbf{0}\\ 0 & 0 & \mathbf{0} & \mathbf{0}\end{bmatrix}+\cdots+\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{A}\end{bmatrix}\right)\\ &= \frac{1}{N-p}\,T_{\mathbf{A}}(\mathbf{A})\end{aligned} \tag{4.31}$$

where $T_{\mathbf{A}}:\mathbb{R}^{(p+1)\times(p+1)}\rightarrow\mathbb{R}^{N\times N}$ is the construction operator that defines the transformation of a $(p+1)\times(p+1)$ matrix $\mathbf{A}$ into a $N\times N$ matrix $(N-p)\,\mathbf{P}$. The ALR is now on the form $\mathcal{L}=D+c$, where D is a quadratic form with known characteristic function

$$\Phi_D^{(i)}(\omega)=\left|\mathbf{I}-2j\omega\mathbf{P}\mathbf{R}_x^{(i)}\right|^{-1/2} \quad : \quad H_i\ . \tag{4.32}$$

From Eqs. (4.29) and (4.31), it is easy to show that the matrix $\mathbf{P}$ is symmetric. Symmetry is also a known property of correlation matrices of stationary processes [Haykin 1996]. It follows that $\mathbf{P}\mathbf{R}_x^{(i)}=\mathbf{P}\mathbf{R}_x^{(i)^T}$. Hence, the matrix product $\mathbf{P}\mathbf{R}_x^{(i)}$ is also symmetric, and the characteristic function may be written on the simpler form [Scharf 1991]

$$\Phi_D^{(i)}(\omega)=\frac{1}{\prod_{k=1}^{N}(1-2j\omega\lambda_k^{(i)})^{1/2}} \quad : \quad H_i \tag{4.33}$$

where the $\{\lambda_k^{(i)}\}_{k=1}^N$ are eigenvalues of $\mathbf{P}\mathbf{R}_x^{(i)}$. The characteristic function of $\mathcal{L}$ follows readily from the definition

$$\Phi_{\mathcal{L}}^{(i)}(\omega)=\frac{e^{j\omega c}}{\prod_{k=1}^{N}(1-2j\omega\lambda_k^{(i)})^{1/2}} \quad : \quad H_i \tag{4.34}$$

and from the Fourier transform relation, the PDF of $\mathcal{L}(\mathbf{x}_n)$ is given by

$$f_{\mathcal{L}}^{(i)}(\mathcal{L})=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{j\omega(\mathcal{L}-c)}}{\prod_{k=1}^{N}(1+2j\omega\lambda_k^{(i)})^{1/2}}\,d\omega \quad : \quad H_i\ . \tag{4.35}$$

The integral form of $f_{\mathcal{L}}^{(i)}(\mathcal{L})$ is of course not the most convenient representation, but it is the best we can obtain, due to the complexity of $\mathcal{L}$. However, the moments of $\mathcal{L}$ can readily be obtained from $\Phi_{\mathcal{L}}^{(i)}(\omega)$. We have e.g. that

$$\mathrm{E}\{\mathcal{L}|H_i\}=\mathrm{tr}\left(\mathbf{P}\mathbf{R}_x^{(i)}\right)=\sum_{k=1}^{N}\lambda_k^{(i)} \quad : \quad H_i \tag{4.36}$$

and

$$\mathrm{Var}\{\mathcal{L}|H_i\} = 2\mathrm{tr}\left[(\mathbf{PR}_x^{(i)})^2\right] \quad : \quad H_i \ . \tag{4.37}$$

The P_e for the ALR detector can also be calculated quite conveniently from the procedure that was derived for the Neyman-Pearson detector in section 4.3. For the ALR detector, we have

$$P_e = \frac{1}{2}\left[1 - F_{\mathcal{L}}^{(0)}(0) + F_{\mathcal{L}}^{(1)}(0)\right] \ . \tag{4.38}$$

From Eq. (4.22) it follows that

$$\begin{aligned} F_{\mathcal{L}}^{(i)}(0) &= \frac{\Phi_{\mathcal{L}}^{(i)}(0)}{2} - \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\{\Phi_{\mathcal{L}}^{(i)}(\omega)\}}{\omega}\, d\omega \\ &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Im}\left[\frac{e^{-j\omega c}}{\prod_{k=1}^N (1+2j\omega\lambda_k^{(i)})^{1/2}}\right] /\omega \, d\omega \ : \ H_i \ . \end{aligned} \tag{4.39}$$

Inserted into Eq. (4.38), this yields

$$\begin{aligned} P_e = \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \Bigg\{ &\mathrm{Im}\left[\frac{e^{-j\omega c}}{\prod_{k=1}^N (1+2j\omega\lambda_k^{(1)})^{1/2}}\right] \\ &-\mathrm{Im}\left[\frac{e^{-j\omega c}}{\prod_{k=1}^N (1+2j\omega\lambda_k^{(0)})^{1/2}}\right]\Bigg\} /\omega \, d\omega \ . \end{aligned} \tag{4.40}$$

The integral can be evaluated through numerical integration. We can also decompose $\Phi_{\mathcal{L}}^{(i)}(\omega)$ into a magnitude component and a phase component, as was done with $\Phi_L^{(i)}(\omega)$. We then have

$$\Phi_{\mathcal{L}}^{(i)}(\omega) = \prod_{k=1}^N M_k^{(i)}(\omega) \exp\left[j\sum_{k=1}^N \Theta_k^{(i)}(\omega)\right] \tag{4.41}$$

where the magnitude component function

$$M_k^{(i)}(\omega) = [1 + (2\lambda_k^{(i)}\omega)^2]^{-1/4} \tag{4.42}$$

and the phase component function

$$\Theta_k^{(i)}(\omega) = \tan^{-1}(2\lambda_k^{(i)}\omega)/2 + \omega c/N \ . \tag{4.43}$$

are seen to have a similar form to the respective counterparts defined in Eqs. (4.17) and (4.18) for $\Phi_L^{(i)}(\omega)$. With reference to [Fukunaga and Krile 1969], the CDF of $\Phi_{\mathcal{L}}^{(i)}(\omega)$ evaluated at $\omega = 0$ can now be written as

$$F_{\mathcal{L}}^{(i)}(0) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\prod_{k=1}^N [1+(2\lambda_k^{(i)}\omega)^2]^{-1/4}}{\omega} \sin\left\{-\sum_{k=1}^N \left[\frac{\tan^{-1}(2\lambda_k^{(i)}\omega)}{2} + \frac{\omega c}{N}\right]\right\} d\omega \ . \tag{4.44}$$

By comparison of the exact and approximated log-likelihood ratios in Eqs. (4.6) and (4.26), respectively, we further see that

$$\mathbf{x}_n^T \left\{ \left[\mathbf{R}_x^{(0)}\right]^{-1} - \left[\mathbf{R}_x^{(1)}\right]^{-1} \right\} \mathbf{x}_n + \ln \frac{|\mathbf{R}_x^{(0)}|}{|\mathbf{R}_x^{(1)}|} \simeq N \left\{ \mathbf{x}_n^T \left[\mathbf{P}_0 - \mathbf{P}_1\right] \mathbf{x}_n + \ln \left(\frac{\sigma_{\epsilon_0}^2}{\sigma_{\epsilon_1}^2} \right) \right\} \tag{4.45}$$

where the substitution $\mathbf{P} = \mathbf{P}_0 - \mathbf{P}_1$ is defined by $\mathbf{P}_i = T_{\mathbf{A}}(\mathbf{A}_i)/(N-p)$, $i \in [0,1]$. We note that $T_{\mathbf{A}}(\mathbf{A}_i)$ has only $2p+1$ non-zero diagonals and that the entries on those diagonals do not have constant value. I.e., the matrices $\mathbf{P}_0$ and $\mathbf{P}_1$ are not Toeplitz like the inverse correlation matrices on the left hand side [Haykin 1996].

4.5 Detection with Additive White Noise

We shall now examine how additive white Gaussian noise affects the performance of the receiver. Assume that the transmitted ARPSK signal is contaminated by white, Gaussian and zero-mean observation noise $v(n) \sim N[0, \sigma_v^2]$. The received signal is modelled as

$$\begin{aligned} y(n) &= x(n) + v(n) \\ &= -\sum_{k=1}^{p} a_k x(n-k) + \epsilon(n) + v(n). \end{aligned} \tag{4.46}$$

Define the observation noise vector of N samples up to and including time n as $\mathbf{v}_n = [v(n-N+1), \ldots, v(n)]^T$. The received data vector is then

$$\mathbf{y}_n = \mathbf{x}_n + \mathbf{v}_n \tag{4.47}$$

From the assumptions on $v(n)$, we know that $\mathbf{y}_n$ is also zero-mean and Gaussian, with correlation matrix

$$\begin{aligned} \mathbf{R}_y^{(i)} &= \mathbf{R}_x^{(i)} + \mathbf{R}_v \\ &= \mathbf{R}_x^{(i)} + \sigma_v^2 \mathbf{I} \quad : \quad \text{under } H_i \,. \end{aligned} \tag{4.48}$$

The variance of the received signal is $\sigma_y^2 = \sigma_x^2 + \sigma_v^2$ and the power spectral density is

$$\begin{aligned} S_{yy}^{(i)}(\omega) &= S_{xx}^{(i)}(\omega) + S_{vv}(\omega) \\ &= \frac{\sigma_{\epsilon_i}^2}{|A^{(i)}(\omega)|^2} + \sigma_v^2 \quad : \quad \text{under } H_i \,. \end{aligned} \tag{4.49}$$

From Eq. (4.46) we can derive that

$$\begin{aligned} y(n) &= -\sum_{k=1}^{p} a_k[x(n-k)+v(n-k)] + \sum_{k=0}^{p} a_k v(n-k) + \epsilon(n) \\ &= -\sum_{k=1}^{p} a_k y(n-k) + \sum_{k=0}^{p} a_k v(n-k) + \epsilon(n) \end{aligned} \tag{4.50}$$

where $a_0 = 1$. The additive noise thus changes the process model from an autoregressive model (AR(p)) to an autoregressive-moving average with exogeneous input model (ARMAX(p,p,1)) [Box et al. 1994]. This result has an impact on the performance of the proposed detectors.

The Neyman-Pearson detector is model independent, and will still be optimum in a minimum detection error probability sense, provided that it knows the exact correlation matrices of the processes with noise. This requires knowledge of the noise variance, information which is not directly available. Estimation of σ_v^2 deteriorates the performance of the Neyman-Pearson detector.

The approximate likelihood ratio detector does not need any information about the exact autocorrelation functions of the transmitted processes, which is an advantage. On the other hand, it is designed specifically for detection of AR-processes, and does not take the model change into account. Thus, the accuracy of the ALR detector decreases as the noise increases and the contribution of the moving average (MA) part of the process becomes more significant.

4.6 Estimation of Additive White Noise Variance

In order to use the Neyman-Pearson detector, we need to estimate the variance σ_v^2 of the additive white noise (In the following, we do not make the assumption that the noise is Gaussian). Since the processes $x(n)$ and $v(n)$ are uncorrelated, we have from Eq. (4.46) that

$$\sigma_y^2 = \sigma_x^2 + \sigma_v^2 \tag{4.51}$$

which is equivalent to

$$\sigma_v^2 = r_{yy}(0) - r_{xx}(0)\ . \tag{4.52}$$

Because of the average process power equalisation constraint, we have $r_{xx}^{(0)}(0) = r_{xx}^{(1)}(0)$, where $\sigma_{x_i}^2 = r_{xx}^{(i)}(0)$ denotes the variance or average power of process X_i, $i \in [0,1]$. Hence,

the average power of the transmission processes with additive noise are also equal under the two hypotheses, $r_{yy}^{(0)}(0) = r_{yy}^{(1)}(0)$. Since $r_{xx}(0)$ is known a priori, the additive noise variance can be estimated from

$$\widehat{\sigma_v^2} = \hat{r}_{yy}(0) - r_{xx}(0) \; . \tag{4.53}$$

With the ACF estimator given in Eq. (2.45), this becomes

$$\widehat{\sigma_v^2} = \frac{1}{N} \sum_{n=1}^{N} y^2(n) - r_{xx}(0) \; . \tag{4.54}$$

From this equation, an estimate of σ_v^2 is obtained at each receival of a source bit, represented by the noisy process realisation $\mathbf{y}_n$. If the additive noise is stationary, then we can average over several such estimates to obtain an improved estimate for every bit that is received. The estimate average can be computed recursively as

$$\widehat{\sigma_v^2}(m) = \left(\frac{m-1}{m} \right) \hat{\sigma}_v^2(m-1) + \left(\frac{1}{m} \right) \left[\frac{1}{N} \sum_{n=1}^{N} y^2(n) - r_{xx}(0) \right] \tag{4.55}$$

where $\widehat{\sigma_v^2}(m)$ is the averaged estimate obtained at the receival of the mth source bit. From Eq. (4.54), we find that $\mathrm{E}\{\widehat{\sigma_v^2}\} = \sigma_v^2$, so the estimator is unbiased. From Eq. (4.53) we further see that the variance of $\widehat{\sigma_v^2}$ is equal to the variance of $\hat{r}_{yy}(0)$, which is given by [Kay 1993] as

$$\mathrm{Var}\{\widehat{\sigma_v^2}\} = \frac{2}{N^2} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} r_{yy}^2(m-n) \; . \tag{4.56}$$

Since $\lim_{N\to\infty} \mathrm{Var}\{\widehat{\sigma_v^2}\} = 0$, the estimator is consistent [Scharf 1991]. Hence, if the noise variance can be estimated recursively for a long sequence of received source bits, the performance of the suboptimal Neyman-Pearson detector (with noise variance estimation) will approach the performance of an ideal Neyman-Pearson detector (where the noise variance is assumed known), which cannot be realised.

If the additive noise is non-stationary, the noise variance can be estimated by a Kalman filter [Scharf 1991, Haykin 1996] that tracks the changes of the time-varying environment. The Kalman filter estimate is optimal in minimum mean squared error sense for Gaussian processes, both when the noise is stationary and non-stationary. Other estimators with different degree of memory can be designed using various window functions, but we shall use the simple estimator defined in Eq. (4.54).

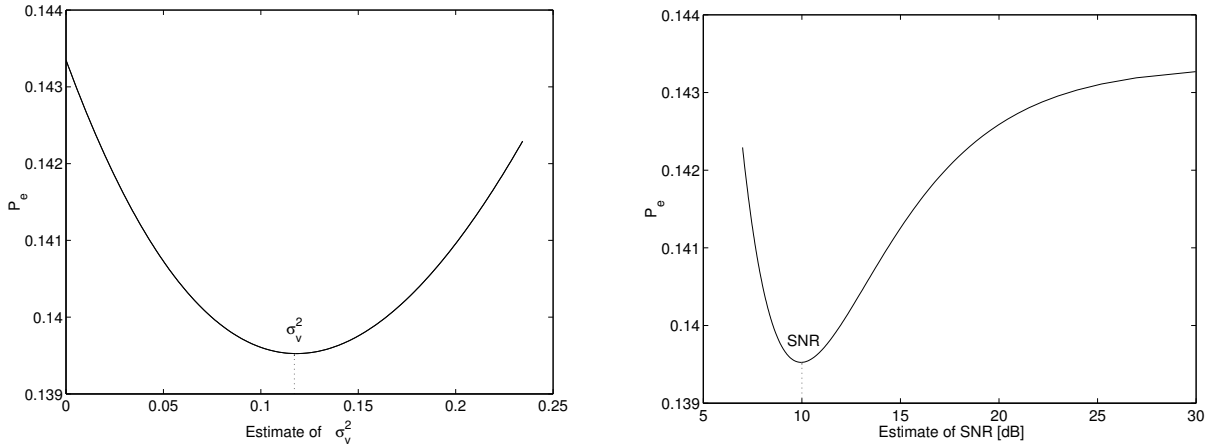


Figure 4.1: Detection error probabilities $P_e(\widehat{\sigma^2_{v_+}})$ (left panel) and $P_e(\widehat{\mathrm{SNR}})$ (right panel) as a function of the non-negative estimate of σ_v^2 and the corresponding SNR estimate (measured in dB). The ARPSK processes have parameter vectors $\mathbf{a}_0 = [0.4, 0.2]^T$, $\mathbf{a}_1 = [0.4, -0.2]^T$ and pulse length $N = 32$. The true $\sigma_v^2 = 0.1172$, which corresponds to SNR = 10.

There is one problem associated with practical use of the noise estimate in Eq. (4.53). The estimator might actually give negative values, which has no physical meaning. Therefore, we replace Eq. (4.53) with the non-negative measure

$$\begin{aligned} \widehat{\sigma^2_{v_+}} &= \frac{1}{2}\left(\widehat{\sigma^2_v} + |\widehat{\sigma^2_v}|\right) \\ &= \frac{1}{2}\Big[\,(\hat{r}_{yy}(0) - r_{xx}(0)) + |\hat{r}_{yy}(0) - r_{xx}(0)|\,\Big]\,. \end{aligned} \tag{4.57}$$

This estimator is implemented in the Neyman-Pearson detector.

It is not trivial to evaluate the effect of variance estimation on the detection error probability, but the exact P_e can be found. The estimator $\widehat{\sigma^2_{v_+}}$ is a stochastic variable which takes on different values for different process realisations. One specific estimate value can be produced by many different process realisations, but it is not certain that the same set of realisations are all correctly classified, regardless of which hypotheses they represent.

Let $P_e(\widehat{\sigma^2_{v_+}})$ be the probability that a process realisation that produces the estimate $\widehat{\sigma^2_{v_+}}$ is wrongly classified. Hence, the overall detection error probability is found as

$$P_e = E\{P_e(\widehat{\sigma^2_{v_+}})\} = \int_0^\infty P_e(\widehat{\sigma^2_{v_+}}) f_{\widehat{\sigma^2_{v_+}}}(\widehat{\sigma^2_{v_+}})\, d\widehat{\sigma^2_{v_+}} \tag{4.58}$$

where $f_{\widehat{\sigma_{v+}^2}}(\widehat{\sigma_{v+}^2})$ is the PDF of the modified estimator $\widehat{\sigma_{v+}^2}$. Hence, $P_e(\widehat{\sigma_{v+}^2})$ is the detection error probability associated with the decision rule

$$\mathbf{x}_n^T \left\{ \left[\hat{\mathbf{R}}_y^{(0)} \right]^{-1} - \left[\hat{\mathbf{R}}_y^{(1)} \right]^{-1} \right\} \mathbf{x}_n \underset{\Omega_0}{\overset{\Omega_1}{\gtrless}} \ln \frac{|\hat{\mathbf{R}}_y^{(1)}|}{|\hat{\mathbf{R}}_y^{(0)}|} \tag{4.59}$$

where $\hat{\mathbf{R}}_y^{(i)} = \mathbf{R}_x^{(i)} + \widehat{\sigma_{v+}^2}\mathbf{I}$, $i \in [0,1]$. An example of such a $P_e(\widehat{\sigma_{v+}^2})$ is shown in figure 4.1. The ARPSK communications system uses two AR(2)-processes with parameter vectors $\mathbf{a}_0 = [0.4, 0.2]^T$ and $\mathbf{a}_1 = [0.4, -0.2]^T$. We have a pulse length of $N = 32$ and SNR = 10, which corresponds to an additive noise variance of $\sigma_v^2 = 0.1172$ for $\sigma_{\epsilon_0}^2 = 1$.

The $P_e(\widehat{\sigma_{v+}^2})$ is displayed both as a function of the estimate $\widehat{\sigma_{v+}^2}$ (left panel) and as a function of the estimated SNR (right panel), given by $\widehat{\text{SNR}} = 10\log_{10}(\sigma_x^2/\widehat{\sigma_{v+}^2})$. We see that $P_e(\widehat{\sigma_{v+}^2})$ and $P(\widehat{\text{SNR}})$ are convex functions, and that their minima naturally occur at the true value of the σ_v^2 and the SNR, respectively.

The PDF of $\widehat{\sigma_{v+}^2}$ can be found by considering $\widehat{\sigma_y^2} = \hat{r}_{yy}(0)$. The noisy process $y(n)$ is zero-mean and Gaussian with $\sigma_y^2 = \sigma_x^2 + \sigma_v^2$. The estimator $\widehat{\sigma_y^2}$ can be seen as a scaled sum of N quadratic terms, $S = (\sigma_y^2/N)\sum_{n=1}^{N} s^2(n)$, where $s(n)$ is standardised Gaussian ($s(n) \sim N[0,1]$). The sum S is known to follow the χ^2 PDF given by Eq. (2.5). We also need the result that for a linear transformation $Y = aX + b$ of a continuous random variable X, the PDF of Y is [Larsen and Marx 1986]

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \tag{4.60}$$

where a and b are real constants and $a \neq 0$. Together with Eq. (2.5), this is used to show that $\widehat{\sigma_y^2}$ has a χ^2-like PDF given by

$$f_{\widehat{\sigma_y^2}}(\widehat{\sigma_y^2}) = \frac{1}{\Gamma(N/2)\widehat{\sigma_y^2}} \left(\frac{N\widehat{\sigma_y^2}}{2\sigma_y^2}\right)^{N/2} \exp\left(-\frac{N\widehat{\sigma_y^2}}{2\sigma_y^2}\right) . \tag{4.61}$$

From this expression, we can show that $E\{\widehat{\sigma_y^2}\} = \sigma_y^2$ (which we already know) and $Var\{\widehat{\sigma_y^2}\} = 2\sigma_y^4/N$. The estimators $\widehat{\sigma_y^2}$ and $\widehat{\sigma_v^2}$ have equal variance and differ only in the mean value by the constant σ_x^2. Hence, it follows from Eq. (4.60) that

$$f_{\widehat{\sigma_v^2}}(\widehat{\sigma_v^2}) = \frac{1}{\Gamma(N/2)(\widehat{\sigma_v^2} + \sigma_x^2)} \left[\frac{N(\widehat{\sigma_v^2} + \sigma_x^2)}{2\sigma_y^2}\right]^{N/2} \exp\left[-\frac{N(\widehat{\sigma_v^2} + \sigma_x^2)}{2\sigma_y^2}\right] . \tag{4.62}$$

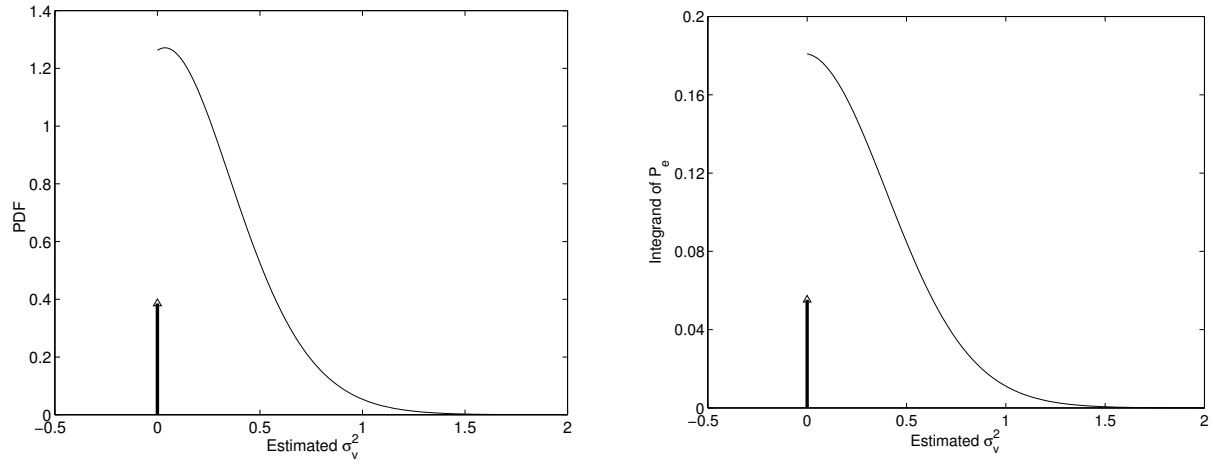


Figure 4.2: The PDF $f_{\widehat{\sigma^2_{v+}}}(\widehat{\sigma^2_{v+}})$ of the non-negative estimator of σ_v^2 (left panel) and the integrand $P_e(\widehat{\sigma^2_{v+}}) f_{\widehat{\sigma^2_{v+}}}(\widehat{\sigma^2_{v+}})$ of the integral defining the total P_e for a Neyman-Pearson detector that uses $\widehat{\sigma^2_{v+}}$ to estimate σ_v^2 (right panel). The ARPSK processes have parameter vectors $\mathbf{a}_0 = [0.4, 0.2]^T$, $\mathbf{a}_1 = [0.4, -0.2]^T$, a pulse length of $N = 32$ and SNR $= 10$.

The modified estimator $\widehat{\sigma^2_{v+}}$ maps all negative values of $\widehat{\sigma^2_v}$ to zero. This creates a Dirac delta function at $\widehat{\sigma^2_{v+}} = 0$ in the PDF of $\widehat{\sigma^2_{v+}}$. Hence, we have

$$f_{\widehat{\sigma^2_{v+}}}(\widehat{\sigma^2_{v+}}) = \begin{cases} \delta(\widehat{\sigma^2_{v+}}) \int_{-\sigma_x^2}^{0} f_{\widehat{\sigma^2_v}}(\widehat{\sigma^2_v})\, d\widehat{\sigma^2_v} & , \widehat{\sigma^2_{v+}} = 0 \\ f_{\widehat{\sigma^2_v}}(\widehat{\sigma^2_{v+}}) & , \widehat{\sigma^2_{v+}} > 0 \end{cases} . \tag{4.63}$$

This expression is inserted into Eq. (4.58) together with $P_e(\widehat{\sigma^2_{v+}})$, which can be calculated from the same equations as the P_e for the ALR detector, that is Eqs. (4.38) and (4.44). In Eq. (4.44), we only need to substitute the constant c with $\varsigma = \ln(|\hat{\mathbf{R}}_y^{(0)}|/|\hat{\mathbf{R}}_y^{(1)}|)$ and the eigenvalues $\{\lambda_k^{(i)}\}_{k=1}^N$ with the set $\{\tilde{\lambda}_k^{(i)}\}_{k=1}^N$ obtained from the generalised eigenvalue problem

$$\left[(\hat{\mathbf{R}}_y^{(0)})^{-1} - (\hat{\mathbf{R}}_y^{(1)})^{-1}\right] \mathbf{R}_y^{(i)} \mathbf{u}_k = \tilde{\lambda}_k^{(i)} \mathbf{u}_k \ : \ H_i \, . \tag{4.64}$$

The PDF of the non-negative additive noise variance estimator, $f_{\widehat{\sigma^2_{v+}}}(\widehat{\sigma^2_{v+}})$, is shown in the left panel of figure 4.2. The processes and parameters used in the example are the same as for figure 4.1. We observe the previously described delta function at $\widehat{\sigma^2_{v+}} = 0$, while the rest of the function follows the χ^2-like PDF of $f_{\widehat{\sigma^2_v}}(\widehat{\sigma^2_v})$. The maximum of a χ^2_N PDF occurs at $N-2$. From what we know about the χ^2_N PDF and linear transformations, we can show that the maximum of $f_{\widehat{\sigma^2_{v+}}}(\widehat{\sigma^2_{v+}})$ occurs at $[(N-2)\sigma_v^2 - \sigma_x^2]/N = 0.0732$, which is confirmed

by the figure.

When $f_{\widehat{\sigma^2_{v+}}}(\widehat{\sigma^2_{v+}})$ is multiplied with $P(\widehat{\sigma^2_{v+}})$, we get the integrand of the integral in Eq. (4.58). This product is plotted in the right panel of figure 4.2. The plot shows how different values of $\widehat{\sigma^2_{v+}}$ contribute to the total P_e for the Neyman-Pearson detector that estimates the additive noise variance. In this example we have $P_e = 0.1483$, while the ideal Neyman-Pearson detector (assuming known additive noise variance) provides $P_e = 0.1394$.

In section 6, the P_e derived assuming known σ_v^2 is used as a bound for detector performance and will be compared with theoretical results and simulation results obtained when σ_v^2 is estimated.

The requirement that $\widehat{\sigma^2_{v+}}$ should be calculated from the samples of only one source symbol is conservative, since this asserts that the aditive white noise varies very fast. In practice, it is more reasonable to assume that the additive noise is piecewise stationary, such that the noise variance estimates can be averaged over the samples of M symbols. The variance of $\widehat{\sigma^2_{v+}}$ will then decrease by a factor M. The appropriate choice of M will depend on the channel, and must be subject to a test for each specific application.

4.7 Detection with Synchronisation Error

Before efficient decoding of the ARPSK communications signal can take place, it is imperative that we obtain perfect synchronisation at the receiver [Meyr et al. 1998]. From the discrete stream of received samples, we have to extract the segments that correspond to distinct process realisations. That is, we have to identify the discrete time instants $n - N, n - 2N, \ldots$ when the initial samples $x(n - kN)$ of the data vector $\mathbf{x}_n$ is received.

Synchronisation algorithms for ARPSK communications will not be addressed in this thesis. In the previous sections, we have assumed that perfect synchronisation has already been achieved. We will now assess the effect of synchronisation errors on the performance of the Neyman-Pearson detector. In the analysis, we assume zero additive noise.

Let the incorrectly synchronised data vector be denoted by

$$\mathbf{x}_{n,d_s} = [x^{(i)}(d_s + 1), \ldots, x^{(i)}(N), x^{(j)}(1), \ldots, x^{(j)}(d_s)]^T \tag{4.65}$$

where d_s is the synchronisation delay measured in sample intervals T/N. Vector element $x^{(i)}(n)$ denotes the nth element of a realisation of process X_i. The data vector consists of

$N - d_s$ samples of process X_i and d_s samples of process X_j, when source bit i is followed by source bit j and $i, j \in [0, 1]$.

The cross-correlation between two samples from consecutive process realisations representing the source bits i and j is

$$E\{x^{(i)}(n)x^{(j)}(m)\} = \begin{cases} r_{xx}(N+m-n) & : \quad i = j \\ 0 & : \quad i \neq j \end{cases} . \tag{4.66}$$

This result occurs since the transmitter consists of two independent signal generators which are continuously producing streams of the respective process realisations. If two consecutive process source bits are equal, then the generator will use the last samples of the first bit realisation to generate the first p samples of the second bit representation. Hence, in this case, the true correlation matrix of $\mathbf{x}_{n,d_s}$ is

$$\mathbf{R}_{x,N}^{(i,j,d_s)} = \mathbf{R}_x^{(i)} \quad : \quad i = j \tag{4.67}$$

where $\mathbf{R}_{x,N}^{(i,j,d_s)}$ is superscripted by source bits i and j and the synchronisation delay d_s, while the second subscript denotes the dimension of the square matrix.

If, on the other hand, the consecutive source bits are different, then there is no dependence between the samples of the first and and the second bit realisation, since the transmitter switches abruptly from one process generator to the other. The correlation matrix of the unsynchronised data vector is thus given by

$$\mathbf{R}_{x,N}^{(i,j,d_s)} = \begin{bmatrix} \mathbf{R}_{x,N-d_s}^{(i)} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{x,d_s}^{(j)} \end{bmatrix} \quad : \quad i \neq j \ . \tag{4.68}$$

Let $P_e^{(i,j)}(d_s)$ denote the detection error probability for the incorrectly synchronised process realisation defined in Eq. (4.65). The total detection error probability for a synchronisation delay d_s is then given by

$$P_e(d_s) = \sum_{i=0}^{1} \sum_{j=0}^{1} p_{ij} P_e^{(i,j)}(d_s) \tag{4.69}$$

where $p_{ij} = 1/4$ is the probability that $\mathbf{x}_n$ contains samples representing source symbol i, followed by samples representing source symbol j. When $i = j$, we have $P_e^{(i,j)}(d_s) = P_e^{(i,j)}(0)$, which is the P_e of an ideal Neyman-Pearson detector under H_i. I.e., the detection error probability of an incorrectly synchronised receiver is equal to that of an perfectly synchronised receiver, as long as only one source symbol is transmitted.

In general, all the $P_e^{(i,j)}(d_s)$ can be calculated from the corresponding characteristic functions $\Phi_L^{(i,j,d_s)}(\omega)$, defined as

$$\begin{aligned}\Phi_L^{(i,j,d_s)}(\omega) &= |\mathbf{I} - 2j\omega\left[(\mathbf{R}_x^{(0)})^{-1} - (\mathbf{R}_x^{(1)})^{-1}\right]\mathbf{R}_{x,N}^{(i,j,d_s)}|^{-1/2}\, e^{j\omega\zeta} \\ &= \prod_{k=1}^{N}[1 + (2\omega\lambda_k^{(i,j)})^2]^{-1/4} \exp\left\{ j\sum_{k=1}^{N}\left[\frac{1}{2}\tan^{-1}(2\omega\lambda_k^{(i,j)}) + \frac{\omega(-\zeta)}{N}\right]\right\}\end{aligned} \tag{4.70}$$

where the eigenvalues $\{\lambda_k^{(i,j)}\}_{k=1}^N$ are obtained from the generalised eigenvalue problem

$$\left[(\mathbf{R}_x^{(0)})^{-1} - (\mathbf{R}_x^{(1)})^{-1}\right]\mathbf{R}_{x,N}^{(i,j,d_s)}\mathbf{u}_k = \lambda_k^{(i,j)}\mathbf{u}_k\ . \tag{4.71}$$

and the constant $\zeta = \ln(|\mathbf{R}_x^{(1)}|/|\mathbf{R}_x^{(0)}|)$ has been previously defined as the threshold of the Neyman-Pearson detector. The relation between $P_e^{(i,j)}(d_s)$ and $\Phi_L^{(i,j,d_s)}(\omega)$ is given by

$$P_e^{(i,j)}(d_s) = \begin{cases} 1 - F^{(i,j,d_s)}(0) & : \quad i = 0 \\ F^{(i,j,d_s)}(0) & : \quad i = 1 \end{cases} \tag{4.72}$$

where $d_s \leq 16$. If $d_s > 16$, then the conditions ($i = 0$ and $i = 1$) on the right-hand side of the above equation should be switched. The CDF $F^{(i,j)}(0)$ can be calculated from

$$F^{(i,j,d_s)}(0) = \frac{1}{2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} -\frac{\mathrm{Im}[\Phi_L^{(i,j,d_s)}(-\omega)]}{\omega}\, d\omega \tag{4.73}$$

or the simplified expression that can be derived in analogy with Eq. (4.25). The detection error probability of an unsynchronised ALR detector can be calculted from the exact same procedure, only substituting the eigenvalues $\{\lambda_k^{(i,j)}\}_{k=1}^N$ with those obtained from the general eigenvalue problem

$$\mathbf{P}\mathbf{R}_{x,N}^{(i,j,d_s)}\mathbf{u}_k = \lambda_k^{(i,j)}\mathbf{u}_k \tag{4.74}$$

and the constant ζ with $-c = \ln(\sigma_{\epsilon_1}^2/\sigma_{\epsilon_0}^2)$.

4.8 A Unifying Framework

We realise that the detection error probability can be calculated from similar procedures in all the cases that have been studied. This is possible because all the detectors can be written as the general expression

$$Q = \mathbf{z}^T\mathbf{M}\mathbf{z} + C \tag{4.75}$$

where $\mathbf{z}$ is the data vector of received process samples, without saying anything about noise or synchronisation, $\mathbf{M}$ is a weighting matrix and C is a threshold constant. The different cases differ by $\mathbf{M}$, C and the correlation matrices $\mathbf{R}_z^{(i)}$, $i \in [0,1]$ of $\mathbf{z}$. The detection error probability can be written as a function of the pulse length N and the eigenvalues $\{\lambda_k^{(i)}\}_{K=1}^N$ of the generalised eigenvalue problem

$$\mathbf{M}\mathbf{R}_z^{(i)}\mathbf{u}_k = \lambda_k^{(i)}\mathbf{u}_k \ : \ H_i \ . \tag{4.76}$$

We have the detection error probability

$$P_e = \frac{1}{2}\left[1 - F_Q^{(0)}(0) + F_Q^{(1)}(0)\right] \tag{4.77}$$

where the cumulative distribution functions $F_Q^{(i)}(0)$ are calculated from

$$F_Q^{(i)}(0) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\prod_{k=1}^N [1 + (2\lambda_k^{(i)}\omega)^2]^{-1/4}}{\omega} \cdot \sin\left\{-\sum_{k=1}^N \left[\frac{\tan^{-1}(2\lambda_k^{(i)}\omega)}{2} + \frac{\omega C}{N}\right]\right\} \, d\omega \ . \tag{4.78}$$

Table 4.1 lists the appropriate expressions used in calculations of the P_e for: (i) The ideal Neyman-Pearson (NP) detector with zero additive noise, (ii) the ideal NP detector with non-zero additive noise, (iii) the suboptimal NP detector implemented with additive white noise estimator $\widehat{\sigma_v^2}$, (iv) the ideal NP detector with synchronisation error, (v) the ALR detector and (vi) the ALR detector with synchronisation error.

However, three of the cases include additional requirements to how P_e is calculated. In case (iii), we must take the expectation value of $P_e(\widehat{\sigma_v^2})$ with respect to the estimate $\widehat{\sigma_v^2}$ to obtain the final P_e value. In case (iv) and (vi), P_e must be calculated for the two values of the process index $j \in [0,1]$ (cf. the true correlation matrix $\mathbf{R}_z = \mathbf{R}_y^{(i,j,d_s)}$), and then averaged.

	Detector	$\mathbf{M}$	C	$\mathbf{R}_z^{(i)}$
(i)	NP($\sigma_v^2 = 0$)	$[\mathbf{R}_x^{(0)}]^{-1} - [\mathbf{R}_x^{(1)}]^{-1}$	$\ln\left(\frac{\lvert\mathbf{R}_x^{(0)}\rvert}{\lvert\mathbf{R}_x^{(1)}\rvert}\right)$	$\mathbf{R}_x^{(i)}$
(ii)	NP($\sigma_v^2 \neq 0$)	$[\mathbf{R}_y^{(0)}]^{-1} - [\mathbf{R}_y^{(1)}]^{-1}$	$\ln\left(\frac{\lvert\mathbf{R}_y^{(0)}\rvert}{\lvert\mathbf{R}_y^{(1)}\rvert}\right)$	$\mathbf{R}_y^{(i)}$
(iii)	NP($\widehat{\sigma_v^2}$)	$[\mathbf{R}_x^{(0)} + \widehat{\sigma_v^2}\mathbf{I}]^{-1} - [\mathbf{R}_x^{(1)} + \widehat{\sigma_v^2}\mathbf{I}]^{-1}$	$\ln\left(\frac{\lvert\mathbf{R}_x^{(0)}+\widehat{\sigma_v^2}\mathbf{I}\rvert}{\lvert\mathbf{R}_x^{(1)}+\widehat{\sigma_v^2}\mathbf{I}\rvert}\right)$	$\mathbf{R}_y^{(i)}$
(iv)	NP(d_s)	$[\mathbf{R}_y^{(0)}]^{-1} - [\mathbf{R}_y^{(1)}]^{-1}$	$\ln\left(\frac{\lvert\mathbf{R}_y^{(0)}\rvert}{\lvert\mathbf{R}_y^{(1)}\rvert}\right)$	$\mathbf{R}_y^{(i,j,d_s)}$
(v)	ALR	$\frac{1}{N-p}\,T_{\mathbf{A}}\left(\frac{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0^T}{\sigma_{\epsilon_0}^2} - \frac{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T}{\sigma_{\epsilon_1}^2}\right)$	$\ln\left(\frac{\sigma_{\epsilon_0}^2}{\sigma_{\epsilon_1}^2}\right)$	$\mathbf{R}_y^{(i)}$
(vi)	ALR(d_s)	$\frac{1}{N-p}\,T_{\mathbf{A}}\left(\frac{\boldsymbol{\alpha}_0\boldsymbol{\alpha}_0^T}{\sigma_{\epsilon_0}^2} - \frac{\boldsymbol{\alpha}_1\boldsymbol{\alpha}_1^T}{\sigma_{\epsilon_1}^2}\right)$	$\ln\left(\frac{\sigma_{\epsilon_0}^2}{\sigma_{\epsilon_1}^2}\right)$	$\mathbf{R}_y^{(i,j,d_s)}$

Table 4.1: Summary of calculation procedures for detection error probabilities.

Chapter 5

Selection of Transmission Processes

It is still a somwhat open question how the parameters of the autoregressive transmission processes representing bit '0' and bit '1' should be chosen. It is difficult to design a cost function that absorbes all logical constraints on the process pair, and it is even harder to find one that can be optimised with respect to the AR-parameters and the model order.

Even if the problem is simplified and subdivided into several stages, many compromises must be made. We initiate the discussion by launching the following criteria that the transmission processes should fulfill.

5.1 Selection Criteria

(i) The processes should provide low detection error probability P_e, in order to meet the demands of a high performance communications system.

(ii) The distance between the processes should be short, in some statistical sense, so that eavesdropping is made as difficult as possible for unauthorised listeners.

(iii) The processes should have similar spectral characteristics, again motivated by security, which is the main objective of the SPSK communications approach.

(iv) The processes should offer the highest possible resistance to additive white noise.

(v) The processes should offer the highest possible resistance to tone jamming and intensional interference from a hostile source.

Some of these criteria conflict, as will be demonstrated shortly. Despite this fact, a selection procedure evolves as we go along and examine and comment on the criteria in more detail.

Criterion (i) is apparently an obvious statement, but the practical significance is that a target P_e must be specified. Since P_e depends strongly on the pulse length N, and also in some way on the process order p, these parameters must be chosen as the very first step. After that, the desired P_e at a given signal-to-noise ratio (SNR) must be decided. The SNR (measured in dB) is defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{\sigma_x^2}{\sigma_v^2} \right) \ . \tag{5.1}$$

A natural choice is to specify P_e at $\sigma_v^2 = 0$ (SNR $= \infty$). Also remark that the choice of p and N may be affected by practical constraints. The pulse length N determines the data rate, while both p and N affect the computational complexity of the transmitter and receiver.

Criterion (i) conflicts with criterion (ii) and (iii), and this illustrates the major compromise that has to made in SPSK communications, the trade-off between performance and security. Processes should be as similar as possible, to avoid eavesdropping, but not indistinguishable for the authorised receiver. However, security is the first priority and this should direct the decision when P_e is specified in the first place.

As soon as P_e is specified, the choices of process pairs are infinite in number. Therefore, one of the processes must be fixed. Criterion (v) can now be used in the selection. An appropriate wide-band process which utilizes the allocated bandwidth to maximum extent should be used. The PSD of two possible choices are shown in figure 5.1. The parameter vectors are $\mathbf{a}_0 = [0.4, 0.2]^T$ for AR(2)-process in the left panel and $\mathbf{a}_0 = [0.4, 0.3, 0.2]^T$ for the AR(3)-process in the right panel. The spectra have highpass characteristics, but the bandwidth is relatively wide.

When process '0' is selected, the other process must be chosen from a surface in p-dimensional parameter space. This is the surface of p-dimensional parameter vectors that provide the specified $P_e(N)$ for the decided pulse length N. The surface is closed, as long as it is not intersected by the region of parameter space which corresponds to unstable processes.

The described parameter surfaces are visualised by examples in figure 5.2 and 5.3. Parameter vector $\mathbf{a}_0$ is fixed in both cases. For the AR(2)-process, the set of allowed parameter vectors at a specified $P_e(N)$: $\{\mathbf{a}_1 | \mathbf{a}_0, P_e(N)\}_{p=2}$, is a curve in the two-dimensional parame-

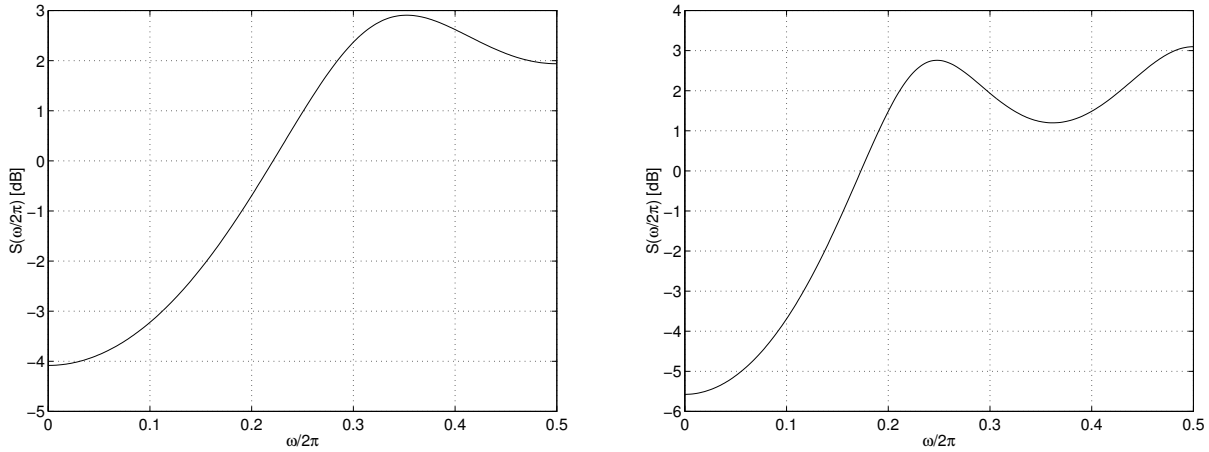


Figure 5.1: Power spectral densities of an AR(2)-process with parameter vector $\mathbf{a}_0 = [0.4, 0.2]^T$ (left) and an AR(3)-process with parameter vector $\mathbf{a}_0 = [0.4, 0, 3, 0.2]^T$ (right) that can be used as transmission process X_0.

ter plane. If the first order parameter of the AR(3)-process is fixed, then $\{\mathbf{a}_1|\mathbf{a}_0, P_e(N)\}_{p=3}$ is also reduced to a two-dimensional curve. But in general, $\{\mathbf{a}_1|\mathbf{a}_0, P_e(N)\}_{p=3}$ has dimension three, as shown in figure 5.3.

All curves are produced by a computer routine which searches the parameter space for a parameter vector

$$\mathbf{a}_1(\theta) = [a_1^{(0)} + r_\theta \cos\theta, a_2^{(0)} + r_\theta \sin\theta, a_3^{(0)}, \ldots, a_p^{(0)}]^T \ , \quad \theta \in \Theta \tag{5.2}$$

that produce the target P_e with a maximum allowed deviation of $P_e/100$ for the predetermined pulse length N. In the examples in figure 5.2 and 5.3, $N = 64$. The two-dimensional search space is defined by a set of direction angles $\Theta = [0, 2\pi/40, \ldots, 2\pi]$ (40 sample values), after all but two AR-parameters are fixed. The search is carried out through variation of r_θ at a fixed θ. For evaluation of P_e, the theoretical expression for the Neyman-Pearson detector without additive noise is chosen, since this defines the lower bound of P_e (optimal detector with respect to P_e).

In figure 5.2, the innermost curve in both panels is the equiprobability curve for $P_e = 10^{-1}$. The P_e is then decreased in steps of 10^{-1}, and the distance between the points on a curve and $\mathbf{a}_0$ naturally increases with decreasing P_e, but at different rate for different direction angles θ in the parameter plane. The shape of the equiprobability curve is seen to converge towards a characteristic shape for the chosen $\mathbf{a}_0$, which can be defined as the

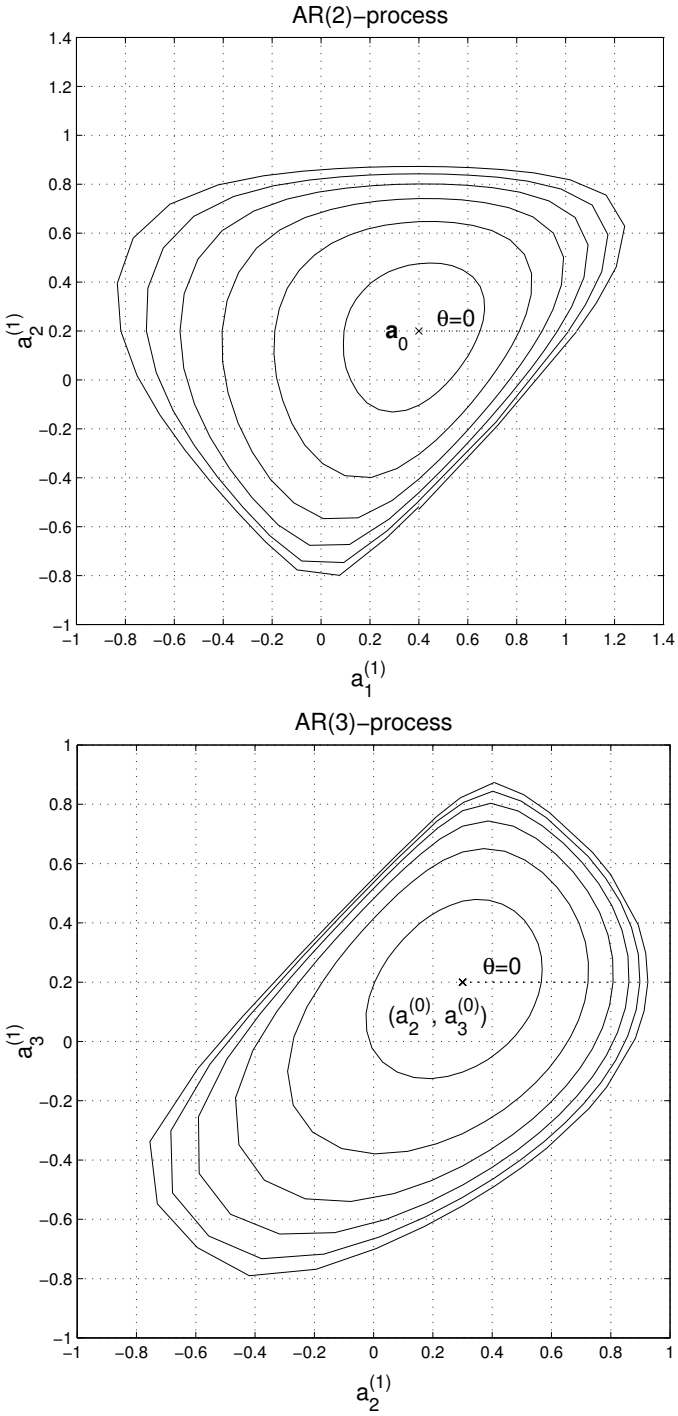


Figure 5.2: Equiprobability curves for P_e for the AR(2)-process $\mathbf{a}_0 = [0.4, 0.2]$ (upper panel) and the AR(3)-process $\mathbf{a}_0 = [0.4, 0.3, 0.2]$ (lower panel). For the AR(3)-process, the first AR-parameter is fixed at $a_1^{(0)} = 0.4$. Curves are plotted for $P_e = [10^{-1}, \ldots, 10^{-6}]$ in steps of 10^{-1}, for a pulse length of $N = 64$.

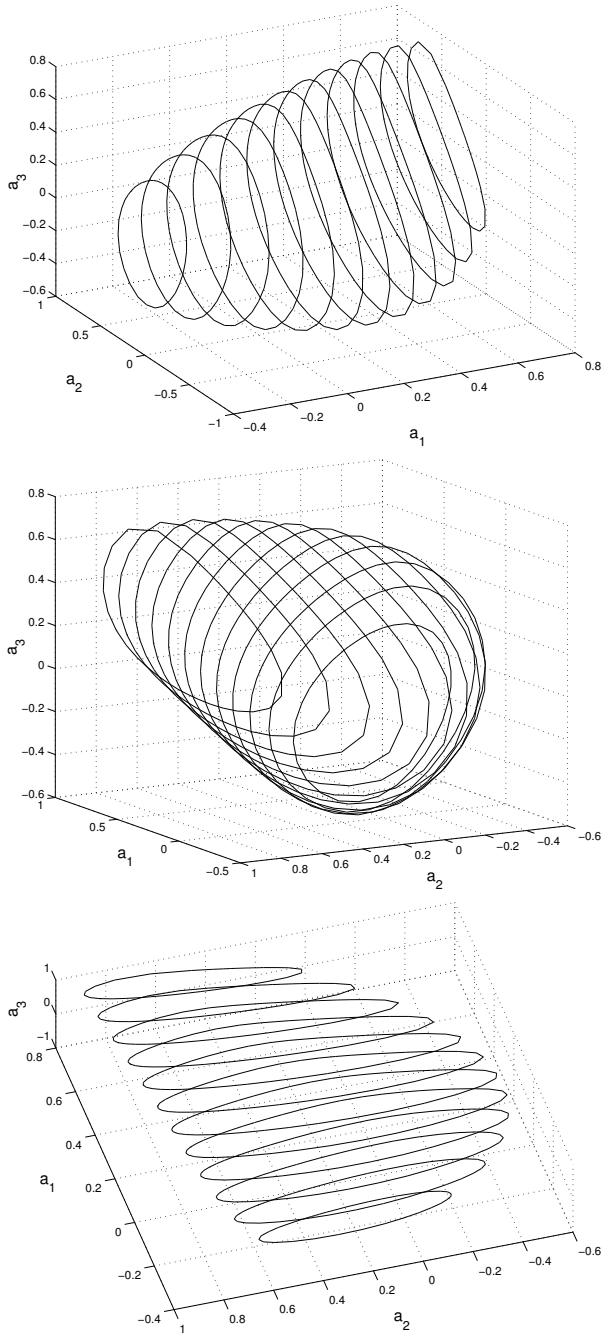


Figure 5.3: Cross-sections of the equiprobability surface for P_e for the AR(3)-process with parameter vector $\mathbf{a}_0 = [0.4, 0.3, 0.2]$, shown at different angles. All vectors $\mathbf{a}_1$ on the surface that is indicated by the cross-sections, satisfy the target $P_e = 10^{-3}$ for a pulse length of $N = 64$. Cross-sections are obtained for fixed values of the parameter $a_1^{(0)}$.

shape of the equiprobability curve for $P_e \to 0$.

With decreasing P_e, we also see that the shape of the equiprobability curves deviate more and more from circular, which would be the shape of curves defining equal Euclidean distance from $\mathbf{a}_0$. This illustrates a point that was made in section 3.1: The Euclidean distance in the AR-parameter plane is not a good statistical distance measure, and it becomes worse with decreasing P_e.

In figure 5.3, the equiprobability surface in the 3-dimensional parameter space is indicated by cross-sections. The cross-sections are equiprobability curves obtained in a 2-dimensional parameter space, after the parameter $a_1^{(0)}$ has been fixed. All curves are thus calculated for the same target $P_e = 10^{-3}$. The indicated surface is shown at three different angles as an aid in the visualisation of the shape of the object, which is impossible to describe in terms of simple geometry.

5.2 Robustness to Additive White Noise

According to criterion (iv), the processes should also be robust to additive noise. On a given surface in p-dimensional space, all parameter vectors provide the prescribed P_e at SNR $= \infty$, but the same processes will produce non-uniform values of P_e at other noise levels. Thus, consulting criterion (iv) alone, the process with the lowest P_e at non-zero noise levels should be chosen.

This is not so simple in practice, though. Unfortunately, processes with good noise resistance properties are observed to be those whose spectral maximum (peak frequency in the PSD) has the largest separation from the spectral maximum of the fixed process X_0. This can be seen from figure 5.4, where P_e is plotted at non-zero noise levels for two sets of processes, $\{\mathbf{a}_1|\mathbf{a}_0, P_e(N)\}_{p=2}$ and $\{\mathbf{a}_1|\mathbf{a}_0, P_e(N)\}_{p=3}$, which are obtained with the AR(2)-process $\mathbf{a}_0 = [0.4, 0.2]^T$ and the AR(3)-process $\mathbf{a}_0 = [0.4, 0.3, 0.2]^T$ that were used in the examples of figure 5.2 and 5.3, respectively. In both cases, we choose $N = 64$ and target $P_e(N) = 10^{-4}$. For the AR(3)-process, the parameter $a_1^{(1)} = 0.4$ is fixed.

In the figure, P_e is displayed as a function of the direction angle θ in the two-dimensional search space. First of all, we see that there is a relatively large variation in $P_e(\theta)$, except for at $SNR = 40$, where the noise is negligible. More interestingly, we note that for the AR(2)-processes, the maximum of $P_e(\theta)$ is found for the θ that provides an $\mathbf{a}_1(\theta)$ such that

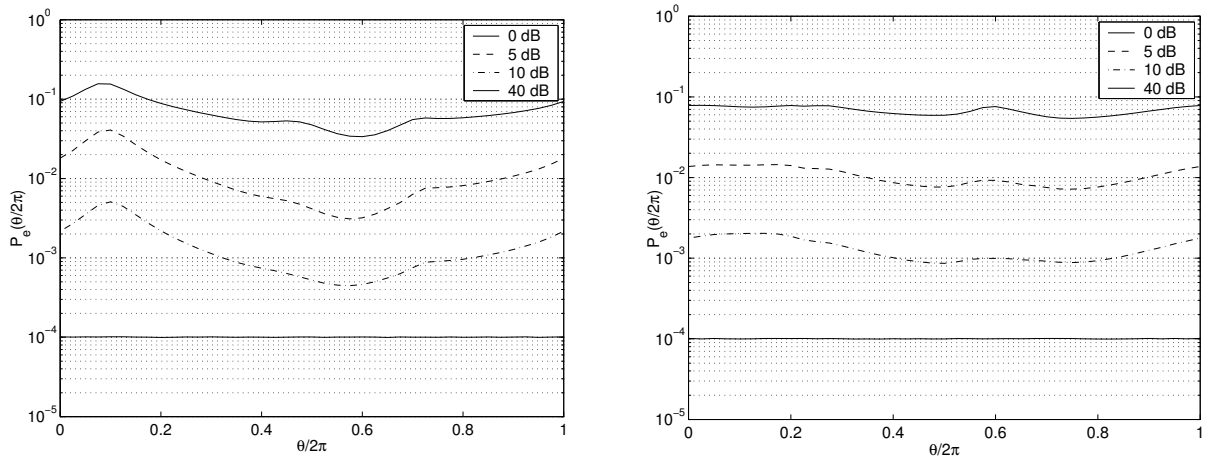


Figure 5.4: Detection error probability calculated at different noise levels for the AR(2)-processes with parameter vectors $\mathbf{a}_1(\theta)$ and $\mathbf{a}_0 = [0.4, 0.2]^T$ (left), and for the AR(3)-processes with $\mathbf{a}_0 = [0.4, 0.3, 0.2]^T$ and $\mathbf{a}_1(\theta)$, where $a_3^{(1)} = 0.2$ is fixed. In both cases, the set $\mathbf{a}_1(\theta)$ is chosen such that $P_e(\theta) = 10^{-4}$ at zero additive noise.

the maxima of $S_{yy}^{(0)}(\omega)$ and $S_{yy}^{(1)}(\omega, \theta)$ match. That is

$$\arg\{\max_{\theta}[P_e(\theta)]\} = \left\{\theta \ : \ \arg\{\max_{\omega}[S_{yy}^{(0)}(\omega)]\} = \arg\{\max_{\omega}[S_{yy}^{(1)}(\omega)]\}\right\} \tag{5.3}$$

However, this is not the exact case for the AR(3)-processes, that have more complex spectra. Note for instance, that the PSD of process X_0 have two peaks of almost the same magnitude, as seen from the right panel of figure 5.1). A similar type of behaviour to that of the AR(2)-processes is indeed observed, but we cannot be as firm about the location the maxima of $P_e(\theta)$ as in the former case.

From the observations, we conclude that there is a conflict between criteria (iii) and (iv). If the peak frequencies of the processes are widely separated, the ARPSK modulation technique turns into a coarse frequency shift keying (FSK) technique, which means that security is compromised. If white noise resistance is associated with this hazard, then criterion (iv) must be rejected from the process selection procedure.

The FSK interpretation gives an intuitive feel of why and how the white noise resistance varies with different process choices. From classical communications we know that the P_e of a communications system with additive noise decreases with increasing distance between two FSK carriers [Gibson 1993]. It is difficult to give an exact mathematical explanation of the observed behaviour for ARPSK communications. Some remarks can be made though,

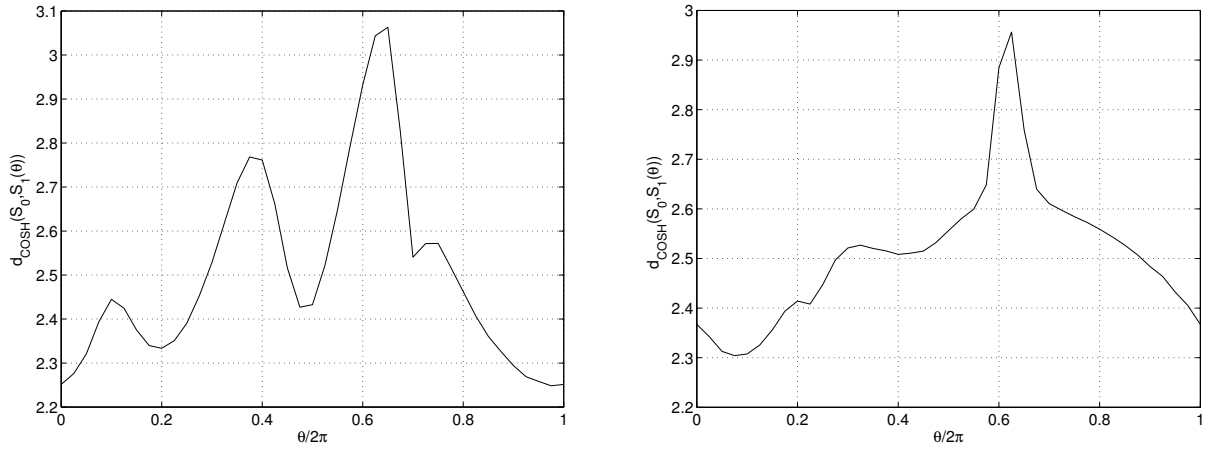


Figure 5.5: Cosh distance between the AR(2)-processes with parameter vectors $\mathbf{a}_0 = [0.4, 0.2]^T$ and $\mathbf{a}_1(\theta)$ (left), and between the AR(3)-processes with $\mathbf{a}_0 = [0.4, 0.3, 0.2]^T$ and $\mathbf{a}_1(\theta)$ (right). In both cases, the set $\mathbf{a}_1(\theta)$ is chosen such that $P_e(\theta) = 10^{-4}$ at zero additive noise.

with aid of the d_{COSH} distance measure, which is repeated here for convenience

$$d_{COSH}(S_0, S_1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\left(S_{xx}^{(1)}(\omega) - S_{xx}^{(0)}(\omega)\right)^2}{S_{xx}^{(0)}(\omega) S_{xx}^{(1)}(\omega)} \, d\omega \; . \tag{5.4}$$

From the derivation in section 3.6, we know that the d_{COSH} measure serves as an approximation to P_e and is a measure of how the log-likelihood ratio distinguishes between the transmission processes [Scharf 1991]. It can tell something about what features that are important for discrimination.

Figure 5.5 shows $d_{COSH}(S_0, S_1(\theta))$ for different $\mathbf{a}_1(\theta)$ at zero noise. The function is not constant like $P_e(\theta)$, since it is not a perfect representation of latter. Nevertheless, we know that the Jeffreys divergence d_J is a convex function of the likelihood ratio [Kailath 1967], a result which is valid for d_{COSH} as well, since d_{COSH} is an asympotic derivative of d_J.

Next consider the normalised measure

$$\begin{aligned} \frac{d_{COSH}(S_0, S_1|\sigma_v^2)}{d_{COSH}(S_0, S_1)} &= \frac{\int_{-\pi}^{\pi} \left(S_{yy}^{(1)}(\omega) - S_{yy}^{(0)}(\omega)\right)^2 / S_{yy}^{(0)}(\omega) S_{yy}^{(1)}(\omega) \, d\omega}{\int_{-\pi}^{\pi} \left(S_{xx}^{(1)}(\omega) - S_{xx}^{(0)}(\omega)\right)^2 / S_{xx}^{(0)}(\omega) S_{xx}^{(1)}(\omega) \, d\omega} \\ &= \frac{\int_{-\pi}^{\pi} D(\omega) G(\omega) \, d\omega}{\int_{-\pi}^{\pi} D(\omega) \, d\omega} \end{aligned} \tag{5.5}$$

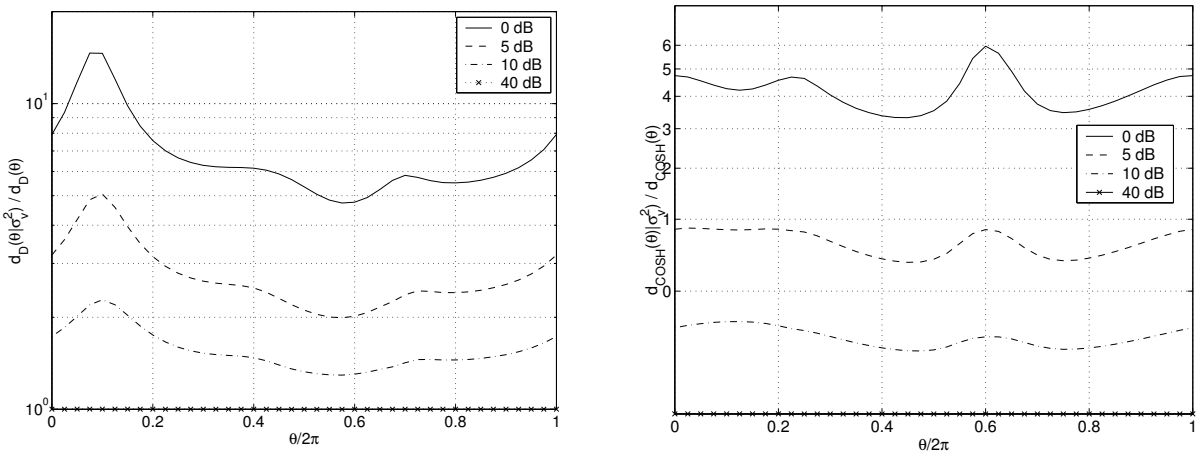


Figure 5.6: Inverse normalised Cosh distance $d_{COSH}(\theta)/d_{COSH}(\theta|\sigma_v^2)$ of AR(2)-processes $\mathbf{a}_1(\theta)$ that satisfy $P_e = 10^{-4}$ for $\mathbf{a}_0 = [0.4, 0.2]^T$ (left) and of AR(3)-processes that satisfy $P_e = 10^{-4}$ for $\mathbf{a}_0 = [0.4, 0.3, 0.2]^T$ with $a_1^{(1)} = 0.4$ fixed (right) with SNR = 10.

where $d_{COSH}(S_0, S_1|\sigma_v^2) = d_{COSH}(S_{yy}^{(0)}(\omega), S_{yy}^{(1)}(\omega))$, $D(\omega)$ is the dispersion spectral density from Eq. (3.25) (the integrand of the Cosh distance) and

$$G(\omega) = \frac{S_{xx}^{(0)}(\omega) S_{xx}^{(1)}(\omega)}{S_{yy}^{(0)}(\omega) S_{yy}^{(1)}(\omega)} \tag{5.6}$$

with the power spectral densities of processes in noise given by $S_{yy}^{(0)}(\omega)$ and $S_{yy}^{(1)}(\omega)$. The Cosh distance at non-zero noise is $d_{COSH}(\theta|\sigma_v^2) \triangleq d_{COSH}(S_{yy}^{(0)}(\omega), S_{yy}^{(1)}(\omega, \theta))$. We attempt to normalise this function by $d_D(\theta) \triangleq d_D(S_{xx}^{(0)}(\omega), S_{xx}^{(1)}(\omega, \theta))$ to correct the effects of the imperfect mapping from P_e to d_D.

It is found as a purely empirical result that the inverse normalised Cosh distance (INCD) $d_D(\theta)/d_D(\theta|\sigma_v^2)$ has the same trends as the noise resistance characteristic $P_e(\theta|\sigma_v^2)$. This is seen by comparison of figure 5.6 with figure 5.4. The inversion is done because distance measures are in general inversely proportional to P_e. From comparing the corresponding curves of the two figures at low SNR values, the INCD could appear to be a monotone function of $P_e(\theta|\sigma_v^2)$. However, this is counterproved by the curves of the AR(3)-process at SNR = 0, since $d_D(\theta)/d_D(\theta|\sigma_v^2)$ has a different maximum from $P_e(\theta|\sigma_v^2)$.

The INCD has a simple analytic form in the spectral domain, which suggest that it could be a tool when we want to assess properties of processes in noise. It is also less computationally expensive than the P_e. Still, it is difficult to draw any consise conclusions

from Eq. (5.5) about the relation between process spectra and resistance to additive white noise. We are left with the remarks that noise resistance depends to some degree on the distance between the spectral maxima of the transmission processes. Moreover, the FSK analogy has shown us that resistance to additive white noise cannot be used a selection criterion.

5.3 Similarity in the Spectral Domain

The remaining criterion which has not been examined properly is (iii). Hence, similarity in the spectral domain is singled out as the key point in process selection. There are different ways of implementing this requirement:

- We can minimise the spectral difference between the processes by minimising one of the presented spectral distance measures, i.e. the Cosh distance d_D or the RMS log-spectral distance measure d_2.
- We may demand that the spectral difference should be evenly distributed over the total bandwidth, such that features like a distinct difference at a certain frequency or separation of the process maxima, do not easily reveal process identities for eavesdroppers. This can be done by minimising the peak log-spectral difference d_∞ or minimising the peak dispersion, defined as $\max D(\omega)$. Another idea is to minimise a flatness index [Kay 1979], calculated on basis of the dispersion spectral density $D(\omega)$ or the squared log-spectral difference $|V(\omega)|^2$.

Experience suggests that the first alternative is the best solution. We choose d_{COSH} as the preferred spectral distance measure with the following argument.

The logarithmic difference in the log-spectral difference measures (or L_P-norms) is introduced by convention to incorporate the knowledge that spectral differences should be more weighted at low power than at high power. The distance measure often appears in speech processing, with the motivation that perceived loudness of an acoustic signal is approximately logarithmic [Rabiner and Juang 1993]. This makes sense in speech recognition, but does not apply to our problem. Moreover, the choice of the parameter P in the L_P-norm is not governed by any rules. There is no optimality criterion between the different d_P measures, only an awareness that the large spectral differences are more heavily weighted with increasing P.

The different weighting of linear spectral differences at different power levels is also a property of the d_{COSH} measure, but here the weighting falls naturally out of the derivation. The starting point (i.e. the motivation of the Jeffreys divergence) is that we want to quantify the dispersion of the likelihood ratio under the different hypotheses. This seems like a reasonable requisite, since maximum likelihood detection and methods derived from the likelihood ratio is attractive choices for both authorised receivers [Dickinson 1981] and eavesdroppers [Basseville 1988].

We could also look at the Cosh distance as the symmetrised Itakura-Saito distance. The d_{IS} measure arised from a study of linear prediction of speech, where speech was modelled by Gaussian AR-processes [Itakura and Saito 1970, Gray and Markel 1976], exactly like our transmission processes. Besides being theoretically appealing [Rabiner and Juang 1993], we know from figure 3.6 that d_{COSH} is approximately equal to d_2 for small spectral difference, while large spectral difference are much more weighted by d_{COSH}. This makes sense, since these are the differences that could be fatal to the security of an ARPSK communications system.

From the discoveries in the discussion on robustness to additive white noise, we should also require that the search for processes X_1 is limited to an area of parameter space such that the separation of the spectral peak frequencies is below a predetermined value, denoted maximum peak separation $\Delta\omega_T$. The maximum peak separation criterion is

$$\Delta\omega_{peak} \triangleq \left|\arg\left\{\max_{\omega}\{S_{xx}^{(0)}\}\right\} - \arg\left\{\max_{\omega}\{S_{xx}^{(1)}\}\right\}\right| < \Delta\omega_T \tag{5.7}$$

where we define $\Delta\omega_{peak}$ as the peak separation. While the targetted $P_e(N)$ determines the transmission quality, the figure $\Delta\omega_T$ determines security of the communications system.

In the same manner, we may determine a threshold that limits the deviation of the transmission process PSDs. This threshold could be a maximum allowed value of the dispersion spectral density, denoted the maximum spectral dispersion D_T. We then have the maximum spectral dispersion criterion

$$\max\{D(\omega)\} < D_T \ . \tag{5.8}$$

Thus, D_T is another figure that determines the security of the system.

5.4 Selection Procedure

The signal selection procedure which has evolved throughout the discussion is summarised below and is also shown schematically in figure 5.7.

1. Choose the order p of the autoregressive processes and the number of samples N in each process realisation.

2. Specify a target detection error probability P_e at zero noise and for pulse length N. Further specify the maximum spectral dispersion D_T and the maximum peak separation $\Delta\omega_T$ allowed.

3. Choose a wideband process X_0 with parameter vector $\mathbf{a}_0$ that utilises the allocated bandwidth to maximum extent.

4. Use a numerical search algorithm to identify the processes that satisfies the constrained P_e at zero noise. From these, find the process X_1 with parameter vector $\mathbf{a}_1$ that minimises the average spectral dispersion d_D.

5. Check if the peak separation of the chosen processes exceeds $\Delta\omega_T$, according to the maximum peak separation criterion. If it does, return to step 4 and choose a X_1 that gives smaller peak separation. Implicitly, the search for X_1 should be restricted to a set such that $\Delta\omega_{peak} < \Delta\omega_T$.

6. Check if the maximum value of the dispersion spectral density exceeds D_T, according to the maximum spectral dispersion criterion. If it does, return to step 1 and increase N or alter the performance parameters in step 2.

One question that has not been answered is how large the difference between the transmission processes are allowed to be before we risk that the transmitted signal can be successfully eavesdropped. As stated in the introductory description of SPSK communications (section 1.4), the security of the technique is based on the existence of the Cramer-Rao lower bound [Larsen and Marx 1986, Scharf 1991], which establishes the lower bound on the variance of any estimator.

The model order p, process parameter vectors $\mathbf{a}_0$ and $\mathbf{a}_1$, and pulse length N are all known to the transmitter and authorised receiver, but must be estimated by an eavesdropper. Hence, it is theoretically possible to calculate the probability that an eavesdropper

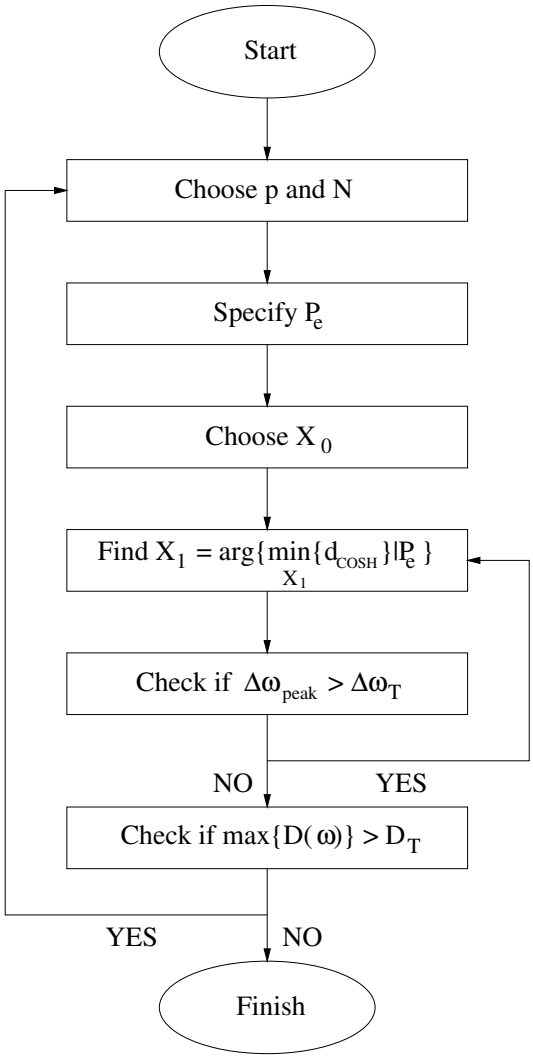


Figure 5.7: Signal selection procedure.

will estimate parameters that provide a detector which has a P_e that is below (or above) a critical value.

For instance, let us assume the worst case scenario that an eavesdropper has managed to obtain the correct value of p and N, and only needs to estimate the process parameter vectors $\mathbf{a}_0$ and $\mathbf{a}_1$. Next assume that the eavesdropper uses an unbiased estimator $\hat{\mathbf{a}}_i$ whose variance touches the Cramer-Rao lower bound (CRLB). If we can give a statistical description of $\hat{\mathbf{a}}_i$, then we can in principle evaluate the statistical properties of the P_e for an ideal Neyman-Pearson detector (assuming known additive white noise variance) implemented with the parameter vector estimates instead of the true parameter vectors. This detection error probability is denoted $P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_0|\mathbf{a}_0, \mathbf{a}_1)$.

For a moderate sample size N_s, [Box et al. 1994] state that the covariance matrix of the Yule-Walker AR-parameter estimate can be approximated by

$$\mathbf{\Sigma}_{\hat{a}} \simeq \frac{\sigma_\epsilon^2}{N_s}\mathbf{R}_x^{-1} \tag{5.9}$$

defining the parameter estimate covariance matrix as $\mathbf{\Sigma}_{\hat{a}}$. The approximation is equal to the asympotic CRLB [Porat and Friedlander 1987, Kay 1993]. Hence, we have

$$Var\{\hat{a}_k^{(i)}\} \geq \frac{\sigma_{\epsilon_i}^2}{N_s}\left[\mathbf{R}_x^{(i)}\right]_{kk}^{-1} \tag{5.10}$$

where $\hat{a}_k^{(i)}$ is the estimate of the kth parameter of process X_i. The sample size N_s does not have to be very large before $Var\{\hat{a}_k^{(i)}\}$ becomes relatively small.

The samples used in the parameter estimation must indeed be taken from a sequence that contains samples of both processes. Moreover, Salberg and Hanssen have shown [Salberg and Hanssen 1999a] that the PSD of ARPSK signal approaches the mean of the individual PSDs as $N \to \infty$. However, the discrete signal can be segmented by use of change detection algorithms [Basseville 1988, Basseville and Nikiforov 1993, Zhang et al. 1994]. Thus, sample sizes in the order of $N_s \simeq N$ or larger can be obtained, and estimates over several segments can be averaged. On the other hand, the estimation problem is more complicated in a practical situation with additive noise [Kay 1979, Wu and Chen 1997, Davila 1998].

To evaluate $E\{P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_0|\mathbf{a}_0, \mathbf{a}_1)\}$ and $Var\{P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_0|\mathbf{a}_0, \mathbf{a}_1)\}$, we have to solve integrals of dimension $2p$ over all estimated parameters. We have e.g.

$$E\{P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_0|\mathbf{a}_0, \mathbf{a}_1)\} = \int P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_0|\mathbf{a}_0, \mathbf{a}_1) f_{\hat{\mathbf{a}}_0,\hat{\mathbf{a}}_1}(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_1)\, d\hat{\mathbf{a}}_0\, d\hat{\mathbf{a}}_1 \tag{5.11}$$

where $f_{\hat{\mathbf{a}}_0,\hat{\mathbf{a}}_1}(\hat{\mathbf{a}}_0,\hat{\mathbf{a}}_1)$ is the joint PDF of the parameter vector estimates. The figures $E\{P_e(\hat{\mathbf{a}}_0,\hat{\mathbf{a}}_0|\mathbf{a}_0,\mathbf{a}_1)\}$ and $Var\{P_e(\hat{\mathbf{a}}_0,\hat{\mathbf{a}}_0|\mathbf{a}_0,\mathbf{a}_1)\}$ can be used to benchmark the security provided by the transmission processes $\mathbf{a}_0$ and $\mathbf{a}_1$. Unfortunately, the required computations are too demanding for practical use. What we can do however, is to assess the $P_e(\hat{\mathbf{a}}_0,\hat{\mathbf{a}}_0|\mathbf{a}_0,\mathbf{a}_1)$ directly. In terms of the general framework in section 4.8, this P_e is calculated from the standard procedure with

$$\mathbf{M} = [\mathbf{R}_{\hat{y}}^{(0)}]^{-1} - [\mathbf{R}_{\hat{y}}^{(1)}]^{-1}, \quad C = \ln\left(\frac{|\mathbf{R}_{\hat{y}}^{(0)}|}{|\mathbf{R}_{\hat{y}}^{(1)}|}\right) \quad \text{and} \quad \mathbf{R}_z^{(i)} = \mathbf{R}_y^{(i)}$$

where $\mathbf{R}_{\hat{y}}^{(i)}$ and $\mathbf{R}_y^{(i)}$ are the true correlation matrices calculated from the estimated parameter vector $\hat{\mathbf{a}}_i$ and the true parameter vector $\mathbf{a}_i$, respectively.

The principal issue addressed here needs further investigation. From a further analysis, we might also be able to find out what influence the choice of p has on the performance and security of the ARPSK system. This has not been discussed in this thesis, because no obvious connections have been discovered.

Chapter 6

Results

The process selection procedure presented above is used to choose three possible pairs of AR(3)-processes that can be used as transmission prosesses. The only part of the procedure that is neglected is point 6. This is done partly because we have not yet obtained a good rule on how to choose the threshold D_T. In addition, the available computing resources limits the parameter choice. For practical reasons, we want to run several of the simulations for a pulse length of $N = 64$. For much larger N, the limited memory capacity of the available computers and the precision in numerical computations cause problems. Hence, we must tolerate that the statistical distance between the processes is relatively large, in order that the resulting P_e should be in a region of interest.

The chosen process pairs are the combinations of a process X_0 with three different choices of process X_1, denoted by $X_1^{(i)}$, $X_1^{(ii)}$ and $X_1^{(iii)}$. These are chosen such that the pairs should produce a P_e of 10^{-3}, 10^{-4} and 10^{-5}, respectively, at zero noise for the pulse length $N = 64$. In digital communications terms, these detection error probabilities are relatively high. However, they are sufficiently low that we can assess the characteristic

Process	AR-parameters	σ_ϵ^2 $P_e(X_0, X)$	$d_{COSH}(X_0, X)$	
X_0	$[0.4, 0.3, 0.2]^T$	1.0	0	0
$X_1^{(i)}$	$[0.7625, 0.2190, 0.3000]^T$	0.4401	10^{-3}	1.464
$X_1^{(ii)}$	$[0.7347, 0.1912, 0.3500]^T$	0.3399	10^{-4}	2.244
$X_1^{(iii)}$	$[0.7351, 0.1673, 0.3500]^T$	0.2744	10^{-5}	3.126

Table 6.1: Parameters of the processes which are used in numerical simulations.

performance of an ARPSK system. The AR-parameters of the processes are given in table 6.1.

In the search for optimal AR-parameters, the third parameter of the X_1 was first fixed, and then varied in small steps to approach a minimum in the AR(3)-parameter plane. The difficulty of the three-dimensional search explains why the precision of the first two parameters is much higher than for the third. The table also lists the driving noise variances determined from the average power equalisation constraint, in addition to the exact P_e and d_{COSH} for the chosen pairs.

6.1 Detection Error Probability as Function of N

Figure 6.1 shows P_e as a function of the pulse length N, assuming zero additive white noise. For each process pair, the P_e is obtained from both the theoretical expression and numerical simulations. This is done for the Neyman-Pearson (NP) detector and the approximate log-likelihood ratio (ALR) detector. Empirical results are obtained from Monte Carlo simulations with 100 000 runs. This number is evidently to low to yield good results at low P_e, since the variance of the simulation result becomes very large. However, the number of runs is limited by practical constraints. Nevertheless, the simulation results are good enough to demonstrate the probability that the theoretical results are correct.

We shall refer to pair (i) as the processes X_0 and $X_1^{(i)}$, and so on for pair (ii) and pair (iii). For each pair, we find that the P_e of the ideal NP detector and the ALR detector are very close. The $P_e(N)$ for the ideal NP detector and the ALR detector are shown as a solid line and a dotted line, respectively. At the end point of the curves, it is indicated which process pair they belong to. Simulation results are plotted on top of the theoretical curves using various symbols, as explained in the caption of figure 6.1.

From all curves, we see that the P_e drops rapidly with increasing N for all process pairs. The P_e approaches zero when N goes to infinity. There is always a finite P_e for finite N. This is in contrast to classical communication systems with deterministic signals in a noiseless environment. We further see that the P_e of the ideal NP detector is lower than the P_e of the ALR detector, but that the difference decreases with increasing N. This is natural, since the NP detector minimises the P_e, the ALR detector is an approximation to the NP detector, and the approximation becomes better as N increases. The simulation results correspond very well to the theoretical results for $P_e > 10^{-4}$, and affirms the theory.

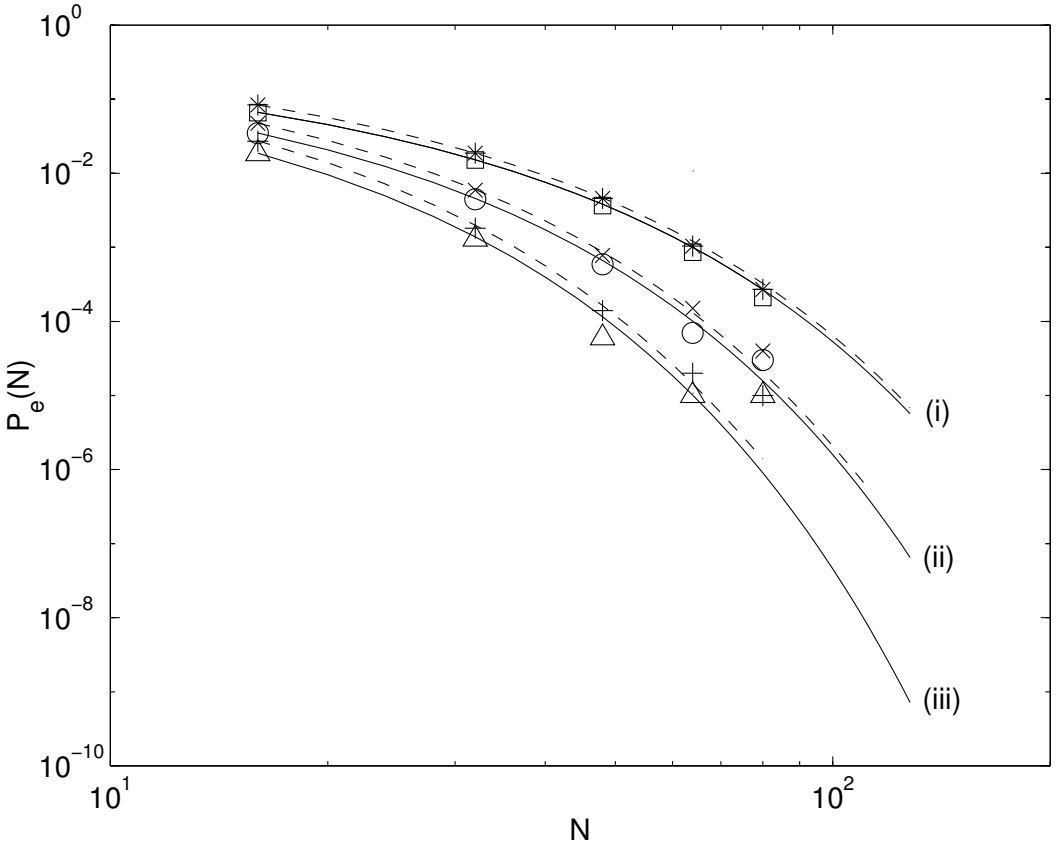


Figure 6.1: Detection error probability $P_e(N)$ as a function of the pulse length N at zero noise. Comparison of the theoretical results obtained with the ideal NP detector (solid lines) and the ALR detector (dashed lines) for each process pair. Empirical results for the NP detector are marked with '$\square$' (process pair (i)), '$\circ$' (ii) and '$\triangle$' (iii), while results for the ALR detector are marked with '$*$' (process pair (i)), '$\times$' (ii) and '$+$' (iii).

6.2 Detection Error Probability as function of SNR

Figure 6.2 shows the P_e as a function of the SNR, when the pulse length is fixed to $N = 64$. The ideal NP detector and the ALR detector are assessed through numerical evaluation of the theoretical results and by virtue of simulation results. The number of runs in Monte Carlo simulations is again 100 000. The line styles and symbols used to designate different results are the same as in the previous figure. These are also specified in the figure caption.

Once more, we see from the figure that the P_e has a lower bound for finite values of N. When the SNR increases (and the additive noise variance becomes negligible), the P_e

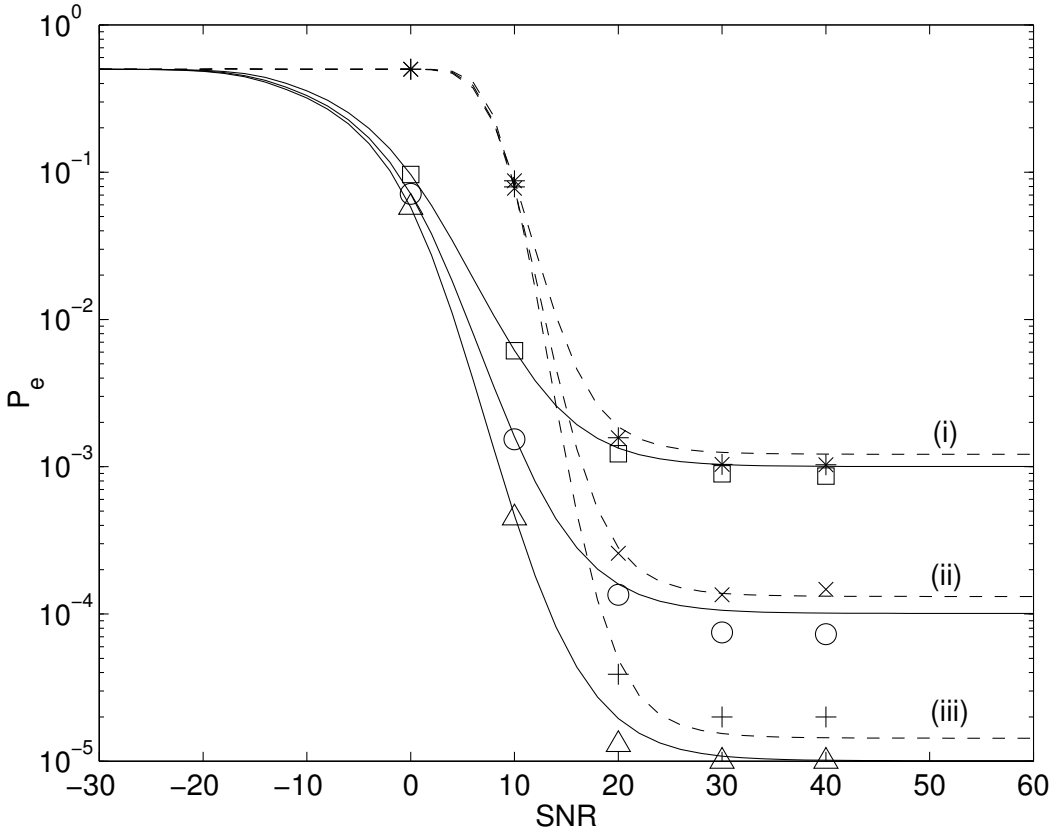


Figure 6.2: Detection error probability P_e(SNR) as a function of the signal-to-noise ratio for a fixed pulse length of $N = 64$. Comparison of the theoretical results obtained with the ideal NP detector (solid lines) and the ALR detector (dashed lines) for each process pair. Empirical results for the NP detector are marked with '$\square$' (process pair (i)), '$\circ$' (ii) and '$\triangle$' (iii), while results for the ALR detector are marked with '$*$' (process pair (i)), '$\times$' (ii) and '$+$' (iii).

tends to a threshold value. We note that the threshold values are equal to the $P_e(\sigma_v^2 = 0)$ specified in the selection procedure for the respective processes. This result is specific for SPSK communications, but has no implications for how applicable the technique is. A noiseless channel is only found in theory. In a practical situation, the important thing to ensure that the communications system provides acceptable P_e values up to a certain SNR. It is in this light we must assess detector candidates.

The NP detector is model independent, while the ALR detector assumes that the received signal is an AR-process. The effect this has on the detector performance difference

is made clear by figure 6.2. At high SNR values, the noise level is negligible and the ALR is a good approximation to the true log-likelihood ratio. However, as the SNR drops below 20-30 dB (which is still very little noise), the P_e of the ALR detector increases dramatically. At 5 dB, the ALR detector is i.e. "guessing" what source symbol is being received. The performance of the NP detector also deteriorates, but not at the same rate. At 0 dB, the NP detector still maintains a $P_e \approx 10^{-1}$. The observed difference is an indication that the distance between the assumed AR-model and the actual ARMAX-model grows too large, and rapidly destroys the capability of the ALR detector, while the NP detector is more robust to noise. Still, we should note that the change in performance occurs at relatively low noise levels for the ideal NP detector as well, and this represents the theoretical bound on detection performance. Simulation results show good resemblance with theoretical results.

6.3 Neyman-Pearson Detector with Additive Noise Variance Estimator

The ideal NP detector cannot be realised due to the unknown variance of the white additive noise, and it must therefore be replaced by an NP detector implemented with an additive noise variance estimator. Hence, the question naturally arising is how close up to the performance of the ideal detector this sub-optimal detector will come. The answer is found in figure 6.3, which compares results for different detectors with process pair (i).

In the upper panel, the figure displays the empirical P_e of the NP detector implemented with the non-negative estimator $\widehat{\sigma}^2_{v+}$. The estimate $\widehat{\sigma}^2_{v+}$ is calculated from the samples representing $M = 1, 4$ and 20 source bits. The pulse length is still $N = 64$, so the estimator uses a total of $NM = 64, 256$ and 1280 signal samples in the respective cases. Only simulation results are shown. Computational complexity and accuracy did not allow the theoretical expression to be successfully evaluated. The results are obtained from Monte Carlo simulations with 10 000 runs.

From figure 6.3 we see the following. The P_e of the ideal NP detector (solid line) and the ALR detector (dotted line) are used as references in the figure. The simulation results for the suboptimal NP detectors are shown as dashed lines with different symbols (refer to the figure legend). We see that the suboptimal detectors track the performance of the ideal NP detector up to a certain SNR level, which depends on M. As the SNR increases,

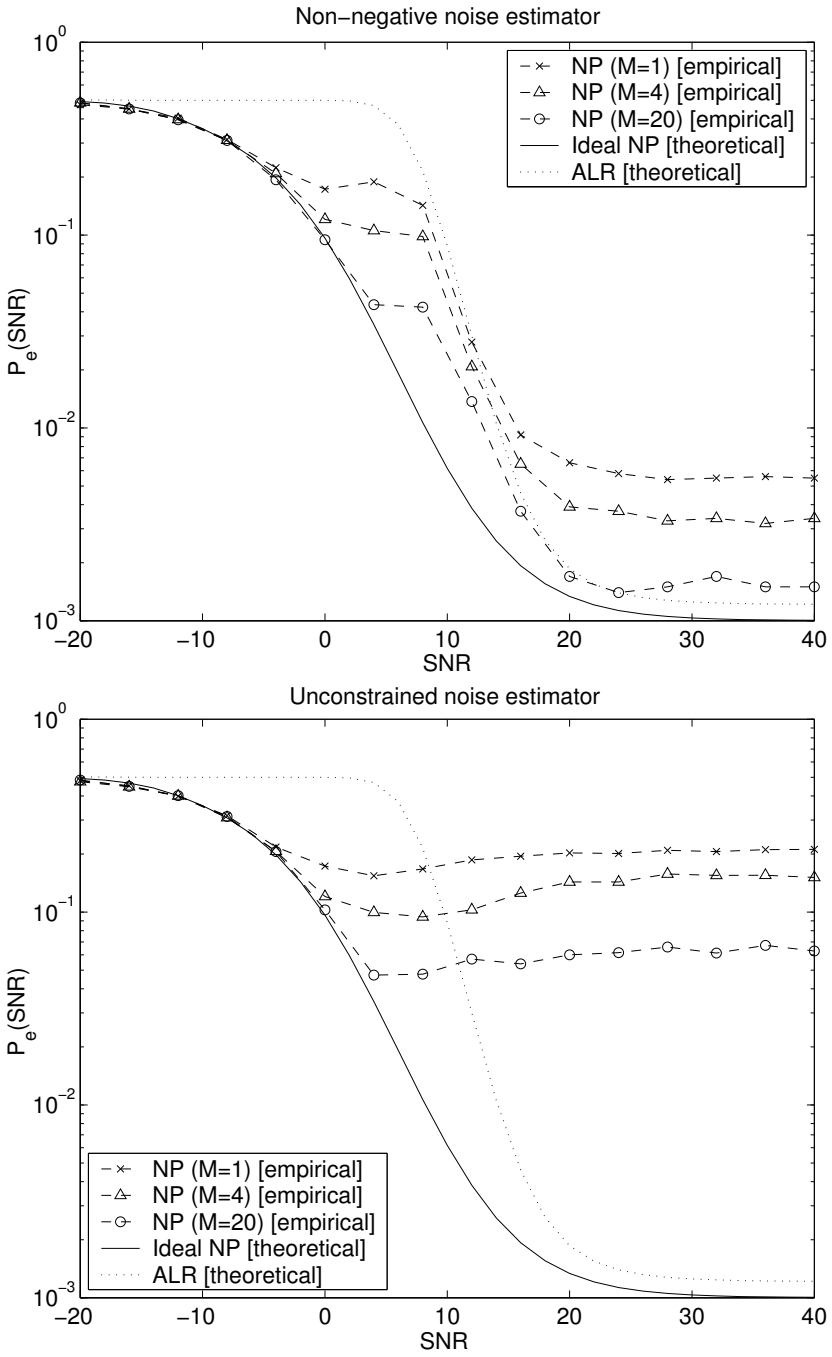


Figure 6.3: Detection error probability $P_e(\text{SNR})$ as a function of the signal-to-noise ratio for process pair (i). Comparison of theoretical performance for the ideal NP detector and the ALR detector, and the empirical performance of the NP detector implemented with a non-negative estimator (upper panel) and with an unconstrained estimator (lower panel), calculated with samples representing M source symbols. The pulse length is $N = 64$.

we observe that the individual realisations of suboptimal detector deviate in performance. The suboptimal NP detector approaches the performance of the ideal NP detector with increasing M. This is expected, since the variance of $\widehat{\sigma}_{v_+}^2$ is inversely proportional to M.

The behaviour of the suboptimal NP detectors between SNR values of 0 dB and 10 dB is an intriguing observation. The fluctuations where P_e(SNR) is not monotonically decreasing are considered as a result of the variance in the simulation results, and are ignored. What we cannot ignore is the intermediate plateau on the transient between high and low P_e. This is a feature that stands out, by comparison with the ideal NP detector. The same trend is observed in attempted evaluations of the theoretical P_e, but these results are not sufficiently accurate to be repeated here.

An explanation is offered, if we look at the results in the lower panel of figure 6.4. Here, the same results are shown for an NP detector implemented with the unconstrained estimator $\widehat{\sigma}_v^2$ (that allows negative estimates of σ_v^2). The different implementations yield identical performance up till the cut-off which is experienced for the constrained estimator implementation at around 10 dB. From here, the P_e of the constrained estimator implementation drops at the same rate as for the ALR detector, while no improvement is found in the P_e of the unconstrained estimator implementation. It is observed in simulations that the cut-off represents the SNR value where the unconstrained estimator starts to produce a significant portion of negative variance estimates.

As the SNR exceeds 25 dB, the P_e of the constrained estimator implementation approaches a lower bound which depends on M. We see that for $M = 20$, the lower bound is still slightly higher than the corresponding bound on the ALR detector. Hence, we expect that the suboptimal NP detector is better than the ALR detector over the whole range of SNR values for some choice of $M > 20$ with the present value of N. For the unconstrained estimator implementation, we see that the P_e has a minimum somewhere between 0 dB and 10 dB, and approaches a steady state value as the SNR increases from there. From this, we conclude that negative values of $\widehat{\sigma}_v^2$ must be associated with a high degree of erroneous decisions in the detector.

In figure 6.4, we once again compare the performance of the ideal NP detector and the ALR detector with the NP detector implemented with the constrained estimator $\widehat{\sigma}_{v_+}^2$. This time, the P_e is displayed as a function of the pulse length N for fixed SNR values. The upper panel shows the results obtained for process pair (i) with fixed SNR of 10 dB (upper panel), and the lower panel shows the same results for an SNR of 20 dB (lower panel).

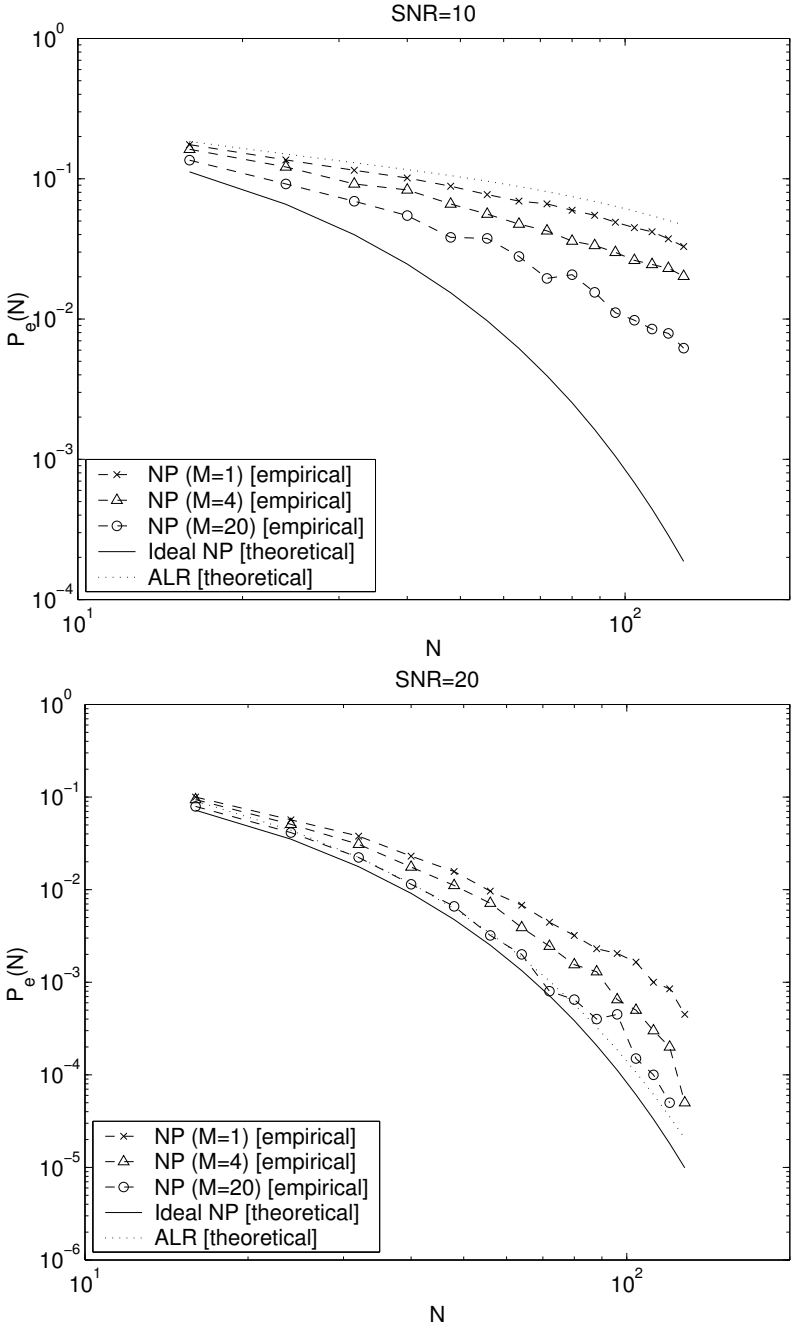


Figure 6.4: Detection error probability $P_e(N)$ as a function of the pulse length N, calculated for process pair (i) at fixed SNR values of 10 dB (upper panel) and 20 dB (lower panel). Comparison of theoretical results for the ideal NP detector and the ALR detector, and empirical results for the NP detector implemented with non-negative estimator for different choices of M, representing the number of process realisations used in the estimator.

There is nothing surprising about the shape of the $P_e(N)$ curves, but we see that the rate of change of P_e versus N depends largely on the detector and the SNR. At 10 dB, the ALR shows the worst performance, and the P_e falls very slowly with increasing N. As in the case of zero additive noise (figure 6.1), the P_e will approach zero as N goes to infinity. The P_e of the NP detectors implemented with the constrained estimator is better then for the ALR, but also decreases relatively slow.

At 20 dB, the ranking of the ALR detector versus the suboptimal NP detectors is no longer uniform over the range of pulse lengths N. The P_e of the suboptimal is only assessed through Monte Carlo simulations with 10 000 runs. The fluctuations of the empirical curves at high values of N are relatively high. Despite the uncertainty implied by the variance in the simulations, the result indicates that the true P_e of the suboptimal NP detector for $M = 20$ probably exceeds the P_e of the ALR detector at some N.

6.4 Detection Error Probability as Function of the Synchronisation Error

Theoretical results and simulation results for detection with unsynchronised data are shown in figure 6.5. The P_e is plotted as a function of the synchronisation delay d_s divided by the pulse length N. The measure $0 \leq (d_s/N) \leq 1,\ d_s = 0, \ldots, N$ gives the synchronisation error as a fraction of the symbol period T. The upper panel shows theoretical results and simulation results for the ideal NP detector (marked with circles and cross, respectively), while the theoretical results for the ALR detector as a reference (solid line). The lower panel shows theoretical results and simulation results for the ALR detector (circles and crosses, respectively), with the theoretical results of the ideal NP detector as a reference (solid line). Process pair (i) is used with a pulse length of $N = 64$. Empirical results are obtained from Monte Carlo simulations with 50 000 runs.

From the transmitter and authorised receiver's point of view, it is beneficial if a synchronisation error causes a significant deterioration of the P_e. Thus, an eavesdropper is less likely to succeed if perfect synchronisation is not achieved. This is under the condition that the authorised receiver posesses a robust method that guarantees perfect synchronisation.

For the NP detector, the $P_e(d_s)$ is symmetric around N. The maximum value is found at $d_s = \lfloor N/2 \rfloor$ (and $d_s = \lceil N/2 \rceil$ if N is odd). Here, $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the nearest integer

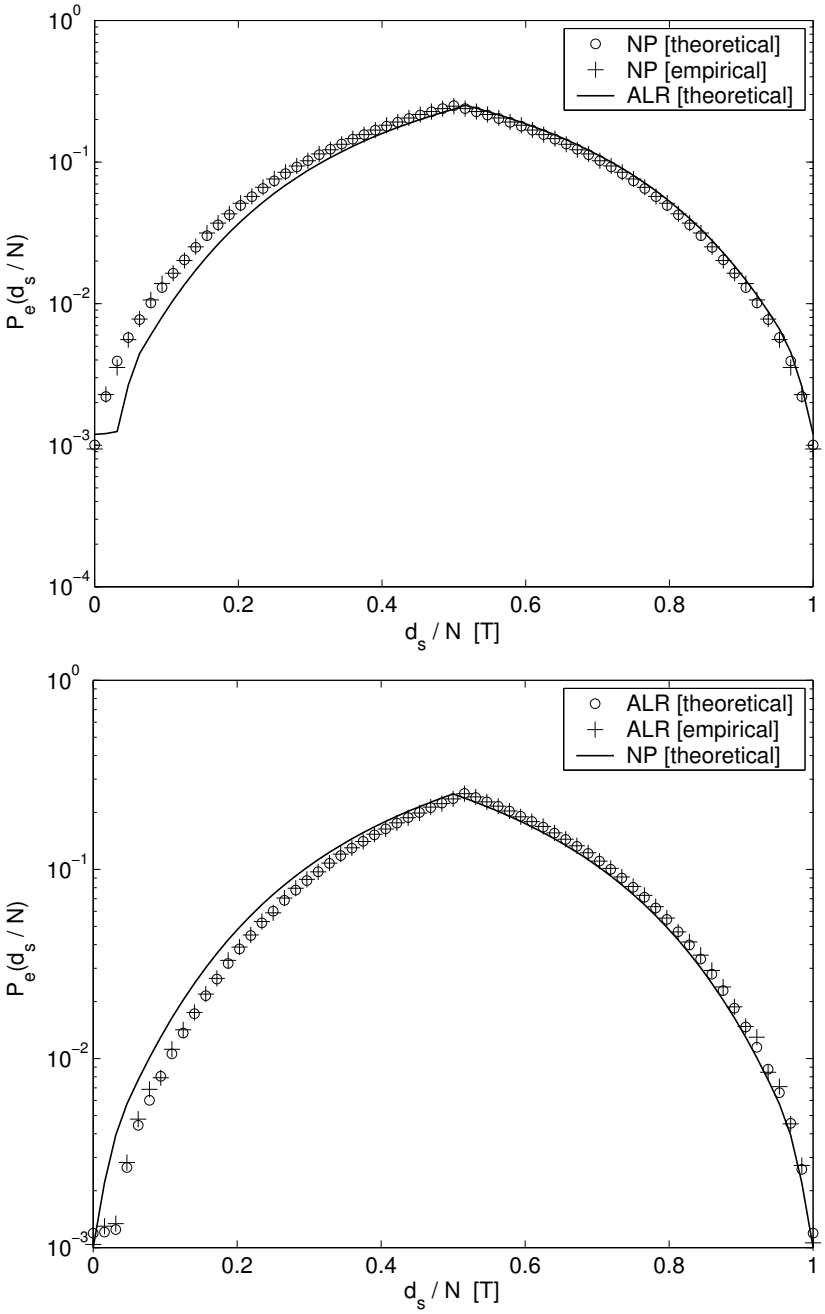


Figure 6.5: Detection error probability $P_e(d_s/N)$ as a function of the normalised synchronisation delay $0 \leq d_s/N \leq 1$ at zero noise. Comparison of theoretical results and simulation results for the ideal NP detector with theoretical results of the ALR detector (upper panel) and vice versa (lower panel). All results obtained with process pair (i) and pulse length $N = 64$.

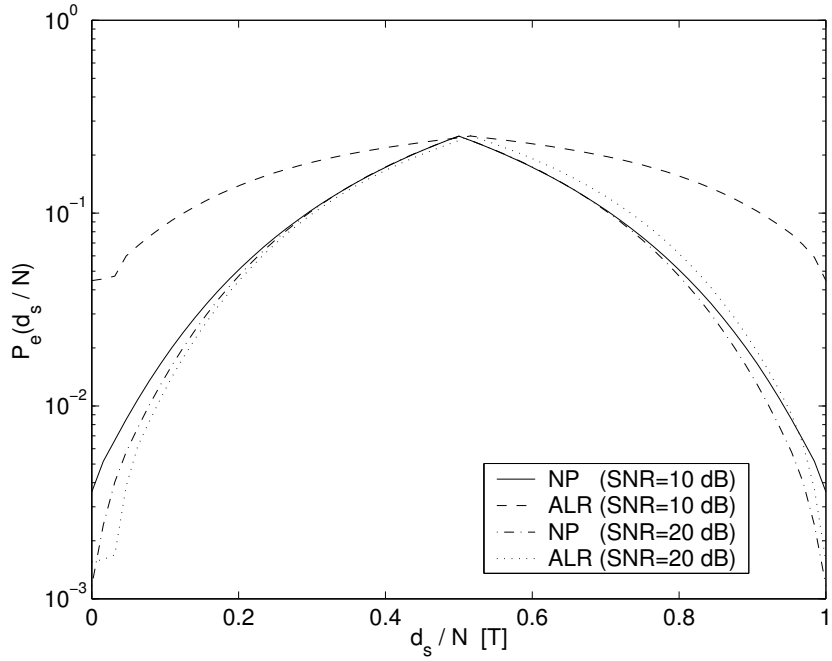


Figure 6.6: Detection error probability $P_e(d_s/N)$ as a function of the normalised synchronisation delay $0 \leq d_s/N \leq 1$ at SNR values of 10 dB and 20 dB. Comparison of theoretical results for the ideal NP detector and the ALR detector. All results obtained with process pair (i) and pulse length $N = 64$.

less than or equal to x, and the nearest integer greater than or equal to x, respectively. From the theoretical expression for the $P_e(d_s)$ in Eq. (4.69), we find that the maximum is

$$P_e(\lfloor N/2 \rfloor) = \frac{1}{2}\left[P_e(0) + \frac{1}{2}\right] \ . \tag{6.1}$$

I.e., it is the mean value between the $P_e(d_s)$ when only one process is transmitted and the $P_e(d_s)$ when alternating process realisations are transmitted.

As seen from the figure, the symmetry of $P_e(d_s)$ does not hold for the ALR detector. On the contrary, we observe a distinct feature at small values of d_s which is not present at the corresponding values $N - d_s$. For the synchronisation delays up to $d_s = 2$, the increase in P_e is very small, but after this it changes at a rate that is similar to what we observe for the NP detector. This behaviour can be explained as follows.

For an ALR detector with perfect synchronisation, we will find that the p first samples of the received process realisation contribute less to detectability than the other samples. For these data points, we do not have access to all of the p precursors which they depend on,

according to the AR-model. The model-dependendent ALR detector implicitly tests how the samples of the process realisation fits with the candidate AR-models. In this respect, the first p samples provide less information. Within the p first samples, $x(1), \ldots, x(p)$, the significance to detectability obviously increases with increasing sample index. This observation applies to the case of an unsynchronised ALR detector as well. If the first p samples are replaced by samples from another process, the performance is not much degraded because the lost samples did not contribute much to the detectability anyway.

Both panels show that the simulation results are very much in agreement with the results obtained through evaluation of the theoretical expressions of $P_e(d_s)$. The $P_e(d_s)$ of the ideal NP detector exceeds the $P_e(d_s)$ for the ALR detector for $d_s \leq N/2$, which clearly shows the assymmetry for the ALR detector.

In the presence of noise, we expect that performance of the ideal NP detector and the ALR detector will deviate according to the results of figure 6.2. This is confirmed by figure 6.6, which shows the theoretical $P_e(d_s)$ of the ideal NP detector and the ALR detector for SNR values of 10 dB and 20 dB. With reference to figure 6.2, we know that the deviation between the P_e(SNR) for the two detectors is small at SNR=20 dB. This is also the case for the $P_e(d_s)$. The deviation between the P_e(SNR) is larger at SNR=10 dB. From figure 6.6, we see that the $P_e(d_s)$ of the ALR detector is much higher than the $P_e(d_s)$ of the ideal NP detector at this noise level. However, the maxima remain essentially constant, since $P_e(d_s{=}0) \ll 1/2$.

6.5 Detection with Estimated AR-parameters

Figure 6.7 demonstrates what happens to the P_e when the detector uses a parameter vector estimate instead of the true parameter vector. Assume that an eavesdropper knows the correct values of p and N and employs a detector which incorporates estimates of the AR-parameters for the two transmission processes. With AR(3) transmission processes, there are 6 parameters to estimate. The $P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_1 | \mathbf{a}_0, \mathbf{a}_1)$ is the detection error probability experienced by the ideal NP detector implemented with parameter vector estimates $\hat{\mathbf{a}}_0$ and $\hat{\mathbf{a}}_1$, given that $\mathbf{a}_0$ and $\mathbf{a}_1$ are the true parameter vectors. It is not possible to visualise how the $P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_1 | \mathbf{a}_0, \mathbf{a}_1)$ varies with all the free parameters in $\hat{\mathbf{a}}_0$ and $\hat{\mathbf{a}}_1$, but we have attempted to show how $P_e(\hat{\mathbf{a}}_0, \hat{\mathbf{a}}_1 | \mathbf{a}_0, \mathbf{a}_1)$ responds when one or two of the estimated parameters deviate from their true values.

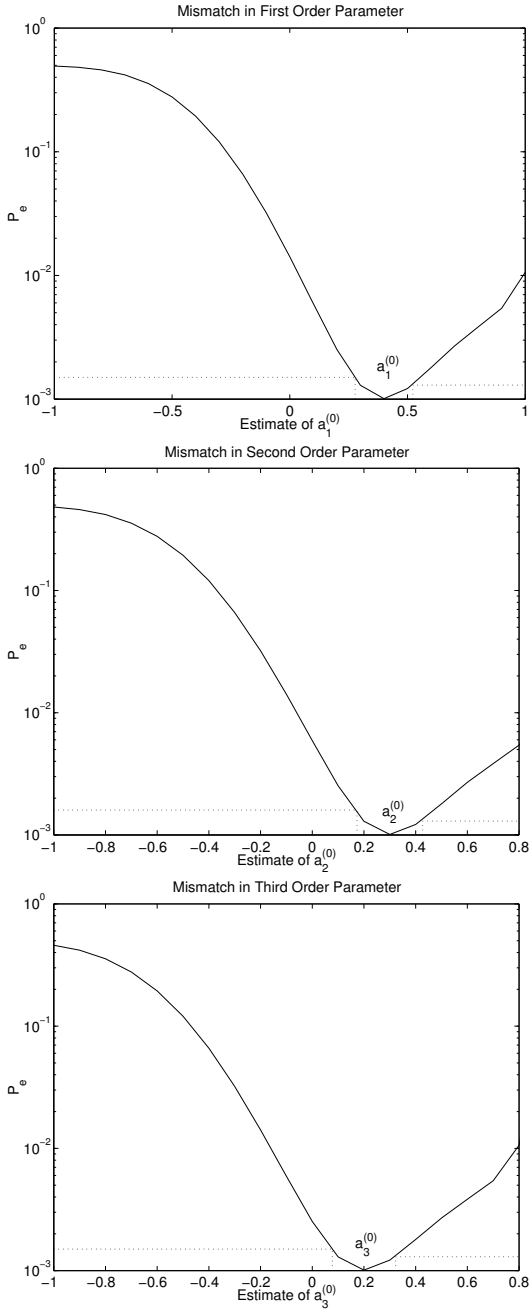


Figure 6.7: Detection error probability $P_e(\hat{a}_k^{(0)}|\mathbf{a}_0, \mathbf{a}_1), k = 1, 2, 3$ of an ideal NP detector as a function of estimates $\hat{a}_1^{(0)}$ (upper panel), $\hat{a}_2^{(0)}$ (middle panel) and $\hat{a}_3^{(0)}$ (lower panel) using process pair (i) and pulse length $N = 64$. The standard deviations from the true parameter value are connected with dotted lines to their corresponding P_e, assuming an optimum estimator using $N_s = 64$ samples.

The figure shows results for process pair (i) and pulse length $N = 64$. At first we assume that perfect estimates are obtained for the AR-parameters of both processes, except for one parameter $a_k^{(0)}$, $k = 1, 2, 3$ of process X_0. Figure 6.7 shows the theoretical $P_e(\hat{a}_k^{(0)}|\mathbf{a}_0, \mathbf{a}_1)$ of an ideal NP detector, as a function of that one parameter which is allowed to vary without constrain. Referring to the figure, the free variable is $a_1^{(0)}$ (upper panel), $a_2^{(0)}$ (middle panel) and $a_1^{(0)}$ (lower panel). The behaviour of P_e is almost identical for all cases. The minimum is found at the true parameter value in each case, and the P_e increases monotonically as we move away from this minimum. The minimum standard deviation of $\hat{a}_k^{(0)}$ can be calculated from Eq. (5.10). In the plots, the values $\hat{a}_k^{(0)} \pm \sigma_{\hat{a}}$ are marked and connected (dotted lines) with the corresponding P_e values. Here, $\sigma_{\hat{a}}$ denotes the standard deviation of an optimum estimator $\hat{a}_k^{(0)}$ which uses $N = 64$ samples.

In the next example, the first parameter of both process X_0 and X_1 is allowed to vary. Figure 6.8 shows a window of the surface defined by $P_e(\hat{a}_1^{(0)}, \hat{a}_1^{(1)}|\mathbf{a}_0, \mathbf{a}_1)$, shown at three different angles. The minimum which occurs at the true values $[a_1^{(0)}, a_1^{(1)}]$ is marked in the plot. The upper panel gives the best perspective on the whole surface. The view of the figure in the middle panel is almost in the direction of the $\hat{a}_1^{(1)}$-axis, and the figure thus visualises how the P_e varies with $\hat{a}_1^{(0)}$. For the same reason, the figure in the lower panel is viewed almost in the direction of the $\hat{a}_1^{(0)}$-axis. The curve in the upper panel of figure 6.7 is equivalent to the curve at $\hat{a}_1^{(1)} = a_1^{(1)} = 0.76$ in the P_e surface of figure 6.8. It can be observed that the P_e increases monotonically along the $\hat{a}_1^{(1)}$-axis as a function of the distance from $a_1^{(1)}$, as expected. This demonstrates how additional uncertainty is included in the estimation problem, with respect to the first example and figure 6.7. Hence, we can imagine how more free parameters will increase the P_e of a detector implemented with estimated parameter values.

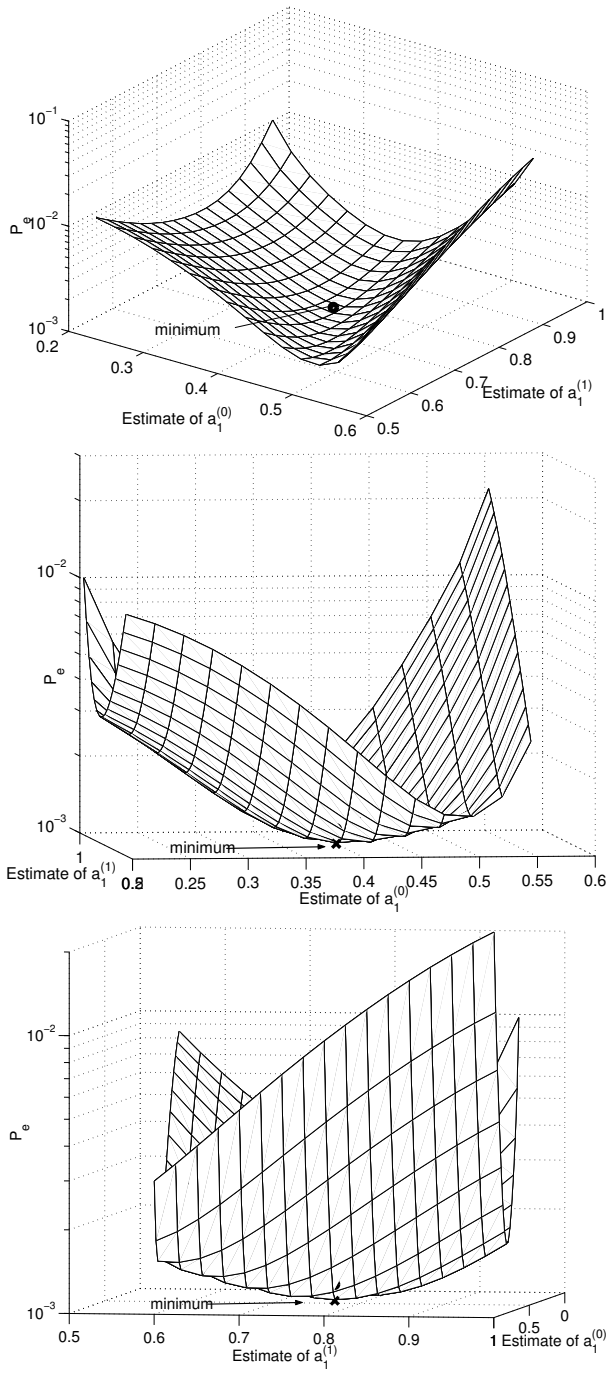


Figure 6.8: Detection error probability $P_e(\hat{a}_k^{(0)}, \hat{a}_k^{(1)}|\mathbf{a}_0, \mathbf{a}_1)$ of an ideal NP detector as a function of the estimates $\hat{a}_k^{(0)}$ and $\hat{a}_k^{(1)}$, shown at different angles. The result is obtained using process pair (i) and pulse length $N = 64$.

Chapter 7

Conclusion and Further Work

7.1 Conclusion

We have in this thesis studied detectors that can be applied in secure digital communications using a modulation technique named autoregressive process shift keying (ARPSK). The theoretical bound on detector performance for such a communications system is established in terms of the detection error probability of the optimum Neyman-Pearson (NP) detector.

The detection error probability P_e of the NP detector decreases rapidly with increasing pulse length N of the stochastic transmission process realisations. As a function of the signal-to-noise ratio (SNR), we find that the P_e approaches a threshold value as $N \rightarrow \infty$. Hence, unlike classical modulation techniques using deterministic signals, ARPSK produces finite P_e for finite values of N. Furthermore, we find that significant degradation of the P_e(SNR) starts at relatively high SNR values. The characteristics of the P_e(SNR) improves with increasing N, but we conclude that $N \gg 100$ for the communications to yield acceptable performance at SNR values that should be tolerated.

The optimum NP detector can not be implemented in a noisy communications channel, since the probabilistic model of the noise is not completely known. If we assume that the channel noise is additive and white, the ideal NP detector can be replaced by a noise compensated version that incorporates an estimator of the additive white noise variance.

In the thesis, the detection error probability is derived for the suboptimal NP detector implemented with a finite memory estimator, assuming stationary noise. From the results we see that it is imperative that a sufficient number of samples can be used in the estimate

of the noise variance. Otherwise, the suboptimal NP yields considerably worse P_e than the optimal detector at high SNR values. If the noise is Gaussian, the optimum estimate of the additive white noise variance is provided by a Kalman prediction filter, both for stationary and non-stationary noise. However, the performance of a real implementation will depend on how fast the channel noise varies.

An approximated NP detector is obtained from a well-known approximation which applies to the log-likelihood ratio for autoregressive Gaussian processes. The P_e of this detector is derived in this thesis, and compared with the NP detector. Since the approximated log-likelihood ratio (ALR) is derived assuming an autoregressive process model, it does not take allowance for additive noise. Hence the performance of the ALR detector is degraded at relatively low noise levels, and the degradation is more severe than for the NP detector. For negligible noise, the ALR detector approaches the NP detector in performance.

To evaluate how vulnerable the ARPSK communications system is to eavesdropping, we have derived the P_e of the NP detector and the ALR detector assuming that the receiver is not perfectly synchronised with the transmitter. The P_e given that the respective detectors are implemented with estimated values of the AR-parameters is also derived. The results show that the detectors are not extremely sensitive to small synchronisation errors, but the sensitivity depends on and increases with the noise level. We further see that the increase in P_e implied by estimated AR-parameters can be made quite small if a sufficient amount of process samples is available.

A selection procedure for transmission processes is formulated, based on a proposed set of criteria that the autoregressive transmission processes should satify. This procedure takes into consideration that the statistical distance between the processes and the difference between their spectra should be as small as possible in order to reduce the posibility of eavesdropping, while at the same time maintaining an acceptable detection error probability. From a review of existing statistical distance measures and spectral distance measures, it is found that the Cosh distance is the appropriate choice for ARPSK communications.

From the previous conclusions, it is questionable whether the ARPSK modulation technique provides the required protection against eavesdropping. However, a definite answer can not be given before we have quantified the allowed statistical distance between the transmission prosesses, given a specified risk that the transmitted message can be successfully eavesdropped.

7.2 Suggestions to Further Work

Even if the described ARPSK technique might not offer the required security, the concept of stochastic process shift keying (SPSK) should be further investigated. One alternative is to employ other transmission processes. Another possibility is to stay with AR-processes, but encode information by means of higher-order statistics. Thus, the pair of transmission processes will have the exact same second-order statistics (e.q. autocorrelation function and power spectral density) and must thus be distinguished by their different higher-order statistics. Estimators of higher-order statistics generally have higher variances and need more samples to provide good estimates [Brillinger 1975, Mendel 1991] than estimators of second-order statistics. Hence, the difference in performance between a detector that knows all system parameters and one that must estimate them might be larger.

Regardless of the choice of transmission processes, the problem of how to obtain synchronisation between transmitter and receiver will be an issue for further work. Another issue one might want to discuss is how multiple access can be built into the SPSK communications system.

Since SPSK is associated with finite detection error probabilities even at zero noise, there is an absolute demand for implementation of error correcting codes. Evaluation of the improvements that such coding will yield is a topic for future research. As a counterpart to frequency-hopping in classic narrowband communications [Gibson 1993, Proakis 1995], an idea would be to implement parameter-hopping in SPSK. This means that system parameters (like the AR-parameters) should be varied cyclically according to a pattern that is known only to the transmitter and authorised receiver.

Bibliography

[Aislam and Edwards 1996] AISLAM, T. and J.A. EDWARDS (1996). "Secure communications using chaotic digital encoding," Electronics Letters, vol. 32, no. 3, pp. 190-191.

[Akansu et al. 1998] AKANSU, A.N., P. DUHAMEL, X. LIN and M. DE COURVILLE (1998). "Orthogonal transmultiplexers in communications: A review," IEEE Transactions on Signal Processing, vol. SP-46, no. 4, pp. 979-995.

[Ali and Silvey 1966] ALI, S.M. and D. SILVEY (1966). "A general class of coefficients of divergence of one distribution from another," Journal of the Royal Statistical Society B, vol. 28, pp. 131-142.

[Basseville 1988] BASSEVILLE, M. (1988). "Detecting changes in signals and systems - A survey," Automatica, vol. 24, pp. 309-326.

[Basseville 1989] BASSEVILLE, M. (1989). "Distance Measures for Signal Processing and Pattern Recognition," Signal Processing (Elsevier), vol. 18, no. 4, pp. 349-369.

[Basseville and Nikiforov 1993] BASSEVILLE, M. and I.V. NIKIFOROV (1993). Detection of Abrupt Changes: Theory and Application, Prentice-Hall, Englewood Cliffs.

[Bellman 1970] BELLMAN, R. (1970). Introduxtion to Matrix Analysis, 2nd ed., McGraw-Hill, New York.

[Bhattacharyya 1943] BHATTACHARYYA, A. (1943). "On a measure of divergence between two statistical populations defined by their probability distributions," Bulletin of Calcutta Mathematical Society, vol. 35, pp. 99-109.

[Box et al. 1994] BOX, G.E.P., G.M. JENKINS and G.C. REINSEL (1994). Time Series Analysis: Forecasting and Control, 3rd ed., Prentice-Hall, Englewood Cliffs.

[Brillinger 1975] BRILLINGER, D.R. (1975). Time Series, Dana Analysis and Theory, Holt, Rinehart and Winston, New York.

[Brownhead et al. 1995] BROWNHEAD, D.S., J.P. HUKE and R. JONES (1995). "Signals in chaos: A method for the cancellation of deterministic noise from discrete signals," Physica D, no. 80, pp. 413-432.

[Chua et al. 1993] CHUA, L.O., M. ITOH, L. KOCAREV and K. ECKERT (1993). "Chaos synchronization in Chua's curcuit," Journal on Curcuits, Systems and Computers, vol. 3, pp. 705-708.

[Cuomo et al. 1993] CUOMO, K.M., A.V. OPPENHEIM and S.H. STROGATZ (1993). "Synchronisation of Lorentz-based curcuits with applications to communications," IEEE Transactions on Curcuits and Systems, vol. CS-40, pp. 626-633.

[Csiszar 1975] CSISZAR, I. (1975). "I-divergence geometry of probability distributions and minimization problems," Ann. Prob., vol. 3, pp. 146-158.

[Davila 1998] DAVILA, C.E. (1998). "A subspace approach to estimation of autoregressive parameters from noisy measurements," IEEE Transactions on Signal Processing, vol. SP-46, pp. 531-534.

[Dickinson 1981] DICKINSON, B.W. (1981). "Properties and Applications of Gaussian Autoregressive Processes in Detection Theory," IEEE Transactions on Information Theory, vol. IT-27, pp. 343-347.

[Ditto and Pecora 1993] DITTO, W. and L.M. PECORA (1993). "Mastering chaos," Scientific American, pp. 78-82.

[Dixon 1994] DIXON, R.C. (1994). Spread Spectrum Systems: With Commercial Applications, John Wiley, New York.

[Drazin 1992] DRAZIN, P.G. (1992). Nonlinear Systems, Cambridge University Press, Cambridge.

[Frey 1993] FREY, D.K. (1993). "Chaotic digital encoding: An approach to secure communication," IEEE Transactions on Curcuits and Systems, vol. CAS-40, no. 10, pp. 660-666.

[Fukunaga and Krile 1969] FUKUNAGA, K. and T.F. KRILE (1969). "Calculation of Bayes' recognition error for two multivariate Gaussian distributions," IEEE Transactions on Computers, vol. C-18, no. 3, pp. 220-229.

[Fukunaga 1990] FUKUNAGA, K. (1990). Introduction to Statistical Pattern Recognition, 2nd ed., Academic Press, San Diego.

[Giannakis 1999] GIANNAKIS, G.B. (Ed.) (1999). "Highlights of signal processing for communications," IEEE Signal Processing Magazine, vol. 16, pp. 14-50.

[Gibson 1993] GIBSON, J.G. (1993). Principles of Digital and Analog Communications, 2nd ed., Prentice-Hall, Upper Saddle River.

[Glisic and Vucetic 1997] GLISIC, S. and B. VUCETIC (1997). Spread Spectrum CDMA Systems for Wireless Communications, Artech House, Boston.

[Golomb 1967] GOLOMB, S.W. (1967). Shift Register Sequences, Holden-Day, San Fransisco.

[Golomb et al. 1994] GOLOMB, S.W., R.E. PEILE and R.A. SCHOLTZ (1994). Basic Concepts in Information Theory and Coding: The Adventures of Secret Agent 00111, Plenum Press, New York.

[Goldreich 1999] GOLDREICH, O. (1999), Modern Cryptography, Probabilistic Proofs and Pseudorandomness, Springer, Berlin.

[Gray and Markel 1976] GRAY, A.H. and J.D. MARKEL (1976). "Distance Measures for Speech Processing," IEEE Transactions on Acoustics, Speech snd Signal Processing, vol. ASSP-24, no.5, pp. 380-391.

[Gray et al. 1980] GRAY, R.M., A. BUZO, A.H. GRAY and Y. MATSUYAMA (1980). "Distortion Measures for Speech Processing," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-28, no. 4, pp. 367-376.

[Haykin 1996] HAYKIN, S. (1996). Adaptive Filter Theory, 3rd ed., Prentice-Hall, Upper Saddle River.

[Hanssen 1997] HANSSEN, A. (1997). "Stochastic process shift keying: a new concept for low-probability-of-intercept digital communications," Internal.

[Itakura and Saito 1970] ITAKURA, F. and S. SAITO (1970). "A statistical method for estimation of speech spectral density and formant frequencies," Electronic Communications Japan, no. 53-A, pp. 36-43.

[Itakura 1975] ITAKURA, F. (1975). "Minimum prediction residual principle applied to speech recognition," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-23, no. 1, pp. 67-72.

[Jeffreys 1946] JEFFREYS, H. (1946). "An invariant form for the prior probability in estimation problems," Proceedings of Royal Society of London A, vol. 186, pp. 453-461.

[Jeffreys 1948] JEFFREYS, H. (1948). Theory of Probability. Oxford University Press, Oxford.

[Kailath 1967] KAILATH, T. (1967). "The divergence and Bhattacharyya distance measures in signal selection," IEEE Transactions on Communication Technology, vol. COM-15, pp. 52-60.

[Kay 1979] KAY, S.M. (1979). "The effects of noise on the autoregressive spectral estimator," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-27, pp. 478-485.

[Kay 1993] KAY, S.M. (1993). Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall, Englewood Cliffs.

[Kazakos and Papantoni 1990] KAZAKOS, D. and P. PAPANTONI-KAZAKOS (1990). Detection and Estimation. Computer Science Press, New Tork.

[Kullback 1959] KULLBACK, S. (1959). Information Theory and Statistics, John Wiley, New York.

[Larsen and Marx 1986] LARSEN, R.J. and M.L. MARX (1986). An Introduction to Mathematical Statistics and Its Applications, 2nd ed., Prentice-Hall, Englewood Cliffs.

[Lee et al. 1997] LEE, C., D.B. WILLIAMS and J. LEE (1997). "A secure communications system using chaotic switching," International Journal of Bifurcation and Chaos, vol. 7, pp. 1383-1394.

[Malakhov and Yakimov 1993] MALAKHOV, A. and A. YAKIMOV (1993). "The physical models and mathematical description of 1/f noise," in Wavelets, Fractals and Fourier Transforms, M. Farge et al. (Eds.), pp. 341-352, Clarendon Press, Oxford.

[Mandelbrot 1999] MANDELBROT, B.B. (1999). Multifractals and 1/f Noise: Wild Self-Affinity in Physics: Selecta Volume N, Springer, New York.

[Mendel 1991] MENDEL, J.M. (1991). "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," Proceedings of the IEEE, vol. 79, pp. 2778-305.

[Meyr et al. 1998] MEYR, H., M. MOENECLAEY and S.A. FECHTEL (1998). Digital Communication Receivers: Synchronization, Channel Estimation and Signal Processing, John Wiley, New York.

[Ojanpera and Prasad 1998] OJANPERA, T. and R. PRASAD (1998). Wideband CDMA for Third Generation Mobile Communications, Artech House, Boston.

[Oppenheim et al. 1983] OPPENHEIM, A.V., A.S. WILLSKY and I.T. YOUNG (1983). Signals and Systems, 1st ed., Prentice-Hall, London.

[Papoulis 1991] PAPOULIS, A. (1991). Probability, Random Variables and Stochastic Processes, 3rd ed., McGraw-Hill, New York.

[Pecora and Caroll 1990] PECORA, L. and T. CAROLL (1990). "Synchronization in chaotic systems," Physical Review Letters, vol. 64, pp. 821-823.

[Pecora and Caroll 1991] PECORA, L. and T. CAROLL (1991). "Driving systems with chaotic signals," Physical Review A, vol. 44, pp. 2374-2383.

[Peebles 1993] PEEBLES, P.Z. (1993). Probability, Random Variables and Random Signal Principles, 3rd ed., McGraw-Hill, Singapore.

[Peterson et al. 1995] PETERSON, R.L., R.E. ZIENER and D.E. BORTH (1995). Introduction to Spread Spectrum Communications, Prentice-Hall, New Jersey.

[Porat and Friedlander 1987] PORAT, B. and B. FRIEDLANDER (1987). "The exact Cramer-Rao bound for Gaussian autoregressive processes," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-34, pp. 537-541.

[Priestley 1988] PRIESTLEY, M. (1988). Non-linear and Non-stationary Time Series Analysis, Academic Press, San Diego.

[Proakis 1995] PROAKIS, J.G. (1995). Digital Communications, 3rd ed., McGraw-Hill, New York.

[Rabiner and Juang 1993] RABINER, L.R. and B.H. JUANG (1993). Fundamentals of Speech Recognition, Prentice-Hall, Englewood Cliffs.

[Salberg and Hanssen 1999a] SALBERG, A.-B. and A. HANSSEN (1999). "Secure digital communications by means of stochastic process shift keying: Principles and properties," Proceedings of NORSIG-99, Asker, Norway, pp. 48-53.

[Salberg and Hanssen 1999b] SALBERG, A.-B. and A. HANSSEN (1999). "Secure digital communications by means of stochastic process shift keying," Proceedings of the 33rd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, 5 pp.

[Salberg and Hanssen 2000] SALBERG, A.-B. and A. HANSSEN (2000). "A secure digital modulation technique," Submitted to IEEE Communications Letters, February 2000, .

[Scharf 1987] SCHARF, L.L. (1987). "Low Rank Detectors for Gaussian Random Vectors," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-35, pp. 1579-1582.

[Scharf 1991] SCHARF, L.L. (1991). Statistical Signal Processing: Detection, Estimation and Time Series Analysis. Addison-Wesley, Reading.

[Simon 1994] SIMON, M.K. et al. (1994). Spread Spectrum Communications Handbook, McGraw-Hill, New York.

[Strogatz 1994] STROGATZ, S.H. (1994). Nonlinear Dynamics and Chaos: With Applications in Physics, Biology, Chemistry and Engineering, Reading.

[Tang et al. 1983] TANG, Y.S., A.I. MEES and L.O. CHUA (1983). "Synchronization and chaos," IEEE Transactions on Curcuits and Systems, vol. CS-30, pp. 620-626.

[Viterbi 1995] VITERBI, A.J. (1995). CDMA: Principles of Spread Spectrum Communication, Addison Wesley, Reading.

[Welsh 1988] WELSH, D.J.A. (1988). Codes and Cryptography, Clarendon Press, Oxford.

[West and Schlesinger 1990] WEST, B.J. and M.F. SCHLESINGER (1990). "The noise in natural phenomena," American Scientist, no. 78, pp. 40-45.

[Wu and Chen 1997] WU, W.R. and P.C. CHEN (1997). "Adaptive AR-modeling in white gaussian noise," IEEE Transactions on Signal Processing, vol. SP-45, pp. 1184-1192.

[Zhang et al. 1994] ZHANG, Q., M. BASSEVILLE and A. BENVENISTE (1994). "Early Warning of Slight Changes in Systems", Automatica, vol. 30, no 1, pp. 95-113.