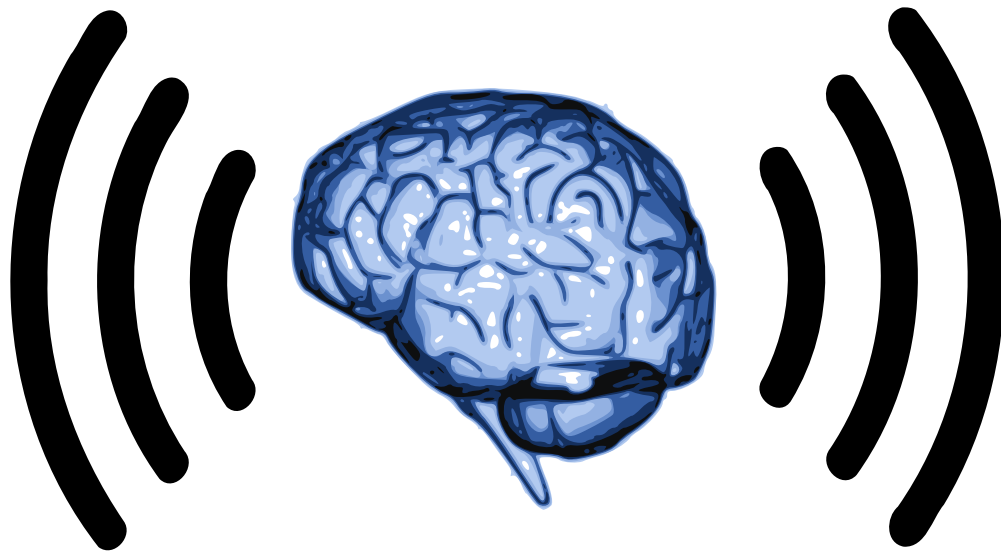# INTERFERENCE AND NETWORK MANAGEMENT IN COGNITIVE COMMUNICATION SYSTEMS

UniversidadeVigo

Gonzalo Vázquez Vilar

University of Vigo
Signal Processing and Communications Department

DOCTORAL DISSERTATION
EUROPEAN MENTION

INTERFERENCE AND NETWORK MANAGEMENT IN COGNITIVE COMMUNICATION SYSTEMS

Author:

Gonzalo Vázquez Vilar

Directed by:

Roberto López Valcarce

Carlos Mosquera Nartallo

2011

DOCTORAL DISSERTATION
EUROPEAN MENTION

INTERFERENCE AND NETWORK MANAGEMENT IN COGNITIVE COMMUNICATION SYSTEMS

Gonzalo Vázquez Vilar

Directed by:
Roberto López Valcarce
Carlos Mosquera Nartallo

EXAMINATION COMMITTEE

President: Fernando Pérez González

Board members: Wilfried Gappmair

                   Santiago Zazo Bello

                   Ashish Pandharipande

Board secretary: Josep Sala Alvarez

Thesis defense: Vigo, June 29, 2011

This edition: July 5, 2011

The wireless channel is a complicated animal.

A. Paulraj

Research is to see what everybody else has seen, and to think what nobody else has thought.

A. Szent-Györgyi

Acknowledgements

First of all, I want to thank my advisors Roberto and Carlos for their guidance and encouragement in the development and preparation of this dissertation. Without them this thesis would not exist.

I would also like to acknowledge all the people I met during these last two years and who participated at any point in the work of this thesis. Specially, I would like to thank Nuria González Prelcic, Josep Sala, Sudharman K. Jayaweera, David Ramírez, and Ashish Pandharipande, as well as the members from the groups of SPCOM (Technical Univ. of Catalonia UPC) and GTAS (University of Cantabria). Additionally I want to acknowledge all the people that, through my blog Spectral Holes, helped me to understand several issues behind Cognitive Radio. I learned from all of them.

Other people who have also contributed to this thesis, but in a different way, are my lab colleagues, Marcos, Norberto, Enrique, Marta, Eli and Debora. I want to thank all of them for doing my stay pleasant and enjoyable. Also my friends Yoli, Borja, Paco and Edu, who shared my lunch-time relax during this period of my life.

And last but not least, I want to acknowledge the support of my family. I thank my parents, Gonzalo and Lidia, my sister, Marta, and my girlfriend Alba for their (not small) patience and (huge) love.

Abstract

The key idea behind Cognitive Radio (CR) is to allow opportunistic access to temporally and/or geographically unused licensed bands, avoiding conflicts with the rightful license owners in those bands. To achieve this, novel interference management algorithms are required to limit the interference seen by the primary (licensed) users. A key aspect of any interference management scheme is spectrum monitoring, that allows to detect and track primary users.

This PhD. Thesis contributes to the field of CR in two different ways. First, we address the problem of primary user monitoring using novel detection schemes which exploit multiple antennas, wideband processing, and the available *a priori* knowledge about primary transmissions. Then, we propose a general framework for interference management in cognitive radio networks in which certain interaction is allowed between primary and secondary systems.

Specifically, the detection problems investigated in this thesis include multiantenna detection exploiting *a priori* spectral information when the noise statistics are assumed known. In this setting we will also derive novel diversity order analysis of the proposed detectors. The case of multiantenna detection under unknown noise statistics is covered under different hypotheses, including both the detection of primary signals with spatial rank larger than one and detection in presence of spatially unstructured noise. Additionally we study the problem of multichannel monitoring. In this context, wideband acquisition can be performed using traditional analog to digital converters or the recently proposed analog to information converters. When the channelization of the primary network is assumed known, we show that guard bands and weak channels can be used to improve detection performance, both when the detection is performed from a set of samples at Nyquist rate, or from a set of compressed measurements.

Finally, we propose a general framework for interference management in cogni-

tive radio networks in which the primary network is allowed to dynamically adjust the tolerable interference margin to be met by the secondary system. In particular, we propose a game theoretical formulation which allows us to study the performance gain which can be expected from this limited interaction between primary and secondary systems. Moreover, we show that certain architectures fulfilling these requirements are implementable in practice and present good performance in both static and dynamic environments.

Publications

The following is a list of journal and conference publications that have been produced as a result of the work on this thesis.

Journal publications

1. Josep Sala, **Gonzalo Vazquez-Vilar**, and Roberto López-Valcarce. Multiantenna detection in unknown spatially correlated noise in cognitive radio networks. *Signal Processing, IEEE Transactions on*, 2011. In review.

2. **Gonzalo Vazquez-Vilar**, Roberto López-Valcarce, and Josep Sala. Multiantenna detection of multicarrier primary signals exploiting spectral a priori information. *Wireless Communications, IEEE Transactions on*, 2011. In review.

3. **Gonzalo Vazquez-Vilar** and Roberto López-Valcarce. Wideband spectrum sensing exploiting guard bands and weak channels. *Signal Processing, IEEE Transactions on*, 2011. Accepted for publication.

4. David Ramírez, **Gonzalo Vazquez-Vilar**, Roberto López-Valcarce, Javier Vía, and Ignacio Santamaría. Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas. *Signal Processing, IEEE Transactions on*, 2011. In press.

5. **Gonzalo Vazquez-Vilar**, Carlos Mosquera, and Sudharman K. Jayaweera. Primary user enters the game: Performance of dynamic spectrum leasing in cognitive radio networks. *Wireless Communications, IEEE Transactions on*, 9(12):3625–3629, December 2010.

6. Sudharman K. Jayaweera, **Gonzalo Vazquez-Vilar**, and Carlos Mosquera. Dynamic Spectrum Leasing (DSL): A new paradigm for spectrum sharing in cognitive radio networks. *Vehicular Technology, IEEE Transactions on*, 59(5):2328–2339, June 2010.

Conference publications

1. **Gonzalo Vazquez-Vilar**, Roberto López-Valcarce, and Ashish Pandharipande. Detection diversity of multiantenna spectrum sensors. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011.

2. David Ramírez, **Gonzalo Vazquez-Vilar**, Roberto López-Valcarce, Javier Vía, and Ignacio Santamaría. Multiantenna detection under noise uncertainty and primary user's spatial structure. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011.

3. Roberto López-Valcarce, **Gonzalo Vazquez-Vilar**, and Josep Sala. Multiantenna spectrum sensing for cognitive radio: overcoming noise uncertainty. In *The 2nd International Workshop on Cognitive Information Processing (CIP 2010)*, Elba Island (Tuscany), Italy, June 2010.

4. Georges El-Howayek, Sudharman K. Jayaweera, Kamrul Hakim, **Gonzalo Vazquez-Vilar**, and Carlos Mosquera. Dynamic spectrum leasing (DSL) in dynamic channels. In *ICC'10 Workshop on Cognitive Radio Interfaces and Signal Processing (ICC'10 Workshop CRISP)*, Cape Town, South Africa, May 2010.

5. **Gonzalo Vazquez-Vilar**, Roberto López-Valcarce, Carlos Mosquera, and Nuria González-Prelcic. Wideband spectral estimation from compressed measurements exploiting spectral a priori information in cognitive radio systems. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010)*, Dallas, U.S.A., March 2010.

6. Roberto López-Valcarce, **Gonzalo Vazquez-Vilar**, and Marcos Álvarez Díaz. Multiantenna detection of multicarrier primary signals exploiting spectral a

priori information. In *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (Crowncom 2009)*, Hannover, Germany, June 2009.

7. Roberto López-Valcarce and **Gonzalo Vazquez-Vilar**. Wideband spectrum sensing in cognitive radio: Joint estimation of noise variance and multiple signal levels. In *2009 IEEE International Workshop on Signal Processing Advances for Wireless Communications (Spawc 2009)*, Perugia, Italy, June 2009.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Contents

1.1 Cognitive Radio: Motivation

In the recent years we have witnessed a constant increase in the price of the spectral resources. The main reason is the rising demand of spectrum as a result of emerging communication standards and services.

However this scarcity of spectral resources happens while most of the allocated spectrum is underutilized. This paradox occurs only due to the inefficiency of traditional static spectrum allocation policies, which translates in a waste of spectral resources (FCC, 2002, 2003). Most of the useful spectrum is allocated to licensed users (e.g. mobile carriers, TV broadcasting companies) that do not transmit at all the geographical locations all the time. If this spectrum is opened for unlicensed use (e.g. private users, short range networks, ...) it is highly likely that a vast array

of new services will appear. One example of this is the huge innovation that has occurred in WiFi and Bluetooth operating in unlicensed bands, even though these two standards share just scraps of spectrum with many other technologies.

The wireless industry has considerable interest in the development of *dynamic spectrum access* (DSA) as a means to improve spectral efficiency (FCC, 2002, 2003). *Cognitive radio* (CR) (Mitola and Maguire Jr., 1999) is receiving considerable attention as the enabling technology to achieve DSA in licensed bands. The key idea behind CR is to create smarter radios which are aware of, and can adapt to, their environment. Hence, in licensed bands CR nodes will monitor primary users in order to transmit in temporally and/or spatially unused slots. For example, the U.S. Federal Communications Commission (FCC) has recently issued a Second Report and Order (FCC, 2010), allowing operation on an unlicensed basis in the TV white spaces of VHF and UHF bands to both fixed and portable devices. While this order requires secondary users to access a database with information of the available resources, it is expected that these first steps start a major change to DSA in most of the spectrum once the CR technology is mature enough.

One of the problems pointed out in FCC (2010) is that the available sensing technology is not reliable enough to guarantee that the interference produced to licensed (primary) users is kept at sufficiently low levels. Wireless propagation phenomena such as shadowing and fading pose significant challenges to the reliable detection of primary users. The received primary signal may be very weak, resulting in very low Signal-to-Noise Ratio (SNR) operation conditions and "hidden node" situations. Hence novel powerful spectrum monitoring techniques are required in order to increase CR network agility (Akyildiz et al., 2008).

On the other hand, CR schemes may lead to very complex networks, in which primary and secondary users coexist in dynamic environments. This may lead to unexpected behavior and/or an impact on system performance. Hence, new schemes and analytical tools are required to control and model the interactions between the different elements of the system.

1.2   Previous work

While CR is a relatively novel area (Mitola's landmark paper appeared in 1999 (Mitola and Maguire Jr., 1999)), it has received significant research interest in the last few years. In this section we present the most relevant previous work directly related

to this thesis.

1.2.1   Spectrum Monitoring

Spectrum monitoring is based on the detection of weak signals from primary trans-
mitters through the local observations of cognitive users, either individually or in a
collaborative fashion. While cooperative sensing has the potential to overcome the
effects of shadowing (Ganesan and Li, 2007a,b), it still relies on standalone detectors
whose performance should be optimized.

Three schemes are generally used for sequential individual sensing of primary
channels, each of them requiring different degrees of knowledge and synchronization
with the primary network:

1. Matched filter detection: If the secondary user is locked[1] to the primary net-
   work, the optimal detection strategy in stationary Gaussian noise is matched
   filtering (Kay, 1998). Note that matched filter detection schemes require full
   synchronization with the primary network and thus they are difficult to im-
   plement in the low SNR conditions, which cognitive networks are expected to
   work in.

2. Feature based detection: Certain properties of the primary signal, such as
   the presence of any pilots or cyclostationary features, could in principle be
   exploited in order to obtain powerful detectors. However, such approaches
   become very sensitive to synchronization errors (Cabric, 2008). With very low
   SNR, the synchronization loops of the monitoring system cannot be expected
   to provide the required accuracy for the carrier frequency and/or clock rate
   estimates.

3. Asynchronous detectors: These detection schemes do not assume any synchro-
   nization with the primary signal. Hence they rely on other signal properties
   such as certain temporal and/or spatial structure. Among these, the most
   popular one is the energy detector, which does not require (or exploit) any
   a priori knowledge about the signal structure. The detector reduces to inte-
   grating received energy in a given frequency band and comparing it to a noise
   level dependent threshold. However, computation of the threshold in energy

---

[1]Meaning that both frequency and timing synchronization loops are locked to a given set of
signals.

detection requires knowledge of the noise variance. Any uncertainty regard-
ing this parameter translates to severe performance degradation, so that the
detection/false alarm requirements may not be satisfied (Tandra and Sahai,
2008).

The reasons exposed motivate the search for asynchronous detectors robust to
noise uncertainty, two possibilities being the use of multiple–antenna sensors and
wideband monitoring covering multiple frequency channels. Moreover, if certain
information about the primary network, such as channelization and modulation,
is available to the spectrum monitor, it should be exploited to increase detection
performance. In this thesis we will consider that this knowledge can be summarized
as the spectrum shape / temporal correlation of the received signal.

Several authors considered the problem of exploiting temporal structure of the
received signal. Under the assumption that the power spectral density (psd) of the
signal is completely known, Zhang et al. (2010b) derive the optimal Neyman-Pearson
detector for both scalar and vector-valued signals. However, in spectrum sensing
applications the propagation channel is unknown, and thus only partial knowledge
of the second-order statistics is available in practice. A possible approach in that
case is to neglect this partial knowledge, and consider test statistics that quantify the
departure of the sample temporal autocorrelation matrix of the observations from the
noise temporal covariance (Zeng and Liang, 2009a,b). Under the assumption that
the signal is bandlimited, while its actual bandwidth, spectral shape and carrier
frequency are unknown, Derakhtian et al. (2009) propose a *generalized likelihood
ratio test* (GLRT) based scheme. Alternatively, metrics quantifying the distance
of the sample correlation matrix from a "candidate" matrix summarizing *a priori*
knowledge can be used: in the single–antenna setting, for example, Perez-Neira et al.
(2009) assume the signal psd known up to a scaling and a shift, respectively modeling
uncertainty about the power level and carrier frequency of the signal. Also assuming
a single antenna, Quan et al. (2011) adopt a similar approach when the carrier
frequency is known, as it often occurs in practice: for instance, for frequency division
multiple access (FDMA) primary networks with public channelization parameters.
However, all these works either assume that temporal correlation of the primary
signal is unknown to the receiver or they do not consider the multiantenna and
wideband settings.

Multiantenna detection

The gain offered by multiantenna processing in energy detection schemes was analyzed in Pandharipande and Linnartz (2007) under the assumption of channel information available to the secondary system. This assumption is not realistic in practice and the channel needs to be estimated. Assuming a temporally white Gaussian model for both signal and noise, spatially white noise with unknown (equal) variance across antennas, and an unknown spatial covariance matrix for the signal, several detectors have been proposed in the literature.

We are particularly interested in the works based on the *generalized likelihood ratio test* (GLRT), since this approach usually results in simple detectors with good performance (Mardia et al., 1979). Under rank-1 spatial covariance for the signal and assuming iid noises, the GLRT is derived in Besson et al. (2006) and its application to CR was presented in Taherpour et al. (2010); Wang et al. (2010). When the signal covariance matrix is unstructured, and the noise assumed iid, the GLRT is the well-known *test for sphericity* (Mauchly, 1940), which was applied to CR in Lim et al. (2008); Zhang et al. (2010a). In these works the authors derived the GLRT for primary signals with spatial rank $P > 1$ under the assumption of iid noises with known variance. In Wilks (1935) the GLRT was derived for the case of an unstructured signal covariance matrix for non-iid noises. This detector was later applied to array signal processing in Leshem and Van der Veen (2001a,b). Other detectors which can handle different (unknown) noise variances have been proposed in Boonstra and Van der Veen (2003); Zeng and Liang (2009b).

However, all of these works either assume rank-1 primary signals or unstructured primary signals. Moreover, they do not exploit any available information about the spectral shape of the primary signals.

Once a multiantenna detector is proposed its performance must be evaluated. In order to quantify and compare the performance gain of multiantenna systems in fading environments, several metrics have been considered, including different concepts of detection diversity. One option is to adopt a definition analogous to the one from the communications literature for a certain performance tradeoff (between the probabilities of detection and false alarm), as proposed in Duan et al. (2010). A similar asymptotic definition based on J-divergence is given in Kim et al. (2009). In the context of radar, diversity order is however a low-SNR concept. For example, Daher and Adve (2010) define diversity order as the slope of the average probability of detection ($\bar{P}_\mathrm{D}$) curve with respect to the SNR at $\bar{P}_\mathrm{D} = 0.5$.

Wideband detection

In order to improve detection performance, the sensing system may also perform simultaneous acquisition of multiple frequency channels. This scheme improves the agility of the detector since multiple channels are processed at once and it provides the spectrum monitor with additional information to estimate the noise statistics as we will see in Chapter 4. This is mainly due to the availability of guard bands between adjacent channels, as well as to the fact that the presence of unused/weak channels within the subband can be exploited for noise variance estimation.

Wideband spectrum sensing has been previously considered by several authors. In Hwang et al. (2010), knowledge of the noise variance is assumed, but the bandwidths and central frequencies of primary transmissions, as well as their number, are assumed unknown and estimated in turn. In the setting of Taherpour et al. (2008, 2009) primary system channelization is known, and the noise variance is regarded as unknown. However, these methods do not exploit *a priori* information about the psd of primary transmissions, and they assume that a minimum number of unused channels exist in the subband under examination.

In a wideband setting, it may not be feasible to acquire the received signal at Nyquist rate. Novel sampling methods allow the reconstruction of the received signals from a set of compressed measurements if certain properties are met (Donoho, 2006). The key technology allowing this is compressive sensing, which is able to construct sparse solutions from a set of underdetermined equations. Several authors have applied compressed sensing to the detection of primary users in cognitive radio systems. Assuming a spectrum model consisting of several flat bandpass signals, and considering the edges between them, the observed signal is sparse in the "spectral edges domain". This fact is used in Tian and Giannakis (2007) to propose a spectrum reconstruction algorithm from compressed samples of the signal autocorrelation estimate. This method was extended in Polo et al. (2009) in order to process directly a compressed version of the received signal (and not of its autocorrelation).

These methods do not assume information about the primary network channelization, so that the spectral edges could occupy any position within the frequency band.

1.2.2 Interference Management

In order to improve spectral efficiency, the wireless industry has prompted proposals for various *dynamic spectrum access* (DSA) approaches. A DSA scheme in which secondary users are allowed to opportunistically access the spectrum on the basis of no-interference to the primary (licensed) users, denoted as hierarchical access, is arguably the method that has received the most attention in recent literature.

Various architectures have been proposed and investigated in recent years to achieve hierarchical dynamic sharing of licensed bands (see Kim et al. (2008); Le and Hossain (2008); Xing et al. (2007); Fattahi et al. (2007); Etkin et al. (2007); Menon et al. (2008) and references therein). A common assumption in these works is that the licensed users which own the spectrum rights are unaware of the presence of secondary users. Hence the burden of interference management relies mainly on the secondary system. In particular, either (i) there is a maximum interference level that the primary system is willing to tolerate, and the secondary powers/activity are to be adjusted within this constraint, or (ii) secondary users are allowed to opportunistically access the spectrum on the basis of no-interference to the primary (licensed) users.

As opposed to this is the the concept of *dynamic spectrum leasing* (DSL), first presented in Jayaweera and Li (2009). A DSL scheme is characterized by the active role of the primary user, which may interact with the secondary system in order to define the allowed interference cap. This scheme allows the system to adapt to changing environmental conditions and may lead to a better spectral utilization.

1.3 Contributions

This thesis treats different aspects of a cognitive radio system. On the one hand, assuming a non-interfering DSA network we propose and analyze novel asynchronous multiantenna detectors and wideband detection schemes. Then, in the last chapter, we will study a DSA system in which certain interference is allowed at primary users, namely a DSL network.

1.3.1   Multiantenna and multichannel detection of primary users

In the previous section we showed the importance of deriving powerful asynchronous detectors for cognitive radio systems. To this end we need to exploit any available information about the primary network. In most of the analysis in this thesis we assume that the modulation / channelization of the primary network is known to the spectral monitor, which translates into *a priori* knowledge on the spectral shape of the primary transmissions.

Additionally, we focus our study on Gaussian signals. The reasons for adopting a Gaussian model for the primary signal are as follows. First, under asynchronous sampling, the actual distribution is unknown; and since the noise is assumed Gaussian as well, the Gaussian pdf for the signal is the least informative one for the detection problem. Second, if the primary system uses multicarrier modulation with a sufficiently large number of subcarriers (which is the case in e.g. broadcasting applications), the Gaussian model is accurate (Tellado, 2000). Third, this model is tractable and leads to useful detectors under other distributions: note that Gaussianity is a common assumption in the development of signal detectors, either explicitly or implicitly, as many *ad hoc* methods that limit themselves to the use of second-order statistics of the observations can often be derived from a Gaussian model (the Energy Detector is the most prominent example).

The main contributions of this thesis in spectrum monitoring are the following:

- **Derivation and analysis of different multiantenna detectors exploiting *a priori* knowledge of the spectral shape of the primary transmissions when the noise statistics are assumed known.** From the (non-implementable) Neyman-Pearson optimal detector we derive a family of practical multiantenna detectors with different levels of complexity. This will allow us to study both the advantages of exploiting spectral information and multiantenna processing under different scenarios.

- **Diversity order analysis of multiantenna detection systems in cognitive radio.** In order to compare and rank the different detectors in fading environments we propose the use of two different performance metrics which reflect the diversity gain obtained by multiantenna systems. The first is analogous to the one used in communications and measures the asymptotic slope of the probability of misdetection with respect to the SNR (in log-log scale) for increasing SNRs. The second is borrowed from the radar community and

is related to the behavior of the probability of detection around the point at which it equals 1/2. These key measures show the advantage of multiantenna processing when detecting primary signals.

- **Multiantenna detection of primary signals with spatial rank larger than one when the noise statistics are assumed unknown.** Under the Gaussian assumption, we derive the GLRT when both signal and noise are assumed temporally white and the primary signal may present an arbitrary spatial rank larger than one, both for spatially iid noises and when the noise is spatially uncorrelated but not necessarily iid.

  We emphasize the practical implications of this scenario. A primary signal with spatial rank larger than one will occur, for example, if multiple independent users (e.g. from adjacent cells) simultaneously access the same frequency channel. Alternatively, many state-of-the-art communication standards consider the simultaneous transmission of different data streams through multiple antennas to achieve multiplexing gain and/or the use of space-time codes to enhance spatial diversity. For these systems, the signal received at the multiantenna sensor will exhibit a spatial rank equal to the number of independent streams or the spatial size of the code, respectively. On the other hand, tolerances in the components of the analog frontends at different antennas will result in deviations of the noise level from antenna to antenna, and as it turns out, detectors derived under the iid assumption are very sensitive to these calibration errors.

- **Derivation and analysis of multiantenna detection of primary signals under strong interference.** Assuming strong interference, modeled as temporally white noise with arbitrary spatial covariance matrix, we derive the GLRT for detection of primary signals with known temporal structure. We additionally propose a low SNR asymptotic analysis of this detector which can be tightened in the SNR range of interest. This analysis shows the existing tradeoff between the spectral shape of the primary signal and detection performance when the spatial structure of the signal is masked by the noise.

  This scenario may occur in the presence of strong cochannel interference generated by other secondary users. In this case, the secondary contributions can be modeled as noise with arbitrary and unknown spatial covariance.

- **Wideband detection in the presence of unknown noise level.** Intuitively, if multiple primary channels are simultaneously acquired and channel-

ization information is available, the guard bands between adjacent channels could be used to estimate the noise power. We will show that when considering the problem of GLRT detection of one of the channels, not only the guard bands but also the empty/weak channels are used to improve the noise estimate. This analysis shows the advantages of performing wideband detection instead of channel-by-channel scanning.

- **Wideband detection from compressed measurements.** We propose a primary user detection scheme from a set of compressed samples based on the GLRT when the channelization of the primary network is assumed known. From a maximum *a posteriori* formulation we establish a connection between the estimation problem of the unknown parameters and certain compressed sensing techniques. Additionally, we propose a simple iterative procedure that conducts to similar detection performance as by using more complex convex optimization schemes.

1.3.2   DSL: an Interference Management Scheme

A DSL based paradigm allows certain amount of secondary interference at the primary system. Then primary user detection becomes less important in comparison to interference management. The main contribution in this section is the study of a family of DSL architectures showing their interference management capabilities.

- **Performance gain of DSL based paradigms.** We present a theoretical analysis of the performance gain obtained by allowing a certain amount of interaction between primary and secondary systems. To this end, we define a family of performance metrics and propose a Stackelberg game formulation for the interactions between primary and secondary systems. We show that the performance gain obtained by allowing this interaction can be indeed large in dynamic environments.

- **Practical DSL scheme.** Finally, we analyze certain practical DSL schemes which are shown to have a unique Nash equilibrium. In the stationary regime, the global performance of the system can be assumed to be the performance at the (unique) Nash equilibrium, which makes its analysis tractable. Moreover, the proposed DSL schemes show a graceful degradation under dynamic conditions and thus perform well in practice.

1.4   Structure of the thesis

This thesis is divided in two different parts. In the first we address different spectrum monitoring schemes, focusing on multiantenna and wideband detectors. Then, in the second part, we propose a general framework for interference management in cognitive radio networks.

In Chapter 2 we will study the problem of multiantenna detection exploiting *a priori* spectral information when the noise statistics are assumed known. In this chapter we will also pose the diversity order analysis of the proposed detectors. The case of multiantenna detection under unknown noise statistics will be covered in Chapter 3, including both the detection of primary signals with spatial rank larger than one, and detection in the presence of spatially unstructured noise. Chapter 4 covers the topic of wideband acquisition and detection, both when the band is acquired at Nyquist rate and when the detection must be performed from a set of compressed measurements.

The analysis of a family of DSL schemes is presented in Chapter 5. Concluding remarks, as well as future lines of research, are included in Chapter 6.

1.5   Notation

Any non-standard notation used in this thesis is defined for the particular chapter at the point where the symbols first occur. For reader's reference, we also include a comprehensive list of the notation in Table 1.1.

Symbol	Description
$\Re(\alpha)$, $\Im(\alpha)$	Real and imaginary parts of α
$\|\alpha\|$, $\arg(\alpha)$	Absolute value and argument of α
$(\cdot)^T$, $(\cdot)^H$	Transpose and conjugate transpose
$\|\cdot\|_\ell$ (resp. $\|\cdot\|$)	norm ℓ (resp. norm 2)
$\det(\mathbf{A})$, $\mathrm{tr}(\mathbf{A})$, $\mathbf{A}^\dagger$	Determinant, trace and pseudoinverse of $\mathbf{A}$
$\mathbf{A}^{1/2}$ (resp. $\mathbf{A}^{-1/2}$)	Hermitian square root matrix of $\mathbf{A}$ (resp. $\mathbf{A}^{-1}$)
$\mathrm{diag}(\mathbf{a})$	Diagonal matrix with diagonal equal to $\mathbf{a}$
$\mathrm{adj}(\mathbf{A})$	Adjugate matrix of $\mathbf{A}$
$\mathrm{vec}\,(\mathbf{A})$	Column-wise vectorization of $\mathbf{A}$
$\odot$	Hadamard product
$\otimes$	Kronecker product
$\mathbf{0}_L$	Zero $L \times 1$ vector or $L \times L$ matrix
$\mathbf{1}_L$	$L \times 1$ all-ones vector
$\mathbf{I}_L$	Identity matrix of size $L \times L$
$\mathbf{a}_k$	k-th column of matrix $\mathbf{A}$
$\mathbf{e}_k$	k-th column of the identity matrix
$\mathrm{E}[\cdot]$	Expectation operator
$\mathrm{var}\{\cdot\}$	Variance operator
$\mathrm{cov}\{\boldsymbol{x}, \boldsymbol{y}\}$	Covariance between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$
$\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$, $\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$	(Complex circular) Gaussian random distribution of mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$
$\mathcal{U}(a, b)$	Uniform random distribution with support $[a, b]$
$Q(x)$	Tail probability of the standard normal distribution
$\mathbb{1}(\cdot)$	Indicator function
$\mathcal{O}(\cdot)$	$f(x) \in \mathcal{O}(g(x))$ iff $\lim_{x\to\infty} f(x)/g(x)$ equals a constant
$o(\cdot)$	$f(x) \in o(x)$ iff $\lim_{x\to 0} f(x)/x = 0$

Table 1.1: Notation used in this Thesis.

Chapter 2

Calibrated Multiantenna Detection

Contents

2.1   Introduction

Primary user monitoring in Cognitive Radio systems is based on the detection of
the signal, generated by a primary transmitter, from the local observations of cog-
nitive users, either individually or in a collaborative fashion. In either case, it
is very likely that future CR terminals will incorporate multiple antennas, given
that multiple-input multiple-output (MIMO) technologies for communications have
reached considerable maturity (Larsson and Stoica, 2003).

In terms of transmission/reception, multiple antennas provide a means to in-
crease channel capacity without bandwidth expansion, as well as to overcome the
effects of fading via space-time coding (Larsson and Stoica, 2003). Several authors
have recently studied the benefits of having multiple antennas in terms of enhanc-
ing detection performance in the context of CR systems, see e.g. Pandharipande
and Linnartz (2007); Taherpour et al. (2010); Lunden et al. (2009). However, these
schemes do not exploit that in several cases certain primary network parameters,
such as channelization, modulation type, etc., are available as *a priori* knowledge. In
this chapter we study the problem of multiantenna detection in the low SNR regime
when some *a priori* information, summarized into knowledge about the spectral
shape of primary transmissions, is available to the spectrum monitor.

To this end, we first pose the Neyman-Pearson (NP) detector for this problem.
This detector is not implementable due to the presence of unknown parameters,
which need to be estimated. The maximum likelihood (ML) estimation of these
parameters in different scenarios leads to a family of multiantenna detectors which
result in an increased diversity gain with respect to single-antenna systems.

2.2   System model

Here we formalize the signal model that will be used in this chapter, and which,
with some additional refinements, will also be useful in the following chapters.

Multiantenna reception

We assume the primary system employs Frequency Division Multiplexing with fixed
channelization. The spectrum monitor is equipped with L antennas with their re-
spective Radio Frequency (RF) chains. A given primary channel is selected and

downconverted to baseband, where it is sampled at f_s samples/s to obtain K complex-valued samples at each antenna ($T = K/f_s$ is the observation time). The samples at the l-th antenna are collected into the $K \times 1$ vector $\mathbf{y}_l$ which can be written as

$$\mathbf{y}_l = h_l \mathbf{s} + \sigma \mathbf{n}_l, \qquad 1 \leq l \leq L, \tag{2.1}$$

where

- $\mathbf{s} = [\, s_0 \, s_1 \, \cdots \, s_{K-1} \,]^T$ comprises the samples of the primary signal.

- h_l is the complex-valued channel gain at antenna l. If the channel is vacant, then $h_l = 0$ for all l.

- $\mathbf{n}_l \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ comprises the noise samples at antenna l.

- $\sigma^2 > 0$ is the background noise power, assumed known and equal at all the antennas[1].

- The noise processes at different antennas are assumed statistically independent, i.e. $\mathrm{E}[\mathbf{n}_l \mathbf{n}_n^H] = \mathbf{I}_K \, \mathbb{1}(l = n)$.

By introducing the vectors

$$\mathbf{y} \doteq \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_L \end{bmatrix}, \quad \mathbf{n} \doteq \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_L \end{bmatrix}, \quad \mathbf{h} \doteq \begin{bmatrix} h_1 \\ \vdots \\ h_L \end{bmatrix}, \tag{2.2}$$

the model (2.1) can be compactly written as

$$\mathbf{y} = \mathbf{h} \otimes \mathbf{s} + \sigma \mathbf{n}. \tag{2.3}$$

Without loss of generality we assume $\mathrm{E}[|s_k|^2] = 1$, since the signal power can be absorbed into the channel vector $\mathbf{h}$. Then, the average SNR per antenna is given by

$$\zeta \doteq \frac{\mathrm{E}[||\mathbf{h} \otimes \mathbf{s}||_2^2]}{\mathrm{E}[||\sigma \mathbf{n}||_2^2]} = \frac{||\mathbf{h}||_2^2}{L\sigma^2}. \tag{2.4}$$

---

[1]Since the noise variance is assumed known at each of the antennas, the derivation can be trivially extended to the case of different noise levels by rescaling the input signals so that all present noise variances equal to one.

Primary signal model

In order to protect primary users from interference, the operational range of spectrum sensors must include primary signals well below decodability levels; in such situations, attempting to synchronize with the potentially present primary signal is unrealistic. Hence, we regard $\{s_k\}$ as a wide-sense stationary random process with power spectral density (psd) $S_{ss}(e^{\jmath\omega})$. We additionally adopt a Gaussian model for the primary signal.

Then we have that $\mathbf{s} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C})$, where $\mathbf{C} \doteq \mathrm{E}[\mathbf{s}\mathbf{s}^H]$. Provided that the channelization and modulation parameters of the primary system are fixed and public (as would be the case, e.g., for broadcast networks), then $S_{ss}(e^{\jmath\omega})$ is known (and so is $\mathbf{C}$). Note that $\mathbf{C}$ is Toeplitz with ones on the diagonal. In general, $\{s_k\}$ will be colored (and $\mathbf{C} \neq \mathbf{I}$) as a result of interchannel guard bands, pulse shaping, etc.

In the sequel we will find useful the following asymptotic eigendecomposition of the primary signal temporal covariance matrix.

Let $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ with $\mathbf{\Lambda} = \mathrm{diag}(\lambda_0\ \lambda_1\ \cdots\ \lambda_{K-1})$ be an eigendecomposition of $\mathbf{C}$, and let $\mathbf{W}$ be the $K \times K$ orthonormal IDFT matrix. As $K \to \infty$ (long observation time) we have the following asymptotic result (Kay, 1998):

$$\lambda_k \to S_{ss}(e^{\jmath\frac{2\pi k}{K}}), \quad 0 \leq k \leq K-1. \tag{2.5}$$

This result is based on the asymptotic equivalence of the sequences of matrices $\{\mathbf{C}\}$ and $\mathbf{W}\mathbf{\Gamma}\mathbf{W}^H\}$, where $\mathbf{\Gamma} = \mathrm{diag}\{S_{ss}(1)S_{ss}(e^{\jmath 2\pi/K})\cdots S_{ss}(e^{\jmath 2\pi(K-1)/K})\}$ for $K = 1$, $2,\ldots$ (Gray, 2006), which has been exploited extensively in the literature; as shown in Zhang et al. (2010c), the loss in detection performance when adopting the approximation

$$\mathbf{C} \approx \mathbf{W}\mathbf{\Lambda}\mathbf{W}^H \tag{2.6}$$

often becomes negligible even for moderate values of K.

The following *spectral shape parameters* will feature in the statistical analysis of the detectors:

$$\bar{b}_n \doteq \frac{1}{K}\,\mathrm{tr}\{\mathbf{C}^n\} \;=\; \frac{1}{K}\sum_{k=0}^{K-1}\lambda_k^n \tag{2.7}$$

$$\approx \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ss}^n(e^{\jmath\omega})\partial\omega \quad \text{for } K \to \infty. \tag{2.8}$$

Note that $\bar{b}_1 = 1$ since $\mathrm{E}[|s_k|^2] = 1$. For white $\{s_k\}$, $\mathbf{C} = \mathbf{I}$ so that $\bar{b}_n = 1$ for all n (in general, one has $\bar{b}_n \geq 1$ by Jensen's inequality).

2.3 Problem formulation

The Neyman-Pearson lemma results in optimal detectors in the sense that the probability of detection is maximized for a given probability of false alarm (Kay, 1998). While in our setup the NP test is not implementable in practice it will lead to a series of practical detectors with a strong connection to the diversity combining techniques employed in communications (Simon and Alouini, 2004).

2.3.1 Neyman-Pearson detector

Based on the $LK \times 1$ vector $\mathbf{y}$ from (2.3), and under the Gaussian model, the corresponding hypothesis test is given by

$$\mathcal{H}_0 : \quad \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathcal{R}_0) \qquad \text{(primary is absent)} \tag{2.9}$$

$$\mathcal{H}_1 : \quad \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathcal{R}_1) \qquad \text{(primary is present)} \tag{2.10}$$

where we have introduced

$$\mathcal{R}_0 \doteq \sigma^2 \mathbf{I}, \tag{2.11}$$

$$\mathcal{R}_1 \doteq \sigma^2 \mathbf{I} + \mathbf{h}\mathbf{h}^H \otimes \mathbf{C}, \quad \text{with} \quad \|\mathbf{h}\|_2^2 > 0. \tag{2.12}$$

This is a *composite test* (Kay, 1998), since $\mathbf{h}$ is unknown.

Let now

$$\mathcal{G} \doteq \mathbf{h}\mathbf{h}^H \otimes \mathbf{C}. \tag{2.13}$$

The NP test for this Gaussian detection problem is an estimator-correlator (Kay, 1998) declaring $\mathcal{H}_1$ true if $\mathbf{y}^H \hat{\mathbf{z}}$ exceeds a threshold, where $\hat{\mathbf{z}}$ is the minimum mean squared error (MMSE) estimator of $\mathbf{z} \doteq \mathbf{h} \otimes \mathbf{s}$ given $\mathbf{y}$ and $\mathbf{h}$. After some straightforward manipulations we obtain

$$\hat{\mathbf{z}} = \mathcal{G}\mathcal{R}_1^{-1}\mathbf{y}, \tag{2.14}$$

so that

$$T_{\mathrm{NP}} \doteq \mathbf{y}^H \hat{\mathbf{z}} \tag{2.15}$$

$$= \mathbf{y}^H \mathcal{G} (\sigma^2 \mathbf{I} + \mathcal{G})^{-1} \mathbf{y} \tag{2.16}$$

$$= \mathbf{y}^H (\mathbf{h}\mathbf{h}^H \otimes \mathbf{C})(\sigma^2 \mathbf{I} + \mathbf{h}\mathbf{h}^H \otimes \mathbf{C})^{-1} \mathbf{y}. \tag{2.17}$$

Note that this test cannot be directly implemented, since it requires knowledge of $\mathbf{h}\mathbf{h}^H$.

Single-antenna case

At this point it is instructive to consider the single-antenna case. If $L = 1$, then $\mathcal{G} = |h|^2 \mathbf{C}$, and the NP test statistic can be written as

$$T_{\mathrm{NP}} = \sum_{k \in \mathcal{B}} \frac{|h|^2 \lambda_k}{\sigma^2 + |h|^2 \lambda_k} |v_k|^2, \tag{2.18}$$

where $\mathbf{v} = [v_0 \, v_1 \, \cdots \, v_{K-1}]^T \doteq \mathbf{U}^H \mathbf{y}$, so that $\mathcal{B} \subset \{0, 1, \ldots, K-1\}$ is the set of indices of nonzero eigenvalues of $\mathbf{C}$. In view of (2.5), for large K one has $\mathbf{v} \approx \mathbf{W}^H \mathbf{y}$ (the DFT of the observations), and $\mathcal{B}$ is the support of $S_{ss}(e^{\jmath\omega})$. In the following asymptotic cases, the NP test becomes independent of $|h|^2$:

- High SNR case: if $|h|^2 \lambda_k \gg \sigma^2$ for all $k \in \mathcal{B}$, then $\mathbf{y}^H \hat{\mathbf{z}} \approx \sum_{k \in \mathcal{B}} |v_k|^2$. Thus the NP test reduces to an Energy Detector (ED) over the spectral support of the primary signal. If $\mathbf{C}$ is full rank, then $\mathbf{y}^H \hat{\mathbf{z}} \approx \mathbf{v}^H \mathbf{v} = \mathbf{y}^H \mathbf{y}$, i.e. the standard energy detector.

- Low SNR case: if $|h|^2 \lambda_k \ll \sigma^2$ for all $k \in \mathcal{B}$, then the NP test declares $\mathcal{H}_1$ true if $\sum_{k \in \mathcal{B}} \lambda_k |v_k|^2 = \mathbf{y}^H \mathbf{C} \mathbf{y}$ exceeds a threshold. This is also the *Locally Most Powerful* (LMP) test for this problem, derived from weak signal detection theory (Kay, 1998). In contrast to the ED test, it makes use of the available information about the primary signal spectrum, since $\mathbf{y}^H \mathbf{C} \mathbf{y}$ can be interpreted as the energy at the output of a filter with frequency response $S_{ss}^{1/2}(e^{\jmath\omega})$ (a *matched filter*) fed by the observations $\mathbf{y}$.

2.3.2 Detection with multiple antennas

However, with $L > 1$ antennas, neither in the high nor low SNR regimes does the dependence of the NP test with $\mathbf{h}\mathbf{h}^H$ disappear. In the following we will focus in the case of asymptotically small SNR, of interest in Cognitive Radio systems.

For asymptotically small SNR, if we make use of the first-order Taylor expansion $\mathcal{R}_1^{-1} \approx \frac{1}{\sigma^2}\mathbf{I}$, one has that the test T_{NP} is proportional to

$$T_0 \doteq \mathbf{y}^H(\mathbf{h}\mathbf{h}^H \otimes \mathbf{C})\mathbf{y} \tag{2.19}$$

$$= \sum_{i=1}^{L}\sum_{j=1}^{L} h_i h_j^* \mathbf{y}_i^H \mathbf{C}\mathbf{y}_j \tag{2.20}$$

$$= \mathbf{g}^H\mathbf{C}\mathbf{g}, \tag{2.21}$$

where we defined

$$\mathbf{g} = \mathbf{g}_{\mathrm{MRC}} \doteq \sum_{l=1}^{L} h_l^* \mathbf{y}_l. \tag{2.22}$$

Here, as in the single-antenna case, T_0 can be interpreted as the energy at the output of a matched filter, which now is fed by a linear combination $\mathbf{g}$ of the signals received at each of the antennas. We use the subscript MRC since this processing is akin to the *Maximal Ratio Combining* technique for multiantenna receivers (Simon and Alouini, 2004), by which the signals collected at each of the antennas are phased-aligned and combined with optimal weighting to maximize the SNR at the combiner output and prior to the demodulation stage. Note that the computation of the NP test statistic for low SNR does not require knowledge of the total channel gain, but only of the spherical component $\bar{\mathbf{h}} \doteq \mathbf{h}/\|\mathbf{h}\|_2$. The threshold can be set to achieve a given false alarm rate under $\mathcal{H}_0$, i.e. under $\|\mathbf{h}\|_2 = 0$.

Now if we neglect the magnitude gains of the channel coefficients in (2.22), then $\mathbf{g}$ can be approximated as

$$\mathbf{g} \approx \mathbf{g}_{\mathrm{EGC}} \doteq \sum_{l=1}^{L} e^{-\jmath\theta_l}\mathbf{y}_l \tag{2.23}$$

where $\theta_l \doteq \arg\{h_l\}$. In this case we correct the phase of the signals received at each of the antennas before the linear combination. This is analogous to the *Equal Gain Combining* (EGC) technique in diversity reception.

The resemblance with different diversity combining techniques suggests a third detector based on *Selection Combining* (SC). In this case $\mathbf{g}$ is approximated by the signal at the branch with highest SNR:

$$\mathbf{g} \approx \mathbf{g}_{\mathrm{SC}} \doteq \mathbf{y}_m \quad \text{with} \quad m = \arg\max_{1 \leq i \leq L} |h_i|^2. \tag{2.24}$$

Note that if the channel gains at all branches have similar magnitudes, then $\mathbf{g}_{\mathrm{EGC}} \approx \mathbf{g}_{\mathrm{MRC}}$. On the other hand, when one of the channel gains is much larger than the remaining ones, then $\mathbf{g}_{\mathrm{SC}} \approx \mathbf{g}_{\mathrm{MRC}}$.

However, as the reader should have noted, none of these three schemes (MRC, EGC and SC) is directly implementable, since they depend on unknown channel parameters. In order to avoid this problem, one option is to replace the unknown parameters by their corresponding estimates. Inspired by the Generalized Likelihood Ratio (GLR) approach, in Section 2.4 we present different scenarios in which the Maximum Likelihood (ML) estimates of the unknown parameters can be obtained; substituting these ML estimates in the corresponding statistics will in turn yield practical detectors. An alternative approach in order to handle the unknown parameters h_i is to disregard antenna crosscorrelation and assume equal weighting for the energy estimates at the different antennas. In this case we obtain the following detector:

2.3.3   Generalized energy detector

By disregarding in T_0 the cross terms depending on $h_i h_j^*$ for $i \neq j$ and assuming $|h_i| \approx |h_j|$ for $i \neq j$, (2.20) reduces to

$$T_{\mathrm{GED}} \;=\; \frac{1}{KL\sigma^2} \sum_{i=1}^{L} \mathbf{y}_i^H \mathbf{C} \mathbf{y}_i \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_{\mathrm{GED}}, \tag{2.25}$$

where the scaling factor $(KL\sigma^2)^{-1}$ was introduced for convenience and γ_{GED} is the decision threshold. We refer to this test as "Generalized Energy Detector" (GED), as it merely collects the (spectrally weighted) energy at all the branches.

Notice that this detector is applicable to distributed settings with L collaborating single-antenna sensors: each node reports its local statistic $\mathbf{y}_i^H \mathbf{C} \mathbf{y}_i$ (scaled by the inverse of the local noise variance, if different nodes are affected by different noise levels) to a Fusion Center, where all such statistics are added together.

The asymptotic performance of this detector is analyzed in Appendix 2.A.1, showing that for sufficiently large K and for a fixed threshold γ_{GED}, the probabilities of false alarm and detection are respectively given by

$$P_{\mathrm{FA}} = Q\left(\sqrt{KL}\frac{\gamma_{\mathrm{GED}} - 1}{\sqrt{\bar{b}_2}}\right), \tag{2.26}$$

$$P_{\mathrm{D}} = Q\left(\sqrt{KL}\frac{\gamma_{\mathrm{GED}} - (1 + \zeta\bar{b}_2)}{\sqrt{L\bar{b}_4\zeta^2 + 2\bar{b}_3\zeta + \bar{b}_2}}\right). \tag{2.27}$$

Note that the performance of the GED test depends only on the average SNR ζ, but not on the spherical component $\bar{\mathbf{h}}$ of the channel vector.

2.4 Parameter estimation and detection

In order to derive the low SNR ML estimates of the unknown parameters under the different models, we first obtain the likelihood function of the estimation problem. The log-likelihood function under $\mathcal{H}_1$ is $\log f(\mathbf{y} \,|\, \mathbf{h}) = -\log\det\mathcal{R}_1 - \mathbf{y}^H\mathcal{R}_1^{-1}\mathbf{y}$, where $\mathcal{R}_1$ depends on $\mathbf{h}$ as per (2.12). In the low SNR regime, using the fact that $\log(1 + x) \approx x$ for small $|x|$, we can approximate

$$\log\det\mathcal{R}_1 \approx KL\log\sigma^2 + \frac{\operatorname{tr}\mathcal{G}}{\sigma^2}. \tag{2.28}$$

On the other hand,

$$\mathcal{R}_1^{-1} = \frac{1}{\sigma^2}\left[\mathbf{I} + \frac{1}{\sigma^2}\mathcal{G}\right]^{-1} \approx \frac{1}{\sigma^2}\left[\mathbf{I} - \frac{1}{\sigma^2}\mathcal{G}\right]. \tag{2.29}$$

Thus, noting that $\operatorname{tr}\mathcal{G} = \operatorname{tr}\mathbf{h}\mathbf{h}^H\operatorname{tr}\mathbf{C} = \|\mathbf{h}\|_2^2 K$, for low SNR one has

$$\log f(\mathbf{y} \,|\, \mathbf{h}) \approx -KL\log\sigma^2 - \frac{K\|\mathbf{h}\|_2^2}{\sigma^2} - \frac{\|\mathbf{y}\|^2}{\sigma^2} + \frac{\mathbf{y}^H\mathcal{G}\mathbf{y}}{\sigma^4}. \tag{2.30}$$

2.4.1 Selection Combining detector

The SC detector is based on the approximation (2.24), and thus requires the estimation of the index l of the antenna with largest SNR. ML estimation of this index in the general case is difficult, and thus we resort to the low SNR approximation (2.30); in addition, we will assume that $\mathbf{h} = h\mathbf{e}_l$, where $\mathbf{e}_l$ is the l-th unit vector. The rea-

son for this is that, as mentioned above, the SC approach is expected to provide close-to-optimal performance in scenarios in which the SNR at one of the antennas is dominant.

Under this assumption, one has $\|\mathbf{h}\|_2^2 = |h|^2$ and $\mathbf{y}^H \mathcal{G} \mathbf{y} = |h|^2 \mathbf{y}_l^H \mathbf{C} \mathbf{y}_l$ in (2.30). Therefore, the ML estimate of l is just $\hat{l} = \arg\max_l \mathbf{y}_l^H \mathbf{C} \mathbf{y}_l$. The resulting decision rule is given by

$$T_{\text{SC}} \doteq \max_{1 \le l \le L} \frac{\mathbf{y}_l^H \mathbf{C} \mathbf{y}_l}{K \sigma^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_{\text{SC}}, \tag{2.31}$$

where the scaling factor $(K\sigma^2)^{-1}$ does not affect the test. Thus, the SC detector picks the antenna with largest spectrally weighted energy and uses that energy as statistic. Note that this amounts to an OR fusion rule, applicable to distributed settings: the channel is declared busy if the spectrally weighted energy at any of the L nodes exceeds a threshold. In that case, only one bit of information has to be sent to the Fusion Center by each node, in contrast with the GED scheme.

In Appendix 2.A.2 the asymptotic performance analysis of the SC detector is given. For large K and for a local threshold γ_{SC}, we obtain the global false alarm rate

$$P_{\text{FA}} = 1 - \left(1 - Q\left(\frac{\gamma_{\text{SC}} - 1}{\sqrt{\bar{b}_2/K}} \right) \right)^L. \tag{2.32}$$

On the other hand, the probability of detection cannot be expressed in closed form, although it can be straightforwardly computed by means of a multivariate Gaussian integration routine; see Appendix 2.A.2. It must be noted that, in contrast with GED, the performance of the SC detector does depend on the spherical component of the channel vector.

2.4.2   Equal Gain Combining detector

For EGC detection, an estimate of the phases $\{\theta_i\}_{i=1}^L$ introduced at the different branches is needed in order to combine the respective signals as per (2.23). Considering again the low SNR approximation (2.30), it is seen that in order to obtain the ML estimates we must maximize the following quantity w.r.t. $\theta_1, \ldots, \theta_M$:

$$\mathbf{y}^H \mathcal{G} \mathbf{y} = \sum_{n=1}^L \sum_{m=1}^L |h_n||h_m| \mathbf{y}_n^H \mathbf{C} \mathbf{y}_m \mathrm{e}^{-\jmath(\theta_m - \theta_n)}. \tag{2.33}$$

Let $a_{nm} \doteq |h_n||h_m|\mathbf{y}_n^H \mathbf{C} \mathbf{y}_m$. Since $a_{nm} = a_{mn}^*$, it is clear that

$$
\begin{aligned}
\mathbf{y}^H \mathcal{G} \mathbf{y} &= \sum_{m=1}^{L} a_{mm} + 2 \sum_{n=1}^{L} \sum_{m=n+1}^{L} \Re\{a_{nm} \mathrm{e}^{-\jmath(\theta_m - \theta_n)}\} \\
&\leq \sum_{m=1}^{L} a_{mm} + 2 \sum_{n=1}^{L} \sum_{m=n+1}^{L} |a_{nm}|,
\end{aligned}
\tag{2.34}
$$

with equality iff $\theta_m - \theta_n = \arg\{a_{nm}\}$ for all (n, m) such that $m > n$. These constitute a set of $L(L-1)/2$ (linear) conditions on our L free parameters, which in general cannot be satisfied if $L > 3$. Nevertheless, careful inspection of the resulting detection statistic $\mathbf{g}^H \mathbf{C} \mathbf{g}$ with $\mathbf{g} = \sum_{l=1}^{L} \mathrm{e}^{-\jmath \hat{\theta}_l} \mathbf{y}_l$ reveals that it is a function of the phase differences $\hat{\theta}_{mn} \doteq \hat{\theta}_m - \hat{\theta}_n$ only. Thus, if we take these phase differences as our free optimization variables and neglect the dependence among them, the corresponding ML estimates are $\hat{\theta}_{mn} = \arg\{\mathbf{y}_n^H \mathbf{C} \mathbf{y}_m\}$. This yields the following EGC detection rule:

$$
T_{\mathrm{EGC}} \doteq \frac{1}{KL\sigma^2} \sum_{n=1}^{L} \sum_{m=1}^{L} |\mathbf{y}_m^H \mathbf{C} \mathbf{y}_n| \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_{\mathrm{EGC}},
\tag{2.35}
$$

which is intuitively satisfying: the lack of knowledge about the phase differences is sidestepped by considering the *modulus* of the crosscorrelation terms.

Unfortunately, finding the distribution of T_{EGC} (under either hypothesis) is intractable. An asymptotic Gaussian approximation is used in Appendix 2.A.3, showing that for large K for a given threshold γ_{EGC},

$$
P_{\mathrm{FA}} \approx Q\left(\sqrt{KL} \frac{\gamma_{\mathrm{EGC}} - \left(1 + \frac{L-1}{2}\sqrt{\frac{\pi \bar{b}_2}{K}}\right)}{\sqrt{\left(2L - 1 + (1-L)\frac{\pi}{2}\right)\bar{b}_2}}\right),
\tag{2.36}
$$

$$
P_{\mathrm{D}} \approx Q\left(\sqrt{K} \frac{\gamma_{\mathrm{EGC}} - (1 + \kappa\zeta\bar{b}_2)}{\sqrt{\bar{b}_4(\kappa\zeta)^2 + 2\bar{b}_3\kappa\zeta + \bar{b}_2}}\right)
\tag{2.37}
$$

where $\kappa \doteq \|\mathbf{h}\|_1^2 / \|\mathbf{h}\|_2^2 = \|\bar{\mathbf{h}}\|_1^2$. Note that (2.37) is a function of the *scaled* average SNR per antenna $\kappa\zeta$, and that the scaling term $\kappa \in [1, L]$ achieves its maximum value when all elements of $\mathbf{h}$ have the same magnitude. Note that it is precisely in such scenarios that one expects the EGC detector to perform best.

2.4.3   Maximal Ratio Combining detector

For MRC detection, an estimate of the spherical component of the channel vector $\bar{\mathbf{h}} = \mathbf{h}/\|\mathbf{h}\|_2$ is needed. Let us introduce the data matrix $\mathbf{Y} \doteq [\ \mathbf{y}_1\ \ \cdots\ \ \mathbf{y}_M\ ]$. Focusing again on the low SNR approximation (2.30), the ML estimate of $\bar{\boldsymbol{h}}$ must maximize

$$
\begin{aligned}
\mathbf{y}^H \mathcal{G} \mathbf{y} &= \mathbf{y}^H (\mathbf{h}\mathbf{h}^H \otimes \mathbf{C}) \mathbf{y} \\
&= \mathbf{h}^H (\mathbf{Y}^H \mathbf{C} \mathbf{Y})^* \mathbf{h} \\
&= \|\mathbf{h}\|_2^2 \bar{\mathbf{h}}^H (\mathbf{Y}^H \mathbf{C} \mathbf{Y})^* \bar{\mathbf{h}}.
\end{aligned}
\tag{2.38}
$$

This is achieved when $\bar{\boldsymbol{h}}$ is the unit-norm eigenvector of $(\mathbf{Y}^H \mathbf{C} \mathbf{Y})^*$ associated to its largest eigenvalue (up to a phase term $\mathrm{e}^{j\phi}$ which does not affect the test). This results in the following scaled MRC detection rule:

$$
T_{\mathrm{MRC}} \doteq \frac{\lambda_{\max}(\mathbf{Y}^H \mathbf{C} \mathbf{Y})}{K\sigma^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_{\mathrm{MRC}}.
\tag{2.39}
$$

Note that neither T_{EGC} nor T_{MRC} lend themselves to distributed implementation, since they require the computation of (spectrally weighted) crosscorrelations across the different antennas.

The statistical analysis of the MRC detector amounts to finding the distribution of the largest eigenvalue of the random matrix $\mathbf{Y}^H \mathbf{C} \mathbf{Y}$ under each hypothesis. For a general covariance matrix $\mathbf{C}$, this remains an open problem. In Appendix 2.A.4 we present the analysis for a special case of practical interest: strictly bandlimited primary signals using a fraction B of the total channel bandwidth, and with flat psd within their passband. In this case, the distribution of T_{MRC} under $\mathcal{H}_0$ asymptotically follows a (shifted and scaled) Tracy-Widom distribution, which can be used to set the threshold γ_{MRC} for a given probability of false alarm. For fixed threshold γ_{MRC} the asymptotic probability of detection is given by

$$
P_{\mathrm{D}}^{\mathrm{EGC}} \approx Q \left(\frac{\gamma_{\mathrm{MRC}} - \left(\delta_1 + \frac{L\delta_1}{K(\delta_1 - 1)} \right)}{\delta_1/\sqrt{K}} \right)
\tag{2.40}
$$

where $\delta_1 \doteq 1 + \bar{b}_2 L \zeta$.

2.5   Detection diversity in fading environments

In the analysis of the previous section we considered that the SNR at each antenna is fixed. In practical conditions this is unlikely to occur. Consider a slow fading scenario in which the channel gains remain constant during the sensing window. Then $\mathbf{h}$ becomes a random variable and for a fixed threshold, the probability of detection P_D is a random variable with expected value given by

$$\bar{P}_\mathrm{D}(\bar{\zeta}) \doteq E_\zeta\{P_\mathrm{D}\} = \int_0^\infty f_\zeta(\zeta)P_\mathrm{D}(\zeta)\,\mathrm{d}\zeta, \qquad (2.41)$$

with $f_\zeta(\zeta)$ the probability density function (pdf) of ζ, and $\bar{\zeta} \doteq E_\zeta\{\zeta\}$ the mean value of the SNR.

In the following we will assume that $\mathbf{h}$ can be modeled with a Ricean distribution (Simon and Alouini, 2004). This accounts for the line-of-sight (LOS) component and for the non-line-of-sight (NLOS) scattering. Hence at each realization we can model the channel vector as

$$\mathbf{h} = \sqrt{\bar{\zeta}\sigma^2}\left(\sqrt{\frac{\eta}{1+\eta}}\bar{\mathbf{h}} + \sqrt{\frac{1}{1+\eta}}\tilde{\mathbf{h}}\right), \qquad (2.42)$$

where $\bar{\zeta}$ denotes the average SNR and η stands for the Rice factor, the LOS channel component $\bar{\mathbf{h}}$ is defined as $\bar{\mathbf{h}} \doteq [\mathrm{e}^{\jmath\theta}\ \mathrm{e}^{\jmath 2\theta}\cdots\mathrm{e}^{\jmath L\theta}]$ with $\theta \sim \mathcal{U}(0,2\pi)$ modeling the relative phase of the antennas of a uniform linear array, $\tilde{\mathbf{h}}$ is a zero-mean complex Gaussian vector modeling the NLOS channel component with iid components $\sim \mathcal{CN}(0,1)$, and independent of θ.

The worst-case scenario in terms of detection performance is given by the NLOS channel, i.e. $\eta = 0$ with pure Rayleigh fading. We will see next that in this scenario the probability of misdetection (asymptotically) decreases only linearly with the SNR (in log scale) with a slope that is given by the asymptotic detection diversity of system. However, this asymptotic measure does not reflect the true detection performance of a detector in the SNR range of interest. To overcome this problem we will also present an analysis based on the diversity measure first proposed by Daher and Adve in the context of radar, related to the probability of detection around the SNR at which $\bar{P}_D = 1/2$.

2.5.1   High SNR diversity order analysis

We restrict here our analysis to the worst case given by iid Rayleigh fading scenarios[2], i.e. (2.42) with $\eta = 0$. In this case, the instantaneous SNR $\zeta = \|h\|^2/(L\sigma^2)$ is gamma-distributed (Simon and Alouini, 2004):

$$f_\zeta(\zeta) = \frac{L^L}{(L-1)!}\frac{\zeta^{L-1}}{\bar{\zeta}^L}\exp\left\{-L\zeta/\bar{\zeta}\right\},\;\; \zeta > 0. \tag{2.43}$$

We will next present the analysis for the GED. This analysis can be extrapolated with minor changes to the MRC detector. First, from (2.27), the probability of misdetection of the GED for a fixed threshold γ_{GED} is given by

$$P_{\mathrm{MD}} \doteq 1 - P_{\mathrm{D}} \tag{2.44}$$

$$= Q\left(\frac{(1 + \zeta\bar{b}_2) - \gamma_{\mathrm{GED}}}{\sqrt{\frac{L\bar{b}_4\zeta^2 + 2\bar{b}_3\zeta + \bar{b}_2}{KL}}}\right). \tag{2.45}$$

Using a first-order Taylor approximation of the argument of the Q-function in (2.44) about $\zeta = 0$, one finds that in the low SNR regime,

$$P_{\mathrm{MD}} \approx Q\left(\frac{\left[\bar{b}_2 + \frac{\bar{b}_3}{\bar{b}_2}(1 - \gamma_{\mathrm{GED}})\right]}{\sqrt{\bar{b}_2/KL}}\zeta + \frac{(1 - \gamma_{\mathrm{GED}})}{\sqrt{\bar{b}_2/(KL)}}\right). \tag{2.46}$$

In a fading environment, P_{MD} in (2.46) becomes a random variable. Its mean value can be upper bounded by noting that

$$Q(x) \le \frac{1}{2}\mathrm{e}^{-x/2}, \quad x \ge 0. \tag{2.47}$$

Note that this bound is looser than the more usual expression $Q(x) \le \mathrm{e}^{-x^2/2}/2$. However, it will allow us to obtain a tight upper bound on the asymptotic slope of the probability of misdetection. By using this bound in (2.46) and then averaging

---

[2]The analysis can be readily extended to Rayleigh fading with a certain correlation matrix $\boldsymbol{\Upsilon}$. The resulting asymptotic diversity will depend on the rank of $\boldsymbol{\Upsilon}$. We refer the interested reader to López-Valcarce et al. (2009).

over $\mathbf{h}$, the average probability of detection $\bar{P}_{\mathrm{MD}} \doteq \mathrm{E}[P_{\mathrm{MD}}]$ can be bounded as

$$\bar{P}_{\mathrm{MD}} \leq \frac{1}{2} \exp\left\{\frac{1 - \gamma_{\mathrm{GED}}}{2\sqrt{\bar{b}_2/(KL)}}\right\}$$
$$\times \mathrm{E}\left[\exp\left\{-\zeta\left(\frac{\bar{b}_2 + \frac{\bar{b}_3}{\bar{b}_2}(1 - \gamma_{\mathrm{GED}})}{2\sqrt{\bar{b}_2/(KL)}}\right)\right\}\right]. \tag{2.48}$$

Using similar steps to those in (Larsson and Stoica, 2003, Sec. 4.4), one finds that

$$\bar{P}_{\mathrm{MD}} \leq C_L \bar{\zeta}^{-L}, \tag{2.49}$$

where

$$C_L \doteq \frac{1}{2} \exp\left\{\frac{1 - \gamma_{\mathrm{GED}}}{2\sqrt{\bar{b}_2/(KL)}}\right\} \left(\frac{2\sqrt{\bar{b}_2 L/K}}{\bar{b}_2 + \frac{\bar{b}_3}{\bar{b}_2}(1 - \gamma_{\mathrm{GED}})}\right)^L \tag{2.50}$$

is a constant independent of the average SNR. Hence the *diversity order,* that is, the slope of $\bar{P}_{\mathrm{MD}}$ versus the SNR when plotted on a log-log scale, is upper bounded by the number of antennas L of the receiver system. This shows the advantage of having multiple antennas for channel sensing under fading conditions even when considering the simple GED detector. By carrying a similar analysis, the MRC detector can be shown to present the same asymptotic diversity order in Rayleigh fading environments.

Note that this analysis cannot be applied to EGC and SC detectors, since their performance depends on the actual SNR at each of the antennas and not on the global instantaneous SNR. If we define

$$\boldsymbol{\zeta} \doteq \frac{1}{\sigma^2}[\,|h_1|^2 \ |h_2|^2 \ \cdots \ |h_M|^2\,]^T, \tag{2.51}$$

so that $\boldsymbol{\zeta} = [\zeta_1 \ \cdots \zeta_M]^T$ is the vector with the instantaneous SNR at each of the antennas, an approximate analysis of the SC detector follows. The average probability of misdetection is in this case given by

$$\bar{P}_{\mathrm{MD}} = \int f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) P_{\mathrm{MD}}(\boldsymbol{\zeta}) \partial\boldsymbol{\zeta} \tag{2.52}$$

where $f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta})$ factorizes given the independent Rayleigh fading assumption. On the other hand, as it was shown in Appendix 2.A.2, $P_{\mathrm{MD}}(\boldsymbol{\zeta})$ does not factorize due to the present signal component. However, in low SNR conditions the correlation present

will be small and we may approximate

$$\bar{P}_{\mathrm{MD}} \approx \prod_{l=1}^{L} \left(\int_0^\infty f_{\zeta_l}(\zeta_l) P_{\mathrm{MD}}^{(l)}(\zeta_l) \partial \zeta_l \right) \tag{2.53}$$

$$= \left(\int_0^\infty f_{\zeta_1}(\zeta_1) P_{\mathrm{MD}}^{(1)}(\zeta_1) \partial \zeta_1 \right)^L \doteq \tilde{P}_{\mathrm{MD}} \tag{2.54}$$

where $P_{\mathrm{MD}}^{(l)}(\zeta_l)$ corresponds to the probability of misdetection of a single-antenna system with the instantaneous SNR given by ζ_l and in (2.54) we used the symmetry between antennas.

Now, the misdetection probability at each of the antennas can be upper bounded using (2.49) with $L = 1$. Then one obtains the approximation

$$\bar{P}_{\mathrm{MD}} \approx \tilde{P}_{\mathrm{MD}} \leq C_1^L \bar{\zeta}^{-L}, \tag{2.55}$$

and, as a result, GED, MRC, and SC detectors cannot have a diversity order larger than L. In fact, as we will numerically show in Section 2.6 the diversity order bound is tight in the high SNR regime for the proposed detectors. Hence the four of them achieve full asymptotic diversity in uncorrelated Rayleigh fading.

2.5.2   Daher-Adve diversity order analysis

The asymptotic diversity order analysis presented in the previous section is a high-SNR concept. However, spectrum sensors for CR systems are expected to provide high detection performance at much lower SNR values. This calls for a different definition of the diversity order better suited to the detection problem. In the context of radar processing, Daher and Adve (2010) define diversity order as the slope of the average probability of detection ($\bar{P}_{\mathrm{D}}$) curve with respect to the SNR at $\bar{P}_{\mathrm{D}} = 0.5$. This notion of diversity is more adequate to CR networks because (i) it indicates a *minimum operational SNR* from which a detection scheme starts working reasonably well (i.e. $\bar{P}_{\mathrm{D}} \geq 0.5$) and (ii) describes how fast $\bar{P}_{\mathrm{D}}$ approaches 1 from this *minimum operational SNR*.

In this section, we characterize different spectrum sensing schemes in independent Rayleigh fading in terms of the Daher-Adve diversity order. As opposed to Daher and Adve (2010), in which the steering vectors are assumed known and a single snapshot is used per sensor for detection, when sensing on wireless channels,

the channel is not known and sensing times are longer in order to acquire several signal samples.

Let the *minimum operational SNR* $\bar{\zeta}^*$ of the detector be defined by $\bar{P}_{\mathrm{D}}(\bar{\zeta}^*) = 0.5$. Following Daher and Adve (2010), the diversity order d is defined as

$$d \doteq \left. \frac{\partial \bar{P}_{\mathrm{D}}(\bar{\zeta})}{\partial \bar{\zeta}} \right|_{\bar{\zeta}=\bar{\zeta}^*}, \qquad \text{with} \quad \bar{P}_{\mathrm{D}}(\bar{\zeta}^*) = \frac{1}{2}. \tag{2.56}$$

Unfortunately, $\bar{P}_{\mathrm{MD}} \doteq \mathrm{E}[P_{\mathrm{MD}}]$ does not admit a closed-form solution for any of the detectors presented. In the high SNR asymptotic analysis we resorted to an upper bound of P_{MD} which allowed us to obtain the analytical integral which is tight for high SNR. However, it is not possible to use a similar approach here since we are interested in intermediate SNRs. In order to obtain an approximation of the diversity order we propose the following first-order piecewise approximation of $P_{\mathrm{MD}}(\zeta)$, where ζ^* is such that $P_{\mathrm{MD}}(\zeta^*) = 0.5$:

$$P_{\mathrm{MD}}(\zeta) \approx \begin{cases} 1, & 0 < \zeta < \zeta_1, \\ \frac{1}{2} - a(\zeta - \zeta^*), & \zeta_1 < \zeta < \zeta_2, \\ 0, & \zeta > \zeta_2, \end{cases} \tag{2.57}$$

where $\zeta_1 \doteq \zeta^* - \frac{1}{2a}$, $\zeta_2 \doteq \zeta^* + \frac{1}{2a}$ and a is the negative of the slope of $P_{MD}(\zeta)$ at $\zeta = \zeta^*$, i.e.

$$a \doteq - \left. \frac{\partial P_{\mathrm{MD}}(\zeta)}{\partial \zeta} \right|_{\zeta=\zeta^*} = \left. \frac{\partial P_{\mathrm{D}}(\zeta)}{\partial \zeta} \right|_{\zeta=\zeta^*}. \tag{2.58}$$

Using (2.43) and (2.57), one obtains

$$\bar{P}_{\mathrm{MD}} = \int_0^\infty f_\zeta(\zeta) P_{\mathrm{MD}}(\zeta) \, \partial\zeta \tag{2.59}$$

$$\approx a \left\{ \zeta_2 \Gamma\left(\frac{L\zeta_2}{\bar{\zeta}}, L \right) - \zeta_1 \Gamma\left(\frac{L\zeta_1}{\bar{\zeta}}, L \right) \right.$$
$$\left. - \bar{\zeta} \left[\Gamma\left(\frac{L\zeta_2}{\bar{\zeta}}, L+1 \right) - \Gamma\left(\frac{L\zeta_1}{\bar{\zeta}}, L+1 \right) \right] \right\} \tag{2.60}$$

where the incomplete Gamma function is defined as

$$\Gamma(x, \alpha) \doteq \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} \mathrm{e}^{-t} \partial t, \tag{2.61}$$

with $\Gamma(\alpha) \doteq \int_0^\infty t^{\alpha-1} \mathrm{e}^{-t} \partial t$ denoting the standard Gamma function.

Taking derivatives in (2.59), and after some algebra, one arrives at the following approximation for the Daher-Adve diversity order

$$d \approx a \left[g_L \left(\frac{\zeta^*}{\bar{\zeta}^*} + \frac{1}{2a\bar{\zeta}^*} \right) - g_L \left(\frac{\zeta^*}{\bar{\zeta}^*} - \frac{1}{2a\bar{\zeta}^*} \right) \right], \tag{2.62}$$

where $g_L(x) \doteq \Gamma(Lx, L+1)$. While (2.62) may look like a *rough* approximation of the diversity order, we will show in Section 2.6 that it effectively captures the behavior of $\bar{P}_{\mathrm{D}}$ in Rayleigh fading environments. We proceed now to compute the parameters ζ^* and a for the different detection schemes.

GED detector performance

Using the asymptotic distribution (2.27), one readily obtains the parameters ζ^* and a for the GED detector:

$$\zeta^*_{\mathrm{GED}} = \frac{1}{\bar{b}_2} \left(\gamma_{\mathrm{GED}} - 1 \right), \tag{2.63}$$

$$a_{\mathrm{GED}} = \sqrt{\frac{KL\bar{b}_2}{2\pi}} \left(L(\gamma_{\mathrm{GED}} - 1)^2 \frac{\bar{b}_4}{\bar{b}_2^3} + 2(\gamma_{\mathrm{GED}} - 1)\frac{\bar{b}_3}{\bar{b}_2^2} + 1 \right)^{-1/2}, \tag{2.64}$$

where we used that the derivative of the Q-function is given by

$$\frac{\partial Q(x)}{\partial x} = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2\right). \tag{2.65}$$

Now, finding the value of $\bar{\zeta}^*$ at which (2.59) equals 0.5 is not straightforward. However, an obvious candidate is $\bar{\zeta}^* \approx \zeta^*_{\mathrm{GED}}$, since the *instantaneous* probability of misdetection satisfies $P_{\mathrm{MD}}(\zeta^*_{\mathrm{GED}}) = 0.5$. With $\epsilon_{\mathrm{GED}} \doteq \frac{1}{2a_{\mathrm{GED}}\zeta^*_{\mathrm{GED}}}$, this yields

$$d_{\mathrm{GED}} \approx a_{\mathrm{GED}} \left[g_L \left(1 + \epsilon_{\mathrm{GED}} \right) - g_L \left(1 - \epsilon_{\mathrm{GED}} \right) \right], \tag{2.66}$$

where both a_{GED} and ζ^*_{GED} depend on the system parameters K, L, $\bar{b}_i$ and P_{FA}. Noting that the bracketed term in (2.66) is less than 1, the following upper bound is obtained:

$$d_{\mathrm{GED}} < a_{\mathrm{GED}} < \sqrt{\frac{KL\bar{b}_2}{2\pi}}. \tag{2.67}$$

As $L \to \infty$ and for any $\epsilon > 0$, we have that[3] $g_L(1+\epsilon) \to 1$ whereas $g_L(1-\epsilon) \to 0$.

---

[3]Intermediate steps can be found in Appendix 2.B.

Thus, for large L, $d_{\text{GED}} \approx a_{\text{GED}}$.

Remark 2.1. From (2.67), Daher-Adve diversity order of the GED detector under uncorrelated Rayleigh fading is asymptotically bounded by $\mathcal{O}\left(\sqrt{KL\bar{b}_2}\right)$. Moreover, for small values of $\bar{\zeta}^*$ this bound becomes asymptotically tight, i.e. for a small *minimum operational SNR*, d_{GED} grows with the square root of the number of antennas L multiplied by the parameter $\bar{b}_2 = \text{tr}\{\mathbf{C}^2\}/K$. Since $\bar{b}_2 \geq 1$ with equality for temporally white primary signals, we have that Daher-Adve diversity increases with the temporal correlation of the primary signals. Moreover, for larger values of $\bar{b}_2$ we attain a lower *minimum operational SNR* (2.63), hence increasing the operational range of the detector.

MRC detection performance

In this case, the parameters for the first-order piecewise approximation of (2.40) are

$$\zeta^*_{\text{MRC}} = \frac{1}{2\bar{b}_2 L}\left(\beta + \sqrt{(2+\beta)^2 - 4\gamma_{\text{MRC}}}\right), \tag{2.68}$$

$$a_{\text{MRC}} = \sqrt{\frac{KL^2\bar{b}_2^2}{2\pi}} \frac{1 - \frac{L}{K(\bar{b}_2 L\zeta)^2}}{1 + \bar{b}_2 L\zeta}, \tag{2.69}$$

where $\beta = \gamma_{\text{MRC}} - \frac{K+L}{K}$, so that

$$d_{\text{MRC}} < a_{\text{MRC}} < \sqrt{\frac{KL^2\bar{b}_2^2}{2\pi}}, \tag{2.70}$$

with $d_{\text{MRC}} \to a_{\text{MRC}}$ as $L \to \infty$.

Remark 2.2. Note however that for small values of $\bar{\zeta}^*$, that is, detectors capable of working in harsh SNR conditions, the Daher-Adve diversity of the MRC increases as $\mathcal{O}(\sqrt{KL^2\bar{b}_2^2})$ while the GED only achieves $\mathcal{O}(\sqrt{KL\bar{b}_2})$, i.e. Daher-Adve diversity does reflect the improved performance of MRC over GED.

SC detection performance

In order to compute the Daher-Adve diversity order for the SC detector, we neglect again the correlation among different antennas under $\mathcal{H}_1$. Then

$$\bar{P}_{\text{MD}} = \int f_{\boldsymbol{\zeta}}(\boldsymbol{\zeta}) P_{\text{MD}}(\boldsymbol{\zeta}) \partial\boldsymbol{\zeta} \approx \left[\int_0^\infty f_\zeta(\zeta) P_{\text{MD}}^{(l)}(\zeta) \partial\zeta\right]^L.$$

We now approximate the integral using the same technique as in the previous points. Using the results obtained for the GED, particularized for $L = 1$, after some algebra, one arrives at

$$\bar{P}_{\mathrm{MD}}(\bar{\zeta}) \approx \left[1 - 2a_{\mathrm{SC}}\bar{\zeta} \sinh\left(\frac{1}{2a_{\mathrm{SC}}\bar{\zeta}}\right) \mathrm{e}^{-\zeta_{\mathrm{SC}}^*/\bar{\zeta}} \right]^L, \tag{2.71}$$

where

$$\zeta_{\mathrm{SC}}^* = \frac{1}{\bar{b}_2}\left(\gamma_{\mathrm{SC}} - 1\right), \tag{2.72}$$

$$a_{\mathrm{SC}} = \sqrt{\frac{K\bar{b}_2}{2\pi}}\left((\gamma_{\mathrm{SC}} - 1)^2\frac{\bar{b}_4}{\bar{b}_2^3} + 2(\gamma_{\mathrm{SC}} - 1)\frac{\bar{b}_3}{\bar{b}_2^2} + 1\right)^{-1/2}. \tag{2.73}$$

Taking the derivative of (2.71), one finds that

$$d_{\mathrm{SC}} \approx \frac{L(\sqrt[L]{2} - 1)}{2a_{\mathrm{SC}}\bar{\zeta}^*}\left[\frac{\zeta_{\mathrm{SC}}^* - \frac{1}{2a_{\mathrm{SC}}}}{\bar{\zeta}^*} + 1\right]. \tag{2.74}$$

One must solve for $\bar{\zeta}^*$ in $\bar{P}_{\mathrm{MD}}(\bar{\zeta}^*) = \frac{1}{2}$ in (2.71), i.e.,

$$1 - \frac{1}{\sqrt[L]{2}} = 2a_{\mathrm{SC}}\bar{\zeta}^* \sinh\left(\frac{1}{2a_{\mathrm{SC}}\bar{\zeta}^*}\right) \mathrm{e}^{-\zeta_{\mathrm{SC}}^*/\bar{\zeta}^*}, \tag{2.75}$$

which can be solved numerically.

Remark 2.3. While we are not able here to obtain a closed-form expression for the Daher-Adve diversity order of the SC detector we conjecture that it grows as $\mathcal{O}(\sqrt{K}\log(L\bar{b}_2))$, similarly to the result obtained in Daher and Adve (2010) for an OR based detector. In fact, in the next section we numerically show that the diversity order of the SC detector is smaller than that of the GED, i.e. $d_{\mathrm{SC}} < \mathcal{O}(\sqrt{KL\bar{b}_2})$.

2.6   Numerical results and discussion

In this section we examine the performance of the proposed detectors via both Monte Carlo simulations and analytical results. The detectors considered are the Generalized Energy Detector (*GED*), the Selection Combining based detector (*SC*), the Equal Gain Combining detector (*EGC*) and the asymptotic GLRT given by the Maximal Ratio Combining detector (*MRC*).
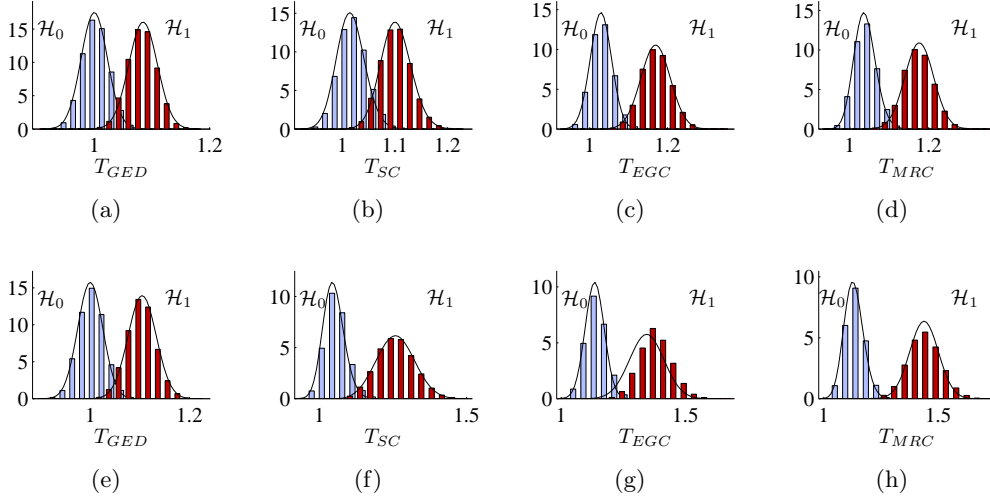
Figure 2.1: Theoretical versus empirical distributions for a DVB-T OFDM signal with $L = 2$: (a)-(d), and for a square root raised cosine signal with rolloff factor $1/2$ and $L = 4$: (e)-(h).

Analytical statistic distributions.

In order to validate the approximations used in the derivation of the analytical results for deterministic channels, we show in Fig. 2.1 the theoretical and the empirical distributions for a set of different scenarios. The first series of plots, namely Figs. 2.1 (a)-(d), were obtained using a primary Digital Video Broadcasting Terrestrial television (DVB-T) signal[4] (ETSI, 2004) with bandwidth $B = 7.61$ MHz quantized to 9-bit precision. The DVB-T physical layer employs orthogonal frequency-division multiplexing (OFDM) with equal power in each of the data subcarriers. Hence its spectrum is bandpass flat. This channel was downshifted to baseband and asynchronously sampled at $f_s = 8$ MHz, thus the multicarrier signal occupying a bandwidth fraction of $\sim 97\%$. The receiver acquires $K = 1024$ samples at each of the $L = 2$ antennas with same instantaneous per antenna SNR equal to -11 dB. From Fig. 2.1 (a)-(d) the good match between analytical and empirical distributions for the four detectors is apparent. Secondly, Fig. 2.1 (e)-(h) have been generated using baseband-equivalent single-carrier 16-QAM primary signals with square-root raised cosine pulse shaping with rolloff factor of $1/2$, sampled at the Nyquist rate (1.5 times the baud rate in this example). We assume here $K = 512$, $L = 4$ and instantaneous SNR at each of the antennas equal to -7, -12, -14 and -18 dB respectively. We

---

[4]8K mode, 64-QAM, guard interval 1/4, inner code rate 2/3.
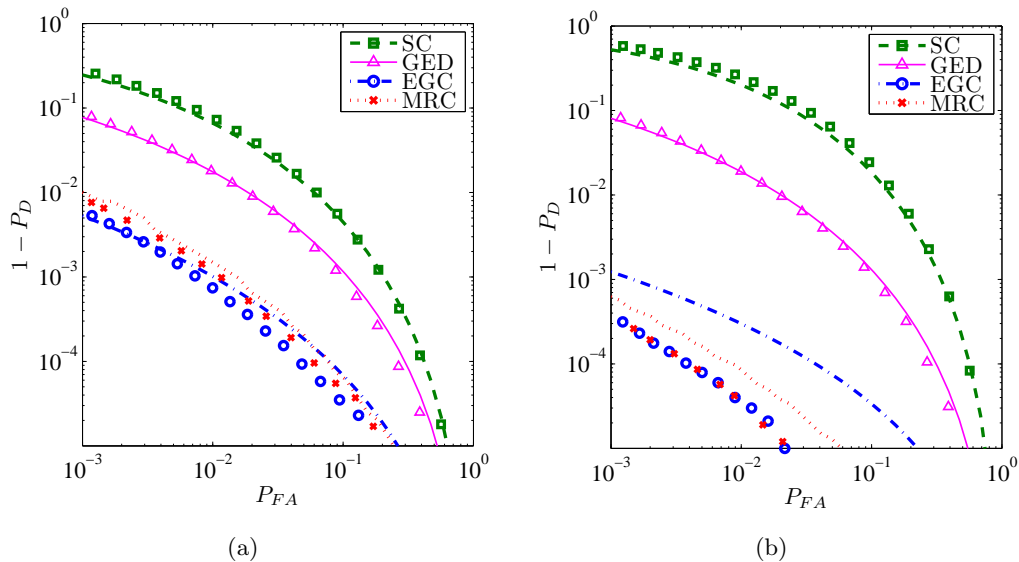
Figure 2.2: Detection ROC curves of the detectors with (a) $L = 2$ and (b) $L = 4$ antennas assuming the same instantaneous per antenna SNR. Lines represent analytical results while markers show simulation results.

can see that under the Hypothesis $\mathcal{H}_0$, the analytical distributions closely match the simulation results for the four detectors. On the other hand, under $\mathcal{H}_1$ we observe a deviation between analytical and theoretical results for the *EGC* and for the *MRC*. The analytical distribution of the *EGC* statistic was obtained under the assumption that the SNR of each of the individual antennas was sufficiently large under $\mathcal{H}_1$. However, in the simulated setup for $K = 512$ the weakest antenna's SNR (-18 dB) is not sufficiently large in order for this approximation to hold. On the other hand, the analytical distribution of the *MRC* was derived under the flat bandpass signal assumption. Here, in the analytical representation we assumed a flat spectrum with frequency support corresponding to the points of the square root raised cosine mask larger than $1/2$, hence the deviation observed.

Effect of the number of antennas.

In order to show the effect of the number of antennas on the proposed detectors, we present here the Receiver Operation Characteristic (ROC) curves of a system with $L = 2$ antennas compared to a system with $L = 4$ for a constant total number of samples $LK = 2^{11}$. The primary signal is assumed to be a flat OFDM signal occupying a bandwidth fraction of $\sim 97\%$ with parameters equal to those in the previous
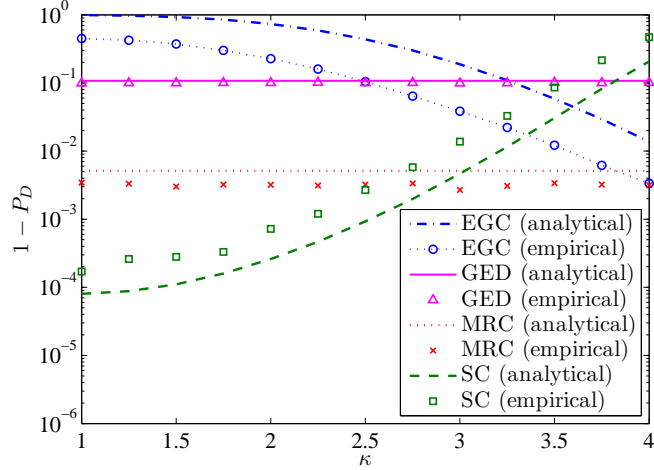
Figure 2.3: Detection performance versus the SNR spread factor κ for $P_{FA} = 0.05$, $L = 4$, $K = 512$ and average SNR equal to -5 dB.

experiment. Figure 2.2 shows the detection performance of the different schemes when the SNR at each of the antennas is fixed to -10 dB. By comparing Fig. 2.2(a) and Fig. 2.2(b) we first note that the performance of the GED is independent of the number of antennas, since it depends only on the received average SNR. For equal SNR at each antenna and equal total number of samples LK MRC and EGC detectors increase their performance with the number of antennas, whereas that of the SC detector worsens. This effect comes from the fact that all the antennas present similar SNR and the number of samples acquired from one antenna decreases with the number of antennas as $1/L$. Since the SC detector uses the information of the best antenna only, as the number of antennas increases its performance decreases. On the other hand, it is interesting to note that for equal per antenna SNR the EGC detector may perform better than the MRC detector. This is apparent from the fact that the EGC detector is derived under the assumption of equal power at each antenna; hence, the EGC detector is exploiting additional *a priori* information in this scenario. Finally, note that for $L = 4$ antennas we observe a mismatch between theoretical and empirical results for the MRC and EGC detectors, especially for the latter. Given the reduced number of samples ($K = 512$), and the order of the probability of misdetection, the asymptotic approximations are not accurate in this setting.

Effect of the instantaneous per antenna SNR.

We now compare the proposed detectors for different instantaneous channels. Figure 2.3 presents the simulation results for different values of $\kappa = \|\bar{\mathbf{h}}\|_1^2$, which is equal to 1 when all the SNR is concentrated at one antenna, while it is equal to L when it is spread over all the antennas. Here we assume that all the antennas have the same SNR, except one of them, which may present a higher SNR. The remaining system parameters are $P_{FA} = 0.05$, $L = 4$, $K = 512$ and average instantaneous SNR equal to -5 dB for primary OFDM signals.

From Fig. 2.3 the advantage of exploiting the available *a priori* information becomes apparent. When the received energy is concentrated in one antenna, the *SC* detector, designed under this model, outperforms the remaining detectors. On the other hand, when the energy is spread across the antennas *EGC* detector shows good performance, even better than the obtained by the *MRC*. It is interesting to note that both *GED* and *MRC* performances depend on the total received SNR and not on how this SNR is spread over the antennas. In this sense the MRC detector offers good performance in the whole range of κ, and thus should be robust to unknown spatial fading. We also note that the asymptotic analytical results, while inaccurate at some points given the small number of samples, show the right global behavior.

Detection in fading environments.

In the previous points we considered that the SNR at each antenna is fixed for the whole experiment. Figure 2.4(a) presents the simulation results for Ricean and Rayleigh fading environments versus the average SNR $\bar{\zeta}$ when $P_{FA} = 0.05$, $L = 4$, $K = 256$ and primary signals are bandpass OFDM. We can see that in Rayleigh environments ($\eta = 0$) the *MRC* detector outperforms other detectors. The reason is that *MRC* performance is robust to different channel realizations which may occur in fading environments (see Fig. 2.3). Interestingly, the *EGC* detector presents only a small degradation in fading environments, and, in fact, when the Rice factor grows it may outperform the *MRC* detector, as it could be expected from the fading model considered in (2.42). As the Rice factor increases we enter in the LOS regime and the SNR received at each antenna is similar. This justifies the use of this detector in practical settings, in which it may be not feasible to compute the largest eigenvalue due to complexity restrictions.
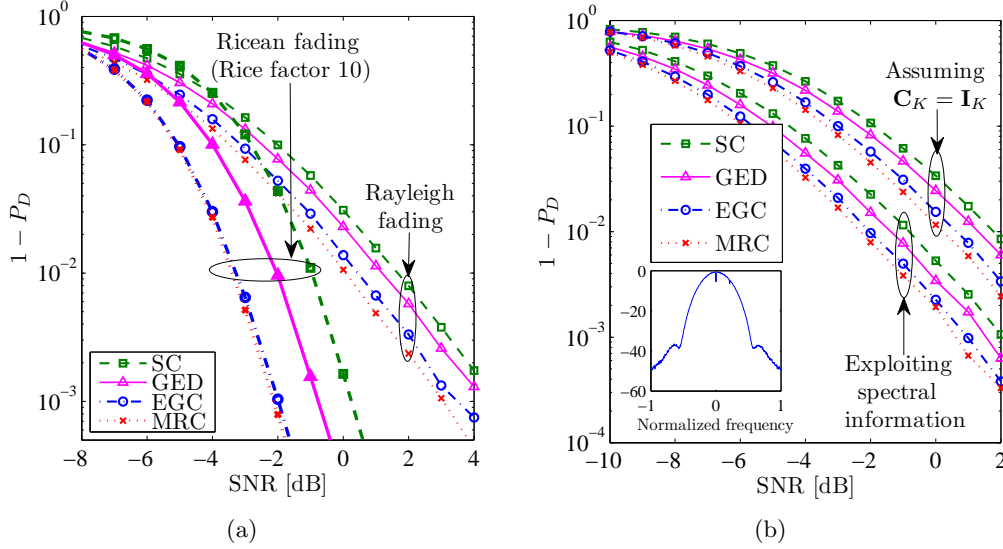
Figure 2.4: Detection performance versus the SNR under spatial fading. $P_{FA} = 0.05$, $L = 4$, $K = 256$. (a) Ricean versus Rayleigh fading. OFDM signal. (b) Exploiting spectral information under Rayleigh fading. GSM signal.

Exploiting *a priori* spectral shape information.

We now study the detection performance gain of exploiting available information about the spectral shape of primary signals. To this end, we consider a Gaussian minimum shift keying (GMSK) signal generated according to the Global System for Mobile communications (GSM) cellular standard (3GPP, 2001), downconverted to baseband and I/Q sampled at 300 ksamples/s. The psd of this signal is shown in the inset in Fig. 2.4(b). The channel from the primary user is assumed to present spatial Rayleigh fading ($\eta = 0$), which is accurate in scenarios without LOS to the primary user. The remaining simulation parameters are $P_{FA} = 0.05$, $L = 4$, $K = 256$.

In Fig. 2.4(b) we compare the performance of the proposed detectors when they exploit the available spectral information against the case they assume a white primary signal, i.e., assuming spectral mask $\mathbf{C}_K = \mathbf{I}_K$. We can see that the performance gain of exploiting spectral information can indeed be large, regardless of the employed detector. In the setup considered here, exploiting spectral information gives around 3 dB detection gain when using the corresponding detector, hence with similar complexity.
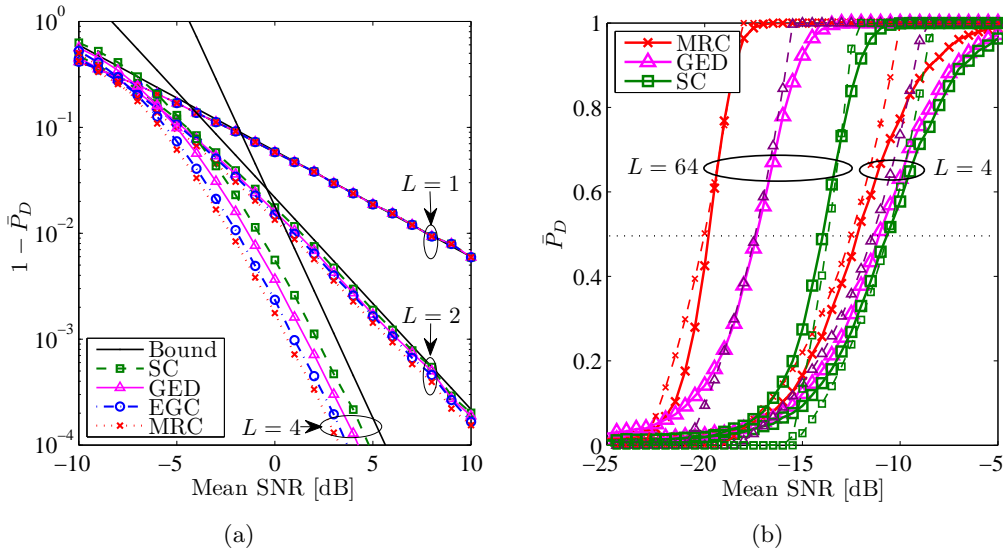
Figure 2.5: High SNR and Daher-Adve diversity orders. (a) High SNR detection performance. (b) Detection performance around $\bar{P}_D = 1/2$. Solid lines: simulation results. Dashed lines: analytical approximations.

Asymptotic diversity order of the different detectors.

Figure 2.5 shows the two different regimes that conduct to the concepts of high SNR asymptotic diversity and Daher-Adve diversity orders in independent Rayleigh fading.

First, in Fig. 2.5(a) we can see the high SNR behavior of the probability of misdetection. The probability of miss has been plotted in the range of the mean SNR $\bar{\zeta} \in [-10, 10]$ dB assuming GMSK signals with parameters as in Fig. 2.4(b) for $K = 256$, $P_{\mathrm{FA}} = 0.01$. We can observe the advantage of having a larger number of antennas in terms of asymptotic detection diversity. As can be seen in Fig. 2.5(a), in Rayleigh fading environments the number of antennas determines the asymptotic slope of the misdetection probability curve versus the SNR. Here we show the bound obtained for the GED. Note that the asymptotic bound presents in all the cases the right slope, though it becomes looser as the number of antennas increases.

On the other hand, assuming white primary signals, Fig. 2.5(b) shows the behavior of the probability of detection (in linear scale) around the point at which it equals $1/2$. The $\bar{P}_{\mathrm{D}}(\bar{\zeta})$ curves for different detectors in Rayleigh fading, obtained by Monte Carlo simulation, are compared to the corresponding piecewise linear approx-
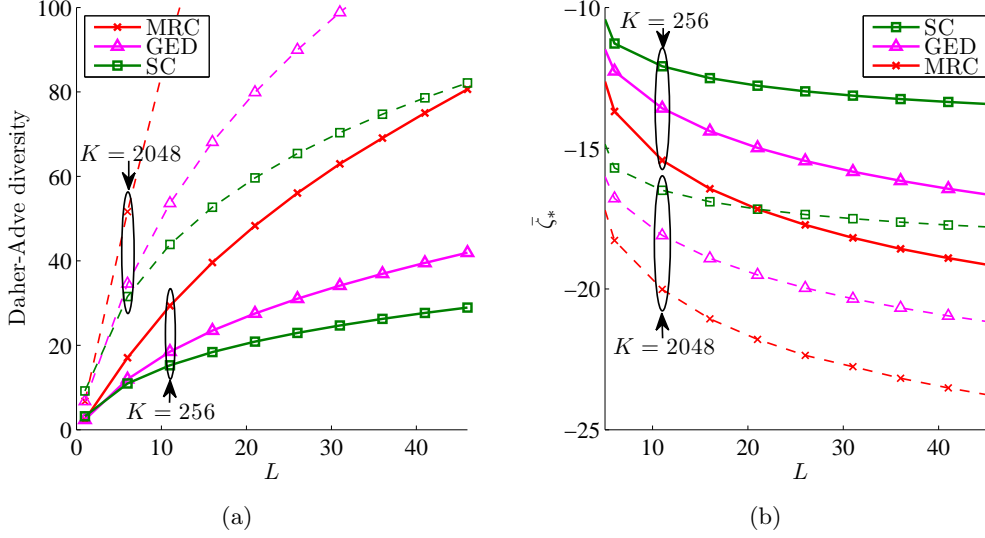
Figure 2.6: Comparison of the different detectors. (a) Daher-Adve diversity order d. (b) Minimum operational SNR.

imations from the previous sections. These match the empirical curves reasonably well around $\bar{P}_\mathrm{D} \approx \frac{1}{2}$.

Daher-Adve diversity.

From Fig. 2.5(b) it is apparent that the detection performance around $\bar{P}_\mathrm{D} \approx \frac{1}{2}$ can be accurately described using two parameters: the minimum operational SNR $\bar{\zeta}^*$ and the Daher-Adve diversity order d.

Figure 2.6 shows the analytical approximations for the diversity order and the *minimum operational SNR* $\bar{\zeta}^*$ as a function of the number of antennas L, for white primary signals in the same set-up as in Fig. 2.5(b). From Figs. 2.6(a) and (b) the performance advantage of the MRC detector is clear. The diversity order of this centralized detector grows almost linearly with the number of antennas, whereas that of the ED is approximately proportional to $\sqrt{L}$. As for the SC detector, it is difficult to find analytical bounds for its diversity order in terms of L. By comparison with ED, it is seen in Fig. 2.6(a) that it increases at a rate no larger than $\sqrt{L}$. It is conjectured that the diversity order of the OR detector is logarithmic in L, similarly to the OR-based detector analyzed in Daher and Adve (2010). It must be noted that the detectors studied in Daher and Adve (2010) are based on a different model

Detector	Test	Asympt. diversity	Daher-Adve diversity
SC / OR	$T_{\mathrm{SC}} = \frac{1}{K\sigma^2} \max_{1 \le m \le L} \mathbf{y}_m^H \mathbf{C} \mathbf{y}_m$	L	$\mathcal{O}\left(\sqrt{K}\log(L\bar{b}_2)\right)^{\ddagger}$
GED	$T_{\mathrm{GED}} = \frac{1}{KL\sigma^2} \sum_{i=1}^{L} \mathbf{r}_i^H \mathbf{C} \mathbf{r}_i$	L	$\mathcal{O}\left(\sqrt{KL\bar{b}_2}\right)$
EGC	$T_{\mathrm{EGC}} = \frac{1}{KL\sigma^2} \sum_{n=1}^{L} \sum_{m=1}^{L} |\mathbf{y}_m^H \mathbf{C} \mathbf{y}_n|$	L	–
MRC	$T_{\mathrm{MRC}} = \frac{1}{K\sigma^2} \lambda_{\max}(\mathbf{Y}^H \mathbf{C} \mathbf{Y})$	L	$\mathcal{O}\left(\sqrt{K}L\bar{b}_2\right)$

Table 2.1: Summary of the proposed multiantenna detectors under known noise statistics in independent Rayleigh fading. ‡ Conjecture.

from the one adopted in this chapter. Their model is adequate for radar systems, but no so for spectrum sensing applications.

2.7   Conclusions

In this chapter we have presented a family of multiantenna detectors designed under different approximations, establishing a connection with diversity combining techniques for multiantenna receivers in communications. Moreover, the proposed detectors exploit *a priori* information about the modulation and channelization schemes employed by the primary network, summarized as the spectral shape of primary transmissions.

In order to analyze the detection performance in fading environments we propose here two different diversity analyses. The first is based on the notion of diversity taken from communications, hence asymptotic in the SNR, whereas the second was borrowed from the radar literature and defines detection diversity as the slope of the probability of detection ($\bar{P}_D$) versus the SNR curve at $\bar{P}_D = 1/2$. In terms of the asymptotic definition of diversity order, all the proposed detectors extract full detection diversity from the receiver antennas in independent Rayleigh fading. However, using the second definition we see how the more sophisticated detectors outperform simpler approaches such as the selection combining detector. A summary with the proposed detectors and their diversity orders is included in Table 2.1.

While some preliminary work on the topic of multiantenna detection was presented in López-Valcarce et al. (2009), the main content of this chapter is a joint work with the Signal Processing for Communications Research Group (SPCOM, Technical

Univ. of Catalonia UPC) under the national research project SPROACTIVE (reference TEC2007-68094-C02-01/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010). This work currently constitutes a joint journal paper submitted to the IEEE Transactions on Wireless Communications (Vazquez-Vilar et al., 2011b). The theoretical results exposed in Section 2.5.2 (Daher-Adve diversity order analysis) have been obtained in collaboration with the Philips Research Depertment (Eindhoven, The Netherlands) and have resulted in a joint conference publication (Vazquez-Vilar et al., 2011a).

Appendix 2.A   Statistical analysis for large data records

Let $t_{ij} \doteq \mathbf{y}_i^H \mathbf{C} \mathbf{y}_j \mathrm{e}^{\jmath\theta_{ij}}$ denote the phase-aligned spectrally weighted crosscorrelation between antennas i and j, where $\theta_{ij} \doteq \arg\{h_i\} - \arg\{h_j\}$. For large K, we can invoke the central limit theorem and assume that t_{ij} is Gaussian distributed. Under this approximation, and using the fact that for zero-mean complex circular Gaussian vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and constant matrices $\boldsymbol{A}$, $\boldsymbol{B}$, it holds that

$$
\begin{aligned}
\mathrm{E}[(\boldsymbol{x}_1^H \mathbf{A} \boldsymbol{x}_2)(\boldsymbol{u}_1^H \boldsymbol{B} \boldsymbol{u}_2)] = \\
\mathrm{tr}\left(\boldsymbol{A}\mathrm{E}[\boldsymbol{x}_2\boldsymbol{x}_1^H]\right)\mathrm{tr}\left(\boldsymbol{B}\mathrm{E}[\boldsymbol{u}_2\boldsymbol{u}_1^H]\right) + \mathrm{tr}\left(\boldsymbol{A}\mathrm{E}[\boldsymbol{x}_2\boldsymbol{u}_1^H]\boldsymbol{B}\mathrm{E}[\boldsymbol{u}_2\boldsymbol{x}_1^H]\right) \quad (2.76)
\end{aligned}
$$

(see e.g. Porat and Friedlander (1986)), then one finds that

$$
\begin{aligned}
\mathrm{E}[t_{ij}] &= K(|h_i||h_j|\bar{b}_2 + \sigma^2 \mathbb{1}(i=j)), & (2.77) \\
\mathrm{var}\{t_{ij}\} &= K\left[|h_i|^2|h_j|^2\bar{b}_4 + \sigma^2(|h_i|^2 + |h_j|^2)\bar{b}_3 + \sigma^4\bar{b}_2\right], & (2.78) \\
\mathrm{E}[t_{ij}^2] - \mathrm{E}^2\{t_{ij}\} &= K\left[|h_i|^2|h_j|^2\bar{b}_4 + \sigma^2(2|h_i||h_j|\bar{b}_3 + \sigma^2\bar{b}_2)\mathbb{1}(i=j)\right] & (2.79)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{E}[t_{ij}t_{kl}^*] - \mathrm{E}[t_{ij}]\mathrm{E}[t_{kl}^*] = K\big[&|h_i||h_j||h_k||h_l|\bar{b}_4 \\
&+ \sigma^2\left(|h_j||h_l|\mathbb{1}(i=k) + |h_i||h_k|\mathbb{1}(j=l)\right)\bar{b}_3 + \sigma^4\bar{b}_2\mathbb{1}(i=k)\mathbb{1}(j=l)\big]. \quad (2.80)
\end{aligned}
$$

From (2.78)-(2.79), it follows that for $i \neq j$ the real and imaginary parts of t_{ij} are uncorrelated ($\mathrm{E}[\Re\{t_{ij}\}\Im\{t_{ij}\}] - \mathrm{E}[\Re\{t_{ij}\}]\mathrm{E}[\Im\{t_{ij}\}] = 0$), with variances given by

$$
\begin{aligned}
\mathrm{var}\{\Re\{t_{ij}\}\} &= K\left[|h_i|^2|h_j|^2\bar{b}_4 + \frac{1}{2}\left((|h_i|^2 + |h_j|^2)\sigma^2\bar{b}_3 + \sigma^4\bar{b}_2\right)\right], & (2.81) \\
\mathrm{var}\{\Im\{t_{ij}\}\} &= \frac{K}{2}\left[(|h_i|^2 + |h_j|^2)\sigma^2\bar{b}_3 + \sigma^4\bar{b}_2\right]. & (2.82)
\end{aligned}
$$

2.A.1   Generalized Energy Detector

The GED statistic can be written as $\mathbf{1}^T\boldsymbol{t}$ up to an irrelevant scaling factor, where $\boldsymbol{t} \doteq [\, t_{11} \; \cdots \; t_{LL} \,]^T$. For large K, $\boldsymbol{t}$ is normally distributed; from (2.77)-(2.78), its

mean and covariance are

$$E[\boldsymbol{t}] = K \left(\bar{b}_2 \mathbf{g} + \sigma^2 \mathbf{1} \right), \tag{2.83}$$

$$\text{cov}\{\boldsymbol{t}, \boldsymbol{t}\} = K \left(\bar{b}_4 \mathbf{g} \mathbf{g}^H + \sigma^4 \bar{b}_2 \mathbf{I} + 2\sigma^2 \bar{b}_3 \, \text{diag}\{\mathbf{g}\} \right), \tag{2.84}$$

where $\mathbf{g} \doteq [\, |h_1|^2 \; |h_2|^2 \; \cdots \; |h_L|^2\,]^T$ is the vector of channel gains. Thus, T_{GED} is asymptotically Gaussian. Using (2.83)-(2.84), its the mean and variance are found to be

$$\mu_{\text{GED}}(\zeta) \doteq E[T_{\text{GED}}] \;\; = \;\; \zeta \frac{\bar{b}_2}{L} + 1, \tag{2.85}$$

$$\alpha_{\text{GED}}^2(\zeta) \doteq \text{var}\{T_{\text{GED}}\} \;\; = \;\; \frac{1}{KL^2} \left(\zeta^2 \bar{b}_4 + 2\zeta \bar{b}_3 + L\bar{b}_2 \right). \tag{2.86}$$

Therefore, for a given threshold γ_{GED} the probabilities of false alarm and detection are respectively given by (2.26)-(2.27).

2.A.2 Selection Combining Detector

The statistic of the SC detector is $T_{\text{SC}} = \frac{1}{K\sigma^2} \max_{1 \le i \le L} t_{ii}$. For a given threshold γ_{SC}, and L antennas, the false alarm probability is

$$P_{\text{FA}}^{\text{SC}}(L) = 1 - \Pr\left\{ t_{ii} \le K\sigma^2 \gamma_{\text{SC}}, \, 1 \le i \le L \mid \mathcal{H}_0 \right\} \tag{2.87}$$

$$= 1 - \prod_{i=1}^{L} \Pr\left\{ t_{ii} \le K\sigma^2 \gamma_{\text{SC}} \mid \mathcal{H}_0 \right\} \tag{2.88}$$

$$= 1 - [1 - P_{\text{FA}}(1)]^L, \tag{2.89}$$

where we have used the fact that under $\mathcal{H}_0$ the t_{ii} are independent, and $P_{\text{FA}}(1)$ denotes the false alarm probability of a single-antenna detector with the same threshold γ_{SC} under the Gaussian approximation, which is found from (2.26):

$$P_{\text{FA}}(1) = Q\left(\frac{\gamma_{\text{SC}} - 1}{\sqrt{\bar{b}_2 K}/K} \right), \tag{2.90}$$

from which (2.32) follows. On the other hand, the probability of detection is

$$P_{\text{D}}^{\text{SC}}(L) = 1 - \Pr\left\{ t_{ii} \le K\sigma^2 \gamma_{\text{SC}}, \, 1 \le i \le L \mid \mathcal{H}_1 \right\} \tag{2.91}$$

Since the random variables t_{ii} are not independent under $\mathcal{H}_1$, (2.91) does not factor out as in (2.88) in general (except if all but one of the channel coefficients are zero, in which case the covariance matrix (2.84) becomes diagonal). Thus, the computation of (2.91) involves the integration of a multivariate Gaussian with mean and covariance given by (2.83)-(2.84). This can be done numerically using, e.g., Matlab's `mvncdf` Gaussian integration package.

2.A.3 Equal Gain Combining detector

The statistic of the EGC detector can be rewritten as

$$T_{\text{EGC}} = \frac{1}{KL\sigma^2} \sum_{i=1}^{L} \sum_{j=1}^{L} |t_{ij}|. \tag{2.92}$$

Note that $|t_{ii}| = t_{ii}$, $1 \le i \le L$, which is asymptotically Gaussian distributed for large K under both hypotheses. On the other hand, for $i \ne j$, t_{ij} is asymptotically complex-valued Gaussian with independent real and imaginary parts.

Distribution of T_{EGC} under $\mathcal{H}_0$

If $h_i = h_j = 0$, then the real and imaginary parts of t_{ij}, $i \ne j$, have both zero mean and the same variance. Therefore $|t_{ij}|$ is Rayleigh distributed with

$$\mathrm{E}[|t_{ij}| \,|\, \mathcal{H}_0] = \frac{\sigma^2}{2} \sqrt{K\pi\bar{b}_2}, \tag{2.93}$$

$$\mathrm{var}\{|t_{ij}| \,|\, \mathcal{H}_0\} = K\sigma^4 \left(1 - \frac{\pi}{4}\right) \bar{b}_2, \tag{2.94}$$

for $i \ne j$. Note from (2.80) that if $(i, j) \ne (k, l)$ then t_{ij} and t_{kl} are uncorrelated (and hence independent for large K) under $\mathcal{H}_0$, and therefore the different terms $|t_{ij}|$ in (2.92) become independent as well. Thus T_{EGC} is the sum of L Gaussian- and $L(L-1)/2$ (since $|t_{ij}| = |t_{ji}|$) Rayleigh-distributed, independent random variables. There is no simple closed-form expression for the resulting distribution, so we propose to use a Gaussian approximation to this end, based on the asymptotic mean

and variance given by

$$\mathrm{E}[T_{\mathrm{EGC}}] = \frac{1}{KL\sigma^2} \left(\sum_{i=1}^{L} K\sigma^2 + \sum_{i=1}^{L} \sum_{j=1, j\neq i}^{L} \frac{\sigma^2}{2} \sqrt{\pi K \bar{b}_2} \right) \tag{2.95}$$

$$= 1 + \frac{L-1}{2} \sqrt{\frac{\pi \bar{b}_2}{K}}, \tag{2.96}$$

$$\mathrm{var}\{T_{\mathrm{EGC}}\} = \frac{1}{(KL\sigma^2)^2} \left(\sum_{m=1}^{L} \mathrm{var}\{|t_{mm}|\} + \sum_{m=1}^{L} \sum_{i=1, i>m}^{L} \mathrm{var}\{2|t_{mi}|\} \right) \tag{2.97}$$

$$= \frac{1}{(KL\sigma^2)^2} \left(\sum_{m=1}^{L} K\sigma^4 \bar{b}_2 + \sum_{m=1}^{L} \sum_{i=1, i>m}^{L} 4K\sigma^4 (1 - \frac{\pi}{4}) \bar{b}_2 \right) \tag{2.98}$$

$$= \frac{1}{KL} \left(\left(\frac{\pi}{2} - 1 \right) + L \left(2 - \frac{\pi}{2} \right) \right) \bar{b}_2, \tag{2.99}$$

from which (2.36) follows.

Distribution of $T_{\mathbf{EGC}}$ under $\mathcal{H}_1$

For $i \neq j$, t_{ij} asymptotically follows a complex normal distribution centered on the real axis; its real and imaginary parts are uncorrelated and have different variances in general. We can write

$$|t_{ij}| = |\Re\{t_{ij}\}| \sqrt{1 + z_{ij}^2}, \tag{2.100}$$

where $z_{ij} \doteq \Im\{t_{ij}\}/\Re\{t_{ij}\}$. Note that z_{ij} is the ratio of two uncorrelated Gaussian random variables, and there is no closed-form expression for its distribution. However, from Hayya et al. (1975), if the coefficient of variation of the denominator (defined as the ratio of its standard deviation to its mean value) is less than 0.39, there exists a transformation $g(\cdot)$ such that the distribution of $g(z_{ij})$ can be accurately approximated by a standard Gaussian $\mathcal{N}(0,1)$. In our case, the coefficient of variation is

$$\beta_{ij} \doteq \frac{\sqrt{\mathrm{var}\{\Re\{t_{ij}\}\}}}{\mathrm{E}[\Re\{t_{ij}\}]} = \frac{1}{K} \frac{\sqrt{|h_i|^2 |h_j|^2 \bar{b}_4 + \frac{1}{2} \left((|h_i|^2 + |h_j|^2) \sigma^2 \bar{b}_3 + \sigma^4 \bar{b}_2 \right)}}{|h_i||h_j|\bar{b}_2}. \tag{2.101}$$

Provided that $h_i \neq 0$, $h_j \neq 0$, then (2.101) goes to zero as $K \to \infty$. Thus, for large enough K, the random variable

$$g(z_{ij}) \doteq \frac{\mathrm{E}[\Re\{t_{ij}\}]z_{ij} - \mathrm{E}[\Im\{t_{ij}\}]}{\sqrt{\mathrm{var}\{\Re\{t_{ij}\}\}z_{ij}^2 + \mathrm{var}\{\Im\{t_{ij}\}\}}} \tag{2.102}$$

$$= \sqrt{K}\frac{|h_i||h_j|\bar{b}_2 z_{ij}}{\sqrt{\|h_i\|^2|h_j|^2\bar{b}_4 z_{ij}^2 + \frac{1}{2}\left((|h_i|^2 + |h_j|^2)\sigma^2\bar{b}_3 + \sigma^4\bar{b}_2\right)\left(1 + z_{ij}^2\right)}} \tag{2.103}$$

is approximately zero-mean Gaussian with unit variance (Hayya et al., 1975). Since the transformation $g(z)$ in (2.103) is one-to-one, it follows that

$$\Pr\{z_{ij}^2 > \epsilon\} = 2Q(g(\sqrt{\epsilon})), \tag{2.104}$$

which approaches zero exponentially fast as $K \to \infty$. Therefore, for K large enough, it is reasonable to approximate $|t_{ij}| \approx |\Re\{t_{ij}\}|$ in (2.100). Moreover, with β_{ij} as in (2.101), one has $\Pr\{\Re\{t_{ij}\} < 0\} = Q(\beta_{ij}^{-1})$, which also goes to zero exponentially with K.

In view of all these, if $|h_m| > 0$ for $m = 1, \ldots, L$, then for sufficiently large K the terms $|t_{ij}|$, $i \neq j$, behave as correlated real-valued Gaussian random variables. Thus T_{EGC} becomes approximately Gaussian, with expected value

$$\mathrm{E}[T_{\mathrm{EGC}}] \approx \frac{1}{KL\sigma^2}\left(\sum_{i=1}^{L}\sum_{j=1}^{L}\mathrm{E}[\Re\{t_{ij}\}]\right) = \frac{\|\mathbf{h}\|_1^2}{L\sigma^2}\bar{b}_2 + 1 \tag{2.105}$$

and variance

$$\mathrm{var}\{T_{\mathrm{EGC}}\} \approx \frac{1}{(KL\sigma^2)^2}\sum_{i,j,k,l}\left(\mathrm{E}[\Re\{t_{ij}\}\Re\{t_{kl}\}] - \mathrm{E}[\Re\{t_{ij}\}]\mathrm{E}[\Re\{t_{kl}\}]\right) \tag{2.106}$$

$$= \frac{1}{2(KL\sigma^2)^2}\sum_{i,j,k,l}\Re\left\{(\mathrm{E}[t_{ij}t_{kl}^*] - \mathrm{E}[t_{ij}]\mathrm{E}[t_{kl}^*])\right.$$

$$\left. + (\mathrm{E}[t_{ij}t_{lk}^*] - \mathrm{E}[t_{ij}]\mathrm{E}[t_{lk}^*])\right\}, \tag{2.107}$$

where we have used the fact that $t_{kl} = t_{lk}^*$. Using (2.80), one readily finds that

$$\mathrm{var}\{T_{\mathrm{EGC}}\} \approx \frac{1}{K}\left(\left(\frac{\|\mathbf{h}\|_1^2}{\sigma^2}\right)^2\frac{\bar{b}_4}{L^2} + 2\frac{\|\mathbf{h}\|_1^2}{\sigma^2}\frac{\bar{b}_3}{L} + \bar{b}_2\right). \tag{2.108}$$

Using (2.105) and (2.108), one obtains (2.37).

2.A.4 Maximal Ratio Combining detector

The statistical analysis of the MRC detector amounts to finding the distribution of the largest eigenvalue of the random matrix $\mathbf{Y}^H \mathbf{C} \mathbf{Y}$ under each hypothesis. For a general covariance matrix $\mathbf{C}$, this remains an open problem. We consider here a special case of practical interest: strictly bandlimited primary signals using a fraction B of the total channel bandwidth, and with flat psd within their passband.

Using the asymptotic diagonalization (2.5) of the covariance matrix $\mathbf{C}$,

$$\lambda_{\max}(\mathbf{Y}^H \mathbf{C} \mathbf{Y}) \approx \lambda_{\max}(\mathbf{Y}^H \mathbf{W} \mathbf{\Lambda} \mathbf{W}^H \mathbf{Y}) \tag{2.109}$$

$$= \lambda_{\max}(\bar{\mathbf{Y}}^H \mathbf{\Lambda} \bar{\mathbf{Y}}) \tag{2.110}$$

where $\bar{\mathbf{Y}} \doteq \mathbf{W}^H \mathbf{Y}$. The ideal bandpass assumption implies that $\mathbf{\Lambda}$ has BK non-zero diagonal entries, which are all equal to $1/B$ (since $\operatorname{tr} \mathbf{C} = K$). Therefore

$$\lambda_{\max}(\mathbf{Y}^H \mathbf{C} \mathbf{Y}) \approx \frac{1}{B} \lambda_{\max}(\bar{\mathbf{Y}}_{\mathcal{B}}^H \bar{\mathbf{Y}}_{\mathcal{B}}) \tag{2.111}$$

where $\bar{\mathbf{Y}}_{\mathcal{B}}$ is a $BK \times L$ matrix comprising the rows of $\bar{\mathbf{Y}}$ corresponding to the non-zero diagonal elements of $\mathbf{\Lambda}$. Note that $\bar{\mathbf{Y}}_{\mathcal{B}}^H \bar{\mathbf{Y}}_{\mathcal{B}}$ is a complex Wishart matrix (Tulino and Verdú, 2004), and thus under $\mathcal{H}_0$ the random variable

$$\Phi \doteq \frac{1}{\nu} \left(\frac{\lambda_{\max}(\mathbf{Y}^H \mathbf{C} \mathbf{Y})}{\sigma^2 K} - \mu \right) \tag{2.112}$$

asymptotically (in K and L) follows a Tracy-Widom distribution (Karoui, 2005), with scale and bias terms given respectively by

$$\mu = (\sqrt{BK - 1} + \sqrt{L})^2, \tag{2.113}$$

$$\nu = (\sqrt{BK - 1} + \sqrt{L}) \left(\frac{1}{\sqrt{BK - 1}} + \frac{1}{\sqrt{L}} \right)^{1/3}. \tag{2.114}$$

On the other hand, under $\mathcal{H}_1$, $\bar{\mathbf{Y}}_{\mathcal{B}}^H \bar{\mathbf{Y}}_{\mathcal{B}}$ follows a spiked population model (Baik and Silverstein, 2006), i.e. only one of the eigenvalues of the true covariance matrix $\mathrm{E}[\bar{\mathbf{Y}}_{\mathcal{B}}^H \bar{\mathbf{Y}}_{\mathcal{B}}]$ is different from 1. If we denote by δ_1 the largest eigenvalue of the true

covariance matrix, given by

$$\delta_1 \doteq \lambda_{\max} \left(\mathrm{E}[\mathbf{Y}^H \mathbf{C} \mathbf{Y}] \right) / (\sigma^2 K) \tag{2.115}$$

$$= 1 + \bar{b}_2 \frac{\|\mathbf{h}\|_2^2}{\sigma^2}, \tag{2.116}$$

we have that for $\delta_1 > 1 + \sqrt{L/K}$ the distribution of $T_{\mathrm{MRC}} = \lambda_{\max}(\mathbf{Y}^H \mathbf{C} \mathbf{Y})/(\sigma^2 K)$ is given by (Baik and Silverstein, 2006; Taherpour et al., 2010)

$$T_{\mathrm{MRC}} \sim \mathcal{N} \left(\delta_1 + \frac{L\delta_1}{K(\delta_1 - 1)}, \ \delta_1^2/K \right), \tag{2.117}$$

asymptotically in both K and L, and (2.40) follows. Even though these results are asymptotic in both K and L, they are remarkably accurate even for reasonably moderate values of these parameters.

Appendix 2.B   Asymptotic analysis of $g_L(x)$

From the definition of the incomplete Gamma function (2.61) and using that $g_L(x) = \Gamma(Lx, L+1)$ we have

$$g_L(x) = \frac{1}{\Gamma(L)} \int_0^x \frac{L^L}{\mathrm{e}^L} \left(t\mathrm{e}^{1-t} \right)^L \partial t, \tag{2.118}$$

where $t\mathrm{e}^{1-t}$ for $t \geq 0$ is a quasi-concave function (Boyd and Vandenberghe, 2004). Taking derivatives we have that

$$\frac{\partial}{\partial t}(t\mathrm{e}^{1-t}) = \mathrm{e}^{1-t} - t\mathrm{e}^{1-t}. \tag{2.119}$$

For $t \geq 0$, (2.119) equals 0 if and only if $t = 1$. Hence, the maximum value of the function $t\mathrm{e}^{1-t}$ is given by

$$\max_t (t\mathrm{e}^{1-t}) = \mathrm{e}^0 = 1. \tag{2.120}$$

Note that $0 \leq t\mathrm{e}^{1-t} < 1$ for all $t \geq 0$, $t \neq 1$.

Let us define now

$$f_L(t) = \frac{1}{\Gamma(L)} \left(\frac{L}{\mathrm{e}} \right)^L (t\mathrm{e}^{1-t})^L, \quad t \geq 0, \tag{2.121}$$

so that $g_L(x) = \int_0^x f_L(t)\partial t$. We show now that $\lim_{L\to\infty} f_L(t) = \delta(t-1)$, where $\delta(x)$ denotes the Dirac delta function. First, note that the integral of $f_L(t)$ over the positive orthant does not depend on L:

$$\int_0^\infty f_L(t)\partial t \;=\; \frac{L^L}{\Gamma(L)} \int_0^\infty t^L e^{-Lt}\partial t \;=\; \frac{\Gamma(L+1)}{L\Gamma(L)} = 1. \tag{2.122}$$

On the other hand, using Stirling's formula $n! \approx \sqrt{2\pi n}(n/e)^n$ for $n \to \infty$, one has, for $0 \le z \le 1$,

$$\lim_{L\to\infty} \left(\frac{L}{e}\right)^L \frac{z^L}{\Gamma(L)} = \lim_{L\to\infty} \left(\frac{L}{e}\right)^L \frac{Lz^L}{\sqrt{2\pi L}} \left(\frac{e}{L}\right)^L \tag{2.123}$$

$$= \frac{1}{\sqrt{2\pi}} \lim_{L\to\infty} \sqrt{L} e^{L\log z}, \tag{2.124}$$

which equals 0 if $0 \le z < 1$ and ∞ if $z = 1$. Hence $\lim_{L\to\infty} f_L(t) = \delta(t-1)$, so that

$$\lim_{L\to\infty} g_L(x) \;=\; \int_0^x \lim_{L\to\infty} f_L(t)\partial t \;=\; \int_0^x \delta(t-1)\partial t \;=\; \begin{cases} 0, & \text{if } x < 1, \\ 1, & \text{if } x > 1, \end{cases} \tag{2.125}$$

as was to be shown.

Chapter 3

Multiantenna Detection under Unknown Noise Statistics

Contents

3.1   Introduction

In the previous chapter we introduced a family of multiantenna detectors that exploit
certain spectral features of the primary signal. These detectors require knowledge of
the noise variance, so that the threshold to which the detection test is compared can
be computed for a given desired probability of false alarm. However, if the actual
value of the noise variance is different from the nominal value, a critical SNR level,
denoted as "SNR wall", appears (Tandra and Sahai, 2008). Primary signals below
this critical value become undetectable, *even if the observation time goes to infinity.*
This serious drawback motivates the search for detectors robust to noise uncertainty.

Exploiting the spectral shape of the primary signal or the availability of multi-
ple antennas constitutes a promising approach to overcome this problem. The basic
idea is to exploit the fact that the primary signal presents either temporal or spatial
correlation which differs from that of the noise process. For example, if the noise
process is assumed uncorrelated across antennas, spatial correlation will only appear
if primary signals are present; on the other hand, if the primary signal has certain
temporal correlation, it can be used to differenciate it from that of the white noise
process. Hence detectors can be designed based on spatial and/or temporal cor-
relation estimates, rather than on the received signal power. Several authors have
explored this strategy in order to enhance detection performance in cognitive radio
systems (see e.g. Taherpour et al. (2010); Wang et al. (2010); Zeng et al. (2008);
Alamgir et al. (2008); Lim et al. (2008); Zhang et al. (2010a)).

In this chapter we extend several of these detectors to a more sophisticated
scenario with important practical implications. In particular, under the Gaussian
hypothesis, we derive the *generalized likelihood ratio test* (GLRT) for the detection
of primary signals with arbitrary spatial rank when the (unknown) spatial noise is
uncorrelated across the different antennas. In a second step, we derive the GLRT
for the detection of spatially rank-1 primary signals with temporal correlation when
the noise covariance matrix is assumed to have arbitrary spatial structure.

The first case models practical scenarios with spatial rank of the received sig-
nals larger than one. This is the case, for example, if multiple independent users
(e.g. from adjacent cells) simultaneously access the same frequency channel. Alter-
natively, many state-of-the-art communication standards consider the simultaneous
transmission of different data streams through multiple antennas to achieve multi-
plexing gain and/or the use of space-time codes to enhance spatial diversity. For

these systems, the signal received at the multiantenna sensor will exhibit a spatial rank equal to the number of independent streams or the spatial size of the code, respectively. Examples range from broadcasting standards, such as the European DVB-T2 (ETSI, 2009) which considers 2-antenna space-time Alamouti codes, to point-to-multipoint standards, such as IEEE 802.11n (IEEE Computer Society, 2009), IEEE 802.16 (IEEE Computer Society and the IEEE Microwave Theory and Techniques Society, 2009) or LTE (3GPP, 2009), which support up to four transmit antennas.

On the other hand, a noise covariance matrix with arbitrary spatial structure can occur in scenarios in which the spectrum sensor experiences strong co-channel interference, originating either in the secondary network it belongs to, or in a different one. In this case, the secondary contributions can be modeled as temporally white noise with arbitrary and unknown spatial covariance. Note that if no structure is imposed to the spatial covariance of the noise process, we must resort to additional information in order to detect the primary signal. To this end we will assume that the temporal correlation matrix of the primary signal is available to the spectral monitor as *a priori* information. As we have seen in the previous chapter, this is a reasonable assumption provided that the channelization and modulation parameters of the primary system are fixed and public.

3.2   Problem formulation

We present here a general model of a multiantenna spectrum monitoring system of primary signals with spatial rank that can be larger than one. This model will be particularized when required.

3.2.1   System model

The sensor has L antennas with their respective RF chains. The same primary channel is selected at all antennas, downconverted to baseband, and asynchronously sampled. Primary transmission is comprised of P independent streams which may present certain temporal correlation. The noise is assumed to have spatial rank equal to the number of antennas L and is assumed to be temporally white.

The spectrum monitor acquires K samples from the lth antenna arranged in a

$K \times 1$ vector $\mathbf{y}_l$, which can be written as

$$\mathbf{y}_l = \mathbf{S}\mathbf{h}_l + \mathbf{N}\mathbf{g}_l, \tag{3.1}$$

where the $K \times P$ matrix $\mathbf{S} \doteq [\mathbf{s}_1 \, \mathbf{s}_2 \cdots \mathbf{s}_P]$ is comprised of P primary signal streams, $\mathbf{h}_l$ denotes the $P \times 1$ channel vector from the primary system to the lth receiver antenna, $\mathbf{N} \doteq [\mathbf{n}_1 \, \mathbf{n}_2 \cdots \mathbf{n}_L]$ represents the $K \times L$ noise matrix, and the $L \times 1$ vector $\mathbf{g}_l$ will conform the noise correlation among the different antennas.

Note that the model above assumes that the channel from the primary transmitter to the spectrum monitor is frequency-flat in the RF channel bandwidth, and that it remains constant for the duration of the sensing time. Because of the reasons presented in the Section 1.3.1 we restrict our analysis to both signal and noise following a zero-mean Gaussian distribution. Without loss of generality, and since any existing spatial correlation can be absorbed into the vectors $\mathbf{h}_l$ and $\mathbf{g}_l$ we assume that the signal streams $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_P$ (respectively noise streams $\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_L$) are mutually independent. Additionally we assume that the primary streams present a certain temporal correlation (equal for all of them) while the noise is temporally white. Then we have that

$$\mathrm{E}[\mathbf{s}_p\mathbf{s}_q^H] = \begin{cases} \mathbf{C} & \text{if } p = q, \\ \mathbf{0} & \text{if } p \neq q, \end{cases} \tag{3.2}$$

$$\mathrm{E}[\mathbf{n}_p\mathbf{n}_q^H] = \begin{cases} \mathbf{I}_K & \text{if } p = q, \\ \mathbf{0} & \text{if } p \neq q. \end{cases} \tag{3.3}$$

Again we will assume that the detector has certain information on the spectral shape of the primary signal, which translates into the *a priori* knowledge of the temporal correlation $\mathbf{C}$, wich is assumed to be normalized so that $\mathrm{tr}\{\mathbf{C}\} = K$. Note that given this model, any noise contribution with spatial rank equal or larger than L can be represented, from the receiver point of view, by just L independent streams. Hence, it suffices to consider L noise streams. Moreover, in this chapter we will assume that the $L \times L$ noise conforming matrix defined as $\mathbf{G} \doteq [\mathbf{g}_1 \, \mathbf{g}_2 \cdots \mathbf{g}_L]$ is nonsingular, so that the received noise is spatially full-rank. This ammounts to requiring that the noise spatial covariance matrix, defined as $\mathbf{\Sigma}^2 \doteq \mathbf{G}^H\mathbf{G}$, is nonsingular. Note that this applies to any system of practical interest since the thermal noise contribution is independent accross antennas. Moreover, if this were not the case we could trivially detect the presence of primary users by monitoring the noise-free dimensions[1].

---

[1]Being one possibility to check the rank of the empirical spatial covariance matrix of the received

The received signal can be compactly written in matrix form as

$$\mathbf{Y} = \mathbf{SH} + \mathbf{NG}, \tag{3.4}$$

where we have defined the $K \times L$ received signal matrix $\mathbf{Y} \doteq [\mathbf{y}_1 \, \mathbf{y}_2 \cdots \mathbf{y}_L]$, and the $P \times L$ channel matrix $\mathbf{H} = [\mathbf{h}_1 \, \mathbf{h}_2 \cdots \mathbf{h}_L]$.

3.2.2 Hypothesis testing problem

If we define $\mathbf{y} \doteq \mathrm{vec}(\mathbf{Y})$, we have that under the Gaussianity assumption, $\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathcal{R})$ with

$$\mathcal{R} \doteq (\mathbf{\Sigma}^2)^T \otimes \mathbf{I}_K + (\mathbf{H}^H \mathbf{H})^T \otimes \mathbf{C}. \tag{3.5}$$

Then, we may write the hypothesis testing problem for primary user detection as

$$\begin{aligned} \mathcal{H}_1 &: \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathcal{R}_1), \\ \mathcal{H}_0 &: \mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathcal{R}_0), \end{aligned} \tag{3.6}$$

where

$$\mathcal{R}_1 = (\mathbf{\Sigma}^2)^T \otimes \mathbf{I}_K + (\mathbf{H}^H \mathbf{H})^T \otimes \mathbf{C}, \tag{3.7}$$

$$\mathcal{R}_0 = (\mathbf{\Sigma}^2)^T \otimes \mathbf{I}_K. \tag{3.8}$$

Generalized likelihood ratio test

As there are unknown parameters under both hypotheses, the Neyman-Pearson detector is not implementable. A sensible approach is to use the Generalized Likelihood Ratio Test (GLRT) since it results in simple detectors with good performance (Mardia et al., 1979). In the GLRT, the unknown parameters are substituted by their Maximum Likelihood (ML) estimates under each hypothesis:

$$T \doteq \frac{\max_{\mathcal{R}_1} f(\mathbf{Y} \,|\, \mathcal{R}_1)}{\max_{\mathcal{R}_0} f(\mathbf{Y} \,|\, \mathcal{R}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{3.9}$$

---

signal. In absence of primary users it will feature a rank smaller than L (given the noise covariance singularity assumption). On the other hand, when a primary signal is present, unless this signal is aligned with the noise subspace, the perceived spatial rank of the received signal will increase.

where γ is a suitable threshold and the parametrized probability density function of the data is given by

$$f(\mathbf{y} \,|\, \mathcal{R}) = \frac{\exp\{-\mathbf{y}^H \mathcal{R}^{-1} \mathbf{y}\}}{\pi^{LK} \det \mathcal{R}}. \tag{3.10}$$

Here it is understood that the maximization operations in (3.9) are with respect to the structure of $\mathcal{R}_1$ and $\mathcal{R}_0$ in (3.7) and (3.8) respectively. Thus, the unknown parameters are $\mathbf{\Sigma}^2$ under $\mathcal{H}_0$ and $\{\mathbf{H}, \mathbf{\Sigma}^2\}$ under $\mathcal{H}_1$.

3.3 Detection of rank-P signals in spatially uncorrelated noise

In this section we consider the case of white primary signals, i.e. $\mathbf{C} = \mathbf{I}_K$. Note that under this assumption both noise and signal are temporally iid, and the likelihood is given by the product of the individual pdfs, i.e.,

$$f(\mathbf{Y} \,|\, \mathbf{R}) = \frac{1}{\pi^{LK} \det(\mathbf{R})^K} \exp\left\{-K \operatorname{tr}\left(\hat{\mathbf{R}} \mathbf{R}^{-1}\right)\right\}, \tag{3.11}$$

where

$$\mathbf{R} \doteq \mathbf{\Sigma}^2 + \mathbf{H}^H \mathbf{H}, \tag{3.12}$$

$$\hat{\mathbf{R}} \doteq \frac{1}{K} \mathbf{Y}^H \mathbf{Y} \tag{3.13}$$

are, respectively, the actual and the sample spatial covariance matrices.

It is clear that, due to the fact that both signal and noise are assumed temporally white, the detectors can only exploit the structure of the spatial statistics.

3.3.1 Spatially uncorrelated iid noise process

If the L analog frontends are perfectly calibrated and the noise is assumed uncorrelated across antennas, we can model the spatial covariance matrix of the noise process as a scaled version of the identity matrix

$$\mathbf{\Sigma}^2 = \sigma^2 \mathbf{I}. \tag{3.14}$$

In order to derive the GLRT under this model, we need the maximum likelihood (ML) estimates of the unknown parameters $\{\sigma^2, \mathbf{H}\}$ under $\mathcal{H}_1$ and the ML estimate of σ^2 under $\mathcal{H}_0$. The ML estimate of the noise variance under $\mathcal{H}_0$ is straightforward to obtain and is given by

$$\hat{\sigma}^2 = \frac{1}{L} \operatorname{tr}\left(\hat{\mathbf{R}}\right). \tag{3.15}$$

In order to obtain the ML estimates under $\mathcal{H}_1$, we consider two cases depending on the rank P.

Lemma 3.1. *If $P \geq L - 1$, the ML estimates of $\mathbf{H}$ and σ^2, given by $\hat{\mathbf{H}}$ and $\hat{\sigma}^2$, satisfy $\hat{\mathbf{H}}\hat{\mathbf{H}}^H + \hat{\sigma}^2 \mathbf{I} = \hat{\mathbf{R}}$.*

Proof. For $P \geq L - 1$, $\mathbf{R} = \mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}$ has no additional structure besides being positive definite Hermitian. In that case, the log-likelihood is maximized for $\mathbf{R} = \hat{\mathbf{R}}$, as shown in Magnus and Neudecker (1999). $\square$

Thus, for $P \geq L - 1$, the GLRT has been derived in Mauchly (1940) and it is the well-known *Sphericity test*

$$\log T = KL \log \left[\frac{\frac{1}{L} \operatorname{trace}\left(\hat{\mathbf{R}}\right)}{\det^{1/L}\left(\hat{\mathbf{R}}\right)} \right]. \tag{3.16}$$

Let $\hat{\mathbf{R}} = \mathbf{V} \operatorname{diag}\left(\delta_1, \ldots, \delta_L\right) \mathbf{V}^H$ be an eigenvalue decomposition (EVD) of the sample covariance matrix, with $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_L$. Note that the argument of the logarithm in (3.16) is the ratio of the arithmetic and geometric means of the eigenvalues $\delta_1, \delta_2, \ldots, \delta_L$.

When $P < L - 1$, the low-rank structure of the primary signal can be further exploited to improve the detection. In that case, to obtain the ML estimates under $\mathcal{H}_1$, let $\mathbf{H}^H \mathbf{H} = \mathbf{U} \mathbf{\Psi}^2 \mathbf{U}^H$ be an eigenvalue decomposition (EVD) of $\mathbf{H}^H \mathbf{H}$, with

$$\mathbf{\Psi}^2 = \operatorname{diag}(\psi_1^2, \psi_2^2, \ldots, \psi_P^2, 0, 0, \ldots, 0), \tag{3.17}$$

with $\psi_1^2 \geq \psi_2^2 \geq \cdots \geq \psi_P^2$.

Lemma 3.2. *For $P < L - 1$, the ML estimates of σ^2, $\mathbf{U}$ and $\mathbf{\Psi}^2$ under $\mathcal{H}_1$ are*

respectively given by

$$\hat{\sigma}^2 = \frac{1}{L-P} \sum_{k=P+1}^{L} \delta_k, \tag{3.18}$$

$$\hat{\mathbf{U}} = \mathbf{V}, \tag{3.19}$$

$$\hat{\psi}_i^2 = \delta_i - \hat{\sigma}^2, \quad i = 1, \dots, P. \tag{3.20}$$

Proof. This result was first proved in Anderson (1963). In Appendix 3.A we include an alternative and, in our opinion, simpler proof based on Majorization Theory. $\square$

Taking into account (3.15) and Lemma 3.2, the log-GLRT for $P < L - 1$ is given, after some straightforward manipulations, by

$$\log T = KL \log \left[\frac{\frac{1}{L}\sum_{i=1}^{L} \delta_i}{\left(\prod_{i=1}^{L} \delta_i\right)^{1/L}} \right] - K(L-P) \log \left[\frac{\frac{1}{L-P}\sum_{i=P+1}^{L} \delta_i}{\left(\prod_{i=P+1}^{L} \delta_i\right)^{1/(L-P)}} \right]. \tag{3.21}$$

Note that the bracketed terms in (3.21) are functions of the ratio of the arithmetic and geometric means of all eigenvalues, and the $L - P$ smallest eigenvalues of $\hat{\mathbf{R}}$, respectively. The first term is the statistic of the sphericity test (3.16), whereas the second term can be seen as a test for the sphericity of the noise subspace, or as a *reference* for sphericity due to finite sample size effects (since as $K \to \infty$, then $\hat{\mathbf{R}} \to \mathbf{R}$ and thus $\delta_i \to \sigma^2$ for $i = P+1, \dots, L$, so that the second term in (3.21) goes to zero). Thus, the log-GLRT may be seen as a *sphericity ratio* (quotient between the sphericity statistics of the sample covariance matrix and its noise subspace).

Remark 3.1. The statistic in (3.21) generalizes the results in Besson et al. (2006); Taherpour et al. (2010); Wang et al. (2010) obtained for the special case of $P = 1$.

3.3.2 Spatially uncorrelated non-iid noise process

When the analog frontends are perfectly calibrated it is possible to assume the same noise variance at each of the antennas. In practice, however, tolerances in the components of the different RF chains will result in deviations of the noise level from antenna to antenna. In this section we derive the GLRT for the more involved model of non-iid noises.

In this case, the only constraint on $\mathbf{\Sigma}^2$ is being diagonal with positive entries.

The ML estimate of $\boldsymbol{\Sigma}^2$ under $\mathcal{H}_0$, was derived in Leshem and Van der Veen (2001a,b)

$$\hat{\boldsymbol{\Sigma}}^2 = \mathrm{diag}\left([\hat{\mathbf{R}}]_{1,1}, \ldots, [\hat{\mathbf{R}}]_{L,L}\right) \doteq \hat{\mathbf{D}}. \tag{3.22}$$

Similarly to the case of iid noises, we study first the effect of the signal rank P on the ML estimate of the covariance matrix under $\mathcal{H}_1$.

Lemma 3.3. *If $P \geq L - \sqrt{L}$, the ML estimates of $\mathbf{H}$ and $\boldsymbol{\Sigma}^2$ under $\mathcal{H}_1$ satisfy* $\hat{\mathbf{H}}^H \hat{\mathbf{H}} + \hat{\boldsymbol{\Sigma}}^2 = \hat{\mathbf{R}}$.

Proof. The proof can be found in Leshem and Van der Veen (2001a); Ramírez et al. (2010). It hinges on the fact that if $P \geq L - \sqrt{L}$, then $\mathbf{H}\mathbf{H}^H + \boldsymbol{\Sigma}^2$ has no further structure beyond being positive definite Hermitian.   □

Using (3.22) and Lemma 3.3, and after some algebra, one finds that for $P \geq L - \sqrt{L}$, the GLRT is given by the *Hadamard ratio* of the sample covariance matrix (Wilks, 1935; Leshem and Van der Veen, 2001a,b):

$$T^{1/K} = \frac{\det\left(\hat{\mathbf{R}}\right)}{\prod_{i=1}^{L}[\hat{\mathbf{R}}]_{i,i}}. \tag{3.23}$$

On the other hand, if $P < L - \sqrt{L}$, the low-rank structure of the signal covariance matrix can be further exploited. In order to simplify the derivation of the ML estimates under $\mathcal{H}_1$, let $\hat{\mathbf{R}}_{\boldsymbol{\Sigma}} \doteq \boldsymbol{\Sigma}^{-1}\hat{\mathbf{R}}\boldsymbol{\Sigma}^{-1}$ (the *whitened* sample covariance matrix) and $\mathbf{H}_{\boldsymbol{\Sigma}} \doteq \mathbf{H}\boldsymbol{\Sigma}^{-1}$. We can rewrite the log-likelihood as

$$\log f\left(\mathbf{Y} \,|\, \mathbf{H}_{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}^2\right) = -LK\log\pi - K\log\det\left(\mathbf{H}_{\boldsymbol{\Sigma}}^H \mathbf{H}_{\boldsymbol{\Sigma}} + \mathbf{I}\right)$$
$$- K\log\det\left(\boldsymbol{\Sigma}^2\right) - K\mathrm{tr}\left[\hat{\mathbf{R}}_{\boldsymbol{\Sigma}}\left(\mathbf{H}_{\boldsymbol{\Sigma}}^H \mathbf{H}_{\boldsymbol{\Sigma}} + \mathbf{I}\right)^{-1}\right]. \tag{3.24}$$

Let $\mathbf{H}_{\boldsymbol{\Sigma}}^H \mathbf{H}_{\boldsymbol{\Sigma}} = \mathbf{G}\boldsymbol{\Phi}^2\mathbf{G}^H$ be the EVD of $\mathbf{H}_{\boldsymbol{\Sigma}}^H \mathbf{H}_{\boldsymbol{\Sigma}}$. The ML estimates of $\mathbf{G}$ and $\boldsymbol{\Phi}^2$ are given next.

Lemma 3.4. *Let*

$$\hat{\mathbf{R}}_{\boldsymbol{\Sigma}} = \mathbf{Q}\,\mathrm{diag}\left(\gamma_1, \ldots, \gamma_L\right)\mathbf{Q}^H \tag{3.25}$$

be the EVD of $\hat{\mathbf{R}}_{\boldsymbol{\Sigma}}$, with $\gamma_1 \geq \cdots \geq \gamma_L$. The ML estimates of $\mathbf{G}$ and $\boldsymbol{\Phi}^2 =$

diag $(\phi_1, \ldots, \phi_L)$ *(which are functions of* $\boldsymbol{\Sigma}^2$*) are*

$$\hat{\mathbf{G}} = \mathbf{Q}, \tag{3.26}$$

$$\hat{\phi}_i^2 = \begin{cases} \gamma_i - 1, & i = 1, \ldots, P, \\ 0, & i = P+1, \ldots, L. \end{cases} \tag{3.27}$$

Proof. Once $\hat{\mathbf{R}}$ and $\mathbf{H}$ have been prewhitened, the problem reduces to the iid case and, therefore, the proof follows the same lines as those in Lemma 3.2. $\qquad\square$

Finally, replacing the ML estimate of $\mathbf{H}_{\boldsymbol{\Sigma}}^H \mathbf{H}_{\boldsymbol{\Sigma}}$ into (3.24) we obtain

$$\log f\left(\mathbf{Y} \mid \boldsymbol{\Sigma}^2\right) = -K\left(+L\log\pi + P + \log\det\left(\hat{\mathbf{R}}\right) + \sum_{i=P+1}^{L}\left[\gamma_i - \log\gamma_i\right]\right). \tag{3.28}$$

To the best of our knowledge, the maximization of (3.28) with respect to $\boldsymbol{\Sigma}^2$ does not admit a closed-form solution if $P < L - \sqrt{L}$. We present two different approaches: an alternating optimization scheme and a closed-form GLRT detector obtained in the limit of asymptotically small SNR.

Alternating optimization

The ML estimation problem in (3.24) can be written as

$$\underset{\mathbf{H}_{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}}{\text{minimize}} \ \operatorname{tr}\left(\hat{\mathbf{R}}\boldsymbol{\Sigma}^{-1}\mathbf{R}_{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}^{-1}\right) - \log\det\left(\boldsymbol{\Sigma}^{-2}\right) + \log\det\mathbf{R}_{\boldsymbol{\Sigma}}, \tag{3.29}$$

$$\text{subject to } \mathbf{R}_{\boldsymbol{\Sigma}} = \mathbf{I}_L + \mathbf{H}_{\boldsymbol{\Sigma}}^H\mathbf{H}_{\boldsymbol{\Sigma}},$$

$$\left[\boldsymbol{\Sigma}^2\right]_{i,i} > 0.$$

While this optimization problem is non-convex, it is possible to partition the free variables in two different sets to obtain an alternating optimization scheme. Then, we will alternatively perform the minimization over each set of parameters while the remaining ones are held fixed. Since at each step the value of the cost function can only decrease, the method is guaranteed to converge to a (local) minimum (Bezdek and Hathaway, 2003).

From (3.29), we note that the individual minimization with respect to $\boldsymbol{\Sigma}$ (considering $\mathbf{H}_{\boldsymbol{\Sigma}}$ fixed) and with respect to $\mathbf{H}_{\boldsymbol{\Sigma}}$ (considering $\boldsymbol{\Sigma}$ fixed) can be easily written as convex problems individually, and, therefore, they can be efficiently solved.

Algorithm 1: Iterative estimation of $\mathbf{H_\Sigma}$ and $\mathbf{\Sigma}$ via alternating optimization.

Input: Starting point $\boldsymbol{\alpha}_{(0)}$ and $\hat{\mathbf{R}}$.
Output: ML estimates of $\mathbf{H_\Sigma}$ and $\mathbf{\Sigma}$.
Initialize: $n = 0$
repeat

  Compute $\mathbf{\Sigma}_{(n)}^{-1} = \mathrm{diag}(\boldsymbol{\alpha}_{(n)})$

  Obtain $\hat{\mathbf{R}}_{\mathbf{\Sigma}}^{(n+1)} = \mathbf{\Sigma}_{(n)}^{-1}\hat{\mathbf{R}}\mathbf{\Sigma}_{(n)}^{-1}$ and its EVD

  Compute $\mathbf{H_\Sigma}^{(n+1)}$ from (3.30) (fixed $\mathbf{\Sigma}_{(n)}^{-1}$)

  Solve (3.32) to obtain $\boldsymbol{\alpha}_{(n+1)}$ (fixed $\mathbf{H_\Sigma}^{(n+1)}$)

  Update $n = n + 1$
until Convergence

Minimization with respect to $\mathbf{H_\Sigma}$. For fixed $\mathbf{\Sigma}$, the optimal $\mathbf{H_\Sigma}$ minimizing (3.29) is given (up to a right multiplication by a unitary matrix) by Lemma 3.4:

$$\hat{\mathbf{H}}_{\mathbf{\Sigma}} = [\mathbf{q}_1 \ \cdots \ \mathbf{q}_P]\left(\mathrm{diag}\left(\gamma_1, \ldots, \gamma_P\right) - \mathbf{I}_P\right)^{1/2}, \tag{3.30}$$

where $[\mathbf{q}_1 \ \cdots \ \mathbf{q}_P]$ are the first P columns of the matrix $\mathbf{Q}$ featuring in the EV decomposition (3.25) of $\hat{\mathbf{R}}_{\mathbf{\Sigma}} = \mathbf{\Sigma}^{-1}\hat{\mathbf{R}}\mathbf{\Sigma}^{-1}$, and $\gamma_1, \ldots, \gamma_P$ are the corresponding eigenvalues.

Minimization with respect to $\mathbf{\Sigma}$. For fixed $\mathbf{H_\Sigma}$, the minimization problem in (3.29) reduces to

$$\underset{\mathbf{\Sigma}}{\mathrm{minimize}} \ \mathrm{tr}\left(\hat{\mathbf{R}}\mathbf{\Sigma}^{-1}\mathbf{R}_{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}^{-1}\right) - \log\det\left(\mathbf{\Sigma}^{-2}\right) \tag{3.31}$$

$$\text{subject to } [\mathbf{\Sigma}]_{i,i} \geq 0.$$

Defining the vector $\boldsymbol{\alpha} \doteq \left[[\mathbf{\Sigma}^{-1}]_{1,1}, \ldots, [\mathbf{\Sigma}^{-1}]_{L,L}\right]^T$, the trace term in (3.31) can be reorganized to obtain an equivalent minimization problem given by

$$\underset{\boldsymbol{\alpha}}{\mathrm{minimize}} \ \boldsymbol{\alpha}^T(\hat{\mathbf{R}}^T \odot \mathbf{R}_{\mathbf{\Sigma}}^{-1})\boldsymbol{\alpha} - \sum_{i=1}^{L} \log \alpha_i^2 \tag{3.32}$$

$$\text{subject to } \alpha_i \geq 0.$$

Note that, given the trace term in (3.31), the matrix $\hat{\mathbf{R}}^T \odot \mathbf{R}_{\mathbf{\Sigma}}^{-1}$ must be positive semidefinite since the trace of the product of two positive (semi)definite matrices is nonnegative. Hence, the problem (3.32) is convex with respect to the parameter

vector $\boldsymbol{\alpha}$ and, therefore, it can be efficiently solved using any convex optimization solver.

The proposed alternating minimization algorithm is summarized in Alg. 1. Once the estimates of $\hat{\mathbf{H}}_{\boldsymbol{\Sigma}}$ and $\hat{\boldsymbol{\Sigma}}$ under $\mathcal{H}_1$ are available, together with $\hat{\boldsymbol{\Sigma}}$ under $\mathcal{H}_0$ (given by (3.22)), they can be substituted into the GLRT in order to obtain the desired test statistic. Although the alternating minimization approach does not guarantee that the global maximizer of the log-likelihood is found, we will show through numerical experiments how the proposed iterative scheme outperforms other detectors that can be applied under the model considered in this section.

Low SNR approximation of the GLRT

The usefulness of the detector given in Alg. 1 in practical settings may be hindered by its complexity. In this context, simpler closed-form detectors become of practical interest. Now, we derive a closed-form expression for the GLRT in the low SNR regime, of particular interest in CR applications. As the SNR goes to zero, the covariance matrix will become close to diagonal, and thus it is possible to approximate the ML estimate of $\boldsymbol{\Sigma}^2$ as $\hat{\boldsymbol{\Sigma}}^2 \approx \hat{\mathbf{D}}$ defined in (3.22). Substituting this back into (3.28), we obtain the final compressed log-likelihood:

$$\log f\left(\mathbf{Y}\right) \;=\; -LK\log\pi \;-\; KP \;-\; K\log\det\left(\hat{\mathbf{R}}\right) \;-\; K\sum_{i=P+1}^{L}\left[\beta_i - \log\beta_i\right], \quad (3.33)$$

where β_i is the i-th largest eigenvalue of the sample spatial coherence matrix $\hat{\mathbf{C}} = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{R}}\hat{\mathbf{D}}^{-1/2}$. Then, the asymptotic log-GLRT is

$$\log T \approx K\sum_{i=1}^{P}\left[\beta_i - \log\beta_i\right] - KP. \quad (3.34)$$

Alternatively, (3.34) can be rewritten as

$$\log T \approx -KP - K\log\prod_{i=1}^{P}\beta_i e^{-\beta_i} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \eta, \quad (3.35)$$

and, thus, the test statistic is seen to be given by the product of the P largest eigenvalues of $\hat{\mathbf{C}}$, each equalized by an exponential term. Note that $\beta e^{-\beta}$ is maximum at $\beta = 1$. Hence, the statistic $\prod_{i=1}^{P}\beta_i e^{-\beta_i}$ measures, in some sense, how far the

vector of the P largest eigenvalues $[\beta_1 \cdots \beta_P]$ is from the vector of all ones. Note that (3.34) yields a closed-form test, in contrast with the iterative scheme presented in the previous section.

Remark 3.2. When $P = 1$, the test statistic depends only on $\beta_1 = \lambda_{\max}(\hat{\mathbf{C}}) \geq \mathrm{tr}\{\hat{\mathbf{C}}\}/K = 1$. Since the function xe^{-x} is monotonous for $x \geq 1$ it follows than an equivalent test statistic is given by

$$\lambda_{\max}(\hat{\mathbf{C}}) \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \eta', \qquad (3.36)$$

which recovers the detector proposed in López-Valcarce et al. (2010) for $P = 1$.

3.3.3 Numerical results and discussion

In this section we evaluate the performance of the proposed algorithms under different scenarios, by means of Monte Carlo simulations. Unless otherwise specified, the noise level at each antenna is fixed for each experiment, and for each Monte Carlo realization the entries of the channel matrix $\mathbf{H}$ are independently drawn from a Gaussian distribution (thus obtaining a Rayleigh fading scenario) and scaled so that the SNR is constant during the experiment:

$$\text{SNR (dB)} = 10 \log_{10} \frac{\mathrm{tr}(\mathbf{H}^H \mathbf{H})}{\mathrm{tr}(\mathbf{\Sigma}^2)}. \qquad (3.37)$$

We evaluate two detectors derived under the iid noise assumption: the proposed GLRT statistic in (3.21) denoted here by *iid-GLRT*, and the sphericity test or GLRT for non-structured primary signals (Mauchly, 1940) (denoted as *Sphericity*). In addition, three detectors derived for uncalibrated receivers ($\mathbf{\Sigma}^2$ diagonal with positive entries) are also evaluated: the proposed alternating optimization scheme from Algorithm 1 denoted here as *alterntng-GLRT*[2], the asymptotic closed-form detector in (3.34) (*asympt-GLRT*), the Hadamard ratio test (Wilks, 1935) or GLRT for unstructured primary signals (3.23) (*Hadamard*). Additionally, we also include two heuristic detectors for comparison: the detector based on statistical covariances

---

[2]Given the observed convergence properties, the iterations are stopped when the cost improvement between iterations is less than 10^{-5} with a maximum of 100 allowed iterations. As starting point we use an estimate given by the scaled low SNR asymptotic solution $\boldsymbol{\alpha}_0 = \sqrt{L/(L-P)} \left[[\hat{\mathbf{D}}]_{1,1}^{-1}, \ldots, [\hat{\mathbf{D}}]_{L,L}^{-1}\right]^T$.
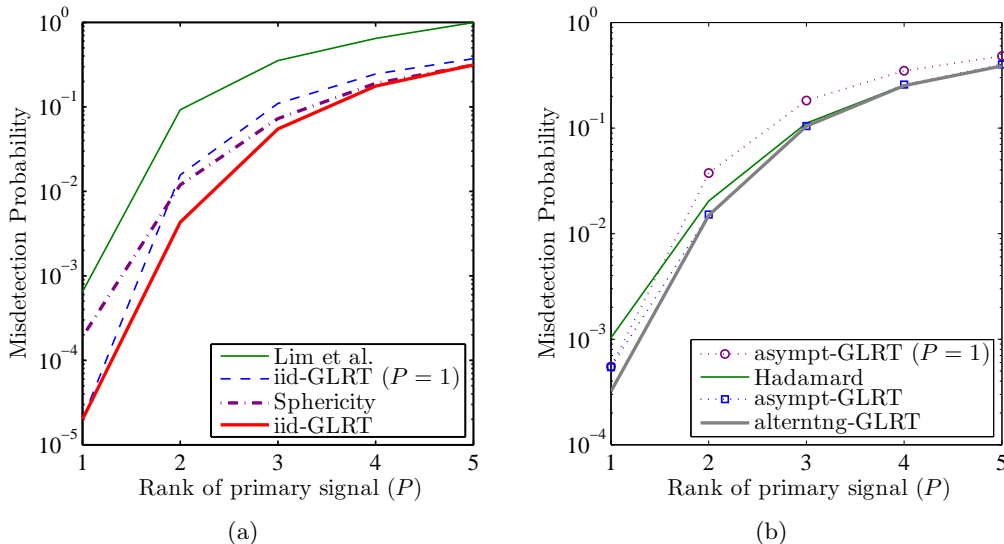
Figure 3.1: Misdetection probability versus P assuming (a) iid noise and (b) non-iid noise.

(Zeng and Liang, 2009b, Alg. 1) (*Covariance*) given by

$$\frac{\sum_{i=1}^{L}\sum_{j=1}^{L}\left|[\hat{\mathbf{R}}]_{ij}\right|}{\sum_{i=1}^{L}\left|[\hat{\mathbf{R}}]_{ii}\right|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_{Cov}, \tag{3.38}$$

where η_{Cov} is a suitably chosen threshold, and that of Eqn. (32) in Lim et al. (2008) given by

$$T_{Lim} = \log\left[\prod_{i=1}^{P}\frac{\delta_i}{\delta_L}\right]^{\frac{1}{P}} - \frac{1}{P}\sum_{i=1}^{P}\frac{\delta_i}{\delta_L} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta_{Lim}, \tag{3.39}$$

where $\delta_1 \geq \cdots \geq \delta_L$ are the eigenvalues of the sample covariance matrix $\hat{\mathbf{R}}$ and η_{Lim} denotes the threshold.

Detection performance for rank-P primary signals

First we compare the performance of the different schemes in terms of the spatial rank of the signal. Figure 3.1 shows the misdetection probability for fixed $P_{FA} = 0.01$ in a scenario with $L = 6$ antennas for primary signals with rank $P = 1, \ldots, 5$, for iid and non-iid noises. The iid scenario assumes SNR $= -6$ dB, $K = 128$ and noise level at all the antennas fixed to 0 dB while the non-iid scenario considers

SNR $= -4$ dB, $K = 64$ and noise power at each antenna fixed to $0, -1, 1, 0, -1$ and 0 dB respectively. Note that in both scenarios, as P increases the covariance matrix becomes less structured. This effect translates into the performance degradation for all the detectors under study with increasing P.

From Figs. 3.1 (a) and (b) we can see that for both iid and non-iid noises the corresponding GLRT detectors consistently provide the best performance for all values of P. While the GLRTs for $P = 1$ present poor performance if the actual rank of the signal is larger than one, the Sphericity and Hadamard ratio tests (which do not assume any structure on the primary signal) degrade for strong structure, i.e. small P. It is interesting to note, however, that as the rank of the signal grows (for $P \geq 4$) the Sphericity and Hadamard ratio tests offer similar performance to that of the rank-based detectors at a lower computational cost. Regarding the heuristic detectors, the covariance based detector (Zeng and Liang, 2009b) presents virtually the same performance as the Hadamard ratio test and it was not included in the plot for clarity. On the other hand, the poor performance of the detector of (Lim et al., 2008) for all values of P is likely rooted in the heuristic estimation of the noise variance.

Finally, it is interesting to note that for $P > 1$, the advantage of the iterative scheme *alterntng-GLRT* over the asymptotic GLRT decreases. This can be explained from the fact that, as the total SNR is divided among a growing number of dimensions, the effective SNR per dimension decreases and one gets closer to the asymptotic regime for which *asympt-GLRT* was derived.

Noise mismatch effect on detection performance

We now investigate the effect of a noise level mismatch at the different antennas on the different detectors. In order to focus on this effect we fix $P = 1$. Figure 3.2 shows the corresponding receiver operating characteristic (ROC) curves in a scenario with iid noises and with non-iid noises. In Fig. 3.2(a) we can see that for an scenario with iid noises (noise powers at each antenna equal to 0 dB) the *iid-GLRT* test, corresponding to the GLRT under this model, yields the best detection performance, whereas the detectors designed for disparate noise variances suffer a noticeable penalty. From the detectors designed for uncalibrated receivers, it is seen that the GLRT based schemes, both asymptotic and iterative, behave similarly and outperform the Hadamard ratio detector. The heuristic detector based on statistical covariances (Zeng and Liang, 2009b) presents almost the same performance as
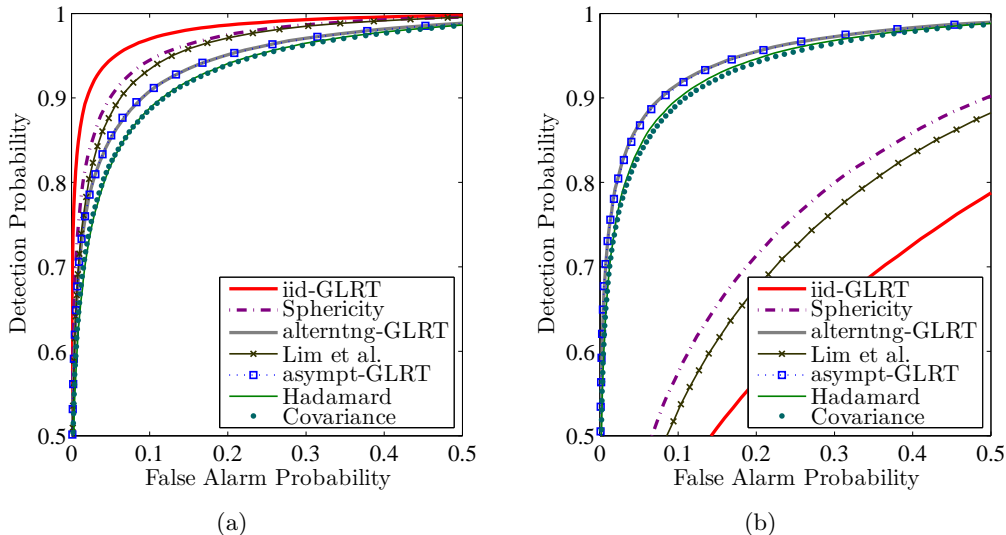
Figure 3.2: ROC curves (SNR$= -8$ dB, $P = 1$, $L = 4$, $K = 128$) (a) without noise power mismatch and (b) with noise power mismatch.

the Hadamard ratio test, while the detector of (Lim et al., 2008) suffers a penalty compared to the GLRT for the same model. Fig. 3.2(b) shows the ROC curves for a similar scenario, but with different noise variances across the antennas, now given by 0, -1, 1.5 and -0.5 dB. Note that the performance of the detectors designed for uncalibrated receivers has not changed with respect to that in Fig. 3.2(a), whereas that of the detectors based on the iid noise assumption is severely degraded.

Asymptotic GLRT performance for finite SNR values

Although the asymptotic GLRT detector (*asympt-GLRT*) given by (3.34) is appealing due to its computational simplicity, it is not clear how much can be gained when the iterative scheme (*alterntng-GLRT*) is used in order to implement the exact GLRT. Fig. 3.3 shows the missed detection probability of the detectors versus the SNR in a scenario similar to that of the previous subsection ($P = 1$, $L = 4$, $K = 128$, different noise levels at each of the antennas fixed to 0, -1, 1.5 and -0.5 dB respectively). The probability of false alarm is fixed to $P_{\text{FA}} = 0.01$ and 0.1. In Fig. 3.3 it is seen that, as expected, for very low SNR values the asymptotic detector presents the same performance as the alternating minimization scheme. However, as the SNR increases, the GLRT outperforms the detector derived for asymptotically
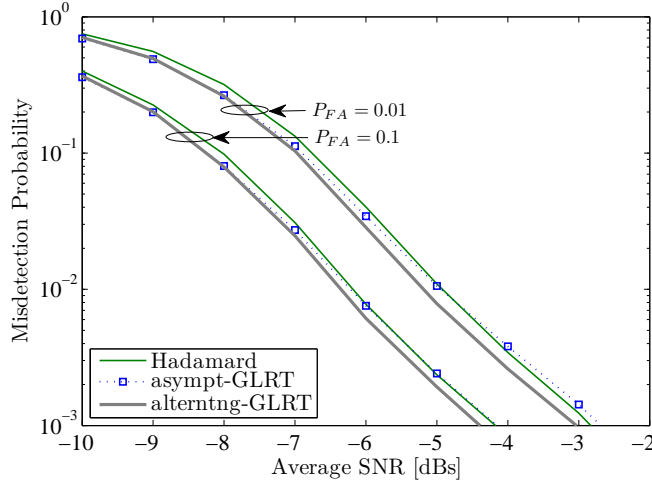
Figure 3.3: Misdetection probability versus SNR for different detectors. Same scenario as in Fig. 3.2(b), with $P_{\mathrm{FA}} = 0.01$ and 0.1.

low SNR, as it could be expected. Note, however, that the performance loss of the asymptotic detector is rather small, and therefore it offers a good tradeoff between performance and complexity.

3.4 Detection of rank-1 signals in spatially correlated noise

In the previous section we assumed the noise process uncorrelated across antennas. While this accurately models the thermal noise effects, other contributions such as cochannel interference will present spatial correlation. In this section no assumptions are made about $\mathbf{\Sigma}^2$ other than to have full rank.

Under the Gaussianity assumption, if we do not impose any constraint on the the spatial covariance of the noise process, we must resort to additional information (other than spatial) in order to detect the primary signal. To this end, in this section, we assume that the primary signal presents certain temporal correlation which translates into $\mathbf{C} \neq \mathbf{I}_K$. The matrix $\mathbf{C}$ is assumed to be available to the spectral monitor as *a priori* information. In this section we will restrict our study to primary signals of rank one, thus, $\mathbf{H} = \mathbf{h}^H$ with $\mathbf{h}$ an $L \times 1$ vector and $\mathbf{S} = \mathbf{s}$ is

$K \times 1$. Then we have that the received signal reduces to

$$\mathbf{Y} = \mathbf{s}\mathbf{h}^H + \mathbf{N}\mathbf{G}. \tag{3.40}$$

In the following we will find useful to define the maximum achievable SNR as

$$\rho \doteq \mathbf{h}^H \mathbf{\Sigma}^{-2} \mathbf{h} = \|\mathbf{h}_{\mathbf{\Sigma}}\|_2^2, \tag{3.41}$$

where the prewhitened channel is given by

$$\mathbf{h}_{\mathbf{\Sigma}} \doteq \mathbf{\Sigma}^{-1} \mathbf{h}. \tag{3.42}$$

Note that $\rho = \mathbf{h}^H \mathbf{\Sigma}^{-2} \mathbf{h}$ is the maximum SNR that can be obtained at the output of a linear combiner (beamformer), i.e. if we let $\mathbf{x} = \mathbf{Y}\mathbf{v} = \mathbf{s}\mathbf{h}^H\mathbf{v} + \mathbf{N}\mathbf{G}\mathbf{v}$ with $\mathbf{v}$ the $L \times 1$ beamformer, then the SNR of $\mathbf{x}$ is given by

$$\zeta \doteq \frac{\mathrm{E}[\mathbf{v}^H \mathbf{h}\mathbf{s}^H \mathbf{s}\mathbf{h}^H \mathbf{v}]}{\mathrm{E}[\mathbf{v}^H \mathbf{G}^H \mathbf{N}^H \mathbf{N}\mathbf{G}\mathbf{v}]} = \frac{\mathbf{v}^H \mathbf{h}\mathbf{h}^H \mathbf{v}}{\mathbf{v}^H \mathbf{\Sigma}^2 \mathbf{v}}, \tag{3.43}$$

which attains its maximum value $\zeta = \mathbf{h}^H \mathbf{\Sigma}^{-2} \mathbf{h}$ for $\mathbf{v} = \mathbf{\Sigma}^{-2} \mathbf{h}$.

As in the previous section $\mathbf{h}$ and $\mathbf{\Sigma}^2 = \mathbf{G}^H \mathbf{G}$ are modeled as deterministic and unknown. However, as a preliminary step before the derivation of the GLRT for this problem, we study the optimal detector assuming *a priori* knowledge of $\mathbf{h}$ and $\mathbf{\Sigma}^2$.

3.4.1 Genie-aided detectors

If $\mathbf{h}$ and $\mathbf{\Sigma}^2$ are assumed known the optimal detector is given by the Neyman-Pearson test. Let us consider the whitened received signal, defined as

$$\mathbf{y}_{\mathbf{\Sigma}} \doteq \left(\mathbf{\Sigma}^{-T} \otimes \mathbf{I} \right) \mathbf{y}, \tag{3.44}$$

as the new input to the detector. It is easy to see that in this case the system model corresponds to the one considered in Chapter 2 when $\sigma^2 = 1$. Hence, following similar steps the optimal NP test can be found to be

$$T_{\mathrm{NP}} = \mathbf{y}_{\mathbf{\Sigma}}^H \left[(\mathbf{h}_{\mathbf{\Sigma}}\mathbf{h}_{\mathbf{\Sigma}}^H)^T \otimes \mathbf{C} \right] \left[\mathbf{I}_{LK} + (\mathbf{h}_{\mathbf{\Sigma}}\mathbf{h}_{\mathbf{\Sigma}}^H)^T \otimes \mathbf{C} \right]^{-1} \mathbf{y}_{\mathbf{\Sigma}} \tag{3.45}$$

where we have defined the whitened channel as $\mathbf{h}_{\mathbf{\Sigma}} = \mathbf{\Sigma}^{-1} \mathbf{h}$. Using the definition of the Kronecker product and the Matrix Inversion Lemma, we have that the inverse

$[\mathbf{I}_{LK} + (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \mathbf{C}]^{-1}$ can be rewritten as[3]

$$[\mathbf{I}_{LK} + (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \mathbf{C}]^{-1} = \mathbf{I}_{LK} - (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \mathbf{C}(\rho), \tag{3.46}$$

where we defined

$$\mathbf{C}(\rho) \doteq \mathbf{C}^{1/2}(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{C}^{1/2}. \tag{3.47}$$

Note that $\mathbf{C}(0) = \mathbf{C}$, whereas $\mathbf{C}(\rho) \approx 1/\rho\mathbf{I}_K$ for $\rho \to \infty$. Applying this result to (3.45) we obtain

$$\begin{align}
T_{\mathrm{NP}} &= \mathbf{y}_{\boldsymbol{\Sigma}}^H \left[(\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \mathbf{C}\right] \left[\mathbf{I}_{LK} - (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \mathbf{C}(\rho)\right] \mathbf{y}_{\boldsymbol{\Sigma}} \tag{3.48} \\
&= \mathbf{y}_{\boldsymbol{\Sigma}}^H \left[(\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \mathbf{C} - (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes (\mathbf{C}\mathbf{C}(\rho))\right] \mathbf{y}_{\boldsymbol{\Sigma}} \tag{3.49} \\
&= \mathbf{y}_{\boldsymbol{\Sigma}}^H \left[(\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \mathbf{C} - (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes (\rho\mathbf{C}\mathbf{C}(\rho))\right] \mathbf{y}_{\boldsymbol{\Sigma}} \tag{3.50} \\
&= \mathbf{y}_{\boldsymbol{\Sigma}}^H \left[(\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes (\mathbf{C}(\mathbf{I}_K - \rho\mathbf{C}(\rho)))\right] \mathbf{y}_{\boldsymbol{\Sigma}} \tag{3.51} \\
&= \mathbf{y}_{\boldsymbol{\Sigma}}^H \left[(\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^H)^T \otimes \left(\mathbf{C}(\mathbf{I}_K + \rho\mathbf{C})^{-1}\right)\right] \mathbf{y}_{\boldsymbol{\Sigma}} \tag{3.52}
\end{align}$$

where in (3.50) we have used that $\mathbf{h}_{\boldsymbol{\Sigma}}^H\mathbf{h}_{\boldsymbol{\Sigma}} = \mathbf{h}^H\boldsymbol{\Sigma}^{-2}\mathbf{h} = \rho$ and in (3.52) we made use of the identity $\mathbf{I}_K - \rho\mathbf{C}(\rho) = (\mathbf{I}_K + \rho\mathbf{C})^{-1}$.

We now apply the property

$$\mathrm{tr}(\mathbf{A}^T\mathbf{B}^H\mathbf{C}\mathbf{D}) = \mathrm{vec}(\mathbf{B})^H(\mathbf{A} \otimes \mathbf{C})\mathrm{vec}(\mathbf{D}) \tag{3.53}$$

in (3.52) to obtain, after some straighforward manipulations,

$$\begin{align}
T_{\mathrm{NP}} &= \mathbf{h}_{\boldsymbol{\Sigma}}^H\mathbf{Y}_{\boldsymbol{\Sigma}}^H\mathbf{C}(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}} \tag{3.54} \\
&= \mathbf{h}^H\boldsymbol{\Sigma}^{-2}\mathbf{Y}^H\mathbf{C}(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y}\boldsymbol{\Sigma}^{-2}\mathbf{h} \tag{3.55}
\end{align}$$

where we have defined $\mathbf{Y}_{\boldsymbol{\Sigma}} \doteq \mathrm{vec}(\mathbf{y}_{\boldsymbol{\Sigma}}) = \mathbf{Y}\boldsymbol{\Sigma}^{-1}$.

Note that $\mathbf{v} = \boldsymbol{\Sigma}^{-2}\mathbf{h}$ corresponds to the optimum beamformer, which maximizes the SNR at its output $\mathbf{x} = \mathbf{Y}\mathbf{v}$. Then the NP test can be written as

$$\begin{align}
T_{\mathrm{NP}} &= \mathbf{x}^H\mathbf{C}(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{x} \tag{3.56} \\
&\approx \mathbf{x}^H\mathbf{W}\boldsymbol{\Lambda}(\mathbf{I}_K + \rho\boldsymbol{\Lambda})^{-1}\mathbf{W}^H\mathbf{x}, \tag{3.57}
\end{align}$$

where have made use of the asymptotic EVD of the covariance matrix $\mathbf{C} \approx \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^H$,

---

[3]Intermediate steps can be found in Appendix 3.B.

first introduced in (2.5). By examining (3.57) the structure of the NP test becomes clear. First, the received signal is fed through the optimal spatial beamformer $\mathbf{x} = \mathbf{Y}\mathbf{v}$. This signal is then transformed into the frequency domain by applying the FFT transform $\mathbf{W}^H$ and fed through the optimal Wiener filter

$$F(\mathrm{e}^{\jmath\omega}) = \left(\frac{S_{ss}(\mathrm{e}^{\jmath\omega})}{1 + \rho S_{ss}(\mathrm{e}^{\jmath\omega})} \right)^{1/2}, \tag{3.58}$$

before the energy of the resulting signal is compared against a threshold. That is, the optimal NP detector relies on two key components. The first is the optimum beamformer $\mathbf{v}$ which maximizes the received SNR by discarding the noise-only spatial dimensions. In second place we have an optimum Wiener filter which acts as a matched filter in the temporal domain. At this point it is instructive to consider the following asymptotic cases, in which the structure of this filter simplifies:

- In the high SNR regime, $\rho \to \infty$, i.e., $F(\mathrm{e}^{\jmath\omega}) \to \rho^{-1/2}$ and the detector amounts to comparing the energy at the output of the optimum beamformer $\|\mathbf{x}\|^2 = \mathbf{x}^H\mathbf{x}$ with a given threshold.

- In the low SNR regime, $\rho \to 0$, so that $\mathbf{C}(\mathbf{I}_K + \rho\mathbf{C})^{-1} \to \mathbf{C}$, and the test reduces to comparing $\mathbf{x}^H\mathbf{C}\mathbf{x}$ against a threshold.

However, neither $\mathbf{h}$ nor $\boldsymbol{\Sigma}^2$ can be assumed known and need to be estimated. The ML estimation of these parameters under both Hypothesis conducts to the GLRT detector.

3.4.2 GLRT detector

We now proceed to the derivation of the GLRT based on (3.40), as presented in (3.9).

ML parameter estimation under $\mathcal{H}_0$

Under $\mathcal{H}_0$ the signal covariance is given by $\mathcal{R}_0 = (\boldsymbol{\Sigma}^2)^T \otimes \mathbf{I}_K$ where $\boldsymbol{\Sigma}^2$ is unstructured.

Lemma 3.5. *Under $\mathcal{H}_0$, the ML estimate of the (unstructured) noise covariance matrix is given by*

$$\hat{\boldsymbol{\Sigma}}^2 = \hat{\mathbf{R}} = \frac{1}{K}\mathbf{Y}^H\mathbf{Y}. \tag{3.59}$$

Proof. Under $\mathcal{H}_0$ the log-likelihood reduces to

$$\log f(\mathbf{Y} \mid \boldsymbol{\Sigma}) = -LK \log \pi - K \log \det(\boldsymbol{\Sigma}^2) - K \operatorname{tr}\left(\hat{\mathbf{R}}\boldsymbol{\Sigma}^{-2}\right), \qquad (3.60)$$

which is maximized for $\boldsymbol{\Sigma}^2 = \hat{\mathbf{R}}$ (Magnus and Neudecker, 1999). $\qquad\square$

ML parameter estimation under $\mathcal{H}_1$

Under $\mathcal{H}_1$, the log-likelihood is given by

$$\log f(\mathbf{y} \mid \mathcal{R}_1) = -LK \log \pi - \log \det(\mathcal{R}_1) - \mathbf{y}^H \mathcal{R}_1^{-1} \mathbf{y}, \qquad (3.61)$$

with

$$\mathcal{R}_1 = (\boldsymbol{\Sigma}^2)^T \otimes \mathbf{I}_K + (\mathbf{h}\mathbf{h}^H)^T \otimes \mathbf{C}. \qquad (3.62)$$

Note that (3.61) offers little information about the ML estimates of $\boldsymbol{\Sigma}^2$ and $\mathbf{h}$. However, from the structure of $\mathcal{R}_1$ in (3.62) it is apparent that spatial and temporal correlations are separable in terms of the Kronecker product. Using this fact, we derive next a simpler expression for (3.61) which will allow us to reduce the optimization problem in $\boldsymbol{\Sigma}^2$ and $\mathbf{h}$ to a scalar optimization problem.

First, the inverse of $\mathcal{R}_1$ can be written as

$$\mathcal{R}_1^{-1} = \left[(\boldsymbol{\Sigma}^2)^T \otimes \mathbf{I}_K + (\mathbf{h}\mathbf{h}^H)^T \otimes \mathbf{C}\right]^{-1} \qquad (3.63)$$

$$= \left[(\boldsymbol{\Sigma}^{-1})^T \otimes \mathbf{I}_K\right] \left[\mathbf{I}_{LK} + (\boldsymbol{\Sigma}^{-1}\mathbf{h}\mathbf{h}^H\boldsymbol{\Sigma}^{-1})^T \otimes \mathbf{C}\right]^{-1} \left[(\boldsymbol{\Sigma}^{-1})^T \otimes \mathbf{I}_K\right] \qquad (3.64)$$

$$= (\boldsymbol{\Sigma}^{-2})^T \otimes \mathbf{I}_K - (\boldsymbol{\Sigma}^{-2}\mathbf{h}\mathbf{h}^H\boldsymbol{\Sigma}^{-2})^T \otimes \mathbf{C}(\rho) \qquad (3.65)$$

where in the last step we used (3.46). Now applying the property (3.53) some straightforward algebra yields

$$\mathbf{y}^H \mathcal{R}_1^{-1} \mathbf{y} = \operatorname{tr}\{\boldsymbol{\Sigma}^{-2}\mathbf{Y}^H\mathbf{Y} - \boldsymbol{\Sigma}^{-2}\mathbf{h}\mathbf{h}^H\boldsymbol{\Sigma}^{-2}\mathbf{Y}^H\mathbf{C}(\rho)\mathbf{Y}\}. \qquad (3.66)$$

On the other hand, using a generalization of Sylvester's determinant identity[4]

---

[4] $\det(\mathbf{A} + \mathbf{U}\mathbf{V}^H) = \det(\mathbf{A})\det(\mathbf{I} + \mathbf{V}^H\mathbf{A}^{-1}\mathbf{U})$.

to Kronecker products, we can rewrite the determinant term in (3.61) as

$$\log \det(\mathcal{R}_1) = \log \det \left((\mathbf{\Sigma}^2)^T \otimes \mathbf{I}_K + (\mathbf{h}\mathbf{h}^H)^T \otimes \mathbf{C} \right) \tag{3.67}$$

$$= K \log \det(\mathbf{\Sigma}^2) + \log \det(\mathbf{I}_K + \rho \mathbf{C}). \tag{3.68}$$

Substituting (3.68) and (3.66) into the log-likelihood (3.61) yields

$$\log f(\mathbf{y} \,|\, \mathbf{h}_{\mathbf{\Sigma}}, \mathbf{\Sigma}) = -LK \log \pi - K \log \det \mathbf{\Sigma}^2 - \log \det(\mathbf{I}_K + \rho \mathbf{C})$$
$$- \operatorname{tr}\left\{ \mathbf{\Sigma}^{-2} \mathbf{Y}^H \mathbf{Y} \right\} + \operatorname{tr}\left\{ \mathbf{h}_{\mathbf{\Sigma}} \mathbf{h}_{\mathbf{\Sigma}}^H \mathbf{\Sigma}^{-1} \mathbf{Y}^H \mathbf{C}(\rho) \mathbf{Y} \mathbf{\Sigma}^{-1} \right\}. \tag{3.69}$$

Defining a new parameter space

In order to maximize (3.69) we will find useful to replace the parameter space $\{\mathbf{h}_{\mathbf{\Sigma}}, \mathbf{\Sigma}^{-2}\}$ by a more convenient form. First, using the following identity: $\mathbf{I}_K - \rho \mathbf{C}(\rho) = (\mathbf{I}_K + \rho \mathbf{C})^{-1}$, we may reorganize the terms in (3.69) to obtain

$$\log f(\mathbf{y} \,|\, \rho, \bar{\mathbf{h}}_{\mathbf{\Sigma}}, \mathbf{\Sigma}^{-2}) = -LK \log \pi - K \log \det \mathbf{\Sigma}^2 - \log \det(\mathbf{I}_K + \rho \mathbf{C})$$
$$- \operatorname{tr}\left\{ \left(\mathbf{I}_L - \bar{\mathbf{h}}_{\mathbf{\Sigma}} \bar{\mathbf{h}}_{\mathbf{\Sigma}}^H \right) \mathbf{\Sigma}^{-1} \mathbf{Y}^H \mathbf{Y} \mathbf{\Sigma}^{-1} \right\} - \operatorname{tr}\left\{ \bar{\mathbf{h}}_{\mathbf{\Sigma}} \bar{\mathbf{h}}_{\mathbf{\Sigma}}^H \mathbf{\Sigma}^{-1} \mathbf{Y}^H (\mathbf{I}_K + \rho \mathbf{C})^{-1} \mathbf{Y} \mathbf{\Sigma}^{-1} \right\}.$$
$$\tag{3.70}$$

where we have introduced the unit-norm vector $\bar{\mathbf{h}}_{\mathbf{\Sigma}} \doteq \mathbf{h}_{\mathbf{\Sigma}} / \|\mathbf{h}_{\mathbf{\Sigma}}\|_2$.

This step is important in what follows because it divides the spatial contribution of the prewhitened received signal $\mathbf{Y}\mathbf{\Sigma}^{-1}$ in two different terms: one orthogonal and one collinear to the prewhitened channel $\mathbf{h}_{\mathbf{\Sigma}}$, respectively.

We now define the two quasi-orthogonal contributions to the inverse of the spatial covariance matrix of the noise $\mathbf{\Sigma}^{-2} = \mathbf{W}_{\perp} + \mathbf{W}_{\|}$ as

$$\mathbf{W}_{\perp} \doteq \mathbf{\Sigma}^{-1} \left(\mathbf{I}_L - \bar{\mathbf{h}}_{\mathbf{\Sigma}} \bar{\mathbf{h}}_{\mathbf{\Sigma}}^H \right) \mathbf{\Sigma}^{-1} = \mathbf{U}_{\perp} \mathbf{\Delta}_{\perp} \mathbf{U}_{\perp}^H, \tag{3.71}$$

$$\mathbf{W}_{\|} \doteq \mathbf{\Sigma}^{-1} \bar{\mathbf{h}}_{\mathbf{\Sigma}} \bar{\mathbf{h}}_{\mathbf{\Sigma}}^H \mathbf{\Sigma}^{-1} = \delta_{\|} \mathbf{u}_{\|} \mathbf{u}_{\|}^H, \tag{3.72}$$

where $\mathbf{U}_{\perp} \mathbf{\Delta}_{\perp} \mathbf{U}_{\perp}^H$ denotes the economy-size eigendecomposition of the matrix $\mathbf{W}_{\perp}$ (of rank $L-1$), and the economy-size eigendecomposition of the rank-1 matrix $\mathbf{W}_{\|}$ is given by $\delta_{\|} \mathbf{u}_{\|} \mathbf{u}_{\|}^H$. Then we have the following result:

Lemma 3.6. *The log-likelihood (3.70) can be rewritten in terms of a new parameter*

space given by $\{\bar{\mathbf{h}}, \rho, \mathbf{U}_\perp, \boldsymbol{\Delta}_\perp, \mathbf{u}_\parallel, \delta_\parallel\}$ *as*

$$\log f(\mathbf{y} \mid \rho, \bar{\mathbf{h}}, \delta_\parallel, \mathbf{u}_\parallel, \boldsymbol{\Delta}_\perp, \mathbf{U}_\perp) =$$
$$- LK \log \pi + K \log \det \boldsymbol{\Delta}_\perp + K \log \delta_\parallel + K \log |\bar{\mathbf{h}}^H \mathbf{u}_\parallel|^2 - \log \det(\mathbf{I}_K + \rho\mathbf{C})$$
$$- \operatorname{tr}\{\mathbf{U}_\perp \boldsymbol{\Delta}_\perp \mathbf{U}_\perp^H \mathbf{Y}^H \mathbf{Y}\} - \operatorname{tr}\{\delta_\parallel \mathbf{u}_\parallel \mathbf{u}_\parallel^H \mathbf{Y}^H (\mathbf{I}_K + \rho\mathbf{C})^{-1} \mathbf{Y}\}, \quad (3.73)$$

with $\bar{\mathbf{h}} \doteq \mathbf{h}/\|\mathbf{h}\|_2$.

Proof. We have that $\mathbf{U}_\perp$ is an $L \times (L-1)$ semiunitary matrix. Its nullspace is given by the direction $\bar{\mathbf{h}}$ since

$$\mathbf{W}_\perp \bar{\mathbf{h}} = \boldsymbol{\Sigma}^{-1} \left(\mathbf{I}_L - \bar{\mathbf{h}}_{\boldsymbol{\Sigma}} \bar{\mathbf{h}}_{\boldsymbol{\Sigma}}^H \right) \boldsymbol{\Sigma}^{-1} \bar{\mathbf{h}} \tag{3.74}$$

$$= \left(\boldsymbol{\Sigma}^{-2} - \boldsymbol{\Sigma}^{-2} \frac{\mathbf{h}\mathbf{h}^H}{\mathbf{h}^H \boldsymbol{\Sigma}^{-2} \mathbf{h}} \boldsymbol{\Sigma}^{-2} \right) \frac{\mathbf{h}}{\|\mathbf{h}\|_2} \tag{3.75}$$

$$= \frac{1}{\|\mathbf{h}\|_2} \left(\boldsymbol{\Sigma}^{-2} \mathbf{h} - \boldsymbol{\Sigma}^{-2} \mathbf{h} \frac{\mathbf{h}^H \boldsymbol{\Sigma}^{-2} \mathbf{h}}{\mathbf{h}^H \boldsymbol{\Sigma}^{-2} \mathbf{h}} \right) = \mathbf{0}. \tag{3.76}$$

As a result $[\mathbf{U}_\perp \ \bar{\mathbf{h}}]$ defines a unitary matrix and the full eigenvalue decomposition of $\mathbf{W}_\perp$ is given by $\mathbf{W}_\perp = [\mathbf{U}_\perp \ \bar{\mathbf{h}}]\tilde{\boldsymbol{\Delta}}_\perp [\mathbf{U}_\perp \ \bar{\mathbf{h}}]^H$, where we defined

$$\tilde{\boldsymbol{\Delta}}_\perp \doteq \begin{bmatrix} \boldsymbol{\Delta}_\perp & \\ & 0 \end{bmatrix}. \tag{3.77}$$

Now, the determinant of $\boldsymbol{\Sigma}^{-2}$ can be rewritten as

$$\det \boldsymbol{\Sigma}^{-2} = \det(\mathbf{W}_\perp + \mathbf{W}_\parallel) \tag{3.78}$$

$$= \det \left([\mathbf{U}_\perp \ \bar{\mathbf{h}}]\tilde{\boldsymbol{\Delta}}_\perp [\mathbf{U}_\perp \ \bar{\mathbf{h}}]^H + \mathbf{u}_\parallel \delta_\parallel \mathbf{u}_\parallel^H \right) \tag{3.79}$$

$$= \det \left(\tilde{\boldsymbol{\Delta}}_\perp + [\mathbf{U}_\perp \ \bar{\mathbf{h}}]^H \mathbf{u}_\parallel \delta_\parallel \mathbf{u}_\parallel^H [\mathbf{U}_\perp \ \bar{\mathbf{h}}] \right) \tag{3.80}$$

$$= \det(\tilde{\boldsymbol{\Delta}}_\perp) + \delta_\parallel \mathbf{u}_\parallel^H [\mathbf{U}_\perp \ \bar{\mathbf{h}}] \operatorname{adj}(\tilde{\boldsymbol{\Delta}}_\perp) [\mathbf{U}_\perp \ \bar{\mathbf{h}}]^H \mathbf{u}_\parallel \tag{3.81}$$

$$= \delta_\parallel \mathbf{u}_\parallel^H [\mathbf{U}_\perp \ \bar{\mathbf{h}}] [\mathbf{e}_L \det(\boldsymbol{\Delta}_\perp) \mathbf{e}_L^T] [\mathbf{U}_\perp \ \bar{\mathbf{h}}]^H \mathbf{u}_\parallel \tag{3.82}$$

$$= \delta_\parallel \mathbf{u}_\parallel^H \bar{\mathbf{h}} \det(\boldsymbol{\Delta}_\perp) \bar{\mathbf{h}}^H \mathbf{u}_\parallel \tag{3.83}$$

$$= \delta_\parallel |\bar{\mathbf{h}}^H \mathbf{u}_\parallel|^2 \det(\boldsymbol{\Delta}_\perp), \tag{3.84}$$

where in (3.81) we have used that (see e.g. (Lancaster and Tismenetsky, 1985, p. 65)) $\det(\boldsymbol{X} + \boldsymbol{y}\boldsymbol{z}^T) = \det(\boldsymbol{X}) + \boldsymbol{z}^T \operatorname{adj}(\boldsymbol{X})\boldsymbol{y}$ where $\operatorname{adj}(\cdot)$ denotes the adjugate matrix, and in (3.82) we have used that $\det(\tilde{\boldsymbol{\Delta}}_\perp) = 0$ and that the adjugate matrix of a

diagonal matrix with the i-th diagonal entry equal to zero, is the zero matrix except for the (i,i)-th component.

Substituting (3.71), (3.72), (3.84) in the likelihood expression given in (3.70) we obtain the desired result. □

That is, in order to maximize (3.70) with respect to $\{\rho, \bar{\mathbf{h}}_{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}^{-2}\}$, we expressed the log-likelihood in terms of an alternative set of parameters, namely $\{\bar{\mathbf{h}}, \rho, \mathbf{U}_\perp, \boldsymbol{\Delta}_\perp, \mathbf{u}_\parallel, \delta_\parallel\}$. Note that the number of degrees of freedom of the two sets of parameters is the same. The first set of parameters presents L^2 degrees of freedom in the Hermitian matrix $\boldsymbol{\Sigma}^{-2}$, $2L - 1$ in the unitary vector $\bar{\mathbf{h}}_{\boldsymbol{\Sigma}}$ and 1 in the effective SNR ρ, that is, a total of $L^2 + 2L$ degrees of freedom. On the other hand, $[\mathbf{U}_\perp \ \bar{\mathbf{h}}]$ defines an unitary matrix with $L^2 - L$ degrees of freedom, $\{\mathbf{u}_\parallel, \delta_\parallel\}$ define a complex vector with $2L$ degrees of freedom, and $\boldsymbol{\Delta}_\perp$ and ρ present $L - 1$ and 1 degrees of freedom respectively, adding up to a total of $L^2 + 2L$ degrees of freedom. Additionally, it is possible to establish a biunivocal mapping between the two sets of parameters. Given $\boldsymbol{\Sigma}^{-2}$, $\bar{\mathbf{h}}_{\boldsymbol{\Sigma}}$, the matrices $\mathbf{W}_\perp$ and $\mathbf{W}_\parallel$ can be computed using (3.71)-(3.72), and viceversa, given $\mathbf{W}_\perp$, $\mathbf{W}_\parallel$, we may compute $\boldsymbol{\Sigma}^{-2}$, $\bar{\mathbf{h}}_{\boldsymbol{\Sigma}}$ as

$$\boldsymbol{\Sigma}^{-2} = \mathbf{W}_\perp + \mathbf{W}_\parallel, \qquad\qquad \bar{\mathbf{h}}_{\boldsymbol{\Sigma}} = \frac{\boldsymbol{\Sigma}\mathbf{u}_\parallel}{\|\boldsymbol{\Sigma}\mathbf{u}_\parallel\|_2}. \qquad (3.85)$$

Hence, the maximization can be carried out over the new set of parameters given by $\{\bar{\mathbf{h}}, \rho, \mathbf{U}_\perp, \boldsymbol{\Delta}_\perp, \mathbf{u}_\parallel, \delta_\parallel\}$.

Compressing the log-likelihood

We must now maximize the log-likelihood with respect to the unknown parameters. Let us define to this end the "orthogonal component" of the observations as

$$(\mathbf{Y}^H\mathbf{Y})_\perp \doteq (\mathbf{I}_L - \bar{\mathbf{h}}\bar{\mathbf{h}}^H)\mathbf{Y}^H\mathbf{Y}(\mathbf{I}_L - \bar{\mathbf{h}}\bar{\mathbf{h}}^H). \qquad (3.86)$$

We recall from (3.76) that $\mathbf{W}_\perp\bar{\mathbf{h}} = 0$, which in turn implies that $\mathbf{U}_\perp\bar{\mathbf{h}} = 0$. It then follows that $\mathbf{U}_\perp^H\mathbf{Y}^H\mathbf{Y}\mathbf{U}_\perp = \mathbf{U}_\perp^H(\mathbf{Y}^H\mathbf{Y})_\perp\mathbf{U}_\perp$. Consider now an economy-size EVD of the rank-$(L-1)$ matrix $(\mathbf{Y}^H\mathbf{Y})_\perp$:

$$(\mathbf{Y}^H\mathbf{Y})_\perp = \mathbf{U}_{\mathbf{Y}\perp}\boldsymbol{\Delta}_{\mathbf{Y}\perp}\mathbf{U}_{\mathbf{Y}\perp}^H. \qquad (3.87)$$

Observe that $(\mathbf{Y}^H\mathbf{Y})_\perp\bar{\mathbf{h}} = 0$, and therefore $\mathbf{U}_\perp\bar{\mathbf{h}} = 0$. Since both $\mathbf{U}_\perp$, $\mathbf{U}_{\mathbf{Y}\perp}$ are $L \times (L-1)$ matrices with orthonormal columns, and both of them are orthogonal to the same vector $\bar{\mathbf{h}}$, it follows that $\mathbf{U}_\perp$, $\mathbf{U}_{\mathbf{Y}\perp}$ span the same subspace. Hence, there must exist an $(L-1) \times (L-1)$ unitary matrix $\mathbf{U}_\Omega$ such that

$$\mathbf{U}_\perp = \mathbf{U}_{\mathbf{Y}\perp}\mathbf{U}_\Omega. \tag{3.88}$$

Using the definition of $\mathbf{U}_\Omega$ in (3.88), the log-likelihood function can be written as

$$\log f(\mathbf{y} \mid \rho, \bar{\mathbf{h}}, \mathbf{u}_\parallel, \delta_\parallel, \boldsymbol{\Delta}_\perp, \mathbf{U}_\Omega) =$$
$$- LK\log\pi + K\log\det\boldsymbol{\Delta}_\perp + K\log\delta_\parallel + K\log|\bar{\mathbf{h}}^H\mathbf{u}_\parallel|^2 - \log\det(\mathbf{I}_K + \rho\mathbf{C})$$
$$- \mathrm{tr}\{\mathbf{U}_\Omega\boldsymbol{\Delta}_\perp\mathbf{U}_\Omega\boldsymbol{\Delta}_{\mathbf{Y}\perp}\} - \delta_\parallel\mathbf{u}_\parallel^H\mathbf{Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y}\mathbf{u}_\parallel. \tag{3.89}$$

The only term depending on $\mathbf{U}_\Omega$ is $-\mathrm{tr}\{\mathbf{U}_\Omega\boldsymbol{\Delta}_\perp\mathbf{U}_\Omega\boldsymbol{\Delta}_{\mathbf{Y}\perp}\}$. For the diagonal elements of $\boldsymbol{\Delta}_{\mathbf{Y}\perp}$ sorted in non-increasing order and the diagonal elements of $\boldsymbol{\Delta}_\perp$ sorted in non-decreasing order (to be checked later) the term $-\mathrm{tr}\{\mathbf{U}_\Omega\boldsymbol{\Delta}_\perp\mathbf{U}_\Omega\boldsymbol{\Delta}_{\mathbf{Y}\perp}\}$ is maximized with respect to $\mathbf{U}_\Omega$ for (Fraikin et al., 2008)

$$\hat{\mathbf{U}}_\Omega = \mathbf{I}_{L-1}. \tag{3.90}$$

Now, the maximization of (3.89) with respect to $\delta_\parallel$ and $\hat{\boldsymbol{\Delta}}_\perp$ is straightforward and yields

$$\hat{\delta}_\parallel = K(\mathbf{u}_\parallel^H\mathbf{Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y}\mathbf{u}_\parallel)^{-1}, \tag{3.91}$$

$$\hat{\boldsymbol{\Delta}}_\perp = K\boldsymbol{\Delta}_{\mathbf{Y}\perp}^{-1}. \tag{3.92}$$

Note that indeed since the diagonal elements of $\boldsymbol{\Delta}_{\mathbf{Y}\perp}$ are sorted in non-increasing order, the diagonal elements of $\boldsymbol{\Delta}_\perp$ are sorted in non-decreasing order. Hence the assumption in obtaining (3.90) was right. Substituting (3.90), (3.91) and (3.92) into (3.89), we obtain the compressed log-likelihood function

$$\log f(\mathbf{y} \mid \rho, \bar{\mathbf{h}}, \mathbf{u}_\parallel) = -LK\log\pi + LK\log K - LK + K\log|\bar{\mathbf{h}}^H\mathbf{u}_\parallel|^2$$
$$- \log\det(\mathbf{I}_K + \rho\mathbf{C}) - K\log\left[\mathbf{u}_\parallel^H\mathbf{Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y}\mathbf{u}_\parallel\right] - K\log\det(\boldsymbol{\Delta}_{\mathbf{Y}\perp}). \tag{3.93}$$

In order to proceed now, the following result will be useful.

Lemma 3.7. *Let $\boldsymbol{A}$ be an $L \times L$ invertible matrix, and let $\boldsymbol{G}$ be an unitary matrix*

partitioned as $\boldsymbol{G} = [\boldsymbol{g}_1 \ \boldsymbol{G}_\perp]$ *where* $\boldsymbol{g}_1$ *is* $L \times 1$ *and* $\boldsymbol{G}_\perp$ *is* $L \times (L-1)$. *Then it holds that*

$$\det(\boldsymbol{G}_\perp^H \boldsymbol{A} \boldsymbol{G}_\perp) = \boldsymbol{g}_1^H \boldsymbol{A}^{-1} \boldsymbol{g}_1 \det(\boldsymbol{A}). \tag{3.94}$$

Proof. The proof can be found in Appendix 3.C. □

Observe that, since $\mathbf{U}_{\mathbf{Y}\perp}^H \mathbf{U}_{\mathbf{Y}\perp} = \mathbf{I}_{L-1}$, one has

$$\boldsymbol{\Delta}_{\mathbf{Y}\perp} = \mathbf{U}_{\mathbf{Y}\perp}^H (\mathbf{Y}^H \mathbf{Y})_\perp \mathbf{U}_{\mathbf{Y}\perp} \tag{3.95}$$

$$= \mathbf{U}_{\mathbf{Y}\perp}^H (\mathbf{Y}^H \mathbf{Y}) \mathbf{U}_{\mathbf{Y}\perp}, \tag{3.96}$$

where the second step follows from the fact that $\mathbf{U}_{\mathbf{Y}\perp} \bar{\mathbf{h}} = 0$. Since the matrix $[\bar{\mathbf{h}} \ \mathbf{U}_{\mathbf{Y}\perp}]$ is unitary, we can apply the result from Lemma 3.7 to obtain

$$\det(\boldsymbol{\Delta}_{\mathbf{Y}\perp}) = (\bar{\mathbf{h}}^H (\mathbf{Y}^H \mathbf{Y})^{-1} \bar{\mathbf{h}}) \det(\mathbf{Y}^H \mathbf{Y}). \tag{3.97}$$

Therefore, the log-likelihood (3.93) reads as

$$\begin{aligned}
\log f(\mathbf{y} \,|\, \rho, \bar{\mathbf{h}}, \mathbf{u}_\parallel) = & \\
& -LK \log \pi + LK \log K - LK + K \log |\bar{\mathbf{h}}^H \mathbf{u}_\parallel|^2 - \log \det(\mathbf{I}_K + \rho \mathbf{C}) \\
& - K \log \left[\mathbf{u}_\parallel^H \mathbf{Y}^H (\mathbf{I}_K + \rho \mathbf{C})^{-1} \mathbf{Y} \mathbf{u}_\parallel \right] - K \log \left[\bar{\mathbf{h}}^H (\mathbf{Y}^H \mathbf{Y})^{-1} \bar{\mathbf{h}} \right] - K \log \det(\mathbf{Y}^H \mathbf{Y}).
\end{aligned} \tag{3.98}$$

Maximizing (3.98) with respect to $\bar{\mathbf{h}}$ and $\mathbf{u}_\parallel$ amounts to minimizing

$$\frac{\bar{\mathbf{h}}^H (\mathbf{Y}^H \mathbf{Y})^{-1} \bar{\mathbf{h}} \, \mathbf{u}_\parallel^H \mathbf{Y}^H (\mathbf{I}_K + \rho \mathbf{C})^{-1} \mathbf{Y} \mathbf{u}_\parallel}{|\bar{\mathbf{h}}^H \mathbf{u}_\parallel|^2}. \tag{3.99}$$

To this end, we can apply the following result:

Lemma 3.8. *Let* $\boldsymbol{A}_1$ *and* $\boldsymbol{A}_2$ *be two* $L \times L$ *Hermitian matrices, and let*

$$J(\boldsymbol{u}_1, \boldsymbol{u}_2) \doteq \frac{\boldsymbol{u}_1^H \boldsymbol{A}_1 \boldsymbol{u}_1 \boldsymbol{u}_2^H \boldsymbol{A}_2 \boldsymbol{u}_2}{|\boldsymbol{u}_1^H \boldsymbol{u}_2|^2}. \tag{3.100}$$

The minimum of J *is given by the smallest eigenvalue of the matrix* $\boldsymbol{A}_1 \boldsymbol{A}_2$ *(which*

is also the smallest eigenvalue of $\boldsymbol{A}_2\boldsymbol{A}_1$), and it is attained, when

$$\boldsymbol{u}_1^\star/c_1 = \textit{eigenvector of } \boldsymbol{A}_2\boldsymbol{A}_1 \textit{ associated to its smallest eigenvalue}, \qquad (3.101)$$

$$\boldsymbol{u}_2^\star/c_2 = \textit{eigenvector of } \boldsymbol{A}_1\boldsymbol{A}_2 \textit{ associated to its smallest eigenvalue}, \qquad (3.102)$$

with c_1 and c_2 two arbitrary nonzero complex scalars.

Proof. The proof can be found in Appendix 3.D. $\qquad\qquad\square$

Therefore, the optimum value of (3.99) is given by

$$\chi \doteq \lambda_{\min}\left((\mathbf{Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y})(\mathbf{Y}^H\mathbf{Y})^{-1}\right), \qquad (3.103)$$

whereas the ML estimates of $\bar{h}$ and $\mathbf{u}_\parallel$, up to a complex scaling factor, are respectively given by

$$\hat{\bar{\mathbf{h}}} = \text{eigenvector of } (\mathbf{Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y})(\mathbf{Y}^H\mathbf{Y})^{-1} \text{ associated to } \lambda_{\min}, \qquad (3.104)$$

$$\hat{\mathbf{u}}_\parallel = \text{eigenvector of } (\mathbf{Y}^H\mathbf{Y})^{-1}(\mathbf{Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y}) \text{ associated to } \lambda_{\min}, \qquad (3.105)$$

where with some abuse of notation we used $\lambda_{\min}$ to denote the smallest eigenvalue of the corresponding matrix. Consider now the singular value decomposition (SVD) of the data matrix

$$\mathbf{Y} = \mathbf{U_Y}\mathbf{S_Y}\mathbf{V_Y}^H \qquad (3.106)$$

where $\mathbf{U_Y}$ is $K \times L$ with orthonormal columns; $\mathbf{S_Y}$ is $L \times L$ diagonal, with the singular values of $\mathbf{Y}$; and $\mathbf{V_Y}$ is $L \times L$ unitary. Then, it is easily seen that χ can be written as

$$\chi = \lambda_{\min}\left(\mathbf{U_Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{U_Y}\right). \qquad (3.107)$$

With this, the compressed log-likelihood under $\mathcal{H}_1$ can be finally written as

$$\ell_1 \doteq -LK(1 + \log\frac{\pi}{K}) - K\log\det\mathbf{Y}^H\mathbf{Y}$$
$$+ K\max_{\rho\geq 0}\left(\log\frac{\lambda_{\min}^{-1}\left((\mathbf{Y}^H(\mathbf{I}_K + \rho\mathbf{C})^{-1}\mathbf{Y})(\mathbf{Y}^H\mathbf{Y})^{-1}\right)}{\det^{1/K}(\mathbf{I}_K + \rho\mathbf{C})}\right), \qquad (3.108)$$

which must be maximized only over the scalar parameter ρ.

GLRT detector

From Lemma 3.5 the compressed log-likelihood under the hypothesis 0 is given by

$$\ell_0 = \max_{\boldsymbol{\Sigma}^2} \log f(\mathbf{y} \,|\, \mathcal{R}_0) \tag{3.109}$$

$$= -LK(1 + \log \frac{\pi}{K}) - K \log \det (\mathbf{Y}^H \mathbf{Y}). \tag{3.110}$$

Therefore, the GLRT statistic (3.9) is found to be

$$T = e^{\ell_1 - \ell_0} = \max_{\rho \geq 0} \frac{\lambda_{\min}^{-K} \left(\mathbf{U}_{\mathbf{Y}}^H (\mathbf{I}_K + \rho \mathbf{C})^{-1} \mathbf{U}_{\mathbf{Y}} \right)}{\det(\mathbf{I}_K + \rho \mathbf{C})}, \tag{3.111}$$

whose computation involves a maximization with respect to a scalar parameter only. It is interesting to note that the dependence of the GLRT statistic T with the data is via $\mathbf{U}_{\mathbf{Y}}$ only. Hence this is a sufficient statistic for this problem, and the GLRT discards the spatial information contained in $\{\mathbf{S}_{\mathbf{Y}}, \mathbf{V}_{\mathbf{Y}}\}$.

To the best of our knowledge, there is no closed-form solution to the maximization of

$$t(\rho) \doteq \frac{\lambda_{\min}^{-K}(\mathbf{U}_{\mathbf{Y}}^H (\mathbf{I}_K + \rho \mathbf{C})^{-1} \mathbf{U}_{\mathbf{Y}})}{\det(\mathbf{I}_K + \rho \mathbf{C})}, \quad \rho \geq 0, \tag{3.112}$$

in the general case. We conjecture that $t(\rho)$ is a quasi convex function on ρ, and therefore its maximization can be efficiently carried out by numerical means (Boyd and Vandenberghe, 2004).

In the next subsections we focus on two particular cases of practical significance for which the maximization with respect to ρ (and therefore the GLRT statistic) can be obtained in closed form. First, we consider the case of bandpass signals with constant psd within the passband. Then we will study the case of arbitrary spectra in the low SNR regime, i.e. when the SNR goes to zero.

GLRT for ideally flat bandpass signals

Using the asymptotic EVD of the covariance matrix $\mathbf{C} \approx \mathbf{W}\boldsymbol{\Lambda}\mathbf{W}^H$ that we introduced in (2.5), we have that, for large K the GLRT statistic is given by

$$T = \max_{\rho \geq 0} \frac{\lambda_{\min}^{-K}(\mathbf{U}_{\mathbf{Y}}^H \mathbf{W}(\mathbf{I}_K + \rho \boldsymbol{\Lambda})^{-1} \mathbf{W}^H \mathbf{U}_{\mathbf{Y}})}{\det(\mathbf{I}_K + \rho \boldsymbol{\Lambda})}. \tag{3.113}$$

We say that a signal is ideally flat bandpass if its psd takes only two values, either zero or a certain constant. This translates into the diagonal of the matrix $\mathbf{\Lambda}$ which presents BK nonzero values, with B denoting the occupied bandwidth fraction. Moreover, due to the normalization of $\mathbf{C}$, namely $\operatorname{tr} \mathbf{C} = K$, we have that the nonzero elements of $\mathbf{\Lambda}$ are equal to $1/B$. Using these properties of the matrix $\mathbf{\Lambda}$, it is easy to see that, for ideally flat bandpass signals, the test can be written as

$$T_{\text{flat}}^{1/K} = \max_{\rho \geq 0} \frac{\lambda_{\min}^{-1}(\mathbf{U}_{\mathbf{Y}}^H \mathbf{W}(I_K - \frac{\rho}{1+\rho/B}\mathbf{\Lambda})\mathbf{W}^H \mathbf{U}_{\mathbf{Y}})}{(1+\rho/B)^B} \tag{3.114}$$

$$= \left[\min_{\rho \geq 0} \left(1 - \frac{\rho}{1+\rho/B}\lambda_{\max}(\mathbf{U}_{\mathbf{Y}}^H \mathbf{W}\mathbf{\Lambda}\mathbf{W}^H \mathbf{U}_{\mathbf{Y}}) \right)(1+\rho/B)^B \right]^{-1} \tag{3.115}$$

$$= \left[\min_{\rho \geq 0} \left(1 - \frac{\rho}{1+\rho/B}\lambda_{\max} \right)(1+\rho/B)^B \right]^{-1}, \tag{3.116}$$

where with some abuse of notation we defined $\lambda_{\max} \doteq \lambda_{\max}(\mathbf{U}_{\mathbf{Y}}^H \mathbf{C}\mathbf{U}_{\mathbf{Y}})$. By taking the derivative of the function to minimize, it is straightforward to find that the minimum in $\rho \geq 0$ is obtained at

$$\rho^\star = \begin{cases} 0, & \text{if } 0 < \lambda_{\max} \leq 1; \\ \frac{\lambda_{\max}-1}{1/B-\lambda_{\max}}, & \text{if } 1 < \lambda_{\max} \leq 1/B; \\ \infty, & \text{if } \lambda_{\max} > 1/B. \end{cases} \tag{3.117}$$

However, it is easily shown that $\lambda_{\max} \leq 1/B$. To see this, write $\mathbf{\Lambda} = \frac{1}{B}\mathbf{J}^2$, where $\mathbf{J}$ is a diagonal matrix with ones in the positions where $\mathbf{\Lambda}$ has nonzero values. Then

$$\lambda_{\max}(\mathbf{U}_{\mathbf{Y}}^H \mathbf{C}\mathbf{U}_{\mathbf{Y}}) = \frac{1}{B} \max_{\mathbf{x} \neq 0} \frac{\|\mathbf{J}\mathbf{W}^H \mathbf{U}_{\mathbf{Y}}\mathbf{x}\|^2}{\|\mathbf{x}\|^2} \tag{3.118}$$

But it is clear that $\|\mathbf{J}\mathbf{W}^H \mathbf{U}_{\mathbf{Y}}\mathbf{x}\|^2 \leq \|\mathbf{W}^H \mathbf{U}_{\mathbf{Y}}\mathbf{x}\|^2 = \mathbf{x}^H \mathbf{U}_{\mathbf{Y}}^H \mathbf{W}\mathbf{W}^H \mathbf{U}_{\mathbf{Y}}\mathbf{x} = \mathbf{x}^H \mathbf{x} = \|\mathbf{x}\|^2$. Hence $\lambda_{\max} \leq 1/B$.

Therefore the test results in

$$T_{\text{flat}}^{1/K} = \begin{cases} \frac{B}{1-B}\left(\frac{1}{B} - \lambda_{\max}\right)\left(\frac{1-B}{\frac{1}{\lambda_{\max}}-B}\right)^B & \text{for } 1 \leq \lambda_{\max} \leq 1/B, \\ 1 & \text{otherwise,} \end{cases} \tag{3.119}$$

which in the region $1 \leq \lambda_{\max} \leq 1/B$ can be shown to be non-decreasing in $\lambda_{\max}$. Hence, an equivalent asymptotic GLRT detector for flat bandpass signals is given

by

$$T' \doteq \lambda_{\max}(\mathbf{U_Y^H C U_Y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma'. \tag{3.120}$$

Note that this is a closed-form detector that can be implemented without much complexity, and it is thus adequate for practical systems.

Asymptotic GLRT in the low SNR regime

We study now the behavior of the GLRT statistic T from (3.111) when the SNR is small. This low SNR regime is of interest in cognitive radio scenarios, in which spectrum sensors must provide reliable decisions regarding the presence of primary transmissions which may be very weak, due to signal fading and shadowing.

Using the definition of $t(\rho)$ in (3.112) we have that the GLRT statistic can be written as

$$T = \max_{\rho \geq 0} t(\rho). \tag{3.121}$$

Consider now the following first-order Taylor approximations around $\rho = 0$:

$$(\mathbf{I}_K + \rho \mathbf{C})^{-1} = \mathbf{I}_K - \rho \mathbf{C} + o(\rho), \tag{3.122}$$

$$\det(\mathbf{I}_K + \rho \mathbf{C}) = 1 + \rho \operatorname{tr}\{\mathbf{C}\} + o(\rho), \tag{3.123}$$

where the "little-o" notation indicates a function that goes to zero faster than ρ as $\rho \to 0$, i.e. $f(\rho) \in o(\rho)$ if $\lim_{\rho \to 0} \frac{f(\rho)}{\rho} = 0$. Now note that

$$\lambda_{\min}(\mathbf{U_Y^H}(\mathbf{I}_K + \rho \mathbf{C})^{-1}\mathbf{U_Y}) = \lambda_{\min}(\mathbf{U_Y^H}(\mathbf{I}_K - \rho \mathbf{C})\mathbf{U_Y} + o(\rho)) \tag{3.124}$$

$$= \lambda_{\min}(\mathbf{U_Y^H}(\mathbf{I}_K - \rho \mathbf{C})\mathbf{U_Y}) + o(\rho) \tag{3.125}$$

$$= 1 - \rho\, \lambda_{\max}(\mathbf{U_Y^H C U_Y}) + o(\rho), \tag{3.126}$$

where the second step follows from the fact that, if $\boldsymbol{A}(\rho) \notin o(\rho)$, then

$$\lim_{\rho \to 0} \frac{1}{\rho} \left[\lambda_{\min}(\boldsymbol{A}(\rho) + o(\rho)) - \lambda_{\min}(\boldsymbol{A}(\rho)) \right]$$

$$= \min_{\mathbf{x} \neq 0} \left[\lim_{\rho \to 0} \frac{\mathbf{x}^H \boldsymbol{A}(\rho)\mathbf{x}}{\rho \mathbf{x}^H \mathbf{x}} + \lim_{\rho \to 0} \frac{\mathbf{x}^H o(\rho)\mathbf{x}}{\rho \mathbf{x}^H \mathbf{x}} \right] - \min_{\mathbf{x} \neq 0} \left[\lim_{\rho \to 0} \frac{\mathbf{x}^H \boldsymbol{A}(\rho)\mathbf{x}}{\rho \mathbf{x}^H \mathbf{x}} \right] \tag{3.127}$$

$$= 0. \tag{3.128}$$

Therefore one has

$$\log t(\rho) = -K \log \left(1 - \rho \lambda_{\max}(\mathbf{U}_\mathbf{Y}^H \mathbf{C} \mathbf{U}_\mathbf{Y}) + o(\rho)\right) - \log \left(1 + \rho \operatorname{tr}\{\mathbf{C}\} + o(\rho)\right) \tag{3.129}$$

$$= K\rho \left[\lambda_{\max}(\mathbf{U}_\mathbf{Y}^H \mathbf{C} \mathbf{U}_\mathbf{Y}) - \frac{\operatorname{tr}\{\mathbf{C}\}}{K}\right] + o(\rho), \tag{3.130}$$

where we have made use of the fact that $\log(1 + a\rho) = a\rho + o(\rho)$. Consider the hypothetical scenario in which the parameter ρ were known. In that case, the GLRT statistic is directly $t(\rho)$. In low SNR, (3.130) shows that this GLRT is equivalent to the test

$$\lambda_{\max}(\mathbf{U}_\mathbf{Y}^H \mathbf{C} \mathbf{U}_\mathbf{Y}) - \frac{\operatorname{tr}\{\mathbf{C}\}}{K} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma'', \tag{3.131}$$

which does not make use of the value of ρ. That is, for sufficiently low SNR, knowledge of ρ becomes irrelevant, and the GLRT can be rephrased as

$$T' = \lambda_{\max}(\mathbf{U}_\mathbf{Y}^H \mathbf{C} \mathbf{U}_\mathbf{Y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma', \tag{3.132}$$

whether ρ is known or unknown. Note that this is the same test as (3.120), which was derived for a specific signal spectrum with general SNR.

Remark 3.3. The fact that the test T' is the GLRT for these two cases of particular interest suggest that it is likely to offer good performance in other scenarios as well. Indeed, in Section 3.4.4 numerical results will be presented that confirm this point.

3.4.3 Asymptotic performance analysis

In this section we derive the analytical performance of the proposed detector in the asymptotic regime when $K \to \infty$. In the Section 3.4.4 numerical results will be presented, showing that this asymptotic distribution approximates well the empirical data even for moderate values of K.

Asymptotic distribution in the weak signal regime

While the GLRT approach to signal detection is not necessarily optimal, it usually offers good detection performance (Mardia et al., 1979). Moreover, the asymptotic distribution of the GLRT statistic under both hypotheses in the weak signal regime

is known under certain regularity conditions (Kay, 1998). If we denote by real-valued θ the set of unknown parameters under the hypothesis $\mathcal{H}_1$ which are fixed to $\theta = \theta_0$ under $\mathcal{H}_0$, and by ϕ the set of real-valued unknown parameters free under both hypotheses, the asymptotic distribution of the GLRT statistic T is given by

$$2 \log T \sim \begin{cases} \chi_Q^2 & \text{under } \mathcal{H}_0, \\ \chi_Q'^2(\lambda) & \text{under } \mathcal{H}_1, \end{cases} \tag{3.133}$$

where χ_Q^2 denotes a central chi-squared distribution with Q degrees of freedom, $\chi_Q'^2(\lambda)$ denotes a non-central chi-squared distribution with Q degrees of freedom and non-centrality parameter λ.

Theorem 3.1. *The asymptotic (as $K \to \infty$) distribution of the GLRT test statistic T given in (3.111), under $\mathcal{H}_0$ and under $\mathcal{H}_1$ as $\theta \to \theta_0$, is given by (3.133), with parameters*

$$Q = 2L - 1, \tag{3.134}$$

$$\lambda = 0. \tag{3.135}$$

Proof. The value of Q is given by (Kay, 1998) the cardinality of the set of (unknown) parameters fixed under $\mathcal{H}_0$, given by θ. In our problem this set corresponds to the real elements of the vector $\mathbf{h}$ fixed to $\mathbf{h} = \mathbf{0}$ under $\mathcal{H}_0$. However, since this vector features in the likelihood function as $\mathbf{h}\mathbf{h}^H$ we may fix one of its components to be real without loss of generality. Hence, we obtain a total number of $Q = 2L - 1$ free real degrees of freedom.

Now, the value of λ is given by (Kay, 1998)

$$\lambda = (\theta_1 - \theta_0)^T \left[[\boldsymbol{F}(\theta_0, \phi)]_{\theta,\theta} - [\boldsymbol{F}(\theta_0, \phi)]_{\theta,\phi} [\boldsymbol{F}(\theta_0, \phi)]_{\phi,\phi}^{-1} [\boldsymbol{F}(\theta_0, \phi)]_{\phi,\theta} \right] (\theta_1 - \theta_0), \tag{3.136}$$

where θ_1 is the true value of the parameter vector θ, θ_0 corresponds to the fixed value of the parameter vector θ under the hypothesis $\mathcal{H}_0$, $\boldsymbol{F}(\theta', \phi')$ denotes the Fisher Information Matrix of the estimation problem evaluated at (θ', ϕ') and $[\cdot]_{\alpha',\beta'}$ denotes the submatrix corresponding to rows and the columns relative to the parameters α' and β' respectively.

We may rearrange the L^2 real components of $\mathbf{\Sigma}^{-2}$ in a $L \times L$ matrix $\mathbf{K}$ as

$$[\mathbf{K}]_{ij} \doteq \begin{cases} \Re\{[\mathbf{\Sigma}^{-2}]_{ij}\} & \text{for } j > i, \\ \frac{1}{2}[\mathbf{\Sigma}^{-2}]_{ij} & \text{for } i = j, \\ \Im\{[\mathbf{\Sigma}^{-2}]_{ij}\} & \text{for } j < i. \end{cases} \qquad (3.137)$$

Then, we may define the spaces for nuisance and non-nuisance parameters, respectively, as

$$\phi \doteq \{\text{vec}(\mathbf{K})\}, \qquad (3.138)$$

$$\theta \doteq \{a_1, \ldots, a_L, b_1, \ldots, b_{L-1}\}, \qquad (3.139)$$

with $\mathbf{a} = [a_1 \ \cdots \ a_L]^T \doteq \Re\{\mathbf{h_\Sigma}\}$ and $\mathbf{b} = [b_1 \ \cdots \ b_{L-1} \ 0]^T \doteq \Im\{\mathbf{h_\Sigma}\}$.

Using this set of parameters, tedious but straightforward algebra yields

$$\mathbf{F}_{\theta_i, \theta_j} \doteq \mathrm{E}\left[-\frac{\partial^2 \log f}{\partial \theta_i \partial \theta_j}\right] = 0, \qquad (3.140)$$

$$\mathbf{F}_{\theta_i, \phi_j} \doteq \mathrm{E}\left[-\frac{\partial^2 \log f}{\partial \theta_i \partial \phi_j}\right] = 0, \qquad (3.141)$$

and from (3.136) finally we obtain $\lambda = 0$. $\qquad \square$

Remark 3.4. Theorem 3.1 applies under the assumptions that (i) the data record K is long enough, and (ii) the value of the set of parameters θ under $\mathcal{H}_1$ is close to θ_0. Under the model considered here, the asymptotic result under $\mathcal{H}_1$ is too coarse in order to offer an useful approximation, since for $\lambda = 0$ the distributions under both $\mathcal{H}_0$ and $\mathcal{H}_1$ are equal. However, as we will see in Section 3.4.4 by means of simulations Theorem 3.1 models accurately the statistic under $\mathcal{H}_0$ and can be used to approximate the distribution under $\mathcal{H}_1$ as we will see next.

Asymptotic analysis under $\mathcal{H}_1$

We present now a result concerning the asymptotic value of the statistic T as $K \to \infty$ regardless of the value of θ, which will then be used to approximate the distribution of the statistic under $\mathcal{H}_1$.

Theorem 3.2. *Consider the GLRT statistic T from (3.111). Then one has*

$$\lim_{K \to \infty} \mathrm{E}\left[|T - \tilde{T}(\rho_0)|^2\right] = 0, \qquad (3.142)$$

where $\rho_0 \doteq \mathbf{h}^H \mathbf{\Sigma}^{-2} \mathbf{h}$ *the true value of the SNR, and*

$$\tilde{T}(\rho_0) \doteq \left(\frac{\operatorname{tr}\{\mathbf{I}_K + \rho_0 \mathbf{C}\}/K}{(\det(\mathbf{I}_K + \rho_0 \mathbf{C}))^{1/K}} \right)^K. \tag{3.143}$$

Proof. The proof is included in Appendix 3.E. □

Remark 3.5. The asymptotic behavior of T given by $\tilde{T}(\rho_0)$ in (3.143) corresponds to the K-th power of the sphericity ratio of the temporal covariance matrix $\mathbf{I}_K + \rho_0 \mathbf{C}$, i.e., a metric which measures how far the received signal is from temporally white noise.

Note that, by virtue of the asymptotic approximation (2.5) of the covariance matrix of a wide-sense stationary process, the limit of $(\tilde{T}(\rho_0))^{1/K}$ is seen to be

$$\lim_{K \to \infty} (\tilde{T}(\rho))^{\frac{1}{K}} = \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} [1 + \rho_0 S_{ss}(\mathrm{e}^{\jmath\omega})] \, \partial\omega}{\exp\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log[1 + \rho_0 S_{ss}(\mathrm{e}^{\jmath\omega})] \, \partial\omega \right\}} \tag{3.144}$$

where $S_{ss}(\mathrm{e}^{\jmath\omega})$ is the psd of the signal $\{s_k\}$. The right hand-hand side of (3.144) is the inverse of the spectral flatness measure (SFM) associated to the power spectrum $1 + \rho_0 S_{ss}(\mathrm{e}^{\jmath\omega})$ (Gray Jr. and Markel, 1974). Its minimum value is 1 for $\rho_0 = 0$ (no signal) and it increases monotonically with ρ_0 towards its asymptotic value, given by the inverse of the SFM associated to $S_{ss}(\mathrm{e}^{\jmath\omega})$ (Dugre et al., 1980). It follows that the primary signal maximizing $\tilde{T}(\rho)$ for a given $\rho > 0$ concentrates its energy in a single frequency point, since in this case the spectral flatness measure (SFM) is minimum. This is not surprising, since these peaky signals are easier to detect in the presence of noise.

From (3.133) we have that under $\mathcal{H}_1$, the asymptotic mean of the statistic is given by $\mathrm{E}[2 \log T] = Q + \lambda$. On the other hand, using the result in Theorem 3.2 it is easy to see that

$$\lim_{K \to \infty} \mathrm{E}[2 \log T] = \lim_{K \to \infty} 2 \log \tilde{T}(\rho) \tag{3.145}$$

$$= \lim_{K \to \infty} 2K \log \frac{\operatorname{tr}\{\mathbf{I}_K + \rho_0 \mathbf{C}\}/K}{(\det(\mathbf{I}_K + \rho_0 \mathbf{C}))^{1/K}}. \tag{3.146}$$

Combining these two results, we have that for sufficiently large K, $\lambda >> Q$ and we
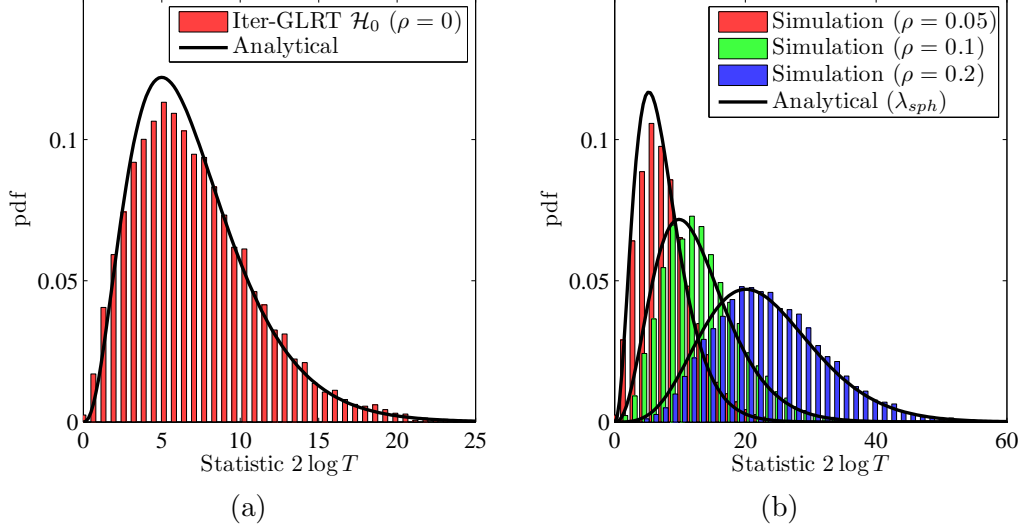
Figure 3.4: Distribution of the statistic $-2\log T$ for $L = 4$ and $K = 128$. (a) Under $\mathcal{H}_0$ (b) Under $\mathcal{H}_1$.

may approximate $\lambda \approx 2\log\tilde{T}(\rho)$. This fact suggests the use of

$$\lambda_{sph} \doteq 2\log\tilde{T}(\rho) \tag{3.147}$$

$$= 2K\log\frac{\text{tr}\{\mathbf{I}_K + \rho_0\mathbf{C}\}/K}{(\det(\mathbf{I}_K + \rho_0\mathbf{C}))^{1/K}} \tag{3.148}$$

as centrality parameter of the distribution under $\mathcal{H}_1$. In Section 3.4.4 it will be shown that this approximation offers a very accurate characterization of the distribution of the statistic for the SNR range of interest.

3.4.4 Numerical results and discussion

In this section we numerically study the performance of the proposed detectors via Monte Carlo simulations and test the accuracy of the analytical approximations. In each experiment we fix the empirical SNR $\rho = \mathbf{h}^H\mathbf{\Sigma}^{-2}\mathbf{h}$, while the actual channel $\mathbf{h}$ and the noise spatial covariance matrix $\mathbf{\Sigma}^2$ are randomly generated at each Monte Carlo realization[5].

First, Fig. 3.4 compares the asymptotic analytical distributions with the simu-

---

[5]At each realization $\mathbf{\Sigma}^2 = \mathbf{G}^H\mathbf{G}$ with $\mathbf{G}$ an $L \times L$ matrix with independent circular Gaussian entries. The vector channel $\mathbf{h}$ presents independent circular Gaussian components scaled to obtain a fixed value of $\rho = \mathbf{h}^H\mathbf{\Sigma}^{-2}\mathbf{h}$.

lation results of the exact GLRT T given in (3.111) optimized[6] over ρ and denoted here as *iterative GLRT*. The setup considered here is $L = 4$, $K = 128$, for the detection of a flat bandpass signal occupying half of the band. The covariance matrix is normalized as $\text{tr}\{\mathbf{C}\} = K$. We observe an excellent agreement between the empirical results and the asymptotic distributions even for moderate values of K.

In the next subsections, we compare the exact GLRT T, *iterative GLRT*, with the closed-form GLRT T' from (3.120), denoted here *asymptotic GLRT*, which has been shown to coincide with the GLRT for vanishing SNR or for rectangular psd of the primary signal.

It is interesting to note that most of the proposed detectors in the literature cannot deal with the strong interference model considered in this section, since the full-rank spatial covariance matrix of the noise always masks the presence of the primary signal when the temporal correlation structure of the latter is ignored. In order to compare the proposed detectors against a suitable benchmark, and inspired by the GED derived in Chapter 2, we consider a generalized energy detector which assumes the noise level, i.e., the trace of $\mathbf{\Sigma}^2$, to be available to the spectral monitor. Then, the detector compares

$$T_{ED} = \frac{\text{tr}\{\mathbf{Y}^H \mathbf{C} \mathbf{Y}\}}{\text{tr}\{\mathbf{\Sigma}^2\}} \tag{3.149}$$

against a threshold. Since any primary signal will increase the energy observed by the system, the expected T_{ED} under $\mathcal{H}_1$ will also be increased.

Detection performance in low SNR.

In Fig. 3.5 we show the empirical performance of the proposed detectors versus the analytical curves for a scenario with $L = 4$, $K = 512$, $\rho = 0.2$ (natural units) and two different primary psd. One corresponds to a DVB-T television signal[7] (ETSI, 2004) with bandwidth $B = 7.61$ MHz quantized to 9-bit precision. This channel was downshifted to baseband and asynchronously sampled at $f_s = 16$ MHz, thus in this case the occupied bandwidth fraction is 48%. A second curve shows the same scenario with a QAM primary signal shaped as a square root of a raised cosine filter with rolloff factor equal to 1 and sampled at twice the baud rate, thus occupying

---

[6]Implemented using a gradient descent algorithm, initialized at $\rho = 1$, initial stepsize $\mu = 100$, decreased to $\mu = \mu/4$ when the descent direction changes sign and stop condition $|\partial\kappa(\rho)/\partial\rho| < 1e-5$ (with a maximum number of iterations equal to 100).

[7]8K mode, 64-QAM, guard interval 1/4, inner code rate 2/3.
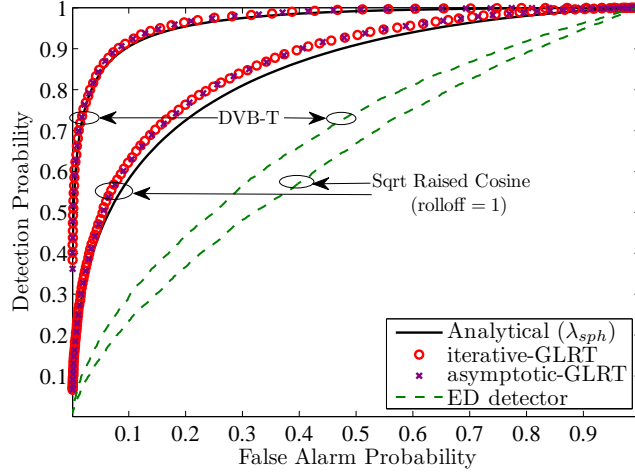
Figure 3.5: ROC curve showing the detection performance for OFDM and square root raised cosine signals when $\rho = 0.15$, $L = 2$ and $K = 512$.

the whole Nyquist bandwidth.

As can be seen from Fig. 3.5, in these harsh SNR conditions, the energy detector offers a poor detection performance compared to the proposed schemes. Moreover, since we are in the low SNR regime, both the iterative and the asymptotic GLRT detectors show virtually the same performance, for both rectangular and squared root raised cosine primary signals. It is interesting to note the good match between the empirical results and the analytical curves, even for this moderate value of K.

Fig. 3.5 also shows the effect of primary signal shaping. As it could be expected from the analytical results in the results in (3.133) and Theorem 3.2, the proposed detectors work noticeably better for less spectrally flat primary signals. The reason is that for fixed ρ, the DVB-T signal presents a smaller spectral flatness measure than the squared root raised cosine primary signals.

Detection performance versus SNR.

We now turn to study the loss incurred by the closed-form asymptotic GLRT detector T' from (3.120) with respect to the exact GLRT T given in (3.111) as the SNR increases. To this end, Fig. 3.6 shows the probability of detection achieved[8] by the two detectors versus the average per antenna SNR $= 10 \log_{10}(\rho/L)$ for fixed $P_{FA} = 0.05$ and the two primary signal types presented in the previous section.

---

[8]Each point was computed using 10^6 Monte Carlo iterations.
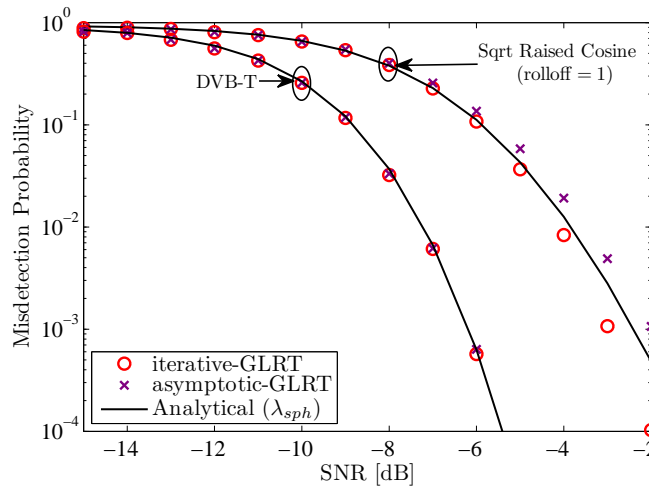
Figure 3.6: P_D performance versus SNR for fixed $P_{FA} = 0.05$, $L = 4$ and $K = 128$.

The remaining system parameters are $L = 4$ and $K = 128$. First, we note that the asymptotic analytical results match reasonably well the empirical results for the SNR range of interest. In second place, it can be seen that for the DVB-T signal both the asymptotic GLRT detector and the exact iterative scheme offer the same performance, since the asymptotic GLRT coincides with the GLRT for rectangular spectra. However, with other kinds of power spectra (such as the squared root raised cosine), the asymptotic GLRT presents a performance penalty which increases with the SNR. This penalty, nevertheless, is not large, which makes the asymptotic GLRT detector a good candidate in settings in which computational complexity is an issue.

3.5 Conclusions

In this chapter we have studied the problem of multiantenna detection of primary signals in the presence of noise with unknown statistics. We derived the GLRT detectors under different assumptions on the noise spatial covariance matrix.

First, we considered the problem of detecting vector-valued rank-P signals when the noise is assumed uncorrelated across the antennas. These detectors are robust to a mismatch in noise levels across the antennas and exploit the rank-P structure of the primary signals, including as particular cases several previous schemes derived either for $P = 1$ or for large P. Then, we considered the case of noise with arbitrary spatial structure. In this case a certain temporal structure is required for the primary

signal in order to make it distinguishable from the noise, and indeed the detection performance of the GLRT has been shown to be closely related to the sphericity ratio of the temporal covariance matrix $\mathbf{I}_K + \rho_0 \mathbf{C}$. The different GLRTs for the problem of multiantenna detection under unknown noise statistics are summarized in Table 3.1, with special emphasis on the cases treated in this chapter.

The content of this chapter is mainly based in two journal articles derived from the preliminary work presented at the 2nd International Workshop on Cognitive Information Processing (CIP 2010) (López-Valcarce et al., 2010). The theoretical results exposed in Section 3.3 (rank-P signal detection in presence of uncorrelated noises) have been obtained in collaboration with the Advanced Signal Processing Group (GTAS, University of Cantabria) under the national research project COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010), and constitutes a joint journal paper to appear in IEEE-TSP (Ramírez et al., 2011). The derivation and performance of the GLR detector under correlated noise (Section 3.4) constitute a joint work with the Signal Processing for Communications Research Group (SPCOM, Technical Univ. of Catalonia UPC) under the national research project SPROACTIVE (reference TEC2007-68094-C02-01/TCM) and COMONSENS in preparation to be submitted to IEEE-TSP as a joint paper (Sala et al., 2011).

Noise structure		Signal structure	
Temporally	Spatially	Temporally	Spatially
White	Uncorrelated iid	White	Unstructured

Sphericity Ratio Test
$$T^{\frac{1}{KL}} = \frac{\frac{1}{L}\mathrm{trace}(\hat{\mathbf{R}})}{\det^{1/L}(\hat{\mathbf{R}})}$$

White	Uncorrelated iid	White	Rank-P

‡ Noise subspace sphericity test
$$\log T^{\frac{1}{KL}} = \log\left[\frac{\frac{1}{L}\sum_{i=1}^{L}\lambda_i}{\left(\prod_{i=1}^{L}\lambda_i\right)^{1/L}}\right] - \frac{L-P}{L}\log\left[\frac{\frac{1}{L-P}\sum_{i=P+1}^{L}\lambda_i}{\left(\prod_{i=P+1}^{L}\lambda_i\right)^{1/(L-P)}}\right]$$
with λ_i the ordered eigenvalues of $\hat{\mathbf{R}}$.

White	Uncorrelated non-iid	White	Unstructured

Hadamard Ratio Test
$$T^{\frac{1}{K}} = \frac{\det(\hat{\mathbf{R}})}{\prod_{i=1}^{L}[\hat{\mathbf{R}}]_{i,i}}$$

White	Uncorrelated non-iid	White	Rank-P

‡ No closed-form solution. Alternating optimization scheme in Section 3.3.2.
Asymptotic low SNR GLRT:
$$\log T^{\frac{1}{K}} \approx -P - \log\prod_{i=1}^{P}\beta_i e^{-\beta_i}$$
with β_i the ordered eigenvalues of the sample spatial coherence matrix $\hat{\mathbf{C}}$.

White	Unstructured	$\mathbf{C}$	Rank-1

‡ No closed-form solution:
$$T^{\frac{1}{K}} = \max_\rho \frac{\lambda_{\min}^{-1}(\mathbf{U}_{\mathbf{Y}}^H(\mathbf{I}_K+\rho\mathbf{C})^{-1}\mathbf{U}_{\mathbf{Y}})}{(\det(\mathbf{I}_K+\rho\mathbf{C}))^{1/K}}.$$
Asymptotic low SNR GLRT (exact for square psd):
$$T' \doteq \lambda_{\max}(\mathbf{U}_{\mathbf{Y}}^H\mathbf{C}\mathbf{U}_{\mathbf{Y}}).$$

Table 3.1: Summary of the GLRT for multiantenna detection under unknown noise statistics. ‡ Proposed.

Appendix 3.A   Proof of Lemma 3.2

Since the EVD of $\mathbf{R}$ is $\mathbf{R} = \mathbf{U}(\boldsymbol{\Psi}^2 + \sigma^2\mathbf{I})\mathbf{U}^H$, the log-likelihood is given by

$$\log f\left(\mathbf{Y} \mid \mathbf{H}, \sigma^2\right) = -LK \log \pi - K \log \det\left(\boldsymbol{\Psi}^2 + \sigma^2\mathbf{I}\right) - K \mathrm{tr}\left[\mathbf{U}^H \hat{\mathbf{R}} \mathbf{U}\left(\boldsymbol{\Psi}^2 + \sigma^2\mathbf{I}\right)^{-1}\right],$$
(3.150)

which, letting $\mathbf{A} = \mathbf{U}^H \hat{\mathbf{R}} \mathbf{U}$, can be rewritten as follows:

$$\log f\left(\mathbf{Y} \mid \mathbf{H}, \sigma^2\right) = -LK \log \pi - K \sum_{i=1}^{P} \log\left(\psi_i^2 + \sigma^2\right) - K(L - P) \log \sigma^2$$

$$- K \sum_{i=1}^{P} \frac{[\mathbf{A}]_{i,i}}{\psi_i^2 + \sigma^2} - \frac{K}{\sigma^2} \sum_{i=P+1}^{L} [\mathbf{A}]_{i,i}. \quad (3.151)$$

Now (3.151) is maximized with respect to ψ_i^2 and σ^2 for

$$\hat{\sigma}^2 = \frac{1}{L - P} \sum_{i=P+1}^{L} [\mathbf{A}]_{i,i}, \tag{3.152}$$

$$\hat{\psi}_i^2 = [\mathbf{A}]_{i,i} - \hat{\sigma}^2, \quad i = 1, \ldots, P. \tag{3.153}$$

Now, substituting (3.152) and (3.153) in (3.151), it becomes

$$\log f\left(\mathbf{Y} \mid \mathbf{H}, \sigma^2\right) = -LK\left(\log \pi + 1\right) - K \sum_{i=1}^{P} \log[\mathbf{A}]_{i,i}$$

$$- K(L - P) \log\left(\frac{1}{L - P} \sum_{i=P+1}^{L} [\mathbf{A}]_{i,i}\right), \quad (3.154)$$

which has to be maximized subject to $\mathbf{A} = \mathbf{U}^H \hat{\mathbf{R}} \mathbf{U}$ and $\mathbf{U}^H \mathbf{U} = \mathbf{I}_L$. It is easy to check that (3.154) is a Schur-convex (Jorswieck and Boche, 2007) function of $[\mathbf{A}]_{i,i}$, $i = 1, \ldots, L$. Consequently, (3.154) is upper bounded by the i-th largest eigenvalue of $\mathbf{A}$ (Jorswieck and Boche, 2007), since the vector of eigenvalues majorizes any vector comprised by the diagonal entries of a given matrix. This upper bound is attained by choosing $\hat{\mathbf{U}} = \mathbf{V}$, so that $\mathbf{A}$ is diagonal. From this result, the proof follows.

Appendix 3.B   Detailed computation of $(\mathbf{I}_{LK} + (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^{H})^{T} \otimes \mathbf{C})^{-1}$

Let us define $\mathbf{g} = \mathbf{h}_{\boldsymbol{\Sigma}}^{*}$ and let $\mathbf{C}^{1/2}$ be a Hermitian square root of $\mathbf{C}$, i.e. $\mathbf{C} = \mathbf{C}^{1/2}\mathbf{C}^{1/2}$, $\mathbf{C}^{H/2} = \mathbf{C}^{1/2}$. Making use of the matrix inversion lemma we have that

$$(\mathbf{I}_{LK} + \mathbf{g}\mathbf{g}^{H} \otimes \mathbf{C})^{-1}$$

$$= (\mathbf{I}_{LK} + (\mathbf{g} \otimes \mathbf{C}^{1/2})(\mathbf{g}^{H} \otimes \mathbf{C}^{1/2}))^{-1} \tag{3.155}$$

$$= \mathbf{I}_{LK} - (\mathbf{g} \otimes \mathbf{C}^{1/2})(\mathbf{I}_{K} + (\mathbf{g}^{H} \otimes \mathbf{C}^{1/2})(\mathbf{g} \otimes \mathbf{C}^{1/2}))^{-1}(\mathbf{g}^{H} \otimes \mathbf{C}^{1/2}) \tag{3.156}$$

$$= \mathbf{I}_{LK} - (\mathbf{g} \otimes \mathbf{C}^{1/2})(\mathbf{I}_{K} + \|\mathbf{g}\|^{2}\mathbf{C})^{-1}(\mathbf{g}^{H} \otimes \mathbf{C}^{1/2}) \tag{3.157}$$

Substituting back $\mathbf{g} = \mathbf{h}_{\boldsymbol{\Sigma}}^{*}$ and noting that $\|\mathbf{g}\|^{2} = \mathbf{h}^{H}\boldsymbol{\Sigma}^{-2}\mathbf{h} = \rho$ we obtain

$$(\mathbf{I}_{LK} + (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^{H})^{T} \otimes \mathbf{C})^{-1} = \mathbf{I}_{LK} - (\mathbf{h}_{\boldsymbol{\Sigma}}\mathbf{h}_{\boldsymbol{\Sigma}}^{H})^{T} \otimes (\mathbf{C}^{1/2}(\mathbf{I}_{K} + \rho\mathbf{C})^{-1}\mathbf{C}^{1/2}). \tag{3.158}$$

Appendix 3.C   Proof of Lemma 3.7

Assume that we have a square invertible matrix $\boldsymbol{A}$ and an unitary matrix $\boldsymbol{G}$. We partition the unitary matrix $\boldsymbol{G}$ as $\boldsymbol{G} = [\boldsymbol{g}_{1}\ \boldsymbol{G}_{\perp}]$, where $\boldsymbol{g}_{1}$ represents the first column and $\boldsymbol{G}_{\perp}$ the remaining ones. The inverse of $\boldsymbol{G}^{H}\boldsymbol{A}\boldsymbol{G}$ can be written as a function of its determinant and adjugate matrix as

$$(\boldsymbol{G}^{H}\boldsymbol{A}\boldsymbol{G})^{-1} = \frac{\text{adj}(\boldsymbol{G}^{H}\boldsymbol{A}\boldsymbol{G})}{\det(\boldsymbol{G}^{H}\boldsymbol{A}\boldsymbol{G})}, \tag{3.159}$$

that can be rewritten as

$$\boldsymbol{G}^{H}(\boldsymbol{A})^{-1}\boldsymbol{G} = \frac{\text{adj}(\boldsymbol{G}^{H}\boldsymbol{A}\boldsymbol{G})}{\det(\boldsymbol{A})}. \tag{3.160}$$

Let now $\mathbf{e}_{1} \doteq [1\ 0\ \cdots\ 0]$. Now we multiply both sides of the eq. (3.160) by $\mathbf{e}_{1}^{H}$ on the left and $\mathbf{e}_{1}$ on the right to obtain

$$\boldsymbol{g}_{1}^{H}(\boldsymbol{A})^{-1}\boldsymbol{g}_{1} = \frac{\mathbf{e}_{1}^{H}\,\text{adj}(\boldsymbol{G}^{H}\boldsymbol{A}\boldsymbol{G})\mathbf{e}_{1}}{\det(\boldsymbol{A})}, \tag{3.161}$$

where we have made use that $\boldsymbol{G}\mathbf{e}_1 = \boldsymbol{g}_1$. Note that

$$\mathbf{e}_1^H \operatorname{adj}(\boldsymbol{G}^H \boldsymbol{A} \boldsymbol{G})\mathbf{e}_1 = [\operatorname{adj}(\boldsymbol{G}^H \boldsymbol{A} \boldsymbol{G})]_{11} \tag{3.162}$$

$$= \det(\boldsymbol{G}_\perp^H \boldsymbol{A} \boldsymbol{G}_\perp). \tag{3.163}$$

Plugging (3.163) into (3.161) we obtain the desired result:

$$\det(\boldsymbol{G}_\perp^H \boldsymbol{A} \boldsymbol{G}_\perp) = \boldsymbol{g}_1^H (\boldsymbol{A})^{-1} \boldsymbol{g}_1 \det(\boldsymbol{A}). \tag{3.164}$$

Appendix 3.D   Proof of Lemma 3.8

The function we want to minimize takes the form

$$J(\boldsymbol{u}_1, \boldsymbol{u}_2) \doteq \frac{\boldsymbol{u}_1^H \boldsymbol{A}_1 \boldsymbol{u}_1 \boldsymbol{u}_2^H \boldsymbol{A}_2 \boldsymbol{u}_2}{|\boldsymbol{u}_1^H \boldsymbol{u}_2|^2}, \tag{3.165}$$

where $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ represent any two vectors and $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are two given Hermitian matrices. Then we can formally state the problem as

$$\{\boldsymbol{u}_1^\star, \boldsymbol{u}_2^\star\} = \arg \min_{\{\boldsymbol{u}_1, \boldsymbol{u}_2\}} J(\boldsymbol{u}_1, \boldsymbol{u}_2) \tag{3.166}$$

Note that $J(\boldsymbol{u}_1, \boldsymbol{u}_2)$ is invariant to scalings in $\boldsymbol{u}_1, \boldsymbol{u}_2$. Then, the minimization (3.166) is equivalent to

$$\arg \min_{\{\boldsymbol{u}_1, \boldsymbol{u}_2\}} \boldsymbol{u}_1^H \boldsymbol{A}_1 \boldsymbol{u}_1 \boldsymbol{u}_2^H \boldsymbol{A}_2 \boldsymbol{u}_2, \quad \text{subject to } |\boldsymbol{u}_1^H \boldsymbol{u}_2|^2 = c^2, \tag{3.167}$$

where c^2 is a positive constant. In order to solve (3.167) we construct the Lagrangian

$$\mathcal{L} = \boldsymbol{u}_1^H \boldsymbol{A}_1 \boldsymbol{u}_1 \boldsymbol{u}_2^H \boldsymbol{A}_2 \boldsymbol{u}_2 - \chi(|\boldsymbol{u}_1^H \boldsymbol{u}_2|^2 - c^2), \tag{3.168}$$

where χ denotes the Lagrange multiplier associated to the constraint. Now, from $\nabla_{\boldsymbol{u}_1^H} \mathcal{L} = \boldsymbol{0}$ and $\nabla_{\boldsymbol{u}_2^H} \mathcal{L} = \boldsymbol{0}$ we obtain, respectively:

$$\boldsymbol{u}_2^H \boldsymbol{A}_2 \boldsymbol{u}_2 \cdot \boldsymbol{A}_1 \boldsymbol{u}_1 = \chi \boldsymbol{u}_2^H \boldsymbol{u}_1 \cdot \boldsymbol{u}_2, \tag{3.169}$$

$$\boldsymbol{u}_1^H \boldsymbol{A}_1 \boldsymbol{u}_1 \cdot \boldsymbol{A}_2 \boldsymbol{u}_2 = \chi \boldsymbol{u}_1^H \boldsymbol{u}_2 \cdot \boldsymbol{u}_1. \tag{3.170}$$

Note the symmetry between (3.169) and (3.170). By solving for the Lagrange multiplier from these two equations we have that

$$\chi = \frac{\boldsymbol{u}_1^H \boldsymbol{A}_1 \boldsymbol{u}_1 \boldsymbol{u}_2^H \boldsymbol{A}_2 \boldsymbol{u}_2}{|\boldsymbol{u}_1^H \boldsymbol{u}_2|^2}, \tag{3.171}$$

that is, the Lagrange multiplier coincides with the quantity we are minimizing. We substitute now the value of $\boldsymbol{u}_2$ from (3.170) and χ from (3.171) into (3.169) to obtain the simplified equality

$$\boldsymbol{A}_2 \boldsymbol{A}_1 \boldsymbol{u}_1 = \chi \boldsymbol{u}_1, \tag{3.172}$$

and, by symmetry of the problem, it is easy to see that

$$\boldsymbol{A}_1 \boldsymbol{A}_2 \boldsymbol{u}_2 = \chi \boldsymbol{u}_2. \tag{3.173}$$

Hence the minimum of $J(\boldsymbol{u}_1, \boldsymbol{u}_2)$ is achieved at

$$J(\boldsymbol{u}_1^\star, \boldsymbol{u}_2^\star) = \lambda_{\min}(\boldsymbol{A}_1 \boldsymbol{A}_2) = \lambda_{\min}(\boldsymbol{A}_2 \boldsymbol{A}_1) \tag{3.174}$$

$$\boldsymbol{u}_1^\star = \text{eigenvector of } \boldsymbol{A}_2 \boldsymbol{A}_1 \text{ associated to its smallest eigenvalue,} \tag{3.175}$$

$$\boldsymbol{u}_2^\star = \text{eigenvector of } \boldsymbol{A}_1 \boldsymbol{A}_2 \text{ associated to its smallest eigenvalue.} \tag{3.176}$$

Appendix 3.E   Proof of Theorem 3.2

First, note that the GLRT test statistic (3.111), can be rewritten as

$$T = \max_\rho \frac{\lambda_{\min}^{-K}\left((\mathbf{Y}^H \mathbf{Y})^{-1}(\mathbf{Y}^H \mathbf{D}(\rho)\mathbf{Y})\right)}{\det(\mathbf{I}_K + \rho \mathbf{C})}, \tag{3.177}$$

with $\mathbf{D}(\rho) \doteq (\mathbf{I}_K + \rho \mathbf{C})^{-1}$.

From the signal model (3.40) we have that

$$\frac{1}{K}\mathrm{E}[\mathbf{Y}^H \boldsymbol{A} \mathbf{Y}] = \frac{1}{K}\mathrm{E}[\mathbf{h}\mathbf{s}^H \boldsymbol{A}\mathbf{s}\mathbf{h}^H] + \frac{1}{K}\mathrm{E}[\mathbf{G}^H \mathbf{N}^H \boldsymbol{A}\mathbf{N}\mathbf{G}] \tag{3.178}$$

$$= \mathbf{h}\frac{1}{K}\mathrm{tr}\{\mathrm{E}[\mathbf{s}\mathbf{s}^H]\boldsymbol{A}\}\mathbf{h}^H + \mathbf{G}^H \frac{1}{K}\mathrm{tr}\{\mathrm{E}[\mathbf{N}\mathbf{N}^H]\boldsymbol{A}\}\mathbf{G} \tag{3.179}$$

$$= \frac{\mathrm{tr}\{\mathbf{C}\boldsymbol{A}\}}{K}\mathbf{h}^H \mathbf{h} + \frac{\mathrm{tr}\{\mathbf{I}_K \boldsymbol{A}\}}{K}\mathbf{G}\mathbf{G}^H. \tag{3.180}$$

Now, noting that $\mathbf{Y}^H \boldsymbol{A} \mathbf{Y}/K$ is a consistent estimator of its mean as $K \to \infty$, the following asymptotic equivalence can be established

$$\frac{\mathbf{Y}^H \boldsymbol{A} \mathbf{Y}}{K} \xrightarrow{\text{var}} \frac{\text{tr}(\mathbf{C} \boldsymbol{A})}{K} \mathbf{h}\mathbf{h}^H + \frac{\text{tr}(\boldsymbol{A})}{K} \boldsymbol{\Sigma}^2, \tag{3.181}$$

where we use the symbol $\xrightarrow{\text{var}}$ to denote stochastic convergence in variance, i.e.

$$a_K \xrightarrow{\text{var}} b_K \quad \Longleftrightarrow \quad \lim_{K \to \infty} E\{|a_K - b_K|^2\} = 0, \tag{3.182}$$

so that in the matrix case $\xrightarrow{\text{var}}$ applies componentwise.

Applying this result to the GLRT test statistic (3.177), in the limit we obtain

$$T \xrightarrow{\text{var}} \max_{\rho \geq 0} \frac{\lambda_{\min}^K \left([\text{tr}(\mathbf{C})\mathbf{h}\mathbf{h}^H + \text{tr}(\mathbf{I}_K)\boldsymbol{\Sigma}^2]^{-1} [\text{tr}(\mathbf{C}\mathbf{D}(\rho))\mathbf{h}\mathbf{h}^H + \text{tr}(\mathbf{D}(\rho))\boldsymbol{\Sigma}^2]\right)}{\det(\mathbf{I}_K + \rho\mathbf{C})}. \tag{3.183}$$

The minimum eigenvalue featuring in the numerator of the right-hand side of (3.183) is computed next. For a given eigenvalue δ with associated eigenvector $\mathbf{v}$ the following equality holds

$$\left(\text{tr}(\mathbf{C})\mathbf{h}\mathbf{h}^H + K\boldsymbol{\Sigma}^2\right)^{-1} \left(\text{tr}(\mathbf{C}\mathbf{D}(\rho))\mathbf{h}\mathbf{h}^H + \text{tr}(\mathbf{D}(\rho))\boldsymbol{\Sigma}^2\right) \mathbf{v} = \delta\mathbf{v}. \tag{3.184}$$

Now, multiplying both sides of this equation by $\boldsymbol{\Sigma}^{-2} \left(\text{tr}(\mathbf{C})\mathbf{h}\mathbf{h}^H + K\boldsymbol{\Sigma}^2\right)$ on the left we obtain

$$s \, \text{tr}(\mathbf{C}\mathbf{D}(\rho))\boldsymbol{\Sigma}^{-2}\mathbf{h} + \text{tr}(\mathbf{D}(\rho))\mathbf{v} = \delta \left(s \, \text{tr}(\mathbf{C})\boldsymbol{\Sigma}^{-2}\mathbf{h} + K\mathbf{v}\right), \tag{3.185}$$

with $s \doteq \mathbf{h}^H \mathbf{v}$ an scalar. Note that both sides of (3.185) present the same structure, namely they are the sum of a contribution parallel to $\boldsymbol{\Sigma}^{-2}\mathbf{h}$ with a contribution parallel to the eigen vector $\mathbf{v}$. Then it is easy to see that for $\rho > 0$ one eigenvector corresponds to $\mathbf{v} = \mathbf{v}_1 \doteq \boldsymbol{\Sigma}^{-2}\mathbf{h}/\|\boldsymbol{\Sigma}^{-2}\mathbf{h}\|_2$. Moreover, every other eigenvector $\mathbf{v} = \mathbf{v}_\perp$ must fulfill $s = \mathbf{h}^H \mathbf{v}_\perp = 0$. If not, $s \neq 0$ and the first term of the sum would result in a contribution parallel to $\mathbf{v}_1$, i.e., it would be not orthogonal to $\mathbf{v}_1$. Then, the eigenvalues associated to the eigenvectors $\mathbf{v}_\perp$ are equal and given by $\delta_\perp = \text{tr}(\mathbf{D}(\rho))/K$. On the other hand, the eigenvalue associated to $\mathbf{v}_1$ is given by

$$\delta_1 = \frac{\text{tr}\{(\mathbf{I}_K + \rho_0\mathbf{C})(\mathbf{I}_K + \rho\mathbf{C})^{-1}\}}{\text{tr}\{\mathbf{I}_K + \rho_0\mathbf{C}\}}, \tag{3.186}$$

with $\rho_0 \doteq \mathbf{h}^H \mathbf{\Sigma}^{-2} \mathbf{h}$ the true value of the effective SNR. Since $\mathbf{C}$ is positive definite and $\rho > 0$ it follows that δ_1 is decreasing with ρ_0, i.e. for $\rho_0 \geq 0$, δ_1 is upper bounded by

$$\delta_1 \leq \frac{\text{tr}\{(\mathbf{I}_K + 0\mathbf{C})(\mathbf{I}_K + \rho\mathbf{C})^{-1}\}}{\text{tr}\{\mathbf{I}_K + 0\mathbf{C}\}} = \delta_\perp. \tag{3.187}$$

Hence, $\lambda_{\min} = \delta_1$ is the smallest eigenvalue in (3.184) with $\mathbf{v}_1$ its associated eigenvector. Substituting (3.186) into (3.183) we obtain

$$\lim_{K \to \infty} T = \max_\rho \frac{1}{\det(\mathbf{I}_K + \rho\mathbf{C})} \left(\frac{\text{tr}\{\mathbf{I}_K + \rho_0\mathbf{C}\}}{\text{tr}\{(\mathbf{I}_K + \rho_0\mathbf{C})(\mathbf{I}_K + \rho\mathbf{C})^{-1}\}} \right)^K. \tag{3.188}$$

We will now show that the ρ maximizing the asymptotic GLRT is the true parameter $\rho = \rho_0$. Using the eigendecomposition of $\mathbf{C} = \mathbf{U}_\mathbf{C}^H \mathbf{\Lambda}_\mathbf{C} \mathbf{U}_\mathbf{C}$ we may isolate all the terms in (3.188) depending in ρ as

$$\zeta(\rho) \doteq \log \det(\mathbf{I}_K + \rho\mathbf{\Lambda}_\mathbf{C}) + K \log \text{tr}\{(\mathbf{I}_K + \rho_0\mathbf{\Lambda}_\mathbf{C})(\mathbf{I}_K + \rho\mathbf{\Lambda}_\mathbf{C})^{-1}\}. \tag{3.189}$$

Now, if we denote by $\lambda_1, \ldots, \lambda_K$ the elements of the diagonal matrix $\mathbf{\Lambda}_\mathbf{C}$, we have that the first and second derivatives of $\zeta(\rho)$ with respect to ρ are given by

$$\frac{\partial \zeta(\rho)}{\partial \rho} = \sum_{i=1}^K \frac{\lambda_i}{1 + \rho\lambda_i} - K \left(\sum_{i=1}^K \frac{1 + \rho_0\lambda_i}{1 + \rho\lambda_i} \right)^{-1} \sum_{i=1}^K \frac{\lambda_i(1 + \rho_0\lambda_i)}{(1 + \rho\lambda_i)^2}, \tag{3.190}$$

$$\frac{\partial^2 \zeta(\rho)}{(\partial \rho)^2} = -\sum_{i=1}^K \frac{\lambda_i^2}{(1 + \rho\lambda_i)^2} - K \left(\sum_{i=1}^K \frac{1 + \rho_0\lambda_i}{1 + \rho\lambda_i} \right)^{-2} \times$$
$$\left[\left(\sum_{i=1}^K \frac{\lambda_i(1 + \rho_0\lambda_i)}{(1 + \rho\lambda_i)^2} \right)^2 - 2 \left(\sum_{i=1}^K \frac{1 + \rho_0\lambda_i}{1 + \rho\lambda_i} \right) \left(\sum_{i=1}^K \frac{(1 + \rho_0\lambda_i)\lambda_i^2}{(1 + \rho\lambda_i)^3} \right) \right], \tag{3.191}$$

respectively. Evaluating these derivatives at $\rho = \rho_0$ we obtain:

$$\frac{\partial \zeta(\rho)}{\partial \rho} \bigg|_{\rho=\rho_0} = 0, \tag{3.192}$$

$$\frac{\partial^2 \zeta(\rho)}{(\partial \rho)^2} \bigg|_{\rho=\rho_0} = \sum_{i=1}^K \left(\frac{\lambda_i}{1 + \rho_0\lambda_i} \right)^2 - \frac{1}{K} \left(\sum_{i=1}^K \frac{\lambda_i}{1 + \rho_0\lambda_i} \right)^2 \tag{3.193}$$

$$= \mathbf{x}^T \mathbf{x} - \frac{\mathbf{x}^T \mathbf{1}}{\mathbf{1}^T \mathbf{1}}, \tag{3.194}$$

where we have defined

$$\mathbf{x} \doteq \left[\frac{\lambda_1}{1 + \rho_0 \lambda_1} \ \cdots \ \frac{\lambda_K}{1 + \rho_0 \lambda_K} \right]^T .$$ (3.195)

From (3.192) we have that the first derivative equals zero at $\rho = \rho_0$, while from (3.194), using Cauchy-Schwarz inequality, it follows that $\partial^2 \zeta(\rho)/(\partial \rho)^2 \geq 0$ when evaluated at $\rho = \rho_0$. Hence $\rho = \rho_0$ is a minimum of $\zeta(\rho)$. Finally, substituting the optimum value $\rho = \rho_0$ into (3.188) we obtain the desired result.

Chapter 4

Wideband Spectrum Sensing

Contents

4.1 Introduction

Cognitive Radios must monitor a wide frequency band comprising a large number
of channels (say N_{ch}). In principle, different strategies are possible. For example,
the spectrum monitor may select one channel at a time, downconvert it to baseband
and perform spectral sensing on this single channel. To this end, we may use any of
the multiantenna schemes presented in the previous chapters which enable spectral
sensing in a single primary channel. On the other hand, it is desirable to process
the whole bandwidth of interest simultaneously in order to increment the agility and
detection performance of the system. However, the requirements on dynamic range
and linearity for the analog stage, as well as on the sampling rate and resolution of
the analog to digital converter (ADC) determine the maximum bandwidth that can
be simultaneously processed.

In order to avoid this drawbacks and make wideband processing practical we
may divide the band of interest into subbands comprising $M < N_{\mathrm{ch}}$ primary channels
and process them sequentially. Hence, by suitably choosing M we obtain a trade-off
between detection performance and complexity of the analog stage. However, in
some cases, the large bandwidth involved makes Nyquist-rate wideband monitoring
impractical, due to power consumption and digital processing constraints. In this
scenario an analog to information (A2I) converter could be used in place of the
classical ADC in order to obtain a sub-Nyquist sampling rate.

In this chapter we will assume that the primary network employs FDMA with
known channelization and we will restrict our analysis to single-antenna secondary
systems. Under these assumptions we consider the problem of detecting primary
users when the analog to digital converter acquires a subband comprising $M \leq N_{\mathrm{ch}}$
primary channels. This wideband approach provides more information about the
background noise level, a parameter that must be estimated in practice. We will
first consider a spectral monitor employing a conventional ADC without any further
assumptions on the primary network activity on the band. Under this model we
will derive the GLRT for the detection of a given channel of the band in cases of
practical interest. As a second step, we will further elaborate the model to consider
an A2I converter and certain *a priori* information on the primary activity of the
band. Under this model we will show a connection between a Bayesian approach to
primary user detection and compressed sensing (CS) theory.

4.2   Problem formulation

The primary network uses Frequency Division Multiple Access (FDMA), with fixed channelization known to the spectrum monitor. Several primary channels are sensed simultaneously, by selecting a wide band containing M of such channels, downconverting it to baseband and sampling the resulting analog signal through an analog to digital or through an analog to information converter.

The baseband analog signal at the receiver after wideband filtering is given by

$$r(t) = \sum_{m=1}^{M} \sigma_m s_m(t) + \sigma n(t), \tag{4.1}$$

where $n(t)$ is a zero-mean, circular complex Gaussian noise with unit variance, assumed to be frequency flat in the captured bandwidth; σ^2 is the background noise power; $s_m(t)$ is the (noiseless) primary signal in channel m, normalized to have unit variance $(\mathrm{E}[|s_m(t)|^2] = 1)$; and σ_m^2 is the power of the primary signal in the mth channel.

Following the motivation exposed in the previous chapters, the primary transmissions $\{s_m(\cdot)\}$ will be assumed Gaussian distributed. Note that this assumption applies specially to this wideband set-up, since the secondary synchronization loop cannot be simultaneously locked to the parameters of multiple primary networks. As a result, possibly existing signal structure, such as pilots or cyclostationary features, degrades strongly due to synchronization errors. Then, the signals $\{s_m(\cdot)\}$ will be modeled as wide-sense stationary, zero mean circular Gaussian processes. Since $\{s_m(\cdot)\}$ correspond to different primary transmissions, they are assumed statistically independent.

4.2.1   Wideband acquisition

We restrict our study to linear ADC and A2I converters, that is, converters that can be represented in matrix form from an oversampled version of the analog signal $r(t)$. For compactness we define $s_0(t) \doteq n(t)$ and $\sigma_0^2 \doteq \sigma^2$. The finite discrete representation of (4.1) at Nyquist rate using the obvious vector notation can be written as

$$\mathbf{y} = \sum_{m=0}^{M} \sigma_m \mathbf{s}_m, \tag{4.2}$$

where $\mathbf{y}$ and $\mathbf{s}_m$ with $m = 0, \ldots, M$ are now $N \times 1$ circular Gaussian vectors, with zero mean and covariance matrix $\mathbf{C}_m \doteq \mathrm{E}[\mathbf{s}_m \mathbf{s}_m^H]$. Due to the normalization and stationarity of the original processes $\{s_m(\cdot)\}$, $\mathbf{C}_m$ is Toeplitz with ones on the diagonal; whereas $\mathbf{C}_0 = \mathbf{I}$, since the noise is assumed white.

If we define the $K \times N$ compression matrix $\mathbf{\Phi}$, with $K \leq N$, we can write the signal available to the digital spectrum monitor as

$$\tilde{\mathbf{y}} \doteq \mathbf{\Phi} \mathbf{y} = \sum_{m=0}^{M} \sigma_m \tilde{\mathbf{s}}_m, \qquad (4.3)$$

with $\tilde{\mathbf{s}}_m \doteq \mathbf{\Phi} \mathbf{s}_m$.

4.2.2   Signal model

If the channels from the primary transmitters to the monitor are frequency flat[1], then the $\mathbf{C}_m$ are known, and they summarize the knowledge about the primary network (channelization and spectral shape of transmissions) available to the spectrum monitor. Since $\mathbf{s}_i$, $\mathbf{s}_j$ with $1 \leq i \neq j \leq M$ correspond to different primary transmissions, they are regarded as statistically independent, and also independent of the background noise $\mathbf{s}_0$. Hence, under this model the observation $\tilde{\mathbf{y}}$ is zero-mean circular Gaussian with covariance

$$\tilde{\mathbf{R}}(\boldsymbol{\sigma}) \doteq \mathrm{E}[\tilde{\mathbf{y}} \tilde{\mathbf{y}}^H] = \sum_{i=0}^{M} \sigma_i^2 \tilde{\mathbf{C}}_i \qquad (4.4)$$

where $\tilde{\mathbf{C}}_i \doteq \mathbf{\Phi} \mathbf{C}_i \mathbf{\Phi}^H$. Note that in (4.4) we have made explicit the dependence of $\tilde{\mathbf{R}}$ with the vector of unknown parameters

$$\boldsymbol{\sigma} \doteq \left[\begin{array}{cccc} \sigma_0^2 & \sigma_1^2 & \cdots & \sigma_M^2 \end{array}\right]^T. \qquad (4.5)$$

Under the Gaussian model, second-order statistics capture all relevant information about the problem. In order to ensure identifiability of the parameter vector $\boldsymbol{\sigma}$ from (4.4), it is assumed that the $\{\mathbf{C}_m\}_{m=0}^{M}$ are linearly independent. (As it turns out, this condition amounts to requiring that the psds of the signals $\{s_m(t)\}_{m=0}^{M}$, denoted by $\{S_m(e^{j\omega})\}_{m=0}^{M}$, be linearly independent). Were this not the case, it would

---

[1] The effect of *unknown* frequency selective channels on the proposed detectors will be considered later using a realistic channel model by means of simulations.

be impossible to distinguish among primary users with linearly dependent emission masks. This assumption is clearly valid in FDMA scenarios, which are the focus of the current chapter[2].

4.2.3 Hypothesis testing problem

The problem is to determine the subset of idle channels in $\{1, \ldots, M\}$. This could be cast as an Hypothesis Testing Problem with 2^M different hypotheses; however, multiple hypothesis testing in the presence of unknown parameters is a difficult problem (Kay, 1998), so we consider instead successive detection of the M channels, one by one. For the m-th channel, the problem becomes:

$$\mathcal{H}_0^m : \quad \sigma_m^2 = 0 \qquad \text{(primary is absent in channel } m), \tag{4.6}$$

$$\mathcal{H}_1^m : \quad \sigma_m^2 > 0 \qquad \text{(primary is present in channel } m). \tag{4.7}$$

This is a composite problem, since the probability density function (pdf) f of the observations under the two hypotheses depends on the vector of unknown parameters $\boldsymbol{\sigma}$. We consider the Generalized Likelihood Ratio Test (GLRT)

$$T_{\text{GLRT}} \doteq \frac{\max_{\boldsymbol{\sigma}|\mathcal{H}_0^m} f(\tilde{\mathbf{y}} \,|\, \boldsymbol{\sigma})}{\max_{\boldsymbol{\sigma}|\mathcal{H}_1^m} f(\tilde{\mathbf{y}} \,|\, \boldsymbol{\sigma})} \tag{4.8}$$

$$= \frac{f(\tilde{\mathbf{y}} \,|\, \hat{\boldsymbol{\sigma}}_{\text{ML},0})}{f(\tilde{\mathbf{y}} \,|\, \hat{\boldsymbol{\sigma}}_{\text{ML},1})} \qquad \begin{matrix} \mathcal{H}_0^m \\ \gtrless \\ \mathcal{H}_1^m \end{matrix} \quad \gamma', \tag{4.9}$$

with γ' a threshold, and $\hat{\boldsymbol{\sigma}}_{\text{ML},j}$ the ML estimate of $\boldsymbol{\sigma}$ under $\mathcal{H}_j^m$.

Conditioned on $\boldsymbol{\sigma}$, the observations are Gaussian distributed:

$$f(\tilde{\mathbf{y}} \,|\, \boldsymbol{\sigma}) = \frac{\exp\left\{-\tilde{\mathbf{y}}^H \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma})\tilde{\mathbf{y}}\right\}}{\pi^K \det \tilde{\mathbf{R}}(\boldsymbol{\sigma})}. \tag{4.10}$$

Note that $\hat{\boldsymbol{\sigma}}_{\text{ML},1}$ is the maximizer of (4.10) w.r.t. $\boldsymbol{\sigma}$ subject to $\sigma_m^2 \geq 0$ for $0 \leq m \leq M$, whereas $\hat{\boldsymbol{\sigma}}_{\text{ML},0}$ is obtained by fixing $\sigma_m^2 = 0$ and maximizing (4.10) w.r.t. the remaining parameters in $\boldsymbol{\sigma}$ under the same constraint. Consequently, one has $f(\tilde{\mathbf{y}} \,|\, \hat{\boldsymbol{\sigma}}_{\text{ML},1}) \geq f(\tilde{\mathbf{y}} \,|\, \hat{\boldsymbol{\sigma}}_{\text{ML},0})$, so that the test statistic in (4.8) satisfies $0 \leq T_{\text{GLRT}} \leq 1$.

---

[2]Linear independence of emission masks may not hold if, for example, different primary users share the same bandwidth using Code Division Multiple Access.

4.3   Wideband spectrum sensing at Nyquist rate

In the first place we will assume that a conventional ADC is used and the input signal to the spectral monitor is acquired at Nyquist rate, i.e., $\boldsymbol{\Phi} = \mathbf{I}_K$, $\tilde{\mathbf{y}} = \mathbf{y}$ and $\tilde{\mathbf{C}}_i = \mathbf{C}_i$ for $i = 0, 1, \ldots, M$. Therefore, and for clarity of exposition, we will drop the $\tilde{\ }$ superscript in the rest of this section.

The GLRT under this scenario will allow us to study the signal features which can be exploited for detection purposes. Interestingly, the noise level estimation process will depend not only on the guard bands between channels, but also on the signal level found at channels perceived as "weak".

4.3.1   GLRT detection

We proceed now with the derivation of the GLRT. To this end we first need to obtain the ML estimate of $\boldsymbol{\sigma}$ under both hypotheses.

Note from (4.10) that the unknown parameter vector $\boldsymbol{\sigma}$ appears in the pdf through the covariance matrix $\mathbf{R}(\boldsymbol{\sigma})$ only. Therefore the problem reduces to the estimation of a covariance matrix with structure given by (4.4) with $\sigma_m^2 \geq 0$ for $m = 0, \ldots, M$, and thus it fits in the framework addressed in Burg et al. (1982). Here we follow a slightly different approach to derive the conditions on the unconstrained ML estimate, which will lead to a simplified closed-form estimator which is asymptotically efficient for certain cases of practical interest.

Unconstrained ML estimation

ML estimation amounts to minimizing the negative of the log-likelihood function

$$L(\mathbf{y}; \boldsymbol{\sigma}) \doteq \ln \det \mathbf{R}(\boldsymbol{\sigma}) + \mathbf{y}^H \mathbf{R}^{-1}(\boldsymbol{\sigma}) \mathbf{y}. \tag{4.11}$$

The partial derivatives of $L(\mathbf{y}; \boldsymbol{\sigma})$ w.r.t. σ_i^2 are

$$\frac{\partial L(\mathbf{y}; \boldsymbol{\sigma})}{\partial \sigma_i^2} = -\operatorname{tr}\{\mathbf{R}^{-1}(\boldsymbol{\sigma}) \mathbf{C}_i\} + \mathbf{y}^H \mathbf{R}^{-1}(\boldsymbol{\sigma}) \mathbf{C}_i \mathbf{R}^{-1}(\boldsymbol{\sigma}) \mathbf{y}. \tag{4.12}$$

Neglecting the positivity constraints $\sigma_j^2 \geq 0$, the unconstrained ML estimate of $\boldsymbol{\sigma}$ satisfies $\partial L / \partial \sigma_i^2 = 0$ for $0 \leq i \leq M$. In view of (4.12), the natural approach to solving these equations seems to be the diagonalization of the matrices involved. To

this end we will make use of the asymptotic diagonalization of Toeplitz matrices introduced in (2.5).

As $K \to \infty$, the following approximation holds:

$$\mathbf{C}_m \approx \mathbf{W}\mathbf{\Lambda}_m\mathbf{W}^H, \qquad m = 0, 1, \ldots, M, \tag{4.13}$$

where $\mathbf{W}$ denotes the $K \times K$ orthonormal IDFT matrix, $\mathbf{\Lambda}_m \doteq \mathrm{diag}(\boldsymbol{\lambda}_m)$, and $\boldsymbol{\lambda}_m \doteq [\lambda_0^{(m)} \ \lambda_1^{(m)} \ \cdots \ \lambda_{K-1}^{(m)}]^T$ with

$$\lambda_k^{(m)} \doteq S_m(e^{j\frac{2\pi k}{K}}), \quad 0 \le k \le K - 1. \tag{4.14}$$

Substituting (4.13) into (4.4), it follows that, as $K \to \infty$,

$$\mathbf{R} \approx \mathbf{W}\mathbf{\Delta}(\boldsymbol{\sigma})\mathbf{W}^H, \qquad \text{with} \qquad \mathbf{\Delta}(\boldsymbol{\sigma}) \doteq \sum_{i=0}^{M} \sigma_i^2 \mathbf{\Lambda}_i. \tag{4.15}$$

Note that $\mathbf{\Delta}(\boldsymbol{\sigma}) = \mathrm{diag}\{[\ \delta_0(\boldsymbol{\sigma}) \ \ \delta_1(\boldsymbol{\sigma}) \ \ \cdots \ \ \delta_{K-1}(\boldsymbol{\sigma}) \ ]\}$ contains uniformly spaced samples of the psd of the observations, given by

$$\delta_k(\boldsymbol{\sigma}) \doteq \sum_{j=0}^{M} \sigma_j^2 \lambda_k^{(j)}, \qquad 0 \le k \le K - 1. \tag{4.16}$$

With this asymptotic diagonalization of $\mathbf{y}$, we can substitute (4.15) back into (4.12) to obtain

$$\begin{aligned} \frac{\partial L(\mathbf{y}; \boldsymbol{\sigma})}{\partial \sigma_i^2} &\approx -\mathrm{tr}\left\{\mathbf{\Delta}^{-1}(\boldsymbol{\sigma})\mathbf{\Lambda}_i\right\} \\ &\quad + \mathbf{v}^H \mathbf{\Delta}^{-1}(\boldsymbol{\sigma})\mathbf{\Lambda}_i\mathbf{\Delta}^{-1}(\boldsymbol{\sigma})\mathbf{v}, \end{aligned} \tag{4.17}$$

where

$$\mathbf{v} \doteq \mathbf{W}^H \mathbf{y} = [v_0 \, v_1 \cdots v_{K-1}]^T \tag{4.18}$$

is the DFT of the observations.

Then, equating (4.17) to zero, we find that as $K \to \infty$ the unconstrained ML estimate $\hat{\boldsymbol{\sigma}}_{\mathrm{ML}}$ will satisfy

$$\sum_{k=0}^{K-1} \frac{\lambda_k^{(i)}}{\delta_k(\hat{\boldsymbol{\sigma}}_{\mathrm{ML}})} = \sum_{k=0}^{K-1} \frac{|v_k|^2 \lambda_k^{(i)}}{\delta_k^2(\hat{\boldsymbol{\sigma}}_{\mathrm{ML}})}, \quad 0 \le i \le M. \tag{4.19}$$

While in general it is not possible to obtain $\hat{\boldsymbol{\sigma}}_{\mathrm{ML}}$ in closed form[3] from the conditions (4.19), it is possible to obtain approximate closed-form solutions as we will see next.

Unconstrained Least Squares estimation

The left-hand side of (4.19) can be rewritten as

$$\sum_{k=0}^{K-1} \frac{\lambda_k^{(i)}}{\delta_k} = \sum_{k=0}^{K-1} \frac{\delta_k \lambda_k^{(i)}}{\delta_k^2} \tag{4.20}$$

$$= \sum_{j=0}^{M} \sigma_j^2 \left(\sum_{k=0}^{K-1} \frac{\lambda_k^{(j)} \lambda_k^{(i)}}{\delta_k^2} \right). \tag{4.21}$$

Substituting (4.21) into (4.19), one obtains, in matrix form,

$$\mathbf{L}^H \boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\sigma}}_{\mathrm{ML}}) \mathbf{L} \hat{\boldsymbol{\sigma}}_{\mathrm{ML}} = \mathbf{L}^H \boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\sigma}}_{\mathrm{ML}}) \mathbf{p}, \tag{4.22}$$

where the $K \times (M+1)$ matrix $\mathbf{L}$ and the $K \times 1$ vector $\mathbf{p}$ (the periodogram) are respectively defined as

$$\mathbf{L} \doteq [\ \boldsymbol{\lambda}_0 \quad \boldsymbol{\lambda}_1 \quad \cdots \quad \boldsymbol{\lambda}_M\ ], \tag{4.23}$$

$$\mathbf{p} \doteq [\ |v_0|^2 \quad |v_1|^2 \quad \cdots \quad |v_{K-1}|^2\ ]^T \tag{4.24}$$

$$= [\ p_0 \quad p_1 \quad \cdots \quad p_{K-1}\ ]^T. \tag{4.25}$$

Note that the periodogram $\mathbf{p}$ is an asymptotically unbiased estimate of the psd of the observations (Stoica and Moses, 2005), and therefore $\mathbf{p}_\star \doteq \lim_{K\to\infty} \mathrm{E}[\mathbf{p}] = \mathbf{L}\boldsymbol{\sigma}$, with $\boldsymbol{\sigma}$ the vector of true parameters. Thus, asymptotically, the expected value of $\mathbf{p}$ lies in the subspace spanned by the columns of $\mathbf{L}$. The linear independence assumption on the psds $\{S_i(e^{j\omega}),\ 0 \leq i \leq M\}$ implies that $\mathbf{L}$ has full column rank, so that $\mathbf{L}^\dagger \mathbf{L} = \mathbf{I}_{M+1}$ with $\mathbf{L}^\dagger$ denoting the pseudoinverse of $\mathbf{L}$. Then it holds that

$$\mathbf{L}\mathbf{L}^\dagger \mathbf{p}_\star = \mathbf{L}\mathbf{L}^\dagger \mathbf{L}\boldsymbol{\sigma} = \mathbf{L}\boldsymbol{\sigma} = \mathbf{p}_\star, \tag{4.26}$$

---

[3]These nonlinear equations can be solved numerically by efficient fixed-point iterative algorithms (Burg et al., 1982; López-Valcarce and Vazquez-Vilar, 2009).

which suggests the approximation $\mathbf{p} \approx \mathbf{L}\mathbf{L}^\dagger\mathbf{p}$. Substituting this in (4.22),

$$\begin{aligned}
\hat{\boldsymbol{\sigma}}_{\text{ML}} &\approx \left[\mathbf{L}^H\boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\sigma}}_{\text{ML}})\mathbf{L}\right]^{-1}\mathbf{L}^H\boldsymbol{\Delta}^{-2}(\hat{\boldsymbol{\sigma}}_{\text{ML}})\mathbf{L}\mathbf{L}^\dagger\mathbf{p} \\
&= \mathbf{L}^\dagger\mathbf{p} \doteq \hat{\boldsymbol{\sigma}}_{\text{LS}}.
\end{aligned} \tag{4.27}$$

The LS subscript refers to the fact that this estimate is the solution to the unconstrained Least Squares problem $\min_{\hat{\boldsymbol{\sigma}}} \|\mathbf{L}\hat{\boldsymbol{\sigma}} - \mathbf{p}\|_2^2$. The rows of $\mathbf{L}^\dagger$ can be interpreted as *matched filters* that combine the power in the different frequency bins (the entries of $\mathbf{p}$) in order to estimate the variances in each channel. The LS estimate is asymptotically unbiased, with covariance given by $\text{cov}(\hat{\boldsymbol{\sigma}}_{\text{LS}}) = \mathbf{L}^\dagger \text{cov}(\mathbf{p})(\mathbf{L}^\dagger)^H$. Since the asymptotic covariance of $\mathbf{p}$ is given by $\lim_{K\to\infty} \boldsymbol{\Delta}^2(\boldsymbol{\sigma})$ (Stoica and Moses, 2005), one finds that

$$\lim_{K\to\infty} \text{cov}(\hat{\boldsymbol{\sigma}}_{\text{LS}}) = \lim_{K\to\infty} \mathbf{L}^\dagger\boldsymbol{\Delta}^2(\boldsymbol{\sigma})(\mathbf{L}^\dagger)^H. \tag{4.28}$$

Cramér-Rao Lower Bound

Under the Gaussian model and assuming $\boldsymbol{\Phi} = \mathbf{I}_K$, the elements of the Fisher information matrix (FIM) $\boldsymbol{F}(\boldsymbol{\sigma})$ are given by (see e.g. Kay (1993)):

$$[\boldsymbol{F}(\boldsymbol{\sigma})]_{ij} = \text{tr}\left\{\mathbf{R}^{-1}(\boldsymbol{\sigma})\frac{\partial\mathbf{R}(\boldsymbol{\sigma})}{\partial\sigma_i^2}\mathbf{R}^{-1}(\boldsymbol{\sigma})\frac{\partial\mathbf{R}(\boldsymbol{\sigma})}{\partial\sigma_j^2}\right\}. \tag{4.29}$$

The Cramér-Rao Lower Bound (CRLB) for any unbiased estimator of $\boldsymbol{\sigma}$ is then given by $\text{var}(\hat{\sigma}_i^2) \geq [\boldsymbol{F}^{-1}(\boldsymbol{\sigma})]_{ii}$. In our case, $\partial\mathbf{R}(\boldsymbol{\sigma})/\partial\sigma_i^2 = \mathbf{C}_i$. Then, using the asymptotic approximations (4.13) and (4.15),

$$\begin{aligned}
[\boldsymbol{F}(\boldsymbol{\sigma})]_{ij} &\approx \text{tr}\left\{\boldsymbol{\Delta}^{-1}(\boldsymbol{\sigma})\boldsymbol{\Lambda}_i\boldsymbol{\Delta}^{-1}(\boldsymbol{\sigma})\boldsymbol{\Lambda}_j\right\} \tag{4.30} \\
&= \sum_{k=0}^{K-1}\frac{\lambda_k^{(i)}\lambda_k^{(j)}}{\delta_k^2(\boldsymbol{\sigma})}. \tag{4.31}
\end{aligned}$$

Thus, the asymptotic FIM is given by

$$\lim_{K\to\infty} \boldsymbol{F}(\boldsymbol{\sigma}) = \lim_{K\to\infty} \mathbf{L}^H\boldsymbol{\Delta}^{-2}(\boldsymbol{\sigma})\mathbf{L}. \tag{4.32}$$

Comparing (the inverse of) (4.32) with (4.28), it is seen that in general the LS estimate (4.27) does not necessarily achieve the CRLB. In Section 4.3.2 a particular setting will be discussed for which it can be shown that the LS estimate is

asymptotically efficient, i.e. it achieves the CRLB as $K \to \infty$.

Quasi-GLRT detection

Let $\hat{\boldsymbol{\sigma}}_j = [\ \hat{\sigma}_{0j}^2 \quad \hat{\sigma}_{1j}^2 \quad \cdots \quad \hat{\sigma}_{Mj}^2 \ ]^T$ denote an estimate (ML, LS, or other) of $\boldsymbol{\sigma}$ under $\mathcal{H}_j^m$, $j \in \{0, 1\}$. Using these estimates in the detection test, we obtain an approximation to the GLRT. Using the asymptotic diagonalization (4.15), we can write

$$\det \mathbf{R}(\hat{\boldsymbol{\sigma}}_j) \approx \det \boldsymbol{\Delta}(\hat{\boldsymbol{\sigma}}_j) = \prod_{k=0}^{K-1} \left[\sum_{i=0}^{M} \hat{\sigma}_{ij}^2 \lambda_k^{(i)} \right], \tag{4.33}$$

$$\mathbf{y}^H \mathbf{R}^{-1}(\hat{\boldsymbol{\sigma}}_j)\mathbf{y} \approx \mathbf{v}^H \boldsymbol{\Delta}^{-1}(\hat{\boldsymbol{\sigma}}_j)\mathbf{v} = \sum_{k=0}^{K-1} \frac{p_k}{\sum_{i=0}^{M} \hat{\sigma}_{ij}^2 \lambda_k^{(i)}}. \tag{4.34}$$

The resulting "Quasi-GLRT" (QGLRT) can be written as

$$\log \frac{f(\mathbf{y} \mid \hat{\boldsymbol{\sigma}}_0)}{f(\mathbf{y} \mid \hat{\boldsymbol{\sigma}}_1)} \approx \sum_{k=0}^{K-1} \log \left[\frac{\sum_{i=0}^{M} \hat{\sigma}_{i1}^2 \lambda_k^{(i)}}{\sum_{i=0}^{M} \hat{\sigma}_{i0}^2 \lambda_k^{(i)}} \right] + \sum_{k=0}^{K-1} \left[\frac{p_k}{\sum_{i=0}^{M} \hat{\sigma}_{i1}^2 \lambda_k^{(i)}} - \frac{p_k}{\sum_{i=0}^{M} \hat{\sigma}_{i0}^2 \lambda_k^{(i)}} \right] \tag{4.35}$$

$$\doteq \log T \tag{4.36}$$

This detector can be implemented once all $\hat{\boldsymbol{\sigma}}_j$ are available, either by numerical means (ML) or in closed form (LS). However, it is difficult in general to evaluate the performance of this detector or to obtain some intuition about its operation. In the next section we focus on a particular scenario whose structure will allow further simplification of (4.35).

4.3.2　Orthogonal frequency-flat signals in white noise

For FDMA-based primary networks, the signals in different channels are orthogonal, i.e. their psds have disjoint supports. In addition, the psd of a multicarrier signal is approximately constant within its support. In this section, the QGLRT (4.35) will be particularized to this setting.

Definition 4.1. A signal is *frequency-flat bandpass* if its psd takes only two levels: zero or a given constant value.

Definition 4.2. Two signals $s_k^{(i)}$, $s_k^{(j)}$ are *non-partially overlapping* if either their psds

have disjoint supports, or the support of one of them contains that of the other.

For this class of signals, it turns out that the LS estimate (4.27) is asymptotically efficient:

Theorem 4.1. If the signals $s_k^{(i)}$ and $s_k^{(j)}$ are frequency-flat bandpass and non-partially overlapping for any $0 \leq i, j \leq M$, then the asymptotic covariance matrix (4.28) of the unconstrained LS estimate equals the inverse of the asymptotic FIM (4.32).

Proof. See Appendix 4.A. □

In the following we will assume that $s_k^{(i)}$, $i = 1, \ldots, M$, are frequency-flat bandpass with disjoint frequency supports. Since $s_k^{(0)}$ (white noise) is frequency-flat bandpass covering the whole bandwidth, it follows that $s_k^{(i)}$, $s_k^{(j)}$ are non-partially overlapping for any $0 \leq i, j \leq M$. This is a special case of the broader family of non-partially overlapping frequency-flat bandpass signals, and will be denoted here as orthogonal frequency-flat signals in white noise. For this class of signals, Theorem 4.1 motivates the use of LS estimates in the QGLRT.

QGLRT with unconstrained LS estimates

Let $\mathcal{W}_i$ denote the set of frequency bins within the support of $S_i(e^{j\omega})$, $i = 1, \ldots, M$, and let the set of "noise-only" frequency bins (comprising all guard bands in the captured bandwidth) be

$$\mathcal{W}_0 \doteq \{k : k \in \{0, 1, \ldots, K - 1\} \text{ and } k \notin \cup_{i=1}^{M} \mathcal{W}_i\}. \tag{4.37}$$

We also define the *fractional bandwidths* $w_i \doteq |\mathcal{W}_i|/K$, $0 \leq i \leq M$, such that $0 < w_i < 1$ and $\sum_{i=0}^{M} w_i = 1$. Since the signals are normalized to unit variance, it follows that

$$[\mathbf{L}]_{ki} = \lambda_k^{(i)} = \begin{cases} 1, & i = 0, \\ \frac{1}{w_i}, & k \in \mathcal{W}_i, i = 1, \ldots, M, \\ 0, & \text{otherwise.} \end{cases} \tag{4.38}$$

The pseudoinverse of $\mathbf{L}$ is given in this case by

$$[\mathbf{L}^{\dagger}]_{ik} = \begin{cases} \frac{1}{Kw_0}, & k \in \mathcal{W}_0, i = 0, \\ \frac{-w_i}{Kw_0}, & k \in \mathcal{W}_0, i \neq 0, \\ \frac{1}{K}, & k \in \mathcal{W}_i, i \neq 0, \\ 0, & k \in \mathcal{W}_j, j \neq i, j \neq 0, \end{cases} \tag{4.39}$$

as can be seen by checking that $\mathbf{L}^\dagger \mathbf{L} = \mathbf{I}_{M+1}$. Let us denote the averaged periodogram over $\mathcal{W}_i$ by

$$q_i \doteq \frac{1}{Kw_i} \sum_{k \in \mathcal{W}_i} p_k, \quad 0 \le i \le M. \tag{4.40}$$

The unconstrained LS estimate of $\boldsymbol{\sigma}$ under $\mathcal{H}_1^m$ is just $\mathbf{L}^\dagger \mathbf{p}$, and is given by

$$\hat{\sigma}_{i1}^2 = \begin{cases} q_0, & i = 0, \\ w_i(q_i - q_0), & i \ne 0. \end{cases} \tag{4.41}$$

On the other hand, the unconstrained LS estimate under $\mathcal{H}_0^m$ is such that the subband corresponding to channel m is consolidated into the "noise-only" set:

$$\hat{\sigma}_{i0}^2 = \begin{cases} q_{0m}, & i = 0, \\ 0, & i = m, \\ w_i(q_i - q_{0m}), & i \ne 0, i \ne m, \end{cases} \tag{4.42}$$

where

$$q_{0m} \doteq (w_0 q_0 + w_m q_m)/(w_0 + w_m). \tag{4.43}$$

If the estimates (4.41)-(4.42) are used in the QGLRT, then some straightforward algebra shows that (4.35) reduces to

$$\frac{1}{K} \log T = (w_0 + w_m) \log \frac{\left(q_0^{w_0} q_m^{w_m}\right)^{\frac{1}{w_0+w_m}}}{q_{0m}}. \tag{4.44}$$

The argument of the log in (4.44) is the *weighted geometric to arithmetic mean ratio* of q_0 and q_m, with respective weights w_0, w_m. This ratio, which is a function of q_m/q_0 alone, is always less than or equal to one, with equality iff $q_0 = q_m$. It is monotonically increasing for $q_m/q_0 < 1$, and decreasing for $q_m/q_0 > 1$. Thus, the QGLRT with unconstrained LS estimates decides that channel m is idle if $q_m/q_0 \in [\alpha, \beta]$, for some $\alpha < 1 < \beta$ depending on the threshold, and busy otherwise. This is against intuition, since $q_m < q_0$ is always a reasonable indicator of an idle channel. Note that if $q_m < q_0$, then the unconstrained LS estimate of σ_m^2 under $\mathcal{H}_1^m$ is $\hat{\sigma}_{m1}^2 = w_m(q_m - q_0) < 0$, which is against prior knowledge. This motivates the use of *constrained* estimators in the QGLRT.

QGLRT with constrained LS estimates

The constrained LS estimate of $\boldsymbol{\sigma}$ for orthogonal frequency-flat signals in white noise is given next.

Theorem 4.2. Let $\mathbf{L}$ *be given by (4.38). The minimizer of* $\|\mathbf{L}\hat{\boldsymbol{\sigma}} - \mathbf{p}\|_2^2$ *subject to* $\hat{\sigma}_i^2 \geq 0$, $0 \leq i \leq M$ *(i.e. the constrained LS estimate under* $\mathcal{H}_1^m$*), takes the following form:*

$$\hat{\sigma}_{j1}^2 = \begin{cases} \frac{w_0 q_0 + \sum_{l \in \mathcal{U}_1} w_l q_l}{w_0 + \sum_{l \in \mathcal{U}_1} w_l}, & j = 0, \\ 0, & j \in \mathcal{U}_1, \\ w_j(q_j - \hat{\sigma}_{01}^2), & \text{otherwise,} \end{cases} \tag{4.45}$$

where $\mathcal{U}_1 \doteq \{j : q_j < \hat{\sigma}_{01}^2, j \neq 0\}$.

Analogously, the minimizer of $\|\mathbf{L}\hat{\boldsymbol{\sigma}} - \mathbf{p}\|_2^2$ *subject to* $\hat{\sigma}_m = 0$, $\hat{\sigma}_i^2 \geq 0$, $i \neq m$ *(i.e. the constrained LS estimate under* $\mathcal{H}_0^m$*), is given by:*

$$\hat{\sigma}_{j0}^2 = \begin{cases} \frac{w_0 q_0 + \sum_{l \in \mathcal{U}_0} w_l q_l}{w_0 + \sum_{l \in \mathcal{U}_0} w_l}, & j = 0, \\ 0, & j \in \mathcal{U}_0, \\ w_j(q_j - \hat{\sigma}_{00}^2), & \text{otherwise,} \end{cases} \tag{4.46}$$

where $\mathcal{U}_0 \doteq \{m\} \cup \{j : q_j < \hat{\sigma}_{00}^2, j \neq 0\}$.

Proof. See Appendix 4.B. □

Note that (4.45)-(4.46) are implicit expressions, since they depend on the sets $\mathcal{U}_1, \mathcal{U}_0$ whose definitions are in terms of $\hat{\sigma}_{01}^2$ and $\hat{\sigma}_{00}^2$ respectively. Nevertheless, these estimates and sets can be easily obtained using the Algorithm 2. It is straightforward to verify that this algorithm outputs sets $\mathcal{U}_1, \mathcal{U}_0$ and estimates $\{\hat{\sigma}_{j1}^2\}$, $\{\hat{\sigma}_{j0}^2\}$ satisfying (4.45) and (4.46) respectively. This scheme successively includes the "weakest" channel (i.e., the channel with smallest averaged periodogram over the corresponding frequency support) into the computation of the noise variance estimate, until this estimate falls below the estimated power levels of the remaining channels. Note that the only difference of the estimation algorithm under the two hypotheses comes from the initialization of the sets $\mathcal{U}_1$ and $\mathcal{U}_0$, with the latter always including the m-th channel.

In order to obtain the QGLRT based on the constrained LS estimates above, we distinguish two cases, depending on the strength with which channel m is perceived

---

Algorithm 2: Constrained LS estimate under $\mathcal{H}_k^m$ for $k = 0, 1$.

Input: Measured energies q_j.

Output: Constrained LS estimates $\hat{\sigma}_{jk}^2$ and $\mathcal{U}_k$.

Initialize:

Set $\mathcal{U}_k = \begin{cases} \{m\} & \text{for } k = 0, \\ \emptyset & \text{for } k = 1. \end{cases}$

repeat

     Set $\hat{\sigma}_{jk}^2 = 0$ for all $j \in \mathcal{U}_k$.

     Obtain the unconstrained estimates $\hat{\sigma}_{jk}^2$ for $j \notin \mathcal{U}_k$:

         $\hat{\sigma}_{0k}^2 = \frac{w_0 q_0 + \sum_{l \in \mathcal{U}_k} w_l q_l}{w_0 + \sum_{l \in \mathcal{U}_k} w_l}$

         $\hat{\sigma}_{jk}^2 = w_j(q_j - \hat{\sigma}_{0k}^2), \quad j \notin \mathcal{U}_k \cup \{0\}$

     if *obtained estimate is not feasible* **then**

         Let $j_\star = \arg\min_{j \notin \mathcal{U}_k \cup \{0\}} \{q_j\}$,

             update $\mathcal{U}_k \leftarrow \mathcal{U}_k \cup \{j_\star\}$

until obtained estimate is feasible

---

relative to the noise level. Note that by construction, $m \in \mathcal{U}_0$, whereas m may or may not belong to $\mathcal{U}_1$.

Case 1

$m \in \mathcal{U}_1$, so that channel m is perceived as "weak" under $\mathcal{H}_1^m$. Then we have the following.

Proposition 4.1. If $m \in \mathcal{U}_1$, then $\mathcal{U}_1 = \mathcal{U}_0$, so that the constrained LS estimates under $\mathcal{H}_1^m$ and $\mathcal{H}_0^m$ are the same.

Proof. See Appendix 4.C. $\square$

Therefore, if $m \in \mathcal{U}_1$, then from (4.35) we have $\log T = 0$, i.e. the QGLRT declares channel m as idle.

Case 2

$m \notin \mathcal{U}_1$, so that channel m is perceived as "not weak" under $\mathcal{H}_1^m$. Then one has:

Proposition 4.2. If $m \notin \mathcal{U}_1$, then $\mathcal{U}_1 \subset \mathcal{U}_0$.

Proof. See Appendix 4.D. □

Hence, if $m \notin \mathcal{U}_1$, then $\mathcal{U}_0 = \mathcal{U}_1 \cup \mathcal{S}$ for some set $\mathcal{S}$ with $\mathcal{S} \cap \mathcal{U}_1 = \emptyset$. For $j \notin \mathcal{U}_0 \cup \{0\}$, it turns out that $w_j \hat{\sigma}_{01}^2 + \hat{\sigma}_{j1}^2 = w_j \hat{\sigma}_{00}^2 + \hat{\sigma}_{j0}^2$, and thus these indices do not contribute to the QGLRT (4.35), which after some algebra is found to yield

$$\frac{1}{K} \log T = \log \frac{(\hat{\sigma}_{01}^2)^{w_0 + \sum_{j \in \mathcal{U}_1} w_j} \prod_{j \in \mathcal{S}} q_j^{w_j}}{(\hat{\sigma}_{00}^2)^{w}}, \tag{4.47}$$

where $w \doteq w_0 + \sum_{j \in \mathcal{U}_0} w_j$. Note that if $\mathcal{S} = \{m\}$, then this ratio becomes a monotonically decreasing function of $q_m / \hat{\sigma}_{01}^2 \geq 1$. Hence, if $\mathcal{U}_0 = \mathcal{U}_1 \cup \{m\}$, the QGLRT can be written as

$$\frac{q_m}{\hat{\sigma}_{01}^2} \underset{\mathcal{H}_0^m}{\overset{\mathcal{H}_1^m}{\gtrless}} \gamma \qquad \text{(Test 1).} \tag{4.48}$$

Remark 4.1. Test 1 compares the power measured in channel m to a threshold $\gamma \hat{\sigma}_{01}^2$ proportional to the estimated noise power. Note that the noise power is estimated using not only the guard bands (corresponding to $\mathcal{W}_0$), but also those channels perceived as weak (i.e. each channel j for which the constrained LS power estimate yields a zero value so that $j \in \mathcal{U}_1$).

If $\mathcal{S} \neq \{m\}$, then the sets of weak channels estimated under $\mathcal{H}_0^m$ and $\mathcal{H}_1^m$ are different and it is not possible to reduce (4.47) to a simple ratio of averaged powers. A possible approach is to disregard the influence of the weak channels with indexes $j \in \mathcal{S}$, obtaining (4.48). Another possibility is to take these channels into account in order to obtain a new estimate of the noise power

$$\hat{\sigma}_{02}^2 \doteq \frac{w_0 q_0 + \sum_{j \in \mathcal{U}_0, j \neq m} w_j q_j}{w_0 + \sum_{j \in \mathcal{U}_0, j \neq m} w_j}, \tag{4.49}$$

and then use (4.49) in the following test:

$$\frac{q_m}{\hat{\sigma}_{02}^2} \underset{\mathcal{H}_0^m}{\overset{\mathcal{H}_1^m}{\gtrless}} \gamma \qquad \text{(Test 2),} \tag{4.50}$$

which reduces to (4.48) if $\mathcal{S} = \{m\}$.

4.3.3 Statistical analysis

The detectors from Section 4.3.2 are based on the random variables q_i, $0 \le i \le M$. Note that $\mathbf{v} = \mathbf{W}^H \mathbf{y}$ is Gaussian with a diagonal (asymptotic) covariance matrix $\mathbf{\Delta}(\boldsymbol{\sigma})$. For orthogonal frequency-flat signals in white noise, the diagonal of $\mathbf{\Delta}(\boldsymbol{\sigma})$ is piecewise constant, and in particular, its elements are constant over the bins corresponding to the i-th channel (this also applies to the set of "noise-only" bins). Hence, q_i is the sum of square magnitudes of zero-mean Gaussian random variables, asymptotically uncorrelated and with the same variance. Thus, for large K, q_i becomes chi-squared distributed with Kw_i degrees of freedom; in turn, as $K \to \infty$, this distribution converges to a Gaussian distribution: $q_i \sim \mathcal{N}(\mu_i, \alpha_i^2)$. Moreover, q_i, q_j are asymptotically uncorrelated for $i \ne j$, since the two sets of bins used for their computation are disjoint. In terms of the SNR in channel i, defined as

$$\rho_i \doteq \sigma_i^2/(w_i \sigma_0^2), \tag{4.51}$$

the mean and variance of q_i are given by

$$\mu_i \doteq \begin{cases} \sigma_0^2, & i = 0, \\ \sigma_0^2(1 + \rho_i), & i > 0, \end{cases} \tag{4.52}$$

$$\alpha_i^2 \doteq \begin{cases} \frac{\sigma_0^4}{Kw_0}, & i = 0, \\ \frac{\sigma_0^4}{Kw_i}(1 + \rho_i)^2, & i > 0, \end{cases} \tag{4.53}$$

as can be readily found from the definition of q_i in (4.40).

Single-channel detection with guard bands

As a first step, we analyze the case $M = 1$, for which all of the proposed detectors boil down to the same test. This test can be expressed as $z \doteq q_1 - \gamma q_0 \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} 0$, where $\gamma > 1$ is a threshold and the statistic z follows a Gaussian distribution:

$$z \sim \mathcal{N}\left(\sigma_0^2(1 + \rho_1 - \gamma), \frac{\sigma_0^4}{K}\left[\frac{(1 + \rho_1)^2}{w_1} + \frac{\gamma^2}{w_0}\right]\right). \tag{4.54}$$

Since $\sigma_0^2 > 0$, the probabilities of false alarm and detection can be respectively written as $P_{\mathrm{FA}} = \Pr\{(z/\sigma_0^2) > 0 \,|\, \rho_1 = 0\}$ and $P_{\mathrm{D}} = \Pr\{(z/\sigma_0^2) > 0 \,|\, \rho_1 > 0\}$. These probabilities do not depend on the noise power σ_0^2, as expected. In order to set the threshold γ, two approaches are possible:

Threshold for fixed P_{FA}. In view of (4.54), it is readily found that, for $P_{\text{FA}} \leq 0.5$,

$$\gamma(P_{\text{FA}}) = \frac{1 + \sqrt{1 - \left(1 - \frac{[Q^{-1}(P_{\text{FA}})]^2}{Kw_0}\right)\left(1 - \frac{[Q^{-1}(P_{\text{FA}})]^2}{Kw_1}\right)}}{1 - \frac{[Q^{-1}(P_{\text{FA}})]^2}{Kw_0}}, \tag{4.55}$$

where $Q(\cdot)$ is the complementary Gaussian cumulative distribution function, and $Q^{-1}(\cdot)$ denotes its inverse. The resulting probability of detection for an SNR ρ_1 is then given by

$$P_{\text{D}} = Q\left(\sqrt{Kw_0}\frac{\gamma(P_{\text{FA}}) - (1 + \rho_1)}{\sqrt{\gamma^2(P_{\text{FA}}) + \frac{w_0}{w_1}(1 + \rho_1)^2}}\right). \tag{4.56}$$

Threshold for fixed P_{D}. In the context of cognitive radio systems, a false alarm results in a missed opportunity of using an idle channel, and therefore P_{FA} is related to the throughput efficiency of the secondary system. However, this parameter is irrelevant to the primary network. On the other hand, a missed detection may result in the secondary user accessing a channel in use, thus producing interference to the primary system. Regulatory bodies are likely to require a minimum detection performance to avoid collisions with primary (licensed) users (FCC, 2008), i.e. $P_{\text{D}} \geq P_{\text{D}}^{\star}$ at some target SNR $\rho_1^{\star}$. The threshold γ is then determined for $P_{\text{D}}^{\star} \geq 0.5$ as

$$\gamma(P_{\text{D}}^{\star}; \rho_1^{\star}) = (1 + \rho_1^{\star})\frac{1 - \sqrt{1 - \left(1 - \frac{\kappa}{Kw_0}\right)\left(1 - \frac{\kappa}{Kw_1}\right)}}{1 - \frac{\kappa}{Kw_0}}, \tag{4.57}$$

with $\kappa \doteq [Q^{-1}(P_{\text{D}}^{\star})]^2$. This yields

$$P_{\text{FA}} = Q\left(\sqrt{Kw_0}\frac{\gamma(P_{\text{D}}^{\star}; \rho_1^{\star}) - 1}{\sqrt{\gamma^2(P_{\text{D}}^{\star}; \rho_1^{\star}) + \frac{w_0}{w_1}}}\right). \tag{4.58}$$

Multichannel detection

Single-channel spectrum sensing, as described in the previous subsection, exploits the presence of upper and lower guard bands to estimate the noise power. In practice, these guard bands will likely appear distorted, due to the transition bands of

the analog filter used for channel extraction. This may preclude the use of those frequency components for noise variance estimation. When $M > 1$ channels are simultaneously captured, the guard bands between adjacent channels remain undistorted, and therefore this problem is alleviated.

Without loss of generality, let $m = M$, so that the channel under scrutiny is the M-th one. In order to simplify the presentation, we restrict our analysis to the case in which all channels have the same bandwidth[4]: $w_1 = w_2 = \cdots = w_M$. Whereas finding analytical expressions for the performance of the QGLRT detector (4.47) seems intractable, an analysis of the simplified tests (4.48) and (4.50) is included in the Appendix 4.E.

The resulting distribution of the statistics (4.48) and (4.50) in the multichannel setting does not present a simple expression as when considering a single channel. Nevertheless, the probability of detection and probability of false alarm of the proposed Tests 1 and 2, can be written as the integral of a multivariate Gaussian distribution over the positive orthant. Therefore, for Tests 1 and 2 P_{FA} and P_{D} can be found for a given scenario without resorting to Monte Carlo simulations.

While no *a priori* assumptions have been made about the occupancy of the band in the derivation of these detectors, as it turns out, in a multichannel scenario their performance depends on the *a priori* probability of any given channel being in use by the primary network. This probability, or *activity factor,* will be denoted by a in what follows. In the next section we will see that in practical implementations of the proposed schemes it is possible to deal with unknown values of the *activity factor a* by considering worst case scenarios.

4.3.4   Numerical results and discussion

We evaluate now the performance of the proposed detectors (QGLRT (4.47), Test 1 (4.48) and Test 2 (4.50)), both theoretically and via Monte Carlo simulations. For the primary system we consider a terrestrial digital TV broadcast network using 8K-mode DVB-T modulation[5]. The channel spacing is 8 MHz with a 7.61 MHz signal bandwidth, which is one of the options considered in the DVB-T standard (ETSI, 2004) resulting in $w_1 = \cdots = w_M = 0.95125/M$, $w_0 = 0.04875$. A band

---

[4]The analysis can be readily modified in order to account for channels with different bandwidths, although the notation becomes somewhat cumbersome.

[5]For Monte Carlo simulation, the modulation parameters of the DVB-T signals were: 64-QAM, guard interval 1/4, inner code rate 2/3.
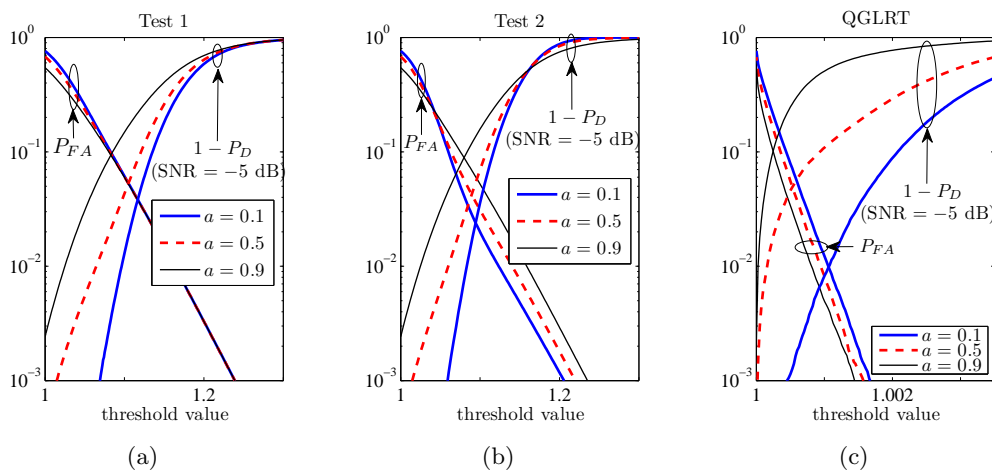
Figure 4.1: False alarm and missed detection performance in a setting with $M = 4$ channels. (a) Test 1 (analytical). (b) Test 2 (analytical). (c) QGLRT (empirical).

comprising M of such DVB-T channels is downconverted to baseband and sampled at Nyquist rate, i.e. $8M$ MHz.

Influence of channel occupancy

In the first experiment we consider a setting with $M = 4$ channels and $K = 2048$ samples. The SNR of the channel to detect is set to -5 dB. The detectors were analyzed for activity factors of $a = 0.1$, 0.5 and 0.9. In the simulations, the SNRs of the active channels (other than that under scrutiny) were generated following a log-normal distribution with mean 0 dB and dB-spread equal to 1 dB.

In order to investigate the issue of threshold selection, we plot in Fig. 4.1 the misdetection and false alarm probabilities of the three schemes, as a function of the detection threshold. For Tests 1 and 2, the analytical method of Appendix 4.E was used, whereas for the QGLRT (4.47), P_{D} and P_{FA} were obtained empirically. It is seen that, in the region of interest (small probability of misdetection), and for fixed thresholds, the detection performance of the three tests improves as a decreases. This is reasonable, since lower primary activity results in more channels perceived as weak and this can be exploited in order to improve the noise variance estimates. Hence, in order to satisfy $P_{\mathrm{D}} \geq P_{\mathrm{D}}^{\star}$ for a given target $P_{\mathrm{D}}^{\star}$ when the activity factor is unknown, the threshold must be set assuming the worst case $a = 1$.

Once the threshold has been fixed in order to satisfy the detection requirements,
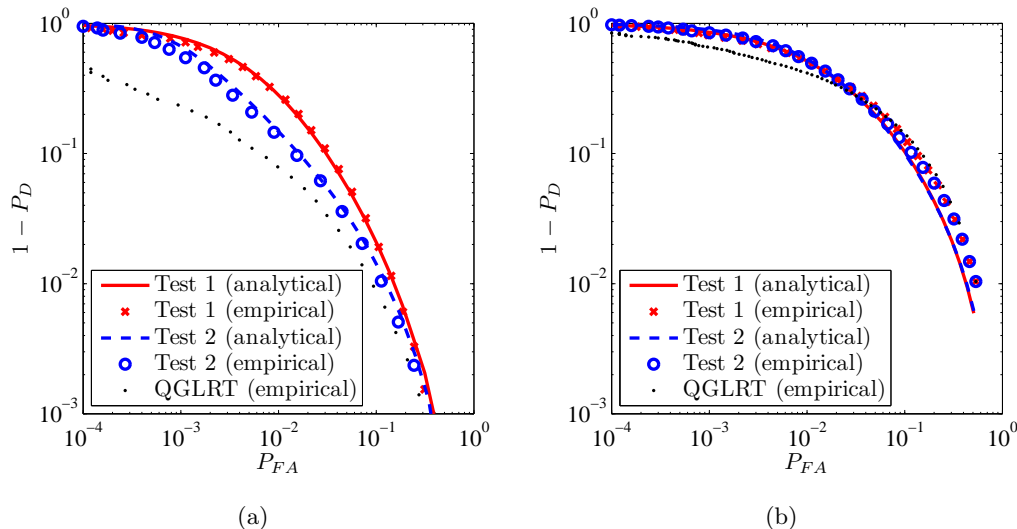
Figure 4.2: Complementary ROC curves in a setting with $M = 8$ channels, for activity factor (a) $a = 0.1$ and (b) $a = 0.9$.

the behavior of P_{FA} in terms of a is different for the three schemes. Whereas the false alarm rate of the QGLRT worsens as a decreases, for Test 1 P_{FA} is almost insensitive to variations in a. Interestingly, for Test 2 a region exists for which *both* P_{D} and P_{FA} improve with decreasing a. Thus, by setting the threshold for a given target $P_{\text{D}} \geq P_{\text{D}}^{\star}$ assuming $a = 1$, performance guarantees in terms of P_{FA} (missed opportunities for transmission) can be given for Tests 1 and 2.

Next we consider a setting with $M = 8$ channels, with the remaining parameters kept at the same values as in the previous experiment. Fig. 4.2 shows the complementary Receiver Operating Characteristics (ROC) curves for the three detectors and different activity factors. As expected, the QGLRT-based detector outperforms the other two suboptimal schemes. Tests 1 and 2 perform similarly for high activity factors, although Test 2 presents an advantage as a decreases. Note that in the extreme case of $a = 1$ there are no idle channels, and thus $\mathcal{S} = \emptyset$ in the context of Sec. 4.3.2, which in turn implies that the three tests become approximately equivalent for $a \to 1$.

In Fig. 4.2 a good agreement is observed between analytical and empirical results for Tests 1 and 2, with just a slight mismatch for high activity settings ($a = 0.9$) which can be explained as follows. In the derivation of the analytical expressions in Appendix 4.E it was assumed that busy channels do not affect the
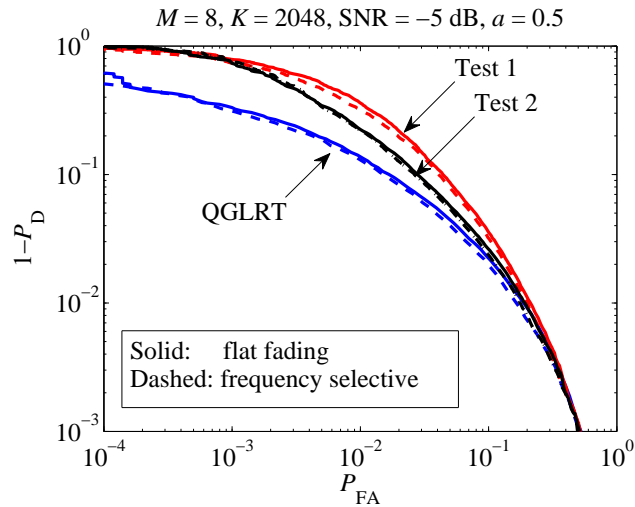
Figure 4.3: Detector performance with frequency selective channels.

distribution of the statistics for these detectors. This assumption is more likely to be violated as the percentage of busy channels (i.e. the activity factor a) increases.

Impact of frequency selectivity

Simulations were carried out in order to gauge the effect of unknown multipath propagation conditions in the performance of the proposed detectors. The multipath channels were generated according to the WINNER Phase II Model (Hentilä et al., 2007) with Profile C1 (Suburban). The central frequency is 800 MHz, and it is assumed that each of the signals at the $M = 8$ different channels arrives from a different transmitter. The locations of the transmitters and of the spectrum monitor were randomly selected on a square of dimension 15×15 km.

Figure 4.3 shows the ROC curves of the three detectors under frequency-flat and frequency-selective channels, for a setting with $K = 2048$, SNR $= -5$ dB and $a = 0.5$. As can be seen, performance remains essentially unaltered under multipath conditions. This can be explained by the structure of the proposed detectors: the linear combinations of different frequency bins effectively averages out the effects of frequency-selective channels.
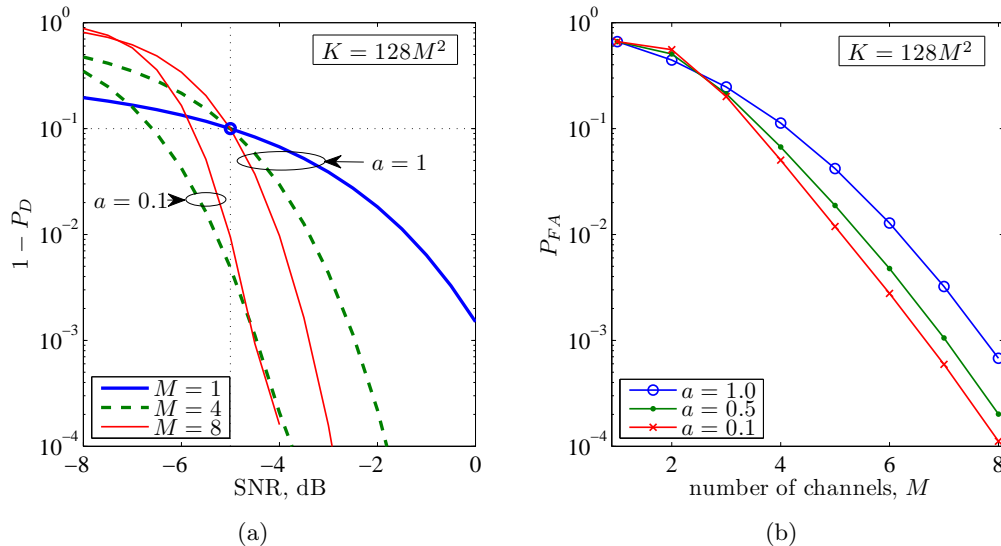
Figure 4.4: Analytical performance of Test 2 as a function of the number of channels M. The sample size is given by $K = 128M^2$. (a) Probability of misdetection. (b) Probability of false alarm.

Influence of the number of channels

Consider a setting in which the operating band consists of N_{ch} channels of B Hz each, which the spectrum monitor must scan in a total time of T s. To this end, the band is subdivided into subbands of $M < N_{\text{ch}}$ channels each, which are sequentially analyzed. The observation time for each of the MB Hz-wide subbands is thus MT/N_{ch} s. Hence, sampling at the Nyquist rate $f_s = MB$ Hz, the number of samples available for processing each subband of M channels is $K = M^2(BT/N_{\text{ch}})$. Thus, at the expense of a linear increase of f_s in terms of M, a *quadratic* increase of K is obtained, so that a favorable trade-off between detection performance and ADC cost/resolution can be achieved.

Assuming $BT/N_{\text{ch}} = 128$, Fig. 4.4(a) shows the analytical probability of misdetection of Test 2 versus SNR for different values of M. For each M, the thresholds are computed in order to achieve $P_{\text{D}} = 0.9$ at a target SNR $= -5$ dB assuming full occupancy (worst case). With this design, having more channels in the subband is seen to improve detection performance for SNRs at and above the target value, for all values of a, thus offering additional interference guarantees to the primary system.

Fig. 4.4(b) shows the corresponding false alarm rate in terms of M. It is seen that P_{FA} decreases *exponentially* with the number of channels included in the subband. A reduction in P_{FA} increases the opportunities of accessing the spectrum and therefore the global throughput of the secondary system.

4.4 Compressed spectrum sensing

In the previous section we have seen that the detection performance in a wideband setting heavily depends on the *activity factor a*, which was assumed unknown. We now consider a more elaborated model that includes general A2I converters and certain *a priori* knowledge on the primary activity summarized in the *activity factor a*. Note that this parameter could be established without much difficulty from empirical measurements in the bands of interest, or estimated online from the observed activity in the band. We recall the signal model from (4.3),

$$\tilde{\mathbf{y}} \doteq \mathbf{\Phi}\mathbf{y} = \sum_{m=0}^{M} \sigma_m \tilde{\mathbf{s}}_m, \tag{4.59}$$

where $\mathbf{\Phi}$ denotes a $K \times N$ compression matrix. Then we can define the set of active channels as

$$\mathcal{A} = \left\{ m \mid \sigma_m^2 > 0, \ 0 \leq m \leq M \right\}. \tag{4.60}$$

It is assumed that the noise is always present, and thus $0 \in \mathcal{A}$ always. For the signal channels, we model the sparsity of the system with each event $m \in \mathcal{A}$ following an independent Bernoulli distribution: $\Pr\{m \in \mathcal{A}\} = a$ for $m = 1,\ldots, M$, with a assumed known to the receiver. In this respect, a gives an indication of the average occupancy of the frequency band and can be estimated beforehand; in a CR context, it is expected that $a \ll 1$. Additionally no assumption is made on σ_m^2 given that channel m is active.

4.4.1 Estimation from compressed measurements

In order to implement the GLRT (4.8) we need to obtain the ML estimate of the vector $\boldsymbol{\sigma}$. However, as mentioned in the previous section there exists no closed-form solution (Burg et al., 1982) for the ML estimation of the structured covariance

matrix

$$\tilde{\mathbf{R}}(\boldsymbol{\sigma}) = \sum_{i=0}^{M} \sigma_i^2 \tilde{\mathbf{C}}_i. \tag{4.61}$$

Instead we will follow a Bayesian approach to the estimation of the parameter vector $\boldsymbol{\sigma}$ which will result in an efficient estimator in the low SNR regime that asymptotically achieves the Cramér-Rao lower bound.

MAP estimation

To explicitly introduce sparsity in our derivation we formulate the joint estimation problem of finding the sparsity pattern $\mathcal{A}$ together with the power vector $\boldsymbol{\sigma}$. Using Bayes' rule we can state the maximum *a posteriori* (MAP) estimation of $\{\boldsymbol{\sigma}, \mathcal{A}\}$ as

$$\{\hat{\boldsymbol{\sigma}}, \hat{\mathcal{A}}\} = \arg\max_{\boldsymbol{\sigma}, \mathcal{A}} f(\mathcal{A}, \boldsymbol{\sigma}|\tilde{\mathbf{y}}) \tag{4.62}$$

$$= \arg\max_{\boldsymbol{\sigma}, \mathcal{A}} f(\tilde{\mathbf{y}}|\mathcal{A}, \boldsymbol{\sigma}) f(\boldsymbol{\sigma}|\mathcal{A}) f(\mathcal{A}) \tag{4.63}$$

$$= \arg\max_{\boldsymbol{\sigma}(\mathcal{A}), \mathcal{A}} f(\tilde{\mathbf{y}}|\mathcal{A}, \boldsymbol{\sigma}(\mathcal{A})) f(\mathcal{A}). \tag{4.64}$$

where in the last step we made use of the fact that the *a priori* distribution $f(\boldsymbol{\sigma}|\mathcal{A})$ is modeled as non-informative for those active components of $\boldsymbol{\sigma}$ with the sparsity pattern imposed by $\mathcal{A}$ (denoted here as $\boldsymbol{\sigma}(\mathcal{A})$). That is, $\boldsymbol{\sigma}(\mathcal{A})$ has zeros at the positions specified by $\{0, \ldots, M\} - \mathcal{A}$, but no prior is assumed for the remaining components.

Substituting (4.10) and the Bernoulli distribution

$$f(\mathcal{A}) = a^{|\mathcal{A}|}(1-a)^{M-|\mathcal{A}|} \tag{4.65}$$

into (4.64), and disregarding constant additive terms, we obtain the following equivalent minimization problem:

$$\{\hat{\boldsymbol{\sigma}}, \hat{\mathcal{A}}\} = \arg\min_{\boldsymbol{\sigma}, \mathcal{A}} \left\{ \log\det(\tilde{\mathbf{R}}(\boldsymbol{\sigma}(\mathcal{A}))) + \tilde{\mathbf{y}}^H \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}(\mathcal{A}))\tilde{\mathbf{y}} - |\mathcal{A}| \log\frac{a}{1-a} \right\}. \tag{4.66}$$

Remark 4.2. It is interesting to note that the weighting factor associated to $|\mathcal{A}|$ depends only on a, that is, the probability that a channel is active. When $a = 1/2$, the cardinality term disappears and the optimization procedure reduces to an ML optimization with respect to $f(\tilde{\mathbf{y}}|\boldsymbol{\sigma})$. On the other hand, when a given band presents

a low (or high) occupancy rate, thus $a \to 0$ ($a \to 1$), the weight of $|\mathcal{A}|$ tends to $-\infty$ ($+\infty$). In this case, as intuition suggests, the optimal estimate $\hat{\mathcal{A}}$ becomes the empty (or full) set of channels, independently of the values of $f(\tilde{\mathbf{y}}|\boldsymbol{\sigma}(\mathcal{A}))$.

The expression (4.66) is a mixed discrete/continuous maximization problem, i.e., $\mathcal{A}$ can only take one out of 2^M values, whereas for fixed $\mathcal{A}$ the maximization is performed over the *continuous* parameter vector $\boldsymbol{\sigma}(\mathcal{A})$. Therefore this optimization problem, as it is, is NP-hard and needs to be simplified in order to conduct to practical spectrum reconstruction algorithms. An approximate MAP estimator that performs close to the constrained CRLB was proposed in Vazquez-Vilar et al. (2010a). However, this approach requires a complex iterative implementation that hinders its practical application.

Here we will follow a slightly different approach, which will make clear the connection between the proposed Bayesian framework and classical compressed sensing theory. As we will see next, a series of simplifications of (4.66) conduct to a psd estimator equivalent to Basis Pursuit (BP) denoising (Tropp, 2006) of the one-shot compressed covariance estimate $\tilde{\mathbf{y}}\tilde{\mathbf{y}}^H$. To this end, we resort first to the Taylor expansion of $\log\det(\tilde{\mathbf{R}}(\boldsymbol{\sigma}))$ and $\tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma})$ around the true vector of power levels $\boldsymbol{\sigma}_\star$, which will then be particularized for the low SNR regime.

Low SNR approximation

To simplify notation, let us define $\tilde{\mathbf{R}}_\star \doteq \tilde{\mathbf{R}}(\boldsymbol{\sigma}_\star) = \sum_{i=0}^{M}(\sigma_i^\star)^2\tilde{\mathbf{C}}_i$, where $\boldsymbol{\sigma}_\star$ is the vector of true power levels. Then for $\boldsymbol{\sigma} \to \boldsymbol{\sigma}_\star$ we have that

$$\log\det(\tilde{\mathbf{R}}(\boldsymbol{\sigma})) \approx \log\det(\tilde{\mathbf{R}}_\star) + \sum_{m=0}^{M}(\sigma_m^2 - (\sigma_m^\star)^2)\operatorname{tr}\{\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_m\}$$
$$-\frac{1}{2}\sum_{i=0}^{M}\sum_{j=0}^{M}(\sigma_i^2 - (\sigma_i^\star)^2)(\sigma_j^2 - (\sigma_j^\star)^2)\operatorname{tr}\{\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_i\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_j\}, \quad (4.67)$$

$$\tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \approx \tilde{\mathbf{R}}_\star^{-1} - \sum_{m=0}^{M}(\sigma_m^2 - (\sigma_m^\star)^2)\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_m\tilde{\mathbf{R}}_\star^{-1}$$
$$+\sum_{i=0}^{M}\sum_{j=0}^{M}(\sigma_i^2 - (\sigma_i^\star)^2)(\sigma_j^2 - (\sigma_j^\star)^2)\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_i\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_j\tilde{\mathbf{R}}_\star^{-1}. \quad (4.68)$$

Substituting (4.67) and (4.68) into (4.66), disregarding the terms not depending

on $\boldsymbol{\sigma}$ we obtain the approximation

$$\hat{\boldsymbol{\sigma}} \approx \arg\min_{\boldsymbol{\sigma}} \left\{ \frac{1}{2} \sum_{i=0}^{M} \sum_{j=0}^{M} \sigma_i^2 \sigma_j^2 \,\mathrm{tr}\{\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_i\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_j\} - \sum_{m=0}^{M} \sigma_m^2 \,\mathrm{tr}\{\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_m\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{y}}\tilde{\mathbf{y}}^H\} \right.$$
$$- \sum_{i=0}^{M} \sum_{j=0}^{M} \sigma_i^2 \sigma_j^2 \,\mathrm{tr}\{\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_i\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_j\mathbf{K}^2\} + 2\sum_{m=0}^{M} \sigma_m^2 \,\mathrm{tr}\{\tilde{\mathbf{R}}_\star^{-1}\tilde{\mathbf{C}}_m\mathbf{K}^2\}$$
$$\left. - \|\boldsymbol{\sigma}_1^M\|_0 \ln\frac{a}{1-a} \right\}, \tag{4.69}$$

where we defined $\mathbf{K}^2 \doteq \tilde{\mathbf{R}}_\star^{-1}(\tilde{\mathbf{R}}_\star - \tilde{\mathbf{y}}\tilde{\mathbf{y}}^H)$, $\boldsymbol{\sigma}_1^M \doteq [\sigma_1^2 \cdots \sigma_M^2]^T$ so that $\|\boldsymbol{\sigma}_1^M\|_0 = |\mathcal{A}|$ and used that $\tilde{\mathbf{R}}_\star = \sum_{m=0}^{M}(\sigma_m^\star)^2\tilde{\mathbf{C}}_m$.

For vanishing SNR at each of the channels $\sigma_m^2 << \sigma_0^2$ for $m = 1, \ldots, M$, we have that asymptotically $\tilde{\mathbf{R}}_\star \to (\sigma_0^\star)^2\mathbf{I}$. Substituting the low SNR approximation $\tilde{\mathbf{R}}_\star \approx (\sigma_0^\star)^2\mathbf{I}$ and rearranging terms we have that the optimization problem (4.69) is equivalent to

$$\hat{\boldsymbol{\sigma}} \approx \arg\min_{\boldsymbol{\sigma}} \frac{1}{2(\sigma_0^\star)^4} \left\| \tilde{\mathbf{y}}\tilde{\mathbf{y}}^H - \sum_{i=0}^{M} \sigma_i^2\tilde{\mathbf{C}}_i \right\|_2^2 - \left\| \left(\mathbf{I} - \sum_{i=0}^{M} \frac{\sigma_i^2}{(\sigma_0^\star)^2}\tilde{\mathbf{C}}_i \right)\mathbf{K} \right\|_2^2 + \gamma_0 \left\| \boldsymbol{\sigma}_1^M \right\|_0,$$
$$\tag{4.70}$$

where we defined the regularization weight $\gamma_0 \doteq \ln((1-a)/a)$ and made use of the definition of the Matrix Frobenius Norm $\|\boldsymbol{A}\|^2 \doteq \mathrm{tr}\,(\boldsymbol{A}\boldsymbol{A}^H)$. Note that the only dependence of (4.70) on the (unknown) actual power vector $\boldsymbol{\sigma}_\star$ is through $(\sigma_0^\star)^2$ and the matrix $\mathbf{K}$. If we neglect the term depending on $\mathbf{K}$, (4.70) reduces to

$$\hat{\boldsymbol{\sigma}} \approx \arg\min_{\boldsymbol{\sigma}} \frac{1}{2(\sigma_0^\star)^4} \left\| \tilde{\mathbf{y}}\tilde{\mathbf{y}}^H - \sum_{i=0}^{M} \sigma_i^2\tilde{\mathbf{C}}_i \right\|_2^2 + \gamma_0 \left\| \boldsymbol{\sigma}_1^M \right\|_0 \tag{4.71}$$

$$= \arg\min_{\boldsymbol{\sigma}} \left\| \tilde{\mathbf{y}}\tilde{\mathbf{y}}^H - \sum_{i=0}^{M} \sigma_i^2\tilde{\mathbf{C}}_i \right\|_2^2 + \gamma_0' \left\| \boldsymbol{\sigma}_1^M \right\|_0, \tag{4.72}$$

where we have defined $\gamma_0' \doteq 2(\sigma_0^\star)^4\gamma_0$.

Remark 4.3. There exists a strong resemblance between the equation (4.72) and the classical compressed sensing theory. The latter considers the problem of estimating a sparse vector from a set of mixed measurements, yielding to a norm-0 regularization similar to (4.72). While this minimization problem is NP-hard, several efficient alternatives have been proposed by relaxing it to obtain a convex formulation. One

example is the Dantzig Selector (DS) first presented in Candes and Tao (2007), which under certain assumptions can be shown to be equivalent (James et al., 2009) to the closely related BP denoising (Tropp, 2006). If we particularize BP denoising in our case we obtain

$$\hat{\boldsymbol{\sigma}} \approx \arg\min_{\boldsymbol{\sigma}} \left\| \tilde{\mathbf{y}}\tilde{\mathbf{y}}^H - \sum_{i=0}^{M} \sigma_i^2 \tilde{\mathbf{C}}_i \right\|_2^2 + \gamma_1 \left\| \boldsymbol{\sigma}_1^M \right\|_1. \tag{4.73}$$

Note that the problem is now convex and can be efficiently solved using any of the existing convex optimization packages. However, while for the pseudo norm ℓ_0 the weighting factor γ_0' can be computed from the actual noise power and from the *a priori* probability of occupancy (which in principle could be estimated), the regularization factor γ_1 corresponding to the ℓ_1-norm needs to be determined numerically for different system parameters. Then, if γ_1 is suitably chosen, basis pursuit denoising performance will be close to the optimal MAP estimator in the low SNR regime.

Bayesian matching pursuit

While the use of the convex formulation in (4.73) may be interesting in some cases, it requires to solve a convex optimization problem at each spectral monitoring iteration. This may not be practical in certain devices which present stringent complexity and power limitations. Here we propose an approximate solution of (4.72) based on a low complexity iterative greedy algorithm.

First note that the optimization (4.72) can be rewritten as

$$\{\hat{\boldsymbol{\sigma}}, \hat{\mathcal{A}}\} \approx \arg\min_{\boldsymbol{\sigma}(\mathcal{A}), \mathcal{A}} \left\| \tilde{\mathbf{y}}\tilde{\mathbf{y}}^H - \sum_{i \in \mathcal{A}} \sigma_i^2 \tilde{\mathbf{C}}_i \right\|_2^2 + \gamma_0'|\mathcal{A}|. \tag{4.74}$$

Then we have that:

Proposition 4.3. For fixed $\mathcal{A} = \{l(1), \ldots, l(|\mathcal{A}|)\}$, so that $[\hat{\boldsymbol{\sigma}}(\mathcal{A})]_m = 0$ for $m \notin \mathcal{A}$, the unconstrained solution to the optimization problem (4.74) is given by

$$[\hat{\boldsymbol{\sigma}}(\mathcal{A})]_{l(i)} = [\hat{\boldsymbol{\theta}}]_i \tag{4.75}$$

with

$$\hat{\boldsymbol{\theta}} = \mathbf{B}_{\mathcal{A}}^{-1}\mathbf{b}_{\mathcal{A}}, \tag{4.76}$$

where

$$[\mathbf{B}_{\mathcal{A}}]_{ij} \doteq \operatorname{tr}\left\{\tilde{\mathbf{C}}_{l(i)}\tilde{\mathbf{C}}_{l(j)}\right\}, \tag{4.77}$$

$$[\mathbf{b}_{\mathcal{A}}]_i \doteq \tilde{\mathbf{y}}^H\tilde{\mathbf{C}}_{l(i)}\tilde{\mathbf{y}}, \tag{4.78}$$

for $i, j = 1, \ldots, |\mathcal{A}|$.

Proof. See Appendix 4.F.   □

Corollary 4.1. Given the true sparsity pattern $\mathcal{A}$, the LS estimator given in (4.76)-(4.78) is asymptotically efficient in the limit as $\boldsymbol{\sigma}_\star \to [(\sigma_0^\star)^2 \ 0 \ \cdots \ 0]^T$, i.e. for low SNR on all channels.

Proof. See Appendix 4.G.   □

This asymptotic efficiency guarantees that, given that the true sparsity pattern is found, the LS estimator will present a good performance in the low SNR regime. Moreover, in the limit the mean squared error of the LS estimator will achieve the CRLB.

Substituting (4.76) back in (4.74) and disregarding the terms not depending on $\mathcal{A}$, we have that the best estimate $\hat{\mathcal{A}}$ is given by

$$\hat{\mathcal{A}} \approx \arg\min_{\mathcal{A}} \mu(\mathcal{A}), \tag{4.79}$$

where $\mu(\mathcal{A}) \doteq \gamma_0'|\mathcal{A}| - \mathbf{b}_{\mathcal{A}}^H\mathbf{B}_{\mathcal{A}}^{-1}\mathbf{b}_{\mathcal{A}}$.

In principle the solution of (4.79) needs to be found by performing an exhaustive search over $\mathcal{A}$. Note that this implies to evaluate (4.77)-(4.78) for each of the 2^M possible combinations. Instead we propose a Bayesian matching pursuit (Schniter et al., 2008) algorithm that iteratively estimates the set of active channels, as described in the Algorithm 3.

This suboptimal greedy solution finds the right set of active channels with high probability, as shown later by means of numerical simulations, offering a performance close to the more complex convex optimization methods. The idea is to construct

---

Algorithm 3: Bayesian matching pursuit.

Input: Measured signal $\tilde{\mathbf{y}}$.
Output: Estimated set of active channels $\hat{\mathcal{A}}$.
Initialize: Set $\hat{\mathcal{A}}_0 = \{0\}$
for $n = 1$ **to** $M_{\max}$ **do**
$\quad\left|\quad\begin{aligned} m^\star &= \underset{m \notin \hat{\mathcal{A}}_{n-1}}{\arg\max}\ \mu(\hat{\mathcal{A}}_{n-1} \cup \{m\}) \\ \hat{\mathcal{A}}_n &= \hat{\mathcal{A}}_{n-1} \cup \{m^\star\} \end{aligned}\right.$
$\hat{\mathcal{A}} = \underset{\hat{\mathcal{A}}_n}{\arg\max}\ \mu(\hat{\mathcal{A}}_n)$

---

the active set estimate $\hat{\mathcal{A}}$ sequentially: starting with the "only noise" set $\hat{\mathcal{A}} = \{0\}$, at each step a new active channel is added to $\hat{\mathcal{A}}$ in order to maximize the corresponding metric $\mu(\hat{\mathcal{A}})$. This procedure is repeated $M_{\max}$ times where $M_{\max} \leq M$ determines the maximum number of active channels which could be possible declared as active by the algorithm. The parameter $M_{\max}$ is then related to the number of iterations, and hence to the complexity, of the algorithm. If no complexity constraints exist we may fix $M_{\max} = M$, which allows the system to declare all channels as busy. On the other hand, if we know that the band is sparsely occupied, $M_{\max}$ can be chosen from the activity factor a so that $\Pr\{|\mathcal{A}| \leq \mathcal{M}_{\max}\}$ is above a certain value. The final estimate $\hat{\mathcal{A}}$ is given by the partial solution $\hat{\mathcal{A}}_n$ with maximum *a posteriori* log-likelihood function.

4.4.2 Quasi-GLRT detection

Once the estimates $\{\hat{\boldsymbol{\sigma}}, \hat{\mathcal{A}}\}$ have been computed under both hypotheses, either by using convex optimization methods or by using the proposed greedy iterative algorithm, the detection can be performed based on the GLRT scheme in (4.8).

If we denote by $\hat{\boldsymbol{\sigma}}_0$, $\hat{\boldsymbol{\sigma}}_1$, the corresponding estimates under the hypothesis $\mathcal{H}_0$, respectively $\mathcal{H}_1$, we define the following Quasi-GLRT statistic

$$T_{\mathrm{cs}} \doteq \frac{f(\tilde{\mathbf{y}} \,|\, \hat{\boldsymbol{\sigma}}_0)}{f(\tilde{\mathbf{y}} \,|\, \hat{\boldsymbol{\sigma}}_1)} \underset{\mathcal{H}_1^m}{\overset{\mathcal{H}_0^m}{\gtrless}} \gamma_{\mathrm{cs}}, \tag{4.80}$$

with γ_{cs} a threshold and $f(\tilde{\mathbf{y}} \,|\, \boldsymbol{\sigma})$ given as in (4.10).

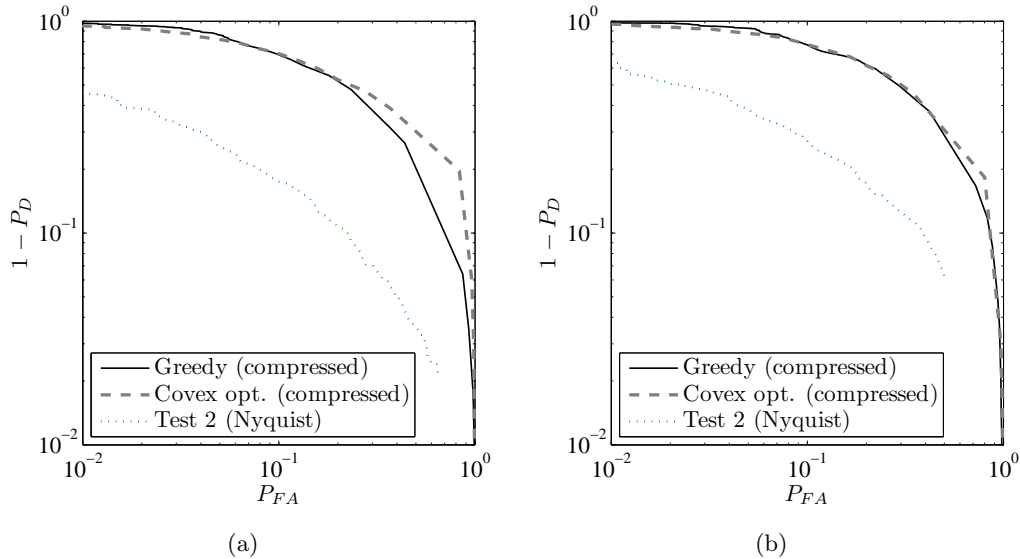(a)   (b)

Figure 4.5: Complementary ROC curves in a setting with $M = 16$ channels, $K/N = 128/2048$, for an activity factor (a) $a = 0.1$ and (b) $a = 0.3$.

4.4.3 Numerical results and discussion

In the following we analyze the performance of two detection schemes based on T_{cs}:

- Greedy detector based on the iteration given in Algorithm 3 and the efficient estimator of Proposition 4.3.

- Convex optimization based detector. The sparsity pattern is computed by solving (4.73) with $\gamma_1 = \gamma_0$. Once the sparsity pattern has been found the actual estimate is taken as the feasible minimizer of (4.73) for $\gamma_1 = 0$. This last step is required to reduce the distortion introduced by the regularization term in the optimization procedure.

To evaluate this detection schemes we consider terrestrial digital TV broadcast networks using 8K-mode DVB-T modulation with parameters as in Section 4.3.4. As compression matrix Φ we consider a random pinning matrix corresponding to K randomly selected rows of the $N \times N$ identity matrix.
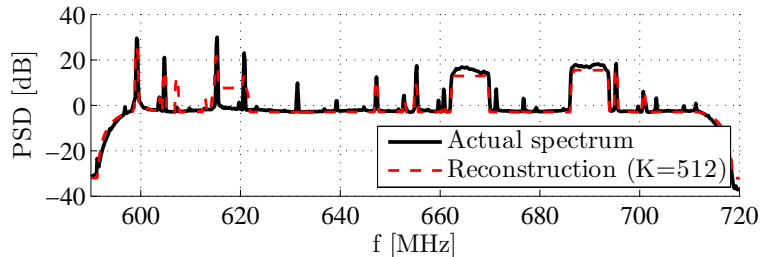
Figure 4.6: Example of reconstruction of a mixed analog/digital broadcasting television band.

Detection performance

Figure 4.5 shows the complementary ROC curves of this two detection schemes for a scenario with $M = 16$ channels, $N = 1024$ and $K = 128$ compressed samples, i.e. compression ratio equal to 8. The channel under scrutiny presents an SNR of 3 dB when active, while the other channels' SNR, when active, follows a lognormal distribution of mean 0 dB and standard deviation of 1 dB. In Fig. 4.5 we can see that the proposed greedy iteration shows a similar performance to the more complex convex optimization scheme for both activity factors $a = 0.1$ and $a = 0.3$.

It is interesting to note that the compressed sampling schemes present a penalty with respect to a detector (Test 2 from (4.50)) using $K = 128$ samples at Nyquist rate. This is due to the fact that the compression process mixes different frequency channels increasing the apparent noise level. However, in some cases, sampling at Nyquist rate may be infeasible and detectors based on CS present practical importance.

Spectrum reconstruction

We now show the spectrum reconstruction capabilities of the proposed Greedy estimation method in a real environment. To this end we captured part of the Spanish TV broadcast band (112 MHz bandwidth, comprising 14 channels with PAL/DVB-T signals just before the analog switch off). The *a priori* covariance matrices were generated using the channelization information of the PAL/DVB-T broadcast network, while the occupancy probability was considered $a = 0.3$. The compression procedure has been simulated in Matlab using a 512×2048 random pinning matrix. No knowledge is fed to the reconstruction algorithm about the particular modula-

tion (PAL or DVB-T) encountered at a given channel. Fig. 4.6 shows the psd of the band (obtained using a large number of uncompressed samples) together with the reconstruction obtained by the proposed method using just $K = 512$ samples. Even this reduced number of samples allows the estimation of 29 power levels needed for the reconstruction (14 DVB-T + 14 PAL + noise level).

4.5   Conclusions

The wideband approach to spectrum sensing provides a means to trade off detection performance and ADC complexity. In practice, primary networks using FDMA exhibit guard bands between adjacent channels which can be used to estimate the noise power to build detectors robust to noise level uncertainty. In addition, the frequency bins of those channels perceived as weak can be used for this task as well. These ideas are exploited by three novel detectors derived starting from a GLRT approach. In this way, the noise uncertainty problem that plagues the popular Energy Detector is largely alleviated.

Assuming a more general A2I converter and that the activity factor of the band can be obtained beforehand, we established a connection between the MAP approach to parameter estimation and classical compressed sensing theory. Moreover, based on the MAP formulation we derive a simpler Greedy iterative algorithm performing close to the convex optimization methods usually employed to solve the compressed sensing formulation.

A key part of the primary user detection schemes derived in this chapter is the estimation of the unknown parameters, namely noise and signal levels. Some preliminary work along this lines was presented in López-Valcarce and Vazquez-Vilar (2009); Vazquez-Vilar et al. (2010a). The first part of this chapter, wideband primary user detection at Nyquist rate, is based on the journal article submitted to IEEE-TSP (Vazquez-Vilar and López-Valcarce, 2011).

Appendix 4.A   Proof of Theorem 4.1

For finite K, consider the matrix

$$\mathbf{A}(\boldsymbol{\sigma}) \doteq [\mathbf{L}^H \boldsymbol{\Delta}^{-2}(\boldsymbol{\sigma})\mathbf{L}] \cdot [\mathbf{L}^\dagger \boldsymbol{\Delta}^2(\boldsymbol{\sigma})(\mathbf{L}^\dagger)^H]. \tag{4.81}$$

We will prove that, under the conditions of the Theorem, $\mathbf{A}(\boldsymbol{\sigma}) = \mathbf{I}$. Then, taking the limit as $K \to \infty$ on both sides of (4.81), the desired result will be obtained, in view of (4.28) and (4.32). To this end, consider the singular-value decomposition $\mathbf{L} = \mathbf{U}\mathbf{D}\mathbf{V}^H$, where $\mathbf{U}$ is $K \times (M+1)$ with orthogonal columns, $\mathbf{D}$ is $(M+1) \times (M+1)$ diagonal with the nonzero singular values, and $\mathbf{V}$ is $(M+1) \times (M+1)$ unitary. The pseudoinverse is thus $\mathbf{L}^\dagger = \mathbf{V}\mathbf{D}^{-1}\mathbf{U}^H$. Then

$$\mathbf{A}(\boldsymbol{\sigma}) = \mathbf{V}\mathbf{D}\mathbf{U}^H \boldsymbol{\Delta}^{-2}(\boldsymbol{\sigma})\mathbf{U}\mathbf{U}^H \boldsymbol{\Delta}^2(\boldsymbol{\sigma})\mathbf{U}\mathbf{D}^{-1}\mathbf{V}^H. \tag{4.82}$$

As seen from (4.82), a sufficient condition for $\mathbf{A}(\boldsymbol{\sigma}) = \mathbf{I}$ is that $\mathbf{U}\mathbf{U}^H$ and $\boldsymbol{\Delta}(\boldsymbol{\sigma})$ commute. This we will show now.

Note that the columns of $\mathbf{U}$ constitute an orthonormal basis of $\mathcal{R}\{\mathbf{L}\}$, the subspace spanned by the columns of $\mathbf{L}$. Without loss of generality (since channel indexing is arbitrary), assume that the columns of $\mathbf{L}$ are sorted such that if the set of indices of nonzero entries of column j contains that of column i, then $i < j$. Additionally, we assume that the rows of $\mathbf{L}$ are arranged such that these sets of indices of nonzero entries contain only contiguous indices (frequency bins). This is also without loss of generality, since one can always apply a permutation to the rows of $\mathbf{L}$ to achieve this.

An orthogonal basis for $\mathcal{R}\{\mathbf{L}\}$ can also be obtained by applying the Gram-Schmidt orthogonalization procedure to the columns of $\mathbf{L}$. It is straightforward to show that this results in a basis $\tilde{\mathbf{U}} = [\ \tilde{\mathbf{u}}_0 \quad \tilde{\mathbf{u}}_1 \quad \cdots \quad \tilde{\mathbf{u}}_M\ ]$ such that (i) any nonzero entries of a given vector $\tilde{\mathbf{u}}_i$ are constant and in contiguous positions, and (ii) for any $\tilde{\mathbf{u}}_i, \tilde{\mathbf{u}}_j$ with $i \neq j$, the two sets of indices of their nonzero entries are disjoint. These properties imply that $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H$ is a block diagonal matrix, with each block on the diagonal having all of its elements equal:

$$\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H = \begin{bmatrix} \alpha_0 \mathbf{1}_{K_0} \mathbf{1}_{K_0}^T & & \\ & \ddots & \\ & & \alpha_M \mathbf{1}_{K_M} \mathbf{1}_{K_M}^T \end{bmatrix} = \mathbf{U}\mathbf{U}^H, \tag{4.83}$$

where α_i are scalars, K_i is the number of nonzero entries in $\tilde{\mathbf{u}}_i$, and the last equality in (4.83) stems from the fact that both $\mathbf{U}\mathbf{U}^H$ and $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H$ are projection matrices onto $\mathcal{R}\{\mathbf{L}\}$.

On the other hand, since the diagonal of $\boldsymbol{\Delta}(\boldsymbol{\sigma})$ is a linear combination of the columns of $\mathbf{L}$, see (4.15), one has that

$$\boldsymbol{\Delta}(\boldsymbol{\sigma}) = \begin{bmatrix} \beta_0 \mathbf{I}_{K_0} & & \\ & \ddots & \\ & & \beta_M \mathbf{I}_{K_M} \end{bmatrix}, \tag{4.84}$$

for some scalars β_i. Given the structure of $\mathbf{U}\mathbf{U}^H$ and $\boldsymbol{\Delta}(\theta)$, it is readily seen that they do commute.

Appendix 4.B   Proof of Theorem 4.2

We shall prove (4.45), as the proof for (4.46) is analogous. The cost $f(\hat{\boldsymbol{\sigma}}) = \|\mathbf{L}\hat{\boldsymbol{\sigma}} - \mathbf{p}\|_2^2$ is convex, and its gradient is given by

$$\frac{1}{K}\nabla f(\hat{\boldsymbol{\sigma}}) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w_1^{-1} & 0 & \cdots & 0 \\ 1 & 0 & w_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & w_M^{-1} \end{bmatrix} \hat{\boldsymbol{\sigma}} - \begin{bmatrix} w_0 q_0 + \cdots + w_M q_M \\ q_1 \\ q_2 \\ \vdots \\ q_M \end{bmatrix}. \tag{4.85}$$

In addition, we have the linear inequality constraints $g_j(\hat{\boldsymbol{\sigma}}) = -\hat{\sigma}_j^2 \le 0$, $0 \le j \le M$, whose gradient is $\nabla g_j(\hat{\boldsymbol{\sigma}}) = -\mathbf{e}_j$, where $\mathbf{e}_j$ is the j-th unit vector. A sufficient condition for $\hat{\boldsymbol{\sigma}}_1 = [\hat{\sigma}_{01}^2 \cdots \hat{\sigma}_{M1}^2]^T$ to be the global optimum is that there exist scalars $\mu_j \ge 0$, $0 \le j \le M$, such that

$$\nabla f(\hat{\boldsymbol{\sigma}}_1) + \sum_{j=0}^{M} \mu_j \nabla g_j(\hat{\boldsymbol{\sigma}}_1) = 0, \tag{4.86}$$

$$\mu_j g_j(\hat{\boldsymbol{\sigma}}_1) = 0, \quad 0 \le j \le M, \tag{4.87}$$

which for this case amounts to saying that $[\nabla f(\hat{\boldsymbol{\sigma}}_1)]_i = 0$ if $\hat{\sigma}_{i1}^2 > 0$ and $[\nabla f(\hat{\boldsymbol{\sigma}}_1)]_i \ge 0$ if $\hat{\sigma}_{i1}^2 = 0$. Now we show that the vector $\hat{\boldsymbol{\sigma}}_1$ given by (4.45) satisfies these conditions.

Note that $\hat{\sigma}_{01}^2 > 0$, and that $\mathcal{U}_1 = \{j : \hat{\sigma}_{j1}^2 = 0\}$. In view of (4.85),

$$\frac{1}{K}[\nabla f(\hat{\boldsymbol{\sigma}}_1)]_0 = \hat{\sigma}_{01}^2 + \sum_{i=1}^{M} \hat{\sigma}_{i1}^2 - w_0 q_0 - \sum_{i=1}^{M} w_i q_i \tag{4.88}$$

$$= -w_0(q_0 - \hat{\sigma}_{01}^2) + \sum_{i=1}^{M}[\hat{\sigma}_{i1}^2 - w_i(q_i - \hat{\sigma}_{01}^2)] \tag{4.89}$$

$$= -w_0(q_0 - \hat{\sigma}_{01}^2) - \sum_{i \in \mathcal{U}_1} w_i(q_i - \hat{\sigma}_{01}^2) \tag{4.90}$$

$$= -\left(w_0 q_0 + \sum_{i \in \mathcal{U}_1}^{M} w_i q_i\right) + \left(w_0 + \sum_{i \in \mathcal{U}_1}^{M} w_i\right)\hat{\sigma}_{01}^2 = 0, \tag{4.91}$$

where the second line follows from $w_0 + \cdots + w_M = 1$; the third, from the definitions of $\mathcal{U}_1$ and $\hat{\sigma}_{j1}^2$, and the last step, from the definition of $\hat{\sigma}_{01}^2$. On the other hand, for $1 \le j \le M$, using again the definitions of $\mathcal{U}_1$ and $\hat{\sigma}_{j1}^2$,

$$\frac{1}{K}[\nabla f(\hat{\boldsymbol{\sigma}}_1)]_j = \hat{\sigma}_{01}^2 + w_j^{-1}\hat{\sigma}_{j1}^2 - q_j \tag{4.92}$$

$$= \begin{cases} \hat{\sigma}_{01}^2 - q_j \ge 0, & j \in \mathcal{U}_1, \\ 0, & j \notin \mathcal{U}_1, \end{cases} \tag{4.93}$$

as was to be shown.

Appendix 4.C   Proof of Proposition 4.1

If the constrained LS estimate under $\mathcal{H}_1^m$ results in $\hat{\sigma}_{m1}^2 = 0$ (i.e. $m \in \mathcal{U}_1$), then it is clear that imposing $\hat{\sigma}_{m0}^2 = 0$ and then minimizing the LS cost under the same constraints for the remaining variables will yield the same result. But this is exactly the constrained LS estimate under $\mathcal{H}_0^m$.

Appendix 4.D   Proof of Proposition 4.2

Let $\mathcal{U}_1 = \{l(1), l(2), \cdots, l(s)\}$ such that $q_{l(1)} \le q_{l(2)} \le \cdots \le q_{l(s)}$. The proof is by induction, and is based on the constructive algorithm given in Algorithm 2. Note that:

- $q_l \le \hat{\sigma}_{01}^2$ for all $l \in \mathcal{U}_1$ (by definition of $\mathcal{U}_1$);

- $\hat{\sigma}_{01}^2 \leq q_0$ (since $\hat{\sigma}_{01}^2$ is a convex combination of q_0 and $\{q_l, \, l \in \mathcal{U}_1\}$);

- $\hat{\sigma}_{01}^2 \leq q_m$ (since $m \notin \mathcal{U}_1$).

The last two facts imply that $\hat{\sigma}_{01}^2 \leq q_{0m} \doteq (w_0 q_0 + w_m q_m)/(w_0 + w_m)$.

Now consider $q_{l(1)}$. In the process of constructing $\mathcal{U}_0$ given in Algorithm 2, the first iteration results in $\hat{\sigma}_{00}^2 = q_{0m}$. The unconstrained estimate with respect to the remaining variables is not feasible, since $q_l < q_{0m}$ for $l \in \mathcal{U}_1$. Therefore, index $l(1)$ is picked so that $l(1) \in \mathcal{U}_0$.

Suppose now that $l(1), \ldots, l(i) \in \mathcal{U}_0$ for $i < s$. This means that the $(i + 1)$-th iteration of the procedure from Algorithm 2 results in

$$\hat{\sigma}_{00}^2 = \frac{w_0 q_0 + w_m q_m + \sum_{t=1}^{i} w_{l(t)} q_{l(t)}}{w_0 + w_m + \sum_{t=1}^{i} w_{l(t)}}. \tag{4.94}$$

Note that, since $l(i+1) \in \mathcal{U}_1$, it holds that

$$\begin{aligned} q_{l(i+1)} &< \frac{w_0 q_0 + \sum_{t=1}^{i} w_{l(t)} q_{l(t)}}{w_0 + \sum_{t=1}^{i} w_{l(t)}} \\ &\leq \frac{w_0 q_0 + \sum_{t=1}^{s} w_{l(t)} q_{l(t)}}{w_0 + \sum_{t=1}^{s} w_{l(t)}} < q_m. \end{aligned} \tag{4.95}$$

But (4.95) implies that $q_{l(i+1)}$ is smaller than the right-hand side of (4.94). Hence, index $l(i + 1)$ is picked so that $l(i + 1) \in \mathcal{U}_0$. By induction, it follows that $\mathcal{U}_1 \subset \mathcal{U}_0$.

Appendix 4.E   Analysis of the detectors *Test 1* and *2*

Test 1 from (4.48) This test is given by $z_1 \gtrless_{\mathcal{H}_0^M}^{\mathcal{H}_1^M} 0$, with $z_1 \doteq q_M - \gamma \hat{\sigma}_{01}^2$, $\gamma > 1$, and $\hat{\sigma}_{01}^2$ a linear combination of q_0 and $\{q_j, \, j \in \mathcal{U}_1\}$, as in (4.45). Denote by I_t the event of having t of the channels $1, \ldots, M - 1$ not in use by the primary network, so that $\Pr\{I_t\} = \binom{M-1}{t} a^{M-1-t}(1 - a)^t$. In addition, denote by U_n the event $|\mathcal{U}_1| = n$.

Then we can write

$$
\begin{aligned}
\Pr\{z_1 > 0\} &= \sum_{t=0}^{M-1} \Pr\{I_t\} \Pr\{z_1 > 0 \,|\, I_t\} \\
&= \sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{M-1} \Pr\{z_1 > 0, U_n \,|\, I_t\} \\
&\approx \sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{t} \Pr\{z_1 > 0, U_n \,|\, I_t\},
\end{aligned}
\tag{4.96}
$$

where in the last step we have neglected the probability of a *busy* channel $j \neq M$ being included in the set $\mathcal{U}_1$ by the constrained LS estimate. This amounts to assuming that busy channels have sufficiently high SNRs. Without this approximation, $\Pr\{z_1 > 0\}$ would depend on the SNRs of the (busy) channels other than channel M, which is clearly undesirable. The accuracy of this assumption will be validated by the simulation results.

Let us define the vectors

$$
\mathbf{x}_n \doteq \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} - \hat{\sigma}_{01}^2 \mathbf{1}_n, \quad \mathbf{x}'_{t-n} \doteq \begin{bmatrix} q_{n+1} \\ \vdots \\ q_t \end{bmatrix} - \hat{\sigma}_{01}^2 \mathbf{1}_{t-n}.
\tag{4.97}
$$

Now, when computing (4.96), we can assume that the t idle channels are channels 1 through t (due to the equal bandwidth assumption), so that

$$
\Pr\{z_1 > 0\} \approx
$$
$$
\sum_{t=0}^{M-1} \Pr\{I_t\} \sum_{n=0}^{t} \binom{t}{n} \Pr\{z_1 > 0, \, \mathbf{x}_n < \mathbf{0}, \, \mathbf{x}'_{t-n} > \mathbf{0} \,|\, I_t\}.
\tag{4.98}
$$

Note that $\mathbf{x}_n < \mathbf{0}, \mathbf{x}'_{t-n} > \mathbf{0}$ imply that $\mathcal{U}_1 = \{1, \cdots, n\}$, so that

$$
\hat{\sigma}_{01}^2 = \frac{\sum_{l=0}^{n} w_l q_l}{\sum_{l=0}^{n} w_l}.
\tag{4.99}
$$

Now one has that

$$
\Pr\{z_1 > 0, \, \mathbf{x}_n < \mathbf{0}, \, \mathbf{x}'_{t-n} > \mathbf{0} \,|\, I_t\} =
$$
$$
\Pr\{[z_1 \ -\mathbf{x}_n^T \ (\mathbf{x}'_{t-n})^T]^T / \sigma_0^2 > \mathbf{0} \,|\, I_t\},
\tag{4.100}
$$

which is the integral of a $(t + 1)$-variate Gaussian distribution over the positive orthant. The mean of this distribution is $\boldsymbol{\mu}_1 = [\,(1-\gamma+\rho_M)\,\mathbf{0}_t^T\,]^T$, and the covariance matrix is found blockwise from the following, where $\bar{w}_n \doteq w_0 + nw_M$:

$$\operatorname{cov}(z_1, z_1) = \frac{\sigma_0^4}{K}\left(\frac{(1+\rho_M)^2}{w_M} + \frac{\gamma^2}{\bar{w}_n}\right), \tag{4.101}$$

$$\operatorname{cov}(\mathbf{x}_n, \mathbf{x}_n) = \frac{\sigma_0^4}{K}\left(\frac{1}{w_M}\mathbf{I} - \frac{1}{\bar{w}_n}\mathbf{1}_n\mathbf{1}_n^T\right), \tag{4.102}$$

$$\operatorname{cov}(\mathbf{x}_{t-n}', \mathbf{x}_{t-n}') = \frac{\sigma_0^4}{K}\left(\frac{1}{w_M}\mathbf{I} + \frac{1}{\bar{w}_n}\mathbf{1}_{t-n}\mathbf{1}_{t-n}^T\right), \tag{4.103}$$

$$\operatorname{cov}(z_1, \mathbf{x}_{t-n}') = \frac{\sigma_0^4}{K}\frac{\gamma}{\bar{w}_n}\mathbf{1}_{t-n}, \tag{4.104}$$

$$\operatorname{cov}(z_1, x_n) = \mathbf{0}, \;\; \operatorname{cov}(\mathbf{x}_n, \mathbf{x}_{t-n}') = \mathbf{0}. \tag{4.105}$$

Thus, $\Pr\{z_1 > 0\}$ is independent of σ_0^2 and can be computed numerically using any multivariate Gaussian integration package, such as Matlab's `mvncdf`. Note that $P_{\mathrm{FA}} = \Pr\{z_1 > 0 \,|\, \rho_M = 0\}$, whereas $P_{\mathrm{D}} = \Pr\{z_1 > 0 \,|\, \rho_M > 0\}$.

Test 2 from (4.50) This test is given by $z_2 \gtrless_{\mathcal{H}_0^M}^{\mathcal{H}_1^M} 0$, with $z_2 \doteq q_M - \gamma\hat{\sigma}_{02}^2$, $\gamma > 1$, and $\hat{\sigma}_{02}^2$ a linear combination of q_0 and $\{q_j,\, j \in \mathcal{U}_0,\, j \neq m\}$, as in (4.49). Denote by $\tilde{U}_n$ the event $|\mathcal{U}_0| = n$. Then, similarly to (4.96),

$$\Pr\{z_2 > 0\} \approx \sum_{t=0}^{M-1}\Pr\{I_t\}\sum_{n=0}^{t}\Pr\{z_2 > 0, \tilde{U}_n \,|\, I_t\} \tag{4.106}$$

$$= \sum_{t=0}^{M-1}\Pr\{I_t\}\sum_{n=0}^{t}\binom{t}{n}\Pr\{z_2 > 0, \tilde{\mathbf{x}}_n < \mathbf{0}, \tilde{\mathbf{x}}_{t-n}' > \mathbf{0} \,|\, I_t\}, \tag{4.107}$$

where now

$$\tilde{\mathbf{x}}_n \doteq \begin{bmatrix} q_1 \\ \vdots \\ q_n \end{bmatrix} - \hat{\sigma}_{00}^2\mathbf{1}, \qquad \tilde{\mathbf{x}}_{t-n}' \doteq \begin{bmatrix} q_{n+1} \\ \vdots \\ q_t \end{bmatrix} - \hat{\sigma}_{00}^2\mathbf{1}. \tag{4.108}$$

In this case, $\tilde{\mathbf{x}}_n < \mathbf{0}$, $\tilde{\mathbf{x}}_{t-n}' > \mathbf{0}$ imply that $\mathcal{U}_0 = \{1, \cdots, n, M\}$, and thus $\hat{\sigma}_{00}^2 = (w_M q_M + \sum_{l=0}^{n} w_l q_l)/(w_M + \sum_{l=0}^{n} w_l)$. The probability in (4.107) can be written again as the integral of a $(t + 1)$-variate Gaussian distribution over the positive orthant:

$$\Pr\{z_2 > 0, \tilde{\mathbf{x}}_n < \mathbf{0}, \tilde{\mathbf{x}}_{t-n}' > \mathbf{0} \,|\, I_t\} =$$
$$\Pr\{[z_2 \;\; -\tilde{\mathbf{x}}_n^T \;\; (\tilde{\mathbf{x}}_{t-n}')^T]^T/\sigma_0^2 > \mathbf{0} \,|\, I_t\}. \tag{4.109}$$

The mean of this distribution is in this case

$$\boldsymbol{\mu}_2 = \left[\begin{array}{ccc} (1 - \gamma + \rho_M) & \frac{w_M \rho_M}{\bar{w}_{n+1}} \mathbf{1}_n^T & \frac{-w_M \rho_M}{\bar{w}_{n+1}} \mathbf{1}_{t-n}^T \end{array}\right]^T, \tag{4.110}$$

whereas the covariance matrix can be found from

$$\text{cov}(z_2, z_2) = \frac{\sigma_0^4}{K} \left(\frac{(1 + \rho_M)^2}{w_M} + \frac{\gamma^2}{\bar{w}_n} \right), \tag{4.111}$$

$$\text{cov}(\tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_n) = \frac{\sigma_0^4}{K} \left[\frac{1}{w_M} \mathbf{I} + \left(\frac{w_M((1 + \rho_M)^2 - 1)}{\bar{w}_{n+1}^2} - \frac{1}{\bar{w}_{n+1}} \right) \mathbf{1}_n \mathbf{1}_n^T \right], \tag{4.112}$$

$$\text{cov}(\tilde{\mathbf{x}}_{t-n}', \tilde{\mathbf{x}}_{t-n}') = \frac{\sigma_0^4}{K} \left[\frac{1}{w_M} \mathbf{I} + \left(\frac{w_M((1 + \rho_M)^2 - 1)}{\bar{w}_{n+1}^2} + \frac{1}{\bar{w}_{n+1}} \right) \mathbf{1}_{t-n} \mathbf{1}_{t-n}^T \right], \tag{4.113}$$

$$\text{cov}(z_2, \tilde{\mathbf{x}}_n) = \frac{\sigma_0^4}{K} \left(\frac{\gamma - (1 + \rho_M)^2}{\bar{w}_{n+1}} - \frac{\gamma}{\bar{w}_n} \right) \mathbf{1}_n, \tag{4.114}$$

$$\text{cov}(\tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_{t-n}') = \frac{\sigma_0^4}{K} \frac{w_M((1 + \rho_M)^2 - 1)}{\bar{w}_{n+1}^2} \mathbf{1}_n \mathbf{1}_{t-n}^T, \tag{4.115}$$

$$\text{cov}(z_2, \tilde{\mathbf{x}}_{t-n}') = \frac{\sigma_0^4}{K} \frac{\gamma - (1 + \rho_M)^2}{\bar{w}_{n+1}} \mathbf{1}_{t-n}. \tag{4.116}$$

Therefore, for Tests 1 and 2 P_{FA} and P_{D} can be found for a given scenario without resorting to Monte Carlo simulation.

Appendix 4.F   Proof of Proposition 4.3

For fixed $\mathcal{A} = \{l(1), \ldots, l(|\mathcal{A}|)\}$, by definition, the components $m \notin \mathcal{A}$ are given by $[\boldsymbol{\sigma}(\mathcal{A})]_m = 0$. The remaining components can be arranged a vector $\boldsymbol{\theta}$ so that

$$[\boldsymbol{\theta}]_i = [\hat{\boldsymbol{\sigma}}(\mathcal{A})]_{l(i)}. \tag{4.117}$$

Then, the optimization problem (4.74) with respect to $\boldsymbol{\theta}$ reduces to a least squares minimization problem. If we make use of the Frobenious Norm definition $\|\mathbf{G}\|^2 \doteq \text{tr}(\mathbf{G}\mathbf{G}^H)$ we have that the error

$$\epsilon_{\mathcal{A}}(\boldsymbol{\theta}) \doteq \left\| \mathbf{y}\mathbf{y}^H - \sum_{l \in \mathcal{A}} \sigma_l^2 \mathbf{C}_l \right\|_2^2 \tag{4.118}$$

becomes:

$$\epsilon_{\mathcal{A}}(\boldsymbol{\theta}) = \text{tr}\left\{(\mathbf{y}\mathbf{y}^H)^2\right\} - 2\sum_{m\in\mathcal{A}}\sigma_m^2\,\text{tr}\left\{\mathbf{y}^H\tilde{\mathbf{C}}_m\mathbf{y}\right\} + \sum_{l\in\mathcal{A}}\sum_{m\in\mathcal{A}}\sigma_l^2\sigma_m^2\,\text{tr}\left\{\tilde{\mathbf{C}}_l\tilde{\mathbf{C}}_m\right\}.$$
(4.119)

Using the definitions in the Proposition 4.3 the Least Squares Estimate of the parameter vector $\boldsymbol{\theta}$ can be compactly written as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \epsilon_{\mathcal{A}}(\boldsymbol{\theta}) \tag{4.120}$$

$$= \arg\min_{\boldsymbol{\theta}} \boldsymbol{\theta}^H \mathbf{B}_{\mathcal{A}}\boldsymbol{\theta} - 2\boldsymbol{\theta}^H \mathbf{b}_{\mathcal{A}} \tag{4.121}$$

$$= \mathbf{B}_{\mathcal{A}}^{-1}\mathbf{b}_{\mathcal{A}}. \tag{4.122}$$

Appendix 4.G   Proof of Corollary 4.1

For fixed $\mathcal{A} = \{l(1),\ldots,l(|\mathcal{A}|)\}$, so that $[\boldsymbol{\sigma}]_m = 0$ for $m \notin \mathcal{A}$ and

$$[\boldsymbol{\sigma}]_{l(i)} = [\boldsymbol{\theta}]_i, \quad \text{for } i = 1,\ldots,|\mathcal{A}|, \tag{4.123}$$

we say that an unbiased estimator $\hat{\boldsymbol{\theta}}$ of the non-zero components of $\boldsymbol{\sigma}$ is efficient at the true parameter vector $\boldsymbol{\theta}_\star$ (see e.g. Scharf (1991)) if and only if

$$\boldsymbol{F}(\boldsymbol{\theta}_\star)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\star) = \boldsymbol{s}(\boldsymbol{\theta}_\star, \tilde{\mathbf{y}}), \tag{4.124}$$

where $\boldsymbol{F}(\boldsymbol{\theta}_\star)$ denotes the Fisher's information matrix of the estimation problem evaluated at $\boldsymbol{\theta}_\star$ and $\boldsymbol{s}(\boldsymbol{\theta}_\star, \tilde{\mathbf{y}})$ stands for the gradient of the log-likelihood function of the estimation problem evaluated at $\boldsymbol{\theta}_\star$ for an input vector $\tilde{\mathbf{y}}$.

Given $\{\mathcal{A}, \boldsymbol{\theta}\}$, the observation $\tilde{\mathbf{y}}$ is zero mean circular Gaussian with covariance $\tilde{\mathbf{R}}(\boldsymbol{\sigma})$ as in (4.4). The elements of the Fisher information matrix (FIM) $\boldsymbol{F}(\boldsymbol{\theta})$, of size $|\mathcal{A}| \times |\mathcal{A}|$, are given by (see e.g. Kay (1998)):

$$[\boldsymbol{F}(\boldsymbol{\theta})]_{ij} = \text{tr}\left\{\tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma})\tilde{\mathbf{C}}_{l(i)}\tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma})\tilde{\mathbf{C}}_{l(j)}\right\}, \tag{4.125}$$

where we used that in our model $\partial\tilde{\mathbf{R}}(\boldsymbol{\sigma})/\partial\sigma_m^2 = \tilde{\mathbf{C}}_m$. On the other hand we have

that

$$[\boldsymbol{s}(\boldsymbol{\theta}, \tilde{\mathbf{y}})]_i = \frac{\partial}{\partial \sigma_{l(i)}^2} \log f(\tilde{\mathbf{y}} | \boldsymbol{\sigma}) \tag{4.126}$$

$$= -\operatorname{tr}\left\{ \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \frac{\partial \tilde{\mathbf{R}}(\boldsymbol{\sigma})}{\partial \sigma_{l(i)}^2} \right\} + \operatorname{tr}\left\{ \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \frac{\partial \tilde{\mathbf{R}}(\boldsymbol{\sigma})}{\partial \sigma_{l(i)}^2} \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \tilde{\mathbf{y}} \tilde{\mathbf{y}}^H \right\} \tag{4.127}$$

$$= -\operatorname{tr}\left\{ \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \tilde{\mathbf{C}}_{l(i)} \right\} + \tilde{\mathbf{y}}^H \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \tilde{\mathbf{C}}_{l(i)} \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \tilde{\mathbf{y}}. \tag{4.128}$$

Then the terms in (4.124) can be reorganized as follows

$$\boldsymbol{F}(\boldsymbol{\theta}_\star) \hat{\boldsymbol{\theta}} = \boldsymbol{s}(\boldsymbol{\theta}_\star, \tilde{\mathbf{y}}) + \boldsymbol{F}(\boldsymbol{\theta}_\star) \boldsymbol{\theta}_\star \tag{4.129}$$

$$= \tilde{\mathbf{y}}^H \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}_\star) \tilde{\mathbf{C}}_{l(i)} \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}_\star) \tilde{\mathbf{y}} \tag{4.130}$$

$$\doteq \boldsymbol{f}(\boldsymbol{\theta}_\star), \tag{4.131}$$

where in (4.130) we made use of the problem structure, that allows us to write

$$[\boldsymbol{F}(\boldsymbol{\theta}_\star) \boldsymbol{\theta}_\star]_i = \operatorname{tr}\left\{ \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \tilde{\mathbf{C}}_{l(i)} \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \tilde{\mathbf{R}}(\boldsymbol{\sigma}) \right\} \tag{4.132}$$

$$= \operatorname{tr}\left\{ \tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}) \tilde{\mathbf{C}}_{l(i)} \right\}. \tag{4.133}$$

Note that for the right sparsity pattern $\mathcal{A}$, the LS estimator given in Proposition 4.3 must fulfill

$$\mathbf{B}_{\mathcal{A}} \hat{\boldsymbol{\theta}} = \mathbf{b}_{\mathcal{A}}, \tag{4.134}$$

which resembles the efficiency condition given in (4.130). In fact, in the asymptotic low SNR regime $\tilde{\mathbf{R}}^{-1}(\boldsymbol{\sigma}_\star) \to \mathbf{I}_K / (\sigma_0^\star)^2$ and

$$\boldsymbol{F}(\boldsymbol{\theta}_\star) \to \frac{1}{(\sigma_0^\star)^4} \boldsymbol{B}_{\mathcal{A}}, \tag{4.135}$$

$$\boldsymbol{f}(\boldsymbol{\theta}_\star) \to \frac{1}{(\sigma_0^\star)^4} \mathbf{b}_{\mathcal{A}}, \tag{4.136}$$

which gives us the desired result.

Chapter 5

Dynamic Spectrum Leasing

Contents

5.1 Introduction

The overlay[1] hierarchical *dynamic spectrum access* (DSA) in which secondary users are allowed to opportunistically access the spectrum on the basis of no-interference

---

[1]In the context of Cognitive Radio overlay/underlay paradigms present different meanings depending on the field. For example, in the Information Theory community underlay transmission usually denotes a transmission causing minimum interference to licensed users while overlay paradigms have the capability of overhearing and/or enhancing primary transmissions. Here we use the term overlay to refer to schemes which avoid to interfere to the primary user, while underlay schemes are allowed to transmit simultaneously to the licensed user once fixed an interference margin.

to the primary (licensed) users presents stringent detection requirements. Hence, powerful detectors, as those studied in the previous chapters become a key aspect of CR systems employing this paradigm.

However, in certain cases primary users may be willing to tolerate certain level of interference when their QoS is not affected and/or when they receive a certain compensation (monetary or of other nature) for this interference. In this case the secondary system may act in an underlay basis, i.e., secondary users transmit simultaneously with the primary system provided that they meet certain requirements on the maximum interference seen by the primary user, denoted here as interference cap (ICAP). In this scenario the detection of primary signals loses importance with respect to interference management tasks, which may become non trivial depending on the network configuration.

Various spectrum underlay and overlay architectures have been proposed and investigated in recent years (see Kim et al. (2008); Le and Hossain (2008); Xing et al. (2007); Fattahi et al. (2007); Etkin et al. (2007); Menon et al. (2008) and references therein). The existing literature in underlay and overlay based secondary networks, however, impose the burden of interference management mainly on the secondary system. In particular, it is assumed that there is a maximum interference level that the primary system is willing to tolerate, and the secondary powers/activity are to be adjusted within this constraint. As opposed to this is the the concept of *dynamic spectrum leasing* (DSL), first presented in Jayaweera and Li (2009).

A DSL scheme is characterized by the active role of the primary user, which may interact with the secondary system in order to define the allowed interference cap. This scheme allows the system to adapt to changing environmental conditions which may lead to a better spectral utilization. In this chapter we will first investigate theoretically the performance improvement that can be expected from a DSL based paradigm with respect to passive spectrum sharing schemes which do not allow dynamic primary-secondary network interaction. The proposed analysis results into a Stackelberg game formulation of the interactions between primary and secondary systems. Once the performance advantage of DSL is demonstrated we will present a game theoretical framework in order to model and analyze practical DSL schemes. We will show certain conditions which guarantee the non-cooperative game to converge and study the behavior of the proposed scheme under dynamic environments.

5.2   System model

We assume that there is one primary wireless communication system that owns the license rights to the spectrum band of interest. The users in this primary system, however, may not be using its spectrum completely all the time, or may be able to tolerate a certain amount of additional co-channel interference without compromising required QoS constraints, leading to an inefficient utilization of radio spectrum. For simplicity of exposition, we focus on a particular channel in the primary system that is allocated to a single primary user (for example, as in FDMA). We assume K secondary transmitters are interested in accessing this spectrum band of interest to the maximum possible extent. The primary user is denoted as user 0, and the secondary users are labeled as users 1 through K. While it is possible to extend this framework to more complex scenarios, for ease of exposition we focus here on the case that we have only one primary and one secondary receiver of interest.

The channel gain between the k-th transmitter (either primary or secondary) and the common secondary receiver is denoted by h_{sk}, and that between the k-th transmitter and the primary receiver is denoted by h_{pk}, for $k = 0, 1, \ldots, K$. Throughout the analysis in this chapter we assume fading to be quasi-static, so that the coefficients stay fixed for a certain duration of time after which they change to a new set of values. It should be mentioned that quasi-static fading model is frequently used in modeling many wireless communications environments (Molisch, 2005). Our model can also be complemented with a channel estimation and tracking algorithm to cope with slowly time-varying situations and as we will show later, the performance of the proposed DSL scheme is fairly robust against such time-varying fading.

Signal model

A discrete-time representation of the received signals at the primary and secondary receivers can be written as

$$r_p[n] = \tilde{h}_{p0} s_0[n] + \sum_{k=1}^{K} \tilde{h}_{pk}[n] \tilde{s}_k[n] + \sigma_p n_p[n]; \tag{5.1}$$

$$r_s[l] = \tilde{h}_{s0} \tilde{s}_0[l] + \sum_{k=1}^{K} \tilde{h}_{sk}[l] s_k[l] + \sigma_s n_s[l] \tag{5.2}$$

where n and l represent the discrete sampling times at primary and secondary receivers respectively, $\tilde{h}_{pk}$ and $\tilde{h}_{sk}$ are the effective channels from k-th transmitter to the primary and secondary receivers respectively. If $s_k(t)$ denotes the signal transmitted by the k-th user, then $s_k[n]$ denotes a synchronously sampled and $\tilde{s}_k[n]$ an asynchronously sampled version of $s_k(t)$. Finally $n_p[n]$ nd $n_s[n]$ are iid Gaussian processes normalized to have variance 1 so that σ_p^2 and σ_s^2 represent the noise power levels at the primary and secondary receivers, respectively.

We denote the transmit power of the k-th user as $p_k \doteq \mathrm{E}[|s_k[n]|^2] \doteq \mathrm{E}[|\tilde{s}_k[n]|^2]$ for $k = 0, 1, \ldots, K$. Note that this assumes that any deviations on the received power due to front-end and bandwidth differences are absorbed into the effective channel coefficients. Then it is straightforward to see that the actual interference power generated by the secondary system at the primary user is given by

$$I_0 \doteq \sum_{k=1}^{K} |\tilde{h}_{pk}|^2 p_k. \tag{5.3}$$

Interference cap

The primary user is assumed to adapt its interference cap (IC), denoted by Q_0, which is the maximum total interference the primary user is willing to tolerate from secondary transmissions at any given time. By adjusting this interference cap Q_0, the primary user can control the total transmit power the secondary users impose on its licensed channel. The motivation for the primary user can be, for instance, the monetary reward obtained by allowing secondary users to access its licensed spectrum. In essence, then, the interference cap determines how much secondary user activity the primary user is ready to allow, and thus its reward should be an increasing function of the interference cap. However, we impose the realistic constraint that the primary user should always maintain a target signal-to-interference-plus-noise ratio (SINR) to ensure its required transmission QoS. Moreover, an unnecessarily large interference cap by the primary user could hinder the performance of both secondary system and other primary transmitters (though, for simplicity, not included in the current model) due to resulting high interference.

Decoding strategy

If the secondary system is equipped with conventional matched-filter receivers at any given time t, the primary user's SINR γ_0 and i-th secondary user SINR γ_i can be respectively defined as

$$\gamma_0 \doteq \frac{|\tilde{h}_{p0}|^2 p_0}{I_0 + \sigma_p^2}, \tag{5.4}$$

$$\gamma_i \doteq \frac{|\tilde{h}_{si}|^2 p_i}{I_i + \sigma_s^2}, \tag{5.5}$$

where the total interference seen by the secondary receiver in detecting the k-th secondary signal is

$$I_i \doteq \sum_{k=0, k \neq i}^{K} |\tilde{h}_{sk}|^2 p_k. \tag{5.6}$$

Then the maximum achievable rate per channel use, assuming secondary interference is treated as noise at the primary system, is given by

$$R_p \leq W_p \log\left(1 + \gamma_0\right) \tag{5.7}$$

$$= W_p \log\left(1 + \frac{|\tilde{h}_{p0}|^2 p_0}{I_0 + \sigma_p^2}\right) \tag{5.8}$$

where W_p represents the bandwidth employed by primary transmissions and the transmitted power p_0 is determined by the required quality of service (QoS) and the interference cap selected.

We impose here the realistic constraint that the primary user should always maintain a target signal-to-interference-plus-noise ratio (SINR) to ensure its required transmission QoS. To that end, we introduce the primary user's target SINR is defined in terms of its assumed worst-case secondary interference:

$$\bar{\gamma}_0 \doteq \frac{|\tilde{h}_{p0}|^2 p_0}{Q_0 + \sigma_p^2}. \tag{5.9}$$

Note that, since Q_0 is the maximum possible interference from secondary users the primary user is willing to tolerate, $\bar{\gamma}_0$ represents the least acceptable transmission quality of the primary user.

Multiuser decoding. While Matched filter (MF) decoding is a popular decoding structure due to its simplicity and performs reasonably well in systems with weak cross-channels, in an interference limited regime it is clearly suboptimal and it is outperformed by joint decoding of multiple users. When evaluating the best theoretical performance a DSL scheme can achieve, we may assume an optimal joint maximum likelihood multiuser decoder (ML MUD) of Verdú (1998) at the secondary user. Note that such decoder will give a fundamental limit against which practical schemes based on matched filtering can be compared.

5.3   Performance gain of DSL based schemes

First we study the maximum performance gain we can expect from a DSL scheme allowing a limited interaction between primary and secondary users. As opposed to previous works we employ here performance metrics based on the multiuser sum-rate attainable by the secondary system. We choose this performance metric because it is a fundamental limit against which practical schemes can be compared, while it is independent of particular DSL implementations.

While in principle the primary signal could be decoded at the secondary receiver, cognitive radio systems are expected to work in harsh SNR conditions. In this regime primary signals cannot be reliably decoded at secondary receivers due to SNR considerations and/or synchronization issues. Therefore we will treat primary transmission purely as noise. Under these assumptions the maximum achievable sum rate at the secondary receiver with total bandwidth W_s treating primary transmissions as noise is, see e.g. (Cover and Thomas, 2006, Sec. 15.3.6),

$$R_s < W_s \log \left(1 + \frac{\sum_{k=1}^{K} |\tilde{h}_{sk}|^2 p_k}{|\tilde{h}_{s0}|^2 p_0 + \sigma_s^2} \right) \tag{5.10}$$

for each of the allowed secondary power assignments p_k with $k = 1, \ldots, K$, which are determined by the maximum interference allowed at the primary user $I_0 \leq Q_0$ and secondary user individual power constraints $p_k < \bar{p}_k$.

The rate region in (5.10) obtained with this decoding scheme is similar to the one obtained in a Gaussian Multiple Access Channel, with the peculiarity that on top of having individual power constraints secondary users have a weighted global power constraint. The individual rates achieved by each secondary user will depend on the particular coding/decoding strategy used.
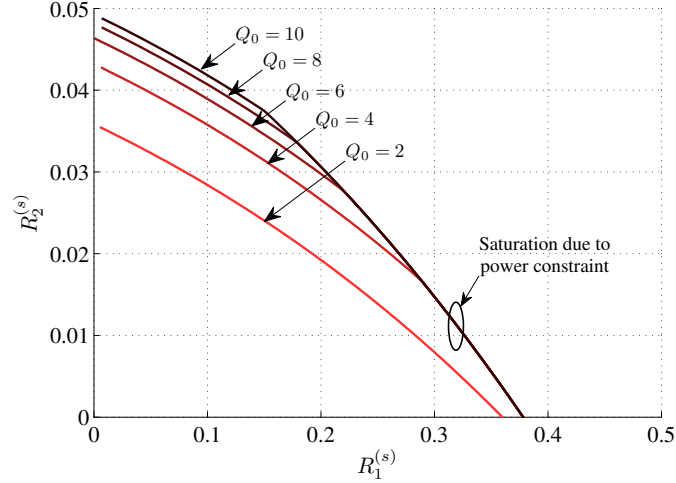
Figure 5.1: Secondary system 2-user rate region for different values of Q_0.

From the constraint $I_0 \leq Q_0$ and given the definition of I_0, we have that the term $\sum_{k=1}^{K} |\tilde{h}_{sk}|^2 p_k$ in (5.10) is upper bounded by a monotonically increasing affine function of Q_0. Then it is apparent from (5.10) that, while the upper bound on the secondary sum-rate is monotonically increasing with Q_0, the growth rate decreases with Q_0 due to the logarithmic relation with $\sum_{k=1}^{K} |\tilde{h}_{sk}|^2 p_k$. Figure 5.1 shows an example of the rate region obtained in a two user secondary system where the channel from user 1 to primary is much weaker than the one from user 2 for different values of Q_0. While in general the region is increasing with Q_0 the effect of the individual power constraints of the secondary nodes translates into the partial saturation of the achievable rate region.

5.3.1   Performance metric

Although performance evaluation of cognitive radio systems is important in comparing and ranking different paradigms, it has received only a limited attention in the existing literature (Zhao et al., 2009). Even for the relatively simple model considered here, there exist several possible evaluation metrics: maximum achievable sum-rates at primary and secondary systems R_p and R_s respectively, power dissipated by a given user p_k, interference generated at the primary user I_0, probability of undesirable primary interference conditions $\Pr\{I_0 > Q_0\}$, fairness among users, and spectral efficiency, among others. Therefore, an adequate utility function must first be defined in order to compare DSL based paradigms with other schemes.

While in our model a natural performance metric for the secondary system should be an increasing function of the attained sum rate R_s, the primary user's utility needs further considerations. Since the primary user suffers from a (permitted) interference I_0 from the secondary system, in order to maintain its QoS the primary user transmitted power p_0 is increased with respect to an exclusive use of the frequency band ($I_0 = 0$). We denote this increment in the transmitted power by Δp_0. Hence the primary user needs an incentive to allow secondary users to use its managed spectrum. We assume here that the secondary system compensates the primary user with a payment (monetary or of other nature) related to the generated interference I_0. As a result, the utility functions for primary and secondary systems can be written as:

$$U_p = u_p(I_0, \Delta p_0), \tag{5.11}$$

$$U_s = u_s(R_s, I_0) \tag{5.12}$$

where primary utility $U_p(\cdot)$ is growing with I_0 and decreasing with Δp_0, while secondary utility $U_s(\cdot)$ increases with R_s. We additionally assume that when the interference constraint is violated, that is, when $I_0 > Q_0$, the penalization imposed by the primary system to the secondary system implies $U_p = \infty$, $U_s = -\infty$. This penalty discourages the secondary system from violating the allowed interference cap.

5.3.2   Performance analysis

For a given interference cap Q_0 the secondary utility U_s is maximized for the secondary power vector $\boldsymbol{p} \doteq [p_1 p_2 \cdots p_K]^T$ provided that

$$\boldsymbol{p}^\star(Q_0) = \arg\max_{\boldsymbol{p}} \{u_s(R_s(\boldsymbol{p}, Q_0), I_0(\boldsymbol{p}))\} \tag{5.13}$$

$$\text{subject to} \quad I_0(\boldsymbol{p}) \leq Q_0, \ \boldsymbol{p} \leq \bar{\boldsymbol{p}}$$

where we defined $\bar{\boldsymbol{p}} \doteq [\bar{p}_1 \bar{p}_2 \cdots \bar{p}_K]^T$ and the operator $\leq$ when applied to vectors denotes element by element comparison. Here we have explicitly shown the dependence of R_s on Q_0. We define the corresponding primary and secondary utilities as $U_p^\star(Q_0) \doteq U_p(\boldsymbol{p}^\star(Q_0), Q_0)$ and $U_s^\star(Q_0) \doteq U_s(\boldsymbol{p}^\star(Q_0), Q_0)$, respectively.

If the primary user fixes *a priori* the interference cap Q_0 in a time varying environment its expected utility is given by $\mathrm{E}[U_p^\star(Q_0)]$, where the expectation is

taken with respect to the channel realizations. On the other hand, in a DSL scheme we allow the primary system to dynamically adjust the allowed interference cap Q_0. We can now compute the maximum achievable utility for both types of schemes:

Schemes with fixed Q_0: If the primary user chooses the value of Q_0 that maximizes the expected utility and uses it for all channel realizations, its utility is given by

$$\bar{U}_p^{\text{fixed}} = \max_{Q_0}\{\mathrm{E}[U_p^\star(Q_0)]\}. \tag{5.14}$$

DSL schemes: On the other hand, in a DSL-based system, the primary will choose the interference cap Q_0 to maximize its own utility for each channel realization. The best expected primary utility achievable in this dynamic environment is

$$\bar{U}_p^{\text{dsl}} = \mathrm{E}[\max_{Q_0}\{U_p^\star(Q_0)\}]. \tag{5.15}$$

It is easy to see from (5.14) and (5.15) that $\bar{U}_p^{\text{dsl}} \geq \bar{U}_p^{\text{fixed}}$, where, for strictly monotonic $U_p^\star(\cdot)$, $\bar{U}_p^{\text{dsl}} = \bar{U}_p^{\text{fixed}}$ if and only if the optimal Q_0 is constant for all channel realizations. In the next section we will use a simple example to show that indeed the gain obtained by a DSL scheme can be significant.

Remark 5.1. In deriving (5.15) we implicitly formulated the interaction between the primary and secondary systems as a Stackelberg game (Fugenberg and Tirole, 1991), in which the primary user acts as Stackelberg *leader* and the secondary system acts as *follower*. While this is a natural model for cognitive radio systems in which the primary can always act unilaterally while secondary users have to adapt their actions to the imposed constraint (Simeone et al., 2008), practical implementations that achieve this behavior are a topic of further research. However, later in this chapter, we will propose a simple theoretical framework based on non-cooperative game theory that conducts to practical DSL schemes.

5.3.3   Example

For illustration purposes, in this section we assume that the utilities associated with primary and secondary users are respectively

$$U_p = I_0 - \mu_P \Delta p_0, \tag{5.16}$$

$$U_s = \mu_R R_s - I_0 \tag{5.17}$$

with the additional restriction that $I_0 \leq Q_0$. That is, the primary system obtains a reward proportional to the suffered interference I_0 and charged to the secondary system. Without loss of generality we assume here the payoff per unit of interference equal to 1. The primary user has a cost associated with the extra power Δp_0 required to maintain its desired QoS, priced at the rate of μ_P. The reward for the secondary system is proportional to the achievable sum rate R_s priced at the rate of μ_R. Note that whereas these utilities keep the spirit of (5.11) and (5.12), they are also simple enough to obtain analytical results.

Assuming equality in (5.10) we may rewrite (5.17) as

$$U_s = \mu_R W_s \log \left(1 + \frac{\sum_{k=1}^{K} \frac{|\tilde{h}_{sk}|^2}{|\tilde{h}_{pk}|^2} \tilde{p}_k}{\sigma_s^2 + |\tilde{h}_{s0}|^2 p_0} \right) - \sum_{k=1}^{K} \tilde{p}_k \tag{5.18}$$

where we have defined $\tilde{p}_k \doteq |\tilde{h}_{pk}|^2 p_k > 0$.

In order to maximize U_s with respect to $\tilde{p}_k$ we first note that for fixed $\sum_{k=1}^{K} \tilde{p}_k = I_0$, U_s is growing with respect to a convex combination of the (positive) ratios $|\tilde{h}_{sk}|^2/(I_0|\tilde{h}_{pk}|^2)$. Hence, for a fixed I_0, U_s is maximized when all the allowed secondary interference I_0 is allocated to the secondary transmitters with the largest ratios $|\tilde{h}_{sk}|^2/|\tilde{h}_{pk}|^2$ up to their individual power constraints. Formally, if we define the indexes of the sorted effective channels as $\{i_1, i_2, \ldots, i_K\}$ such that

$$\frac{|\tilde{h}_{si_1}|^2}{|\tilde{h}_{pi_1}|^2} \geq \frac{|\tilde{h}_{si_2}|^2}{|\tilde{h}_{pi_2}|^2} \geq \cdots \geq \frac{|\tilde{h}_{si_K}|^2}{|\tilde{h}_{pi_K}|^2}, \tag{5.19}$$

the optimal power assignment is given by

$$\tilde{p}_{i_k}^{\star} \doteq \begin{cases} |\tilde{h}_{pi_k}|^2 \bar{p}_{i_k}, & \delta_k < I_0, \\ I_0 - \delta_{k-1}, & \delta_{k-1} \leq I_0 \leq \delta_k, \\ 0, & \text{elsewhere}, \end{cases} \tag{5.20}$$

where we defined $\delta_k \doteq \sum_{l=1}^{k} |\tilde{h}_{pi_l}|^2 \bar{p}_{i_l}$. Then we may define the instantaneous channel ratio η as

$$\eta \doteq \frac{\sum_{k=1}^{K} \frac{|\tilde{h}_{sk}|^2}{|\tilde{h}_{pk}|^2} \tilde{p}_k^{\star}}{I_0}. \tag{5.21}$$

Note that when the secondary individual power constraints are not active $I_0 \leq |\tilde{h}_{pi_1}|^2 \bar{p}_{i_1}$, hence η reduces to the largest channel ratio pair: $\eta = \max_k \{|\tilde{h}_{sk}|^2/|\tilde{h}_{pk}|^2\}$. Otherwise η is a convex combination of the strongest channel ratio pairs.

Remark: While the simple utility (5.17) leads to an opportunistic access scheme that does not take into account fairness among secondary users, in the general setting U_s could take a more complex form in order to guarantee fairness. However this analysis lies out of the scope of the present work.

Using (5.21) and substituting (5.20) in (5.18), we have that

$$U_s = \mu_R W_s \log \left(1 + \frac{\eta I_0}{\sigma_s^2 + |\tilde{h}_{s0}|^2 p_0} \right) - I_0. \tag{5.22}$$

Equating the derivative of (5.22) with respect to I_0 to zero, we obtain the global U_s maximizer. Taking into account the additional constraint $I_0 < Q_0$, one obtains that the optimal I_0 is given by

$$I_0^{\star}(Q_0) = \min \left(Q_0, W_s \mu_R + \frac{\sigma_s^2 + |\tilde{h}_{s0}|^2 p_0}{\eta} \right). \tag{5.23}$$

From (5.9) we have that $p_0 = \bar{\gamma}_0 (Q_0 + \sigma_p^2)/|\tilde{h}_{p0}|^2$. Then it follows

$$U_p^{\star}(Q_0) = I_0^{\star} - \mu_P \Delta p_0 \qquad \text{and} \tag{5.24}$$

$$U_s^{\star}(Q_0) = \mu_R W_s \log \left(1 + \frac{\eta I_0^{\star}}{\sigma_s^2 + |\tilde{h}_{s0}|^2 p_0} \right) - I_0^{\star}. \tag{5.25}$$

It is interesting to note that both primary and secondary utilities depend only on the channel of the secondary user with the smallest channel ratio $|\tilde{h}_{pk}|^2/|\tilde{h}_{sk}|^2$. This comes from the fact that the secondary system's sumrate is maximized when all the allowed interference at the primary user is allocated to this single secondary user.

The maximal primary utility is achieved by a DSL system maximizing $U_p^{\star}(Q_0)$. Given the restriction $I_0 \leq Q_0$ and since $U_p^{\star}$ is growing with $I_0^{\star}$ and decreasing with

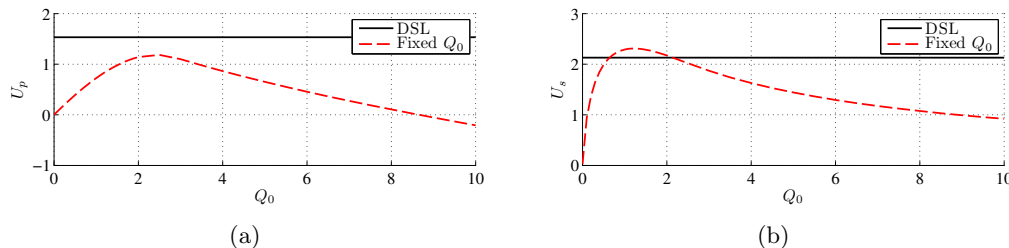(a)   (b)

Figure 5.2: Primary/secondary users average performance in a time varying environment. (a) Primary user performance. (b) Secondary user performance.

Q_0 it can be shown that $U_p^\star(Q_0)$ is maximized when $I_0^\star = Q_0$. Hence the optimal instantaneous Q_0 is given by

$$Q_0^\star = \frac{\eta W_s \mu_R - \frac{|\tilde{h}_{s0}|^2}{|\tilde{h}_{p0}|^2}\bar{\gamma}_0 \sigma_p^2 - \sigma_s^2}{\eta + \frac{|\tilde{h}_{s0}|^2}{|\tilde{h}_{p0}|^2}\bar{\gamma}_0}. \tag{5.26}$$

Note that, as can be seen from (5.26) above, the optimal strategy for the primary user is heavily dependent on the scenario and thus cannot be fixed a priori. In order to compute the expected gain in a dynamic environment for a DSL based scheme over a paradigm with fixed Q_0, given by $\bar{U}_p^{\text{dsl}} - \bar{U}_p^{\text{fixed}}$, we further need to define a channel model and compute the average of (5.24) with respect to all channel realizations. Although, in general, the expected gain cannot be computed in a closed form, it can easily be evaluated numerically for any given set of parameters.

5.3.4   Numerical results

Here we show the performance gain that can be expected when the simple utilities in (5.16)-(5.17) are used. We assume that channels from secondary transmitters $\tilde{h}_{pk}$ and $\tilde{h}_{sk}$ are Rayleigh distributed with $\mathrm{E}[|\tilde{h}_{pk}|^2] = \mathrm{E}[|\tilde{h}_{sk}|^2] = 1$ for $k = 1, \ldots, K$ while the primary transmitter and primary and secondary receivers are assumed stationary so that $|\tilde{h}_{p0}|^2 = |\tilde{h}_{s0}|^2 = 1$. The remaining system parameters are $K = 3$, $\bar{p}_k = 100$, $\bar{Q}_0 = 10$, and $\sigma_p^2 = \sigma_s^2 = 1$. We employ normalized bandwidth $W_p = W_s = 1$, target SINR $\bar{\gamma}_0 = 1$ and resource prizes initially set to $\mu_P = 0.1$ and $\mu_R = 2$.

Figure 5.2 shows the comparison between a DSL based scheme and a scheme in which the allowed interference cap Q_0 is fixed for the given set of system parameters. In Fig. 5.2(a) we can see that even if a fixed system were to use the optimal $Q_0 \approx 2.5$,

the primary utility attainable by a DSL based scheme is about 25% larger than the one of the fixed scheme. On the other hand, if we look at the secondary utility obtained by a DSL based scheme compared to a scheme with fixed Q_0, as shown in Fig. 5.2(b), we can see that while fixed schemes perform better than DSL for a small range of Q_0 values, for the optimal operating point of the fixed scheme ($Q_0 \approx 2.5$) DSL performs slightly better than the fixed scheme. That is, in this setting both primary and secondary users can benefit from the use of a DSL scheme. Moreover in a DSL based scheme the allowed interference at the primary is computed on line, and thus it does not need to be fixed *a priori*. Hence, DSL schemes can be robust against inaccurate knowledge of the system parameters that may degrade both primary and secondary performance at the expense of the extra complexity required for dynamically setting the value of Q_0. Note from Fig. 5.2 that a small change in the Q_0 value for the fixed scheme can significantly degrade the global system performance.

However, as we pointed out above the advantage of DSL based schemes vanishes if the optimal primary user action Q_0 is independent from the channel realization. If we assume high reward for the secondary system sum rate, that is $\mu_R = 100$, the best responses for both primary and secondary users turn out to be $I_0 = Q_0 = \bar{Q}_0$, not depending on the channel realization. In this case DSL and fixed schemes with $Q_0 = \bar{Q}_0$ turn out to be equivalent achieving $U_p \approx 8.99$ and $U_s \approx 235$.

5.4 General formulation for practical DSL schemes

In the previous section we derived the expected performance gain that can be obtained from a DSL based scheme. However in the resulting Stackelberg formulation we assumed that both primary and secondary systems have perfect knowledge of all system parameters, and thus they can optimize their performance by maximizing their own utilities. Hence this scheme cannot be directly implemented in practice.

Next, we develop a general formulation which, we will see, leads to practical DSL schemes. To this end we will make use of non cooperative game theory to model the interactions among primary and secondary systems. We establish the conditions under which the proposed game-theoretic formulation has a unique Nash equilibrium to which both primary and secondary best-response adaptations would converge and the performance of the system will be found to be the performance at this Nash equilibrium.

Naturally, any DSL system requires each system to know a certain amount of information about the other system. However, it may arguably be desirable to minimize the awareness the primary system needs to have on the secondary operation. In this section, we show that indeed successful dynamic spectrum leasing can be achieved still relegating most of the interference management burden to the secondary system, with the primary system just broadcasting two parameter values periodically, namely, its tolerable interference cap and the total interference it is currently experiencing from the secondary transmissions. These are quantities that are readily available at the primary users (or can be easily estimated). Additionally we will assume in the following matched filter decoding in the secondary receiver, which is more adequate for practical implementations.

5.4.1   Non-cooperative game model

Primary and secondary users interact with each other by adjusting their interference cap and transmit power levels, respectively, in order to maximize their own utility. Hence, game theory provides a natural framework to model and analyze this DSL network. At a given time we may formulate the above system as in the following non-cooperative game $(\mathcal{K}, \mathcal{A}_k, u_k(\cdot))$:

1. Players: $\mathcal{K} = \{0, 1, \cdots, K\}$, where we assume that the 0-th user is the primary user and $k = 1, 2, \cdots, K$ represents the k-th secondary user.

2. Action space: $\mathcal{P} = \mathcal{A}_0 \times \mathcal{A}_1 \cdots \times \mathcal{A}_K$, where $\mathcal{A}_0 = \mathcal{Q} = [0, \bar{Q}_0]$ represents the primary user's action set and $\mathcal{A}_k = \mathcal{P}_k = [0, \bar{p}_k]$, for $k = 1, 2, \ldots, K$, represents the k-th secondary user's action set. Note that $\bar{Q}_0$ and $\bar{p}_k$ represent, respectively, the maximum possible interference cap of the primary user and the maximum transmission power of the k-th secondary user (as determined by the system and regulatory considerations). We denote the action vector of all users by $\boldsymbol{a} = [Q_0, p_1, \cdots, p_K]^T$, where $Q_0 \in \mathcal{Q}$ and $p_k \in \mathcal{P}_k$. It is customary to denote the action vector excluding the k-th user, for $k = 0, 1, \ldots, K$, by $\boldsymbol{a}_{-k}$.

3. Utility function: We denote by $u_0(Q_0, \mathbf{a}_{-0})$ the primary user's utility function, and by $u_k(p_k, \mathbf{a}_{-k})$, for $k = 1, 2, \ldots, K$, the k-th secondary user's utility function.

One of the main features of the dynamic spectrum leasing approach is the consideration of the coupling of primary system with the secondary-user system in

terms of mutual interference. The primary user can then define its utility function as:

$$u_0\left(Q_0, \mathbf{a}_{-0}\right) = \left(\bar{Q}_0 - \left(Q_0 - I_0(\mathbf{a}_{-0})\right)\right) Q_0 \tag{5.27}$$

$$\doteq\ u_0\left(Q_0, I_0\right). \tag{5.28}$$

Note that (5.28) assumes that the utility of the primary user is proportional to both demand and its interference cap level Q_0. The demand is taken to be decreasing when the extra interference margin $Q_0 - I_0$ increases. This discourages the primary user to swamp all other transmissions (both primary and secondary), by setting too large an interference cap which would lead to higher transmission power. It is also worth noting that this u_0 is continuous in $\boldsymbol{a}$ and concave in Q_0.

The (selfish) objective of each secondary user is to maximize a given utility function that depends on its own SINR (for example, throughput) without violating the primary user interference cap. Any utility function in a reasonable communication system is of course a monotonically increasing function of the received SINR γ_k. Observe from the expression of γ_0 in (5.4) that as long as the secondary user interference I_0 is below the interference cap Q_0 set by the primary user, the required QoS of the primary user will be guaranteed. To ensure this the utilities of secondary users must be fast decaying functions of $I_0 - Q_0$ when this difference is positive. Motivated by these arguments we propose the following form for the secondary user utility function:

$$u_k\left(p_k, \mathbf{a}_{-k}\right) = \left(Q_0 - \lambda_s I_0\right) g\left(\gamma_k\right) \tag{5.29}$$

$$= \left(Q_0 - \lambda_s I_0\right) g\left(\frac{|\tilde{h}_{sk}|^2 p_k}{I_k + \sigma_s^2}\right) \tag{5.30}$$

where $g(\cdot)$ is a suitable, non-negative reward function, and λ_s is a positive weighting coefficient. Note that in (5.29) the coefficient λ_s essentially controls how strictly secondary users need to adhere to the primary user's interference cap. The proposed utility function (5.29) leaves the performance metrics of the secondary system to be arbitrary by allowing for any reasonable reward function $g(\cdot)$ that will satisfy the conditions to be set forth in the next section. Without loss of generality, we may assume that the reward function $g\left(\gamma_k\right)$ satisfies $g(0) = 0$ and $g'(0) > 0$, since when the received SINR of a user vanishes no useful communication is possible for that user.

5.4.2   Nash equilibrium

In the following we investigate equilibrium strategies on the proposed DSL game $G = (\mathcal{K}, \mathcal{A}_k, u_k)$. The most commonly used equilibrium concept in non-cooperative game theory is the Nash equilibrium:

Definition 5.1. A strategy vector $\boldsymbol{a} = (a_0, a_1, \cdots, a_k)$ is a Nash equilibrium of the primary-secondary user dynamic spectrum leasing game $G = (\mathcal{K}, \mathcal{A}_k, u_k)$ if, for every $k \in \mathcal{K}$, $u_k(a_k, \boldsymbol{a}_{-k}) \geq u_k(a'_k, \boldsymbol{a}_{-k})$ for all $a'_k \in \mathcal{A}_k$.

In essence, at a Nash equilibrium no user has an incentive to unilaterally change its own strategy when all other users keep their strategies fixed. Hence, the Nash equilibrium can be viewed as a stable outcome where a game might end up when non-cooperative users adjust their strategies according to their self-interests. In fact, the best response correspondence of a user gives the best reaction strategy a rational user would choose in order to maximize its own utility, in response to the actions chosen by other users. That is, the user k's best response $r_k : \mathcal{A}_{-k} \longrightarrow \mathcal{A}_k$ is the set

$$r_k(\boldsymbol{a}_{-k}) = \left\{ a_k \in \mathcal{A}_k : u_k(a_k, \boldsymbol{a}_{-k}) \geq u_k(a'_k, \boldsymbol{a}_{-k}) \text{ for all } a'_k \in \mathcal{A}_k \right\}. \qquad (5.31)$$

If we define the total interference from all secondary users to the primary user, excluding that from the k-th user signal as $I_{0,-k} \doteq I_0 - |\tilde{h}_{pk}|^2 p_k$, where $I_{0,-k}$ is now independent of p_k we have the following result.

Theorem 5.1. *A Nash equilibrium exists in game* $G = (\mathcal{K}, \mathcal{A}_k, u_k)$ *if*

1. $g(0) = 0$, $g'(0) > 0$ *and* $\lim_{\gamma \to \infty} \frac{g(\gamma)}{g'(\gamma)} > -\infty$

2. $\frac{g(\gamma)g''(\gamma)}{(g'(\gamma))^2} < 2$ *for all* $\gamma > 0$

3. $0 < \lambda_s \leq \frac{Q_0}{I_{0,-k}}$ *for* $k = 1, 2, \ldots, K$

Proof. See Appendix 5.A. $\square$

Clearly, the above DSL game model is general enough to allow for various secondary reward functions $g(\cdot)$ that may satisfy above conditions. We have seen that choosing the most suitable secondary user performance metric and the associated

reward function in a cognitive radio network can itself be a non-trivial task. For illustrative purposes, we consider in the following two specific reward functions:

$$g_k^{(1)}(\gamma_k) = W_k \log(1 + \gamma_k), \tag{5.32}$$

$$g_k^{(2)}(\gamma_k) = R_k \frac{C_{\texttt{BSC}}(P_e(\gamma_k))}{p_k} \tag{5.33}$$

where W_k and R_k are the bandwidth and data rate of user k, respectively, $P_e(\gamma_k)$ is the probability of bit error with received SINR of γ_k and $C_{\texttt{BSC}}(P_e)$ is the capacity of a binary symmetric channel with cross-over probability P_e which can be written in terms of the binary entropy function $H(P_e) = -P_e \log_2 P_e - (1 - P_e) \log_2(1 - P_e)$ as $C_{\texttt{BSC}}(P_e) = 1 - H(P_e)$.

The reward function $g^{(1)}(\cdot)$ is a measure of user k's capacity in the presence of all other users, while $g^{(2)}(\cdot)$ is a measure of its throughput per unit power. Both these reward functions can be justified in a wide variety of contexts. For example, the reward function $g^{(1)}(\cdot)$ can be justified in a dynamic spectrum leasing application in which the secondary users are mainly concerned with getting access to the spectrum and their power consumption is not a major concern. On the other hand, $g^{(2)}(\cdot)$ is suitable when secondary users are interested in achieving best throughput per unit energy spent. For concreteness, we will assume that $P_e(\gamma_k) = \frac{1}{2} \exp(-\gamma_k)$ (i.e. BPSK modulation with a matched-filter receiver).

5.4.3 Best response adaptations and implementation issues

Primary user. Since the best response by a player in a game is a strategy that maximizes its own utility given all other players actions, the best response of the primary user in the above DSL game is obtained by setting $u_0'(Q_0) = 0$. The unique interior solution is given by

$$Q_0^*(I_0) = \frac{\bar{Q}_0 + I_0}{2}. \tag{5.34}$$

Note that, since $u_0(Q_0)$ is monotonic increasing for $Q_0 < Q_0^*$, if the maximum interference cap is such that $\bar{Q}_0 < Q_0^*$, the best response of the primary user would be to set the interference cap to $Q_0 = \bar{Q}_0$. Hence, the primary user's best response

is given by

$$r_0\left(\boldsymbol{a}_{-0}\right) = r_0\left(I_0\right) \tag{5.35}$$

$$= \min\left\{\bar{Q}_0, Q_0^*(I_0)\right\}. \tag{5.36}$$

We observe that in order to determine its best response for a chosen action vector $\boldsymbol{a}_{-0}$ by the secondary users, the only quantity that the primary user needs to know is the total secondary interference at the primary receiver I_0. Indeed, this total interference can be estimated at the primary receiver without much difficulty.

Secondary user. On the other hand, the best response of the k-th secondary user to the transmit powers of the other secondary users as well as interference cap set by the primary user is given by the solution to $u_k'(p_k) = 0$ that we denote here $p_k^*(\boldsymbol{a}_{-k})$.

Since u_k is quasi-concave in p_k, if $p_k^*(\boldsymbol{a}_{-k}) > \bar{p}_k$ where $\bar{p}_k$ is the k-th user's maximum possible transmit power, its best response is to set its transmit power to $p_k = \bar{p}_k$. Hence, we have the best response of k-th secondary user, for $k = 1, 2, \ldots, K$:

$$r_k\left(\boldsymbol{a}_{-k}\right) = \min\left\{\bar{p}_k, p_k^*(\boldsymbol{a}_{-k})\right\}. \tag{5.37}$$

It can be shown that $p_k^*(\boldsymbol{a}_{-k}) = p_k^*(Q_0, I_{0,-k}, I_k)$, that is, the best response of the k-th secondary user is a function of the primary interference cap Q_0, the residual interference $I_{0,-k}$ from all other secondary users to the primary user, and the total interference from all secondary and primary users to the k-th user's received signal at the secondary receiver I_k. The secondary system can of course estimate the latter quantity.

To obtain the knowledge of Q_0 and $I_{0,-k}$ we assume that the primary system periodically broadcasts Q_0 and I_0. Note that this is the only interaction that the primary system will need to keep with the secondary system. Since these two quantities are readily available to the primary system, we believe that the periodic broadcast of these quantities, informing the secondary system what it needs to know in order to avoid severe conflicts with primary transmissions, is a reasonable expectation for a future cognitive radio system that expects to harvest spectrum leasing gains. Observe that by knowing I_0, each secondary user can compute the residual interference $I_{0,-k} = I_0 - |\tilde{h}_{0k}|^2 p_k$ given that it can estimate the channel state information $|\tilde{h}_{0k}|^2$ to the primary. If reverse link signals are available in the same band this may be feasible. Otherwise the secondary receiver does not necessarily need the CSI of its

link with the primary receiver, as it will be shown in the numerical results section, since the approximation $I_{0,-k} \approx I_0$ performs well in practice, especially when the number of secondary users K is sufficiently large.

5.4.4 Performance analysis

In the following we consider a dynamic spectrum leasing cognitive radio system that fits into the proposed game-theoretic framework. Our goal is to investigate the behavior of the primary and secondary systems at the equilibrium. It is to be noted that the Nash equilibrium can reasonably be expected to be the natural outcome of the system when it reaches steady-state. Thus, the performance of the system is to be considered as its performance at the Nash equilibrium.

Now, to illustrate the characteristics of the Nash equilibrium in this primary-secondary user dynamic spectrum leasing game, we first consider a simplified scenario with identical secondary users. This scenario allows to analytically determine the Nash equilibrium state and its general behavior. We analyze next a more general scenario with non-identical secondary users and fading channels by means of simulations.

Stationary system with identical secondary users

When all secondary users present the same channel to the secondary receiver it is possible to characterize the best response correspondences of primary and secondary users to graphically visualize the Nash equilibrium. If $|\tilde{h}_{sk}|^2 = |\tilde{h}_s|^2$ and $|\tilde{h}_{pk}|^2 = |\tilde{h}_p|^2$ for all k. By symmetry, in this case all secondary users must have the same power $p_k = p^*$ at the Nash equilibrium (equivalently, the same SINR $\gamma_k = \gamma^*$). Thus the Nash equilibrium is characterized by the intersection (Q_0^*, p^*) of the following two curves:

$$Q_0 = r_0(p) = \frac{\bar{Q}_0 + K|\tilde{h}_p|^2 p}{2} \tag{5.38}$$

$$p = \text{(solution to equation } \psi_{Q_0}(p) = 0) \doteq r_s(Q_0) \tag{5.39}$$

where

$$\psi_{Q_0}(p) \doteq Kp + \frac{I_k + \sigma_s^2}{|\tilde{h}_s|^2} \frac{g(\gamma_1(p))}{g'(\gamma_1(p))} - \frac{Q_0}{\lambda_s |\tilde{h}_p|^2}. \tag{5.40}$$
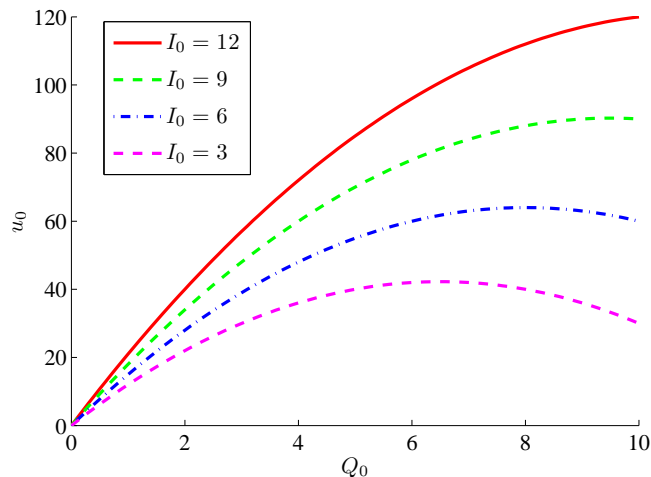
Figure 5.3: Primary user utility u_0 for a fixed secondary interference I_0 in a single-user secondary system.

Combining (5.38) and (5.39), the Nash power p^* of the secondary users is given by the solution to the equation

$$K\left(1 - \frac{1}{2\lambda_s}\right)p + \frac{I_k + \sigma_s^2}{|\tilde{h}_s|^2}\frac{g\left(\gamma_1(p)\right)}{g'\left(\gamma_1(p)\right)} - \frac{\bar{Q}_0}{2|\tilde{h}_p|^2\lambda_s} = 0. \qquad (5.41)$$

Figure 5.3 shows the primary utility function for fixed secondary network actions in a single secondary user system, that is $K = 1$, assuming that $\bar{Q} = Q_{\max} = 10$, $\bar{p}_1 = 12$, $W_1 = 1$, $\lambda_s = 1$, $\bar{\gamma}_0 = 1$, $\tilde{h}_{p1} = \tilde{h}_{p0} = \tilde{h}_{s0} = \tilde{h}_{s1} = 1$ and $\sigma_s^2 = \sigma_p^2 = 1$.

On the other hand, for the setup described, secondary utility and best response depends on the considered reward function $g(\gamma)$. First Figs. 5.4(a) and 5.4(b) assume the secondary reward function $g(\gamma) = g^{(1)}(\gamma) = \log(1+\gamma)$. In Fig. 5.4(a) we can see the concavity of the secondary utility function for fixed primary response, and thus the existence of a best response. The primary and secondary best response curves $Q_0 = r_0(p_1)$ and $p_1 = r_1(Q_0)$ for the setup described are presented in Fig. 5.4(b). Notice that the intersection of these two best response curves specifies the Nash equilibrium for this system: $(Q^*, p_1^*) = (6.505, 3.010)$.

Similarly, Figs. 5.5(a) and (b) show the secondary user utility for a fixed primary interference cap and the best response functions, respectively, when the secondary utility function is chosen to be $g(\gamma) = g^{(2)}(\gamma) = R_1\frac{C_{\text{BSC}}(P_e(\gamma))}{p}$ with $R_1 = 1$ and all other parameters as in the previous scenario. From 5.5(a) we observe that
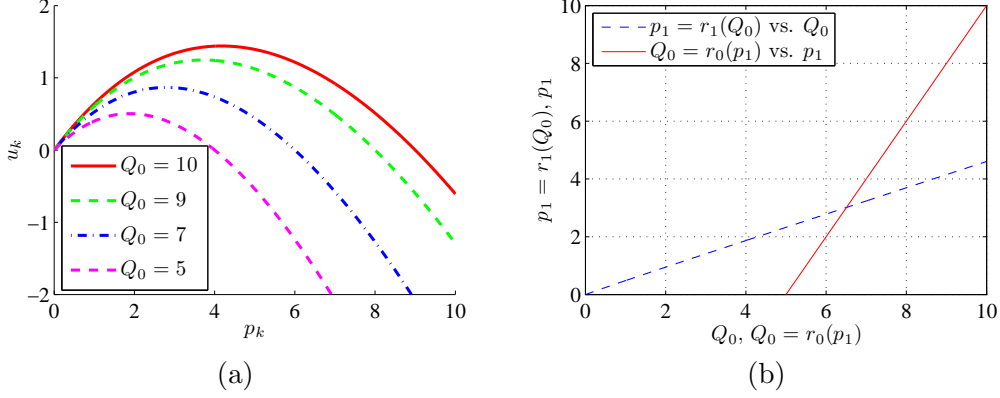
Figure 5.4: System behavior for identical secondary users when $g(\gamma) = g^{(1)}(\gamma)$. (a) Secondary utility. (b) Best-response functions.

the secondary utility function is still concave in secondary power. The best response curves in Fig. 5.5(b) are characterized by (5.39) and (5.38) where, now, $g(\gamma) = g^{(2)}(\gamma)$. Figure 5.5(b) shows that the Nash equilibrium in this system is $(Q^*, p_1^*) = (6.325, 2.650)$. Note that this NE shows that due to the penalty for increasing transmit power in the secondary system, the secondary user now settles for a slightly lower transmit power level compared to the earlier situation in which it was not concerned with power expenditure. As a result, the primary user is also better off by slightly lowering its interference cap so that it keeps the demand high. With this in mind, and for ease of explanation we will restrict our discussion to the secondary utility $g(\gamma) = g^{(1)}(\gamma)$. Extension to $g(\gamma) = g^{(2)}(\gamma)$ is straightforward and it is left to the curious reader.

It is of interest to investigate the equilibrium behavior of this dynamic spectrum leasing system as a function of the secondary system size K. In Fig. 5.6 we show the allowed interference cap Q_0 and the actual secondary interference I_0 at the system equilibrium for a system such that $\bar{Q} = 10$, $\bar{p}_k = 10$, $W_k = 1$, $\bar{\gamma}_0 = 1$, $\tilde{h}_{pk} = \tilde{h}_{sk} = 1$ for all k, and $\sigma_s^2 = \sigma_p^2 = 1$. From Fig. 5.6 we can observe how the total interference I_0 increases with increasing K, and how, in turn, the primary user also increases its interference cap to maximize its utility. It is also of interest to note that the safety margin $Q_0 - I_0$ is large for smaller number of users, and seems to monotonically decrease with increasing K. This, we believe, is essentially due to the fact that the number of degrees of freedom in a multiuser system is being proportional to the number of users. When the number of secondary users K is large, the interference generated by the secondary system I_0 is close to the interference cap
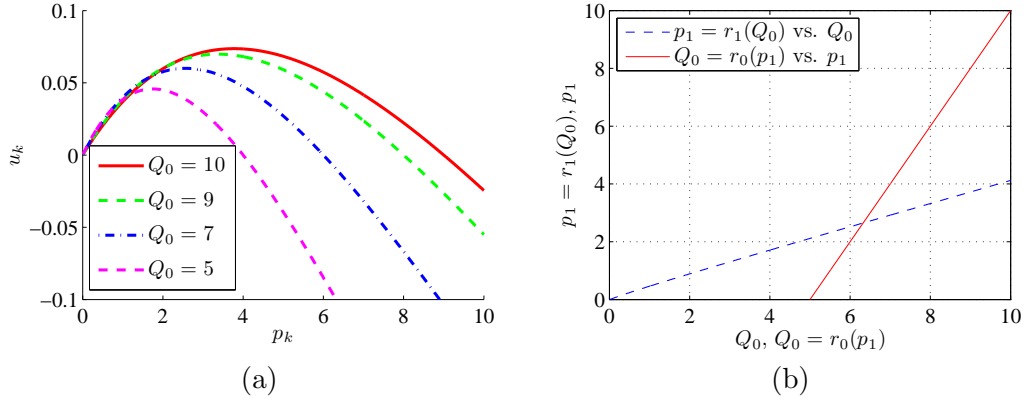
Figure 5.5: System behavior for identical secondary users when $g(\gamma) = g^{(2)}(\gamma)$. (a) Secondary utility. (b) Best-response functions.

Q_0, yet, as desired, is always below it. Figure 5.6 shows the game outcomes when exact channel state information for the primary system is available at each secondary user (via estimation) so the exact $I_{0,-k}$ is used in its best response adaption, as well as when this channel state information to the primary is not available, so that the secondary user employs the approximation $I_{0,-k} \approx I_0$. As we may observe from Fig. 5.6, the system that does not rely on the knowledge of channel state information demonstrates the same performance trends at the equilibrium. In particular, still the DSL game converges to a Nash equilibrium that does not violate the primary interference cap. It seems that the only effect of not having the exact $I_{0,-k}$ is that the safety margin $Q_0 - I_0$ at the equilibrium is slightly larger. This is essentially due to the fact that each secondary user believes an exaggerated residual interference $I_{0,-k}$ making it to decrease its power.

Figure 5.7 shows the primary and secondary utilities at the Nash equilibrium of the system considered in Fig. 5.6 as a function of the secondary system size. In Fig. 5.7 we show the utilities achieved when exact channel state information for the primary system is available at each secondary user (via estimation) so the exact $I_{0,-k}$ is used in its best response adaptation, as well as when this channel state information to the primary is not available, so that the secondary user employs the approximation $I_{0,-k} \approx I_0$. In particular, as seen by Fig. 5.7(a) the primary utility u_0^* at the Nash equilibrium typically increases with the number of secondary users K. However, the rate of increase decreases with increasing K. Thus, from a design point of view we may argue that the primary user might prefer the system to operate at a point where its rate of utility increase is above a certain threshold
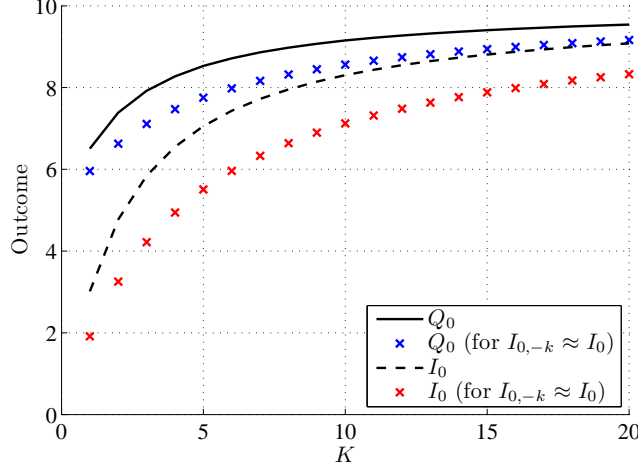
Figure 5.6: Game outcome assuming identical secondary users for a quasi-static scenario versus the number of secondary users K.

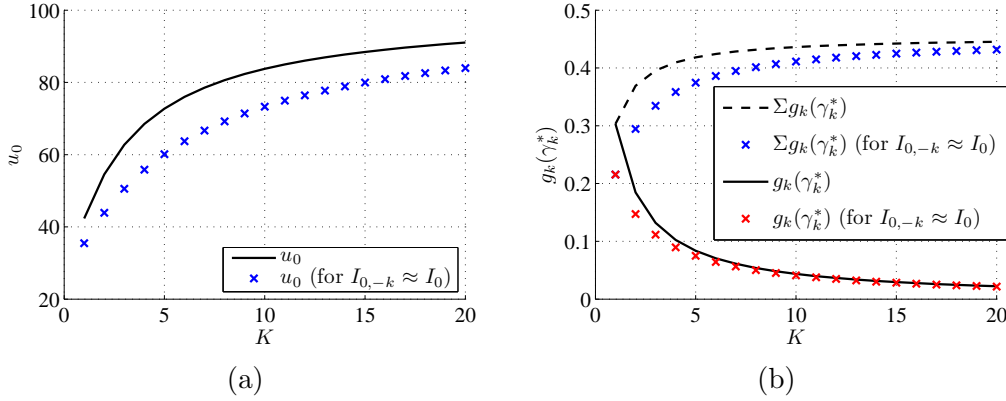(a)   (b)

Figure 5.7: System performance of the DSL game in quasi-static environments assuming identical secondary users. (a) Primary user utility. (b) Secondary user reward function.

value. However, the primary system cannot impose this explicitly on the secondary system and indeed it is not a requirement. The only requirement is that $I_0 \leq Q_0$. However, as we see next from Fig. 5.7(a) the secondary system has the incentive to keep K not too high. It is also observed from Fig. 5.7(a) that the equilibrium utility of the primary user is decreased when exact channel state information is not available at the secondary users.

Figure 5.7(b) shows both the sum-rate $\sum_{k=1}^{K} g_k(\gamma_k^*)$ as well as the per-user rate

$\frac{1}{K} \sum_{k=1}^{K} g_k(\gamma_k^*)$ achieved by the secondary system, with and without exact channel state information. As was the case with primary utility, the secondary utilities are also reduced slightly in the absence of channel state information. However, as we observe from Fig. 5.7(b), this performance degradation seems to be small when the secondary system size is sufficiently large. Note that, from a system point of view the secondary system would prefer to maximize the sum-rate. As we see from Fig. 5.7(b), the sum-rate monotonically increases with K both with and without CSI. Thus, at a first glance, allowing more secondary users to operate simultaneously seems to be the preferred solution. However, Fig. 5.7(b) also shows that the per-user rate is monotonically decreasing in K, leading to decreasing incremental gains in sum-rate as additional secondary users are added to the system. Depending on the application and the QoS requirement of the secondary system, each secondary user will have a minimum required rate (in bits per transmission) below which the transmissions would be useless. Thus we note that this QoS requirement will determine the maximum number of secondary users K the secondary system would want to support at any given time. For example, if the minimum per-user rate required is 0.1 bps, the optimal K would be $K^* = 4$, assuming exact CSI. If, on the other hand, the rate threshold was reduced to 0.025 bps, the secondary system may allow up to $K = 18$ secondary users to simultaneously operate.

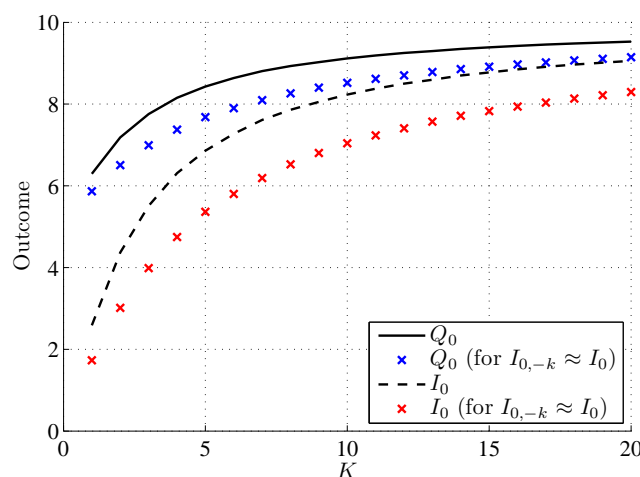### DSL network under quasi-static fading channels

Figure 5.8: Game outcome in the presence for a quasi-static scenario versus the number of secondary users K.
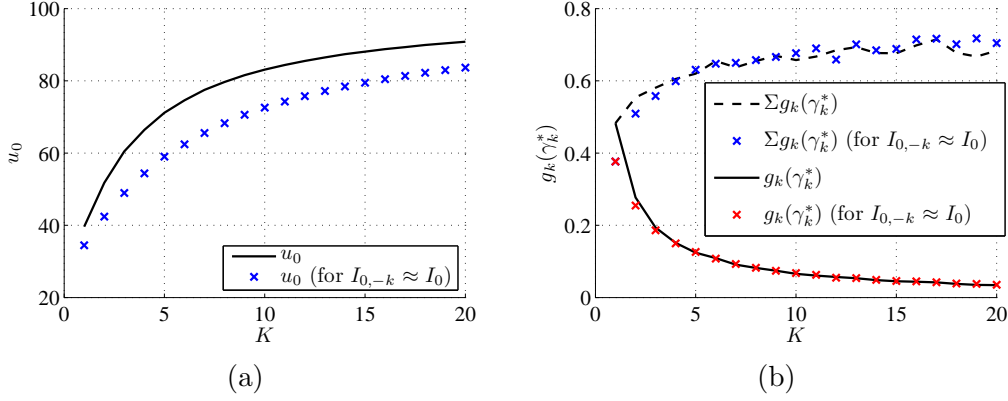
Figure 5.9: System performance of the DSL game in quasi-static environments. (a) Primary user utility. (b) Secondary user reward function.

In the presence of wireless channel fading, the Nash equilibrium power profile of the dynamic spectrum leasing system will depend on the observed channel state realization. In particular, it is expected that in this case the Nash equilibrium transmit powers of individual secondary users will be different for each user. In Fig. 5.8 we have shown the game outcome at the Nash equilibrium in the presence of channel fading as a function of the number of secondary users K, both with and without CSI (when there is no channel state information, again, we use the approximation $I_{0,-k} \approx I_0$). Figures 5.9(a) and 5.9(b) show the corresponding primary and secondary user utilities achieved at the Nash equilibrium in the presence of channel fading. In obtaining Fig. 5.8 and Fig. 5.9 we have assumed all channel gains in the system to be Rayleigh distributed with all channel coefficients normalized so that $\mathrm{E}[h^2] = 1$. The remaining parameters are assumed $\bar{Q} = 10$, $\bar{p}_k = 10$, $W_k = 1$, $\bar{\gamma}_0 = 1$, and $\sigma_s^2 = \sigma_p^2 = 1$ as in the previous section. This essentially allows us to consider, without any loss in generality, the transmit powers p_k to be equal to the average received power (averaged over fading). Note that, due to interference averaging in the presence of fading, in this case the secondary system is able to achieve better sum- and per-user rates compared to those with non-fading channels.

Note that, when the reward function $f = f^{(1)}$, the reward for a secondary user is the capacity (in bps) it can achieve assuming all other transmissions (both primary and secondary) are purely noise. In the presence of channel fading, this capacity is a random quantity determined by the fading coefficients of all users. As we saw earlier with identical users, the per-user reward is typically a decreasing function of the increasing secondary system size. The interpretation is simple: Essentially, all

secondary users in the system must share the allowed interference level set by the primary system. As we mentioned earlier, a secondary user may require a minimum capacity to ensure at least an acceptable QoS for its applications.

DSL network under time-varying environmental conditions

In the above discussion, we have assumed the quasi-static fading in which fading realizations stay fixed for a period of duration and then change to new values. This facilitated the Nash equilibrium analysis without having to deal with time-varying channel coefficients. While quasi-static assumption may be justified in certain channel environments, sometimes it is likely that the channel coefficients may slowly vary in time. It is easy to see that for the best-response adaptations to converge to a Nash equilibrium, the rate of adaptations need to be faster than the time-variations of the channel. One may expect that in the presence of channel variations, the convergence may be slowed, or even not occur. However, as we will demonstrate in this section, the proposed DSL-game has the desired property of being tolerant towards slow time-variations of the channel state. Moreover, the Nash equilibrium of the proposed DSL-game is robust against small channel estimation errors. This is also a desired property since in practice the channel coefficients need to be estimated, and these estimations are almost always not perfect.

In this section we investigate the behavior of the proposed DSL-game based dynamic spectrum sharing networks. We first study the effect of both channel variation rate ϵ and CSI updating interval L on the network performance and how they affect the probability of the secondary system meeting the target interference cap Q_0. Secondly we compare a quasi-static scenario where the system has enough time to converge to its Nash equilibrium with a more realistic time-varying scenario for a different number of secondary users K.

Channel Model. Channel coefficients are assumed Rayleigh distributed and independent across users, that is,

$$h_{.k}[n] \sim \mathcal{CN}\left(0, \sigma_{h_{.k}}^2\right), \tag{5.42}$$

where with some abuse of notation we use a dot in the subscript to denote either primary (p) or secondary (s) systems. We assume that channel coefficients present temporal correlations. Under the slow varying channel assumption we model temporal correlation as a first order Gauss-Markov process (Maybeck, 1979) at sampling

rate:

$$h_{.k}[n] = \sqrt{1 - \epsilon^2} h_{.k}[n-1] + \epsilon w_{.k}[n], \qquad (5.43)$$

where the driving noise $w_{.k}$ follows an iid $\mathcal{CN}\left(0, \sigma^2_{h_{.k}}\right)$, and ϵ, defined here as channel variation rate, is a parameter related to the normalized Doppler spread. First order Markovian assumption for fading correlations has been shown to be accurate both experimentally and analytically (Wang and Chang, 1996).

Additionally we assume that the channel state information (CSI) update at the system is not instantaneous, and the system obtains CSI periodically every L samples. Then system decisions are taken as a function of the last CSI estimate

$$\hat{h}_{.k}[n] \doteq h_{.k}[L\lfloor n/L \rfloor], \qquad (5.44)$$

where $\lfloor \cdot \rfloor$ stands for the operand integer part of the argument.

For each set of parameters the results are averaged over 1000 Monte Carlo repetitions. We divide each realization into a transient period (100 best-response iterations), where the system evolves to its stationary regime and the period in which the results are averaged (500 best-response iterations). Unless specified otherwise, simulation parameters were set to $\bar{\gamma}_0 = 1$, $\sigma^2_p = \sigma^2_s = 1$, $\sigma^2_{\tilde{h}_{pk}} = \sigma^2_{\tilde{h}_{sk}} = 1$ for $k = 0, \ldots, K$. The maximum allowed interference cap was fixed to $\bar{Q}_0 = 10$ while the maximum transmission power of each of the secondary transmitters was set to $\bar{P}_k = 10$ for $k = 1, \ldots, K$. The weighting coefficient was set to $\lambda_s = 1$. As we will see below we can compensate for both channel variations and channel state information inaccuracies by changing the value of the parameter λ_s.

Effect of channel variation rate

Here we fix the size of the secondary system ($K = 10$) and the channel estimation period ($L = 10$), and we study the effect of the channel variation rate ϵ. To this end we vary its value from a static channel $\epsilon = 0$ to a (moderately fast) varying channel $\epsilon = 0.2$.

Fig. 5.10(a) shows the averaged game outcome for different values of the channel variation rate ϵ. This figure must be seen in conjunction with Fig. 5.10(b) that shows the probability of the secondary system outcome not meeting the allowed interference cap Q_0 set by primary user. We compare here two values of the weighting coefficient λ_s.
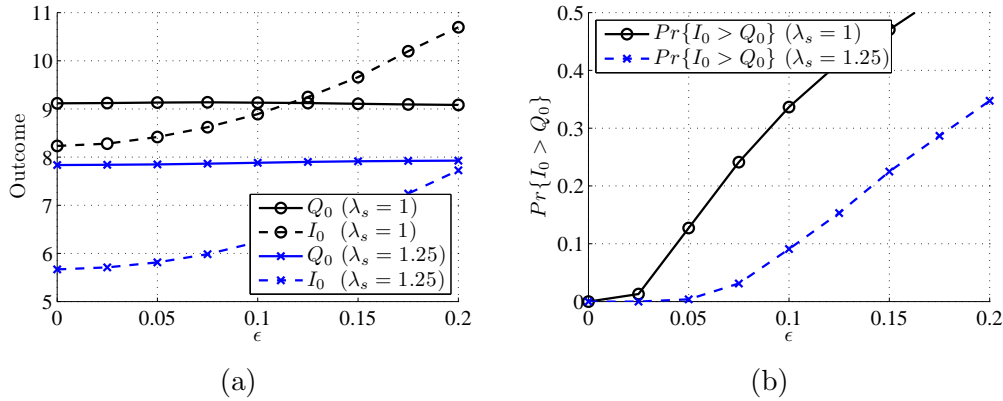
Figure 5.10: Outcome averaged over the fading for two different values of the weighting coefficient λ_s versus a growing channel variation rate ϵ. (a) Game outcome. (b) Probability of undesired operation.

In Fig. 5.10(a) we can see that for both $\lambda_s = 1$ and $\lambda_s = 1.25$ an increase in the rate of variation does not affect the average interference cap Q_0 at the system equilibrium. However the actual interference I_0 at the primary receiver increases on average with ϵ. Fig. 5.10(b) shows that the probability of undesired operation defined as $\Pr\{I_0 > Q_0\}$ turns to be non-zero even for moderate values of ϵ. Moreover, above a given variation rate I_0 exceeds the allowed interference cap on average, as seen from Fig. 5.10(a). Note that by adjusting the weighting coefficient λ_s we can control the safety margin of the system. In Fig. 5.10(b) we can see that by increasing λ_s we shift the probability of undesired operation curve to the right, allowing the system to work satisfactorily even at higher channel variation rates. This increased tolerance to time-variations, however, comes at a price. In our case, it is the smaller equilibrium interference cap chosen by the primary user, which in turn leads to reduced both primary utility and secondary system sum-reward.

Effect of channel estimation period

We set the channel variation rate $\epsilon = 0.1$ for the same secondary system size ($K = 10$). We vary the channel estimation period L from 1 (each best-response iteration carried out with perfect CSI) to $L = 25$, that is, one estimate every 25 best-response round-robin iterations.

In Fig. 5.11 we can see that the effect of the channel estimation period L resembles the channel variation rate effect. For longer channel estimation periods the actual interference at the primary increases although the allowed interference
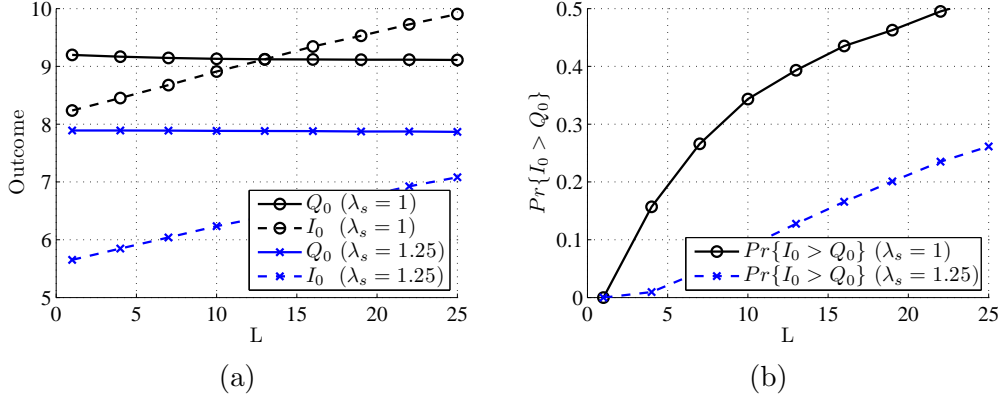
Figure 5.11: Outcome averaged over the fading for two different values of the weighting coefficient λ_s versus a varying channel estimation period L. (a) Game outcome. (b) Probability of undesired operation.

cap remains constant even in the presence of outdated channel information. In Fig. 5.11(b) we can see that by increasing the weighting coefficient λ_s the probability of undesired operation curve gets shifted allowing a larger operating range.

Effect of the number of users

Here we compare the effect of the secondary system size in a quasi-static scenario, corresponding to $\epsilon = 0$, where the system converges to its Nash equilibrium and the channel is known perfectly, with a time-varying situation corresponding to a channel variation rate of $\epsilon = 0.1$ and channel estimation period of $L = 10$ best response adaptations.

Figure 5.12 shows the game outcome versus the secondary system size. We see that in a quasi-static scenario the actual interference I_0 at the primary meets the allowed interference cap Q_0 even for a moderate number of secondary users. For a growing number of secondary users the safety margin $Q_0 - I_0$ decreases. On a time-varying scenario with outdated CSI the safety margin $Q_0 - I_0$ is reduced due to incomplete adaptation to the actual environment that prevents the convergence of the network to its desired Nash equilibrium, as we have seen in the previous results. This effect becomes more pronounced for a growing number of secondary users because of the required longer convergence time of a larger network.

In Figures 5.13(a) and (b) we show the utility and the reward function of the primary and secondary systems, respectively. In the case of time-varying channels the primary utility is larger than for stationary channels, as seen from Fig. 5.13(a).
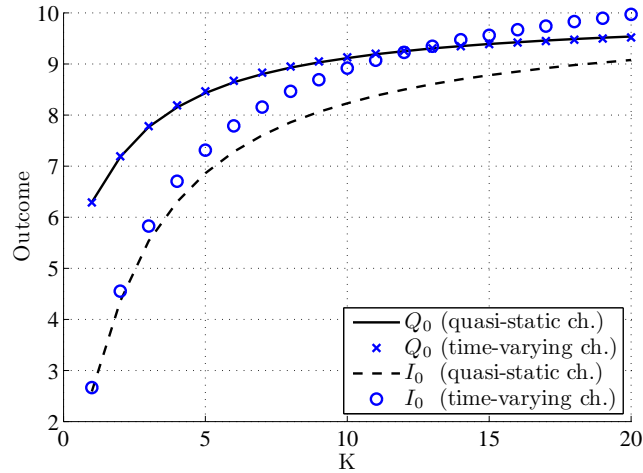
Figure 5.12: Game outcomes averaged over fading for a quasi-static scenario and for time-varying scenario for a varying number of secondary users K.
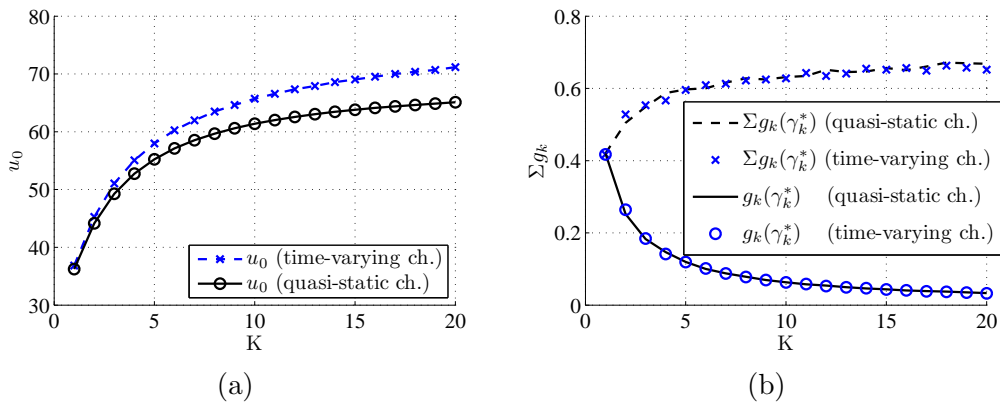

Figure 5.13: System performance averaged over fading for a quasi-static scenario and for time-varying scenario. (a) Primary user utility. (c) Secondary reward function.

This effect comes from the fact that the actual interference seen at the primary is larger than for stationary channels. On the other hand, Fig. 5.13(b) shows that the reward function for the secondary users remains almost unaltered for time-varying channels, even when the interference at primary has been increased. This is an undesirable effect since an increase in transmitted power by secondary users does not translate into a larger reward for the secondary system. The reason is that the increased transmitted power hinder both primary and other secondary users due to the suboptimality of matched filter decoding.

5.5 Conclusions

A formal analysis of Cognitive Radio paradigms which allow different level of awareness between primary and secondary networks shows the potential gain of the interaction between the two heterogeneous systems.

Specifically, in this chapter we have studied the expected utility improvement of a primary system which dynamically adjusts the level of allowed interference. The proposed analysis results into a Stackelberg game formulation of the interactions between primary and secondary systems, which is not implementable in practice. Instead, we propose a general game theoretical framework which results in a family of practical (though suboptimal) DSL communication schemes. The stability and performance of this system has been analyzed under different scenarios showing a good behavior in realistic environments.

The results presented in this chapter are the result of the collaboration between the Signal Processing in Communications Group (GPSC, University of Vigo) and the Communications and Information Sciences Lab (CISL, University of New Mexico). This collaboration conducted to three publications in the topic of DSL based schemes. First, Vazquez-Vilar et al. (2010b), published in the IEEE-TWC, studies the potential gain of a family of DSL schemes and covers Section 5.3. On the other hand, the content in Section 5.4 is mainly based on Jayaweera et al. (2010), published in IEEE-TVT and was extended to time varying environments in El-howayek et al. (2010), presented at the 1st International Workshop on Cognitive Radio Interfaces and Signal Processing (CRISP 2010).

Appendix 5.A   Proof of Theorem 5.1

Note that the primary user action set is of the form of $\mathcal{A}_0 = \mathcal{Q} = [0, \bar{Q}_0]$, where $\bar{Q}_0$ is the maximum interference cap that is determined by the required minimum QoS and the maximum possible transmit power of primary user. Clearly $\mathcal{A}_0$ is both compact and convex. Similarly, for all $k = 1, \cdots, K$, the secondary user strategy sets are of the form of $\mathcal{A}_k = \mathcal{P}_k = [0, \bar{p}_k]$. Again, it is easy to observe that all secondary user action sets are convex and compact (being closed and bounded real intervals). Further, both $u_0(\boldsymbol{a})$ and $u_k(\boldsymbol{a})$ are continuous in the action vector $\boldsymbol{a}$, and u_0 is concave in Q_0. For the existence of a Nash equilibrium we need to additionally ensure the quasi-concavity of u_k's in p_k for $p_k \geq 0$, for $k = 1, \cdots, K$.

Let us define

$$\phi_k(\gamma_k) \doteq \frac{I_{0,-k}}{Q_0} + \frac{|\tilde{h}_{pk}|^2 \sigma_s^2}{Q_0} \left(\gamma_k + \frac{g(\gamma_k)}{g'(\gamma_k)} \right). \tag{5.45}$$

Then, it can be seen that u_k has a local maximum that is indeed a global maximum if $\phi_k(\gamma_k) = \frac{1}{\lambda_s}$ has only one solution for $p_k \in \mathcal{P}_k$. Clearly, $\phi_k(\gamma_k) = \frac{1}{\lambda_s}$ has a solution if $\phi_k(0) \leq \frac{1}{\lambda_s} < \lim_{\gamma_k \to \infty} \phi_k(\gamma_k)$, and, moreover, this solution is indeed a global maximum if in addition $\phi_k'(\gamma_k) > 0$ for $\gamma_k > 0$. It can be easily verified that $\phi_k'(\gamma_k) > 0$ will be true if the reward function is such that $\frac{g(\gamma_k) g''(\gamma_k)}{(g'(\gamma_k))^2} < 2$ for all $\gamma_k > 0$. Note that, this is trivially true for any reward function that is concave in γ_k since in that case $g''(\gamma_k) \leq 0$. Note also that $\phi_k(0) = \frac{I_{0,-k}}{Q_0}$ and $\lim_{\gamma_k \to \infty} \phi_k(\gamma_k) = \infty$ if $\lim_{\gamma_k \to \infty} \frac{g(\gamma_k)}{g'(\gamma_k)} > -\infty$. Hence, if reward function $g(\gamma_k)$ and the coefficient λ_s satisfy the following conditions, u_k indeed has a local maximum that is a global maximum:

1. $g(0) = 0$, $g'(0) > 0$ and $\lim_{\gamma_k \to \infty} \frac{g(\gamma_k)}{g'(\gamma_k)} > -\infty$

2. $\frac{g(\gamma_k) g''(\gamma_k)}{(g'(\gamma_k))^2} < 2$ for all $\gamma_k > 0$

3. $0 < \lambda_s \leq \frac{Q_0}{I_{0,-k}}$

Chapter 6

Conclusions

Contents

This thesis is framed within the field of cognitive radio, a smarter communications paradigm in which radios may learn and adapt to the environment. While this novel scheme promises a better strectrum utilization by allowing dynamic access in certain licensed bands, there exists a series of challenges which need to be addressed before the technology is mature enough for its deployment.

In the first chapters of this thesis we presented different detection schemes exploiting the available information about the primary network. In particular, in Chapters 2 and 3 we presented different multi-antenna detection schemes exploiting the spatial structure of the received signal together with the available information about its spectral shape. In Chapter 4 we addressed the issue of acquisition and detection of wideband signals, both when the sampling is performed at Nyquist rate and when the acquisition is done through novel analog to information converters. Wideband processing is shown to provide additional information that can be used to increase the detection performance.

As a whole, this thesis provides a set of tools that, depending on the known parameters on the primary network and the architecture of the detection system, allows the system designer to construct efficient detectors. For each of the proposed detection schemes we developed a rigurous performance analysis, either analytically when mathematically tractable, or through extensive Monte Carlo simulations otherwise.

The second part of the thesis focuses on the study of a general architecture

for interference management in cognitive radio networks. In this sense, Chapter 5 discusses the advantages of allowing some interaction between the primary system and secondary system, concluding that this advantage can be significant in dynamic environments. Given the current rising trend in the use of mobile devices, the analysis of systems in dynamic environments will have a big impact in the design of the next generation communication standards.

6.1 Future work

The work developed within the framework of this thesis opens multiple lines of research. Here I will present some of the most interesting yet unanswered questions.

- In Chapter 2 we proposed a low SNR diversity analysis for different detection schemes which assume known noise variance at the receiver. Extending this analysis to more sophisticated detectors, for example the ones presented in Chapter 3, is an open problem.

  In fact, this question can be reformulated in a broader scope. Assume that we have a given family of detectors, parametrized by the set of unknown system parameters. If their performance depends on a subset of these parameters, which may suffer from fading, the detection performance is a random variable which needs to be averaged over the parameter realizations. How is this average performance affected by the presence of additional unknown parameters?

- The detection schemes derived in Chapter 3 allow the detection of rank-P signals in spatially uncorrelated noise, and the detection of rank-1 signals in presence of noise with spatial correlation. However, the approach used to derive the latter cannot be directly applied to solve the most general problem of detection of spatially rank-P signals in spatially correlated noise. This detector would have important implications in certain practical scenarios.

- All of the detectors considered in Chapter 3 assume knowledge of the signal rank P. While this may be reasonable in some contexts, for example if the space-time coding scheme used by primary transmitters is known, there are scenarios in which P is unknown, for example if it is related to the number of primary users simultaneously transmitting. Future research should consider estimation of P (Chiani and Win, 2010) and primary signal detection jointly.

- In Chapter 4 we investigated primary signal detection in a wideband setting. This conducted to the estimation of parameters which may present certain sparsity patterns. Exploiting this sparsity has been shown to be fundamental to increase detection peformance in the proposed scenario. An interesting extension would be to derive a general performance analysis of the gain obtained by exploiting sparsity information in general detection / estimation problems.

- The multichannel detection schemes proposed in Chapter 4 have been derived under the assumption of a single-antenna spectrum monitor. Exploiting both multiple antennas and wideband processing opens new interesting posibilities, which, compared to the case of exploiting both properties separately, may increase the detection performance.

- Chapter 5 presents a novel framework which allows certain interaction between primary and secondary networks assuming single-carrier systems. This framework can in principle be extended to the case of multi-carrier multi-antenna systems, which offer multiple degrees of freedom which can be used in the optimization procedure. Additionally, one may think in more complex scenarios in which different secondary networks coexist. This networks may either compete for the resources or collaborate to obtain a better utilization. These scenarios can be studied from a game theoretical point of view by either considering non-cooperative (as the one considered in Chapter 5) or cooperative games (which is left as a line of future research).

- Also in the context of cognitive radio other, somehow more philosophical, questions arise. The scholar community recently proposed a wide set of detection schemes that can be applied to cognitive radio networks. However, much less attention has been given to the integration of these detectors in the CR receiver. In this sense it is clear that in a system with limited resources, say, for instance, number of available dimensions, a trade-off exists between the sensing performance and the spectral efficiency of the network. If several of the available dimensions are used for spectral sensing, there are only a few left for transmission and a small efficiency is achieved. On the other hand, if few resources are allocated to the sensing procedure, the system may not be able to detect transmission opportunities and thus suffers again a reduced spectral efficiency. Somewhere between these two extremes a trade-off exists which maximizes the spectral efficiency obtained. In this thesis we presented a set of multiantenna detectors, and different analyses which allow us to quantify

their detection performance. This, together with novel results on finite length bounds (Tauste et al., 2011), can be used as an starting point to formalize the existing trade-off.

6.2 Concluding remarks

While in this thesis we have focused on the context of CR, several of the fundamental results obtained can be applied in other fields. For example, the GLRT detectors presented in Chapter 3 may have practical importance in radio astronomy or in target detection in the field of radar. The bargain scheme presented in Chapter 5 could be applied with minor changes, for instance, to femtocell interference management in cellular networks.

The work presented in this thesis has led to several articles published in international journals such as IEEE Transactions on Signal Processing, IEEE Transactions on Vehicular Technology, or the IEEE Transactions on Wireless Communications, while other articles are still in review process. Additionally, some preliminary results have been presented in some of the events with major impact in the field of signal processing, such as IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) conference, as well as in specific workshops in the field of Cognitive Radio, such as the ICST Conference on Cognitive Radio Oriented Wireless Networks (CrownCOM) or the IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC).

Some of the results this thesis were produced in collaboration with different research groups worldwide. Specifically, the results exposed in Section 3.3 have been obtained in collaboration with the Advanced Signal Processing Group (GTAS, University of Cantabria) and the theoretical results in Section 3.4 constitute a joint work with the Signal Processing for Communications Research Group (SPCOM, Technical Univ. of Catalonia UPC). The analysis of a new interference management paradigm presented in Chapter 5 has been developed in collaboration with the Communications and Information Sciences Lab (CISL, University of New Mexico). Additionally, a collaboration with Philips Research (Netherlands) has led to a joint conference paper on the topic of detection diversity in cognitive radio systems.

References

3GPP (2001). TS 45.004. Modulation. *3GPP*.

3GPP (2009). LTE Release 9. *3GPP*.

Akyildiz, I., Lee, W.-Y., Vuran, M., and Mohanty, S. (2008). A survey on spectrum management in cognitive radio networks. *IEEE Commun. Magazine*, 46(4):40–48.

Alamgir, M., Faulkner, M., Gao, J., and Conder, P. (2008). Signal detection for cognitive radio using multiple antennas. In *2008 IEEE Int. Symp. Wireless Commun. Systems*, pages 488–492.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis,. *Ann. Math. Statist.*, 34(1):122–148.

Baik, J. and Silverstein, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivar. Anal.*, 97:1382–1408.

Besson, O., Kraut, S., and Scharf, L. L. (2006). Detection of an unknown rank-one component in white noise. *IEEE Trans. Signal Process.*, 54(7):2835–2839.

Bezdek, J. C. and Hathaway, R. J. (2003). Convergence of alternating optimization. *Neural, Parallel Sci. Comput.*, 11:351–368.

Boonstra, A.-J. and Van der Veen, A.-J. (2003). Gain calibration methods for radio telescope arrays. *IEEE Trans. Signal Process.*, 51:25–38.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK.

Burg, J., Luenberger, D., and Wenger, D. (1982). Estimation of structured covariance matrices. *Proc. IEEE*, 70(9):963–974.

Cabric, D. (2008). Addressing the feasibility of cognitive radios. *IEEE Signal Proc. Mag.*, 25(6):85–93.

Candes, E. and Tao, T. (2007). The dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.*, 35(6):2313–2351.

Chiani, M. and Win, M. (2010). In *2nd Proc. Int. Work. Cognitive Inf. Process. (CIP)*, pages 156 –161.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory, 2nd Edition*. Wiley-Interscience, John Wiley and Sons, New York.

Daher, R. and Adve, R. (2010). A notion of diversity order in distributed radar networks. *IEEE Trans. Aerosp. Electron. Syst.*, 46(2):818–831.

Derakhtian, M., Tadaion, A., and Gazor, S. (2009). Detection of a bandlimited signal with unknown parameters. In *IEEE/SP Work. Stat. Signal Process. (SSP)*, pages 145–148.

Donoho, D. (2006). Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289 –1306.

Duan, D., Yang, L., and Principe, J. (2010). Cooperative diversity of spectrum sensing for cognitive radio networks. *IEEE Trans. Signal Process.*, 58(6):3218–3227.

Dugre, J.-P., Scharf, L., and Beex, A. (1980). A note on the measurement of spectral flatness and the calculation of prediction error variances. In *1980 IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, volume 5, pages 607–611.

El-howayek, G., Jayaweera, S., Hakim, K., Vazquez-Vilar, G., and Mosquera, C. (2010). Dynamic spectrum leasing (DSL) in dynamic channels. In *2010 IEEE Int. Conf. Commun. Workshops (ICC)*.

Etkin, R., Parekh, A., and Tse, D. (2007). Spectrum sharing for unlicenced bands. *IEEE Journ. Select. Areas Commun.*, 3(25):517–528.

ETSI (2004). Digital video broadcasting (DVB); framing structure, channel coding and modulation for terrestrial television (dvb-t). *ETSI EN 300 744 V1.5.1.*

ETSI (2009). Digital Video Broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2). *ETSI EN 302 755.*

Fattahi, A. R., Fu, F., Schaar, M. V. D., and Paganini, F. (2007). Mechanism-based resource allocation for multimedia transmission over spectrum agile wireless networks. *IEEE Journ. Select. Areas Commun.*, 3(25):601–612.

FCC (2002). Report of the spectrum efficiency working group. Technical report, FCC spectrum policy task force.

FCC (2003). FCC 03-322, notice of proposed rulemaking and order. ET Docket 03-322.

FCC (2008). FCC 08-260, second report and order and memorandum opinion and order, in the matter of unlicensed operation in the TV broadcast bands and additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band. ET Docket 08-260.

FCC (2010). FCC 10-174, second memorandum opinion and order, in the matter of unlicensed operation in the TV broadcast bands and additional spectrum for unlicensed devices below 900 MHz and in the 3 GHz band. ET Docket 10-174.

Fraikin, C., Nesterov, Y., and Dooren, P. V. (2008). Optimizing the coupling between two isometric projections of matrices. *SIAM J. Matrix Analysis App.*, 30(1):324–345.

Fugenberg, D. and Tirole, J. (1991). *Game Theory.* MIT Press.

Ganesan, G. and Li, Y. (2007a). Cooperative spectrum sensing in cognitive radio, part I: Two user networks. *IEEE Trans. Wireless Commun.*, 6(6):2204–2213.

Ganesan, G. and Li, Y. (2007b). Cooperative spectrum sensing in cognitive radio, part II: Multiuser networks. *IEEE Trans. Wireless Commun.*, 6(6):2214–2222.

Gray, R. M. (2006). *Toeplitz and circulant matrices: a review.* Hanover/Now, New York.

Gray Jr., A. and Markel, J. (1974). A spectral-flatness measure for studying the autocorrelation method of linear prediction of speech analysis. *IEEE Trans. Acoust. Speech Signal Process.*, 22(3):207–217.

Hayya, J., Armstrong, D., and Gressis, N. (1975). A Note on the Ratio of Two Normally Distributed Variables. *Management Science*, 21(11):1338–1341.

Hentilä, L., Kyösti, P., Käske, M., Narandzic, M., and Alatossava, M. (2007). Matlab implementation of the WINNER Phase II channel model v1.1. Online: https://www.ist-winner.org/phase_2_model.html.

Hwang, C.-H., Lai, G.-L., and Chen, S.-C. (2010). Spectrum sensing in wideband OFDM cognitive radios. *IEEE Trans. Signal Process.*, 58(2):709–719.

IEEE Computer Society (2009). IEEE Standard 802.11n-2009. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 5: Enhancements for Higher Throughput.

IEEE Computer Society and the IEEE Microwave Theory and Techniques Society (2009). IEEE Standard 802.16-2009. Part 16: Air Interface for Broadband Wireless Access Systems.

James, G. M., Radchenko, P., and Lv, J. (2009). Dasso: connections between the dantzig selector and lasso. *IEEE Trans. Signal Process.*, 71(1):127–142(16).

Jayaweera, S. K. and Li, T. (2009). Dynamic spectrum leasing in cognitive radio networks via primary-secondary user power control games. *IEEE Trans. Wireless Commun.*, 8(6):3300–3310.

Jayaweera, S. K., Vazquez-Vilar, G., and Mosquera, C. (2010). Dynamic spectrum leasing (DSL): A new paradigm for spectrum sharing in cognitive radio networks. *IEEE Trans. Vehic. Tech.*, 59(5):2328–2339.

Jorswieck, E. and Boche, H. (2007). *Majorization and Matrix-Monotone Functions in Wireless Communications*, volume 3 of *Foundations and Trends in Communications and Information Theory*.

Karoui, N. E. (2005). Recent results about the largest eigenvalue of random covariance matrices and statistical application. *Acta Phys. Pol. B*, 36(9):2681–2697.

Kay, S. M. (1993). *Fundamentals of statistical signal processing: estimation theory.* Prentice-Hall, Englewood Cliffs, NJ.

Kay, S. M. (1998). *Fundamentals of statistical signal processing: detection theory.* Prentice-Hall, Englewood Cliffs, NJ.

Kim, D. I., Le, L. B., and Hossain, E. (2008). Joint rate and power allocation for cognitive radios in dynamic spectrum access environment. *IEEE Trans. Wireless Commun.*, 7(12):5517–5527.

Kim, H.-S., Wang, J., Cai, P., and Cui, S. (2009). Detection outage and detection diversity in a homogeneous distributed sensor network. *IEEE Trans. Signal Process.*, 57(7):2875–2881.

Lancaster, P. and Tismenetsky, M. (1985). *The theory of matrices: With applications.* Academic Press, Orlando.

Larsson, E. G. and Stoica, P. (2003). *Space-Time Block Coding for Wireless Communications.* Cambridge University Press, Cambridge, UK.

Le, L. B. and Hossain, E. (2008). Resource allocation for spectrum underlay in cognitive radio networks. *IEEE Trans. Wireless Commun.*, 7(12):5306–5315.

Leshem, A. and Van der Veen, A.-J. (2001a). Multichannel detection and spatial signature estimation with uncalibrated receivers. In *Proc. 11th IEEE Work. Stat. Signal Process.*, pages 190–193.

Leshem, A. and Van der Veen, A.-J. (2001b). Multichannel detection of Gaussian signals with uncalibrated receivers. *IEEE Signal Process. Lett.*, 8(4):120–122.

Lim, T. J., Zhang, R., Liang, Y.-C., and Zeng, Y. (2008). GLRT-based spectrum sensing for cognitive radio. In *2008 IEEE Global Commun. Conf. (Globecom)*.

López-Valcarce, R. and Vazquez-Vilar, G. (2009). Wideband spectrum sensing in cognitive radio: Joint estimation of noise variance and multiple signal levels. In *2009 IEEE Signal Proc. Workshop on Signal Proc. Advances in Wireless Commun. (SPAWC)*.

López-Valcarce, R., Vazquez-Vilar, G., and Álvarez Díaz, M. (2009). Multiantenna detection of multicarrier primary signals exploiting spectral a priori information. In *4th Int. Conf. Cog. Radio Oriented Wireless. Networks and Commun. (Crowncom)*.

López-Valcarce, R., Vazquez-Vilar, G., and Sala, J. (2010). Multiantenna spectrum sensing for cognitive radio: overcoming noise uncertainty. In *2nd Proc. Int. Work. Cognitive Inf. Process. (CIP)*.

Lunden, J., Koivunen, V., Huttunen, A., and Poor, H. (2009). Collaborative cyclostationary spectrum sensing for cognitive radio systems. *IEEE Trans. Signal Process.*, 57(11):4182 –4195.

Magnus, J. R. and Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics.* John Wiley and Sons, New York, 2nd edition.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis.* New York: Academic.

Mauchly, J. (1940). Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Statist.*, 11:204–209.

Maybeck, P. S. (1979). *Stochastic models, estimation, and control*, volume 141 of *Mathematics in Science and Engineering*.

Menon, R., Buehrer, R., and Reed, J. (2008). On the impact of dynamic spectrum sharing techniques on legacy radio systems. *IEEE Trans. Wireless Commun.*, 7(11):4198–4207.

Mitola, J. and Maguire Jr., G. Q. (1999). Cognitive radio: Making software radios more personal. *IEEE Pers. Commun.*, 6:13–18.

Molisch, A. F. (2005). *Wireless Communications.* John Wiley and Sons, New York.

Pandharipande, A. and Linnartz, J.-P. (2007). Performance analysis of primary user detection in a multiple antenna cognitive radio. *Proc. IEEE Int. Conf. Commun. (ICC)*, pages 6482–6486.

Perez-Neira, A., Lagunas, M., Rojas, M., and Stoica, P. (2009). Correlation matching approach for spectrum sensing in open spectrum communications. *IEEE Trans. Signal Process.*, 57(12):4823–4836.

Polo, Y. L., Wang, Y., Pandharipande, A., and Leus, G. (2009). Compressive wideband spectrum sensing. In *2009 IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*.

Porat, B. and Friedlander, B. (1986). Computation of the exact information matrix of Gaussian time series with stationary random components. *IEEE Trans. Acoust., Speech, Signal Process.*, 34(1):118–130.

Quan, Z., Zhang, W., Shellhammer, S., and Sayed, A. (2011). Optimal spectral feature detection for spectrum sensing at very low SNR. *IEEE Trans. Commun.*, 59(1):201–212.

Ramírez, D., Vazquez-Vilar, G., López-Valcarce, R., Vía, J., and Santamaría, I. (2011). Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas. *IEEE Trans. Signal Process.* In press.

Ramírez, D., Vía, J., Santamaría, I., and Scharf, L. L. (2010). Detection of spatially correlated Gaussian time-series. *IEEE Trans. Signal Process.*, 58(10).

Sala, J., Vazquez-Vilar, G., and López-Valcarce, R. (2011). Multiantenna detection in unknown spatially correlated noise in cognitive radio networks. *IEEE Trans. Signal Process.* In review.

Scharf, L. L. (1991). *Statistical Signal Processing*. Prentice-Hall, Englewood Cliffs, NJ.

Schniter, P., Potter, L. C., and Ziniel, J. (2008). Fast bayesian matching pursuit. In *Inf. Theory and Appl. Workshop (ITA)*, La Jolla, CA, U.S.A.

Simeone, O., Stanojev, I., Savazzi, S., Bar-Ness, Y., Spagnolini, U., and Pickholtz, R. (2008). Spectrum leasing to cooperating secondary ad hoc networks. *IEEE Journ. Select. Areas Commun.*, 26(1):2013–213.

Simon, M. K. and Alouini, M.-S. (2004). *Digital communication over fading channels*. Wiley-Interscience, John Wiley and Sons, Hoboken, NJ.

Stoica, P. and Moses, R. L. (2005). *Spectral analysis of signals*. Prentice-Hall, Englewood Cliffs, NJ.

Taherpour, A., Gazor, S., and Nasiri-Kenari, M. (2008). Wideband spectrum sensing in unknown white Gaussian noise. *IET Commun.*, 2(6):763–771.

Taherpour, A., Nasiri-Kenari, M., and Gazor, S. (2009). Invariant wideband spectrum sensing under unknown variances. *IEEE Trans. Wireless Commun.*, 8(5):2182–2186.

Taherpour, A., Nasiri-Kenari, M., and Gazor, S. (2010). Multiple antenna spectrum sensing in cognitive radios. *IEEE Trans. Wireless Commun.*, 9(2):814–823.

Tandra, R. and Sahai, A. (2008). SNR walls for signal detection. *IEEE J. Sel. Top. Signal Process.*, 2:4–17.

Tauste, A., Vazquez-Vilar, G., Guillen i Fabregas, A., and Martinez, A. (2011). Random-coding joint source-channel bounds. In *2011 IEEE Int. Symp. Inf. Theory (ISIT)*.

Tellado, J. (2000). *Multicarrier modulation with low PAR: Applications to DSL and wireless*. Kluwer Academic, Norwell, MA.

Tian, Z. and Giannakis, G. (2007). Compressed sensing for wideband cognitive radios. In *IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*.

Tropp, J. (2006). Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inf. Theory*, 52(3):1030–1051.

Tulino, A. M. and Verdú, S. (2004). *Random Matrix Theory and Wireless Communications*, volume 1 of *Foundations and Trends in Communications and Information Theory*.

Vazquez-Vilar, G. and López-Valcarce, R. (2011). Wideband spectrum sensing exploiting guard bands and weak channels. *IEEE Trans. Signal Process.* Accepted for publication.

Vazquez-Vilar, G., López-Valcarce, R., Mosquera, C., and González-Prelcic, N. (2010a). Wideband spectral estimation from compressed measurements exploiting spectral a priori information in cognitive radio systems. In *2010 IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*.

Vazquez-Vilar, G., López-Valcarce, R., and Pandharipande, A. (2011a). Detection diversity of multiantenna spectrum sensors. In *2011 IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*.

Vazquez-Vilar, G., López-Valcarce, R., and Sala, J. (2011b). Multiantenna detection of multicarrier primary signals exploiting spectral a priori information. *IEEE Trans. Wireless Commun.* In review.

Vazquez-Vilar, G., Mosquera, C., and Jayaweera, S. K. (2010b). Primary user enters the game: Performance of dynamic spectrum leasing in cognitive radio networks. *IEEE Trans. Wireless Commun.*, 9(12):3625–3629.

Verdú, S. (1998). *Multiuser Detection*. Cambridge University Press, Cambridge, UK.

Wang, H. S. and Chang, P.-C. (1996). On verifying the first-order markovian assumption for a rayleigh fading channel model. *IEEE Trans. Vehic. Tech.*, 45(2):353–357.

Wang, P., Fang, J., Han, N., and Li, H. (2010). Multiantenna-assisted spectrum sensing for cognitive radio. *IEEE Trans. Vehic. Tech.*, 59(4):1791–1800.

Wilks, S. (1935). On the independence of k sets of normally distributed statistical variables. *Econometrica*, 3:309–325.

Xing, Y., Mathur, C. N., Haleem, M. A., Chandramouli, R., and Subbalakshmi, K. P. (2007). Dynamic spectrum access with QoS and interference temperature constraints. *IEEE Trans. Mobile Comp.*, 6(4):423–433.

Zeng, Y. and Liang, Y.-C. (2009a). Spectrum-sensing algorithms for cognitive radio based on statistical covariances. *IEEE Trans. Vehic. Tech.*, 58(4):1804–1815.

Zeng, Y. and Liang, Y.-C. (2009b). Spectrum-sensing algorithms for cognitive radio based on statistical covariances. *IEEE Trans. Vehic. Tech.*, 58(4):1804–1815.

Zeng, Y., Liang, Y.-C., and Zhang, R. (2008). Blindly combined energy detection for spectrum sensing in cognitive radio. *IEEE Signal Process. Lett.*, 15:649–652.

Zhang, R., Lim, T. J., Liang, Y.-C., and Zeng, Y. (2010a). Multi-antenna based spectrum sensing for cognitive radios: a GLRT approach. *IEEE Trans. Commun.*, 58(1):84–88.

Zhang, W., Poor, H., and Quan, Z. (2010b). Frequency-domain correlation: an asymptotically optimum approximation of quadratic likelihood ratio detectors. *IEEE Trans. Signal Process.*, 58(3):969–979.

Zhang, W., Poor, H. V., and Quan, Z. (2010c). Frequency-domain correlation: An asymptotically optimum approximation of quadratic likelihood ratio detectors. *IEEE Trans. Signal Process.*, 58(3):969–979.

Zhao, Y., Mao, S., Neel, J., and Reed, J. (2009). Performance evaluation of cognitive radios: Metrics, utility functions, and methodology. *IEEE Proc.*, 97(4):642–659.