Frugal Sensing and Estimation over Wireless Networks

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Omar Mehanna

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Professor Nicholas D. Sidiropoulos, Advisor
Professor Georgios B. Giannakis, Co-Advisor

April, 2014
Acknowledgements

I would like to express my sincere gratitude to all people who made my experience as a PhD student one of the most rewarding periods of my life. First and foremost, my deepest gratitude goes to my advisor, Professor Nikos Sidiropoulos. His advice has been extraordinary by all means, dedicating extensive amounts of time to guide me through my research, and to improve my technical writing and oral presentations. His constant encouragement and demonstrated confidence in me challenged me to perform to the fullest extent of my capabilities. Professor Sidiropoulos is a great mentor and fatherly figure; it has been a privilege working with him.

I am very grateful to my co-advisor, Professor Georgios B. Giannakis, whose vision, technical guidance, support, and motivation have been invaluable. I would also like to thank Professor Nihar Jindal for giving me the opportunity to embark on this journey as a PhD student, and for his guidance and support during the first years of my PhD. I would like to extend my deep gratitude to Professor Hesham El Gamal who taught me how to be a researcher; he guided me to write and publish my first research paper, and helped me start my PhD journey. I am thankful to Professor Tom Luo for his thoughtful insights in optimization theory which were crucial to achieve many results of this thesis, and also for agreeing to serve on my committee. Due thanks go to Professor Stergios I. Roumeliotis for agreeing to serve on my committee.

The material in this thesis has benefited from numerous additional inputs, discussions, ideas, and insights from several incredibly smart and talented individuals, and I have learned a lot from their knowledge and way of thinking. I would like to extend my gratitude to Balasubramanian Gopalakrishnan, Kejun Huang, Dr. Niranjay Ravindran, Thimios Tsakonas, Prof. Geert Leus, John Marcos, Artem Mosesov, Xiao Fu, and Aritra Konar. I enjoyed our stimulating discussions and friendship.
My sincere gratitude is due to my family and friends, for their continuous generosity in supporting me, for their precious time and sincerity to care about me, and for their pleasant company and sharing of those exciting moments. In particular, I would like to thank my Egyptian friends in Minnesota, Basant Hamouda, Mohamed Sarwat, Mohamed Elbadry, Ahmed Soliman, Mohamed Maher, Shady Taha, Shereen Sabet, Rundah Arafat, and Mohamed Maamoun, whose company made my days more enjoyable. Special thanks to my parents; your love is always the essential strength for me to face up everything coming to me and rejoice upon everything rejoicing me. Finally, I lovingly dedicate this thesis to my wife May and my son Zayd for the love and joy they bring me every day.

Omar Mehanna

Minneapolis, March 24, 2014
Abstract

Spectrum sensing and channel estimation are two important examples of background tasks needed for efficient wireless network operations. Channel and spectrum state communication overheads can become a serious burden, unless appropriate sensing and estimation strategies are designed that can do the job well with very limited, judicious feedback. This thesis considers two ‘frugal’ sensing and estimation problems in this regime: crowdsourced power spectrum sensing using a network of low-end sensors broadcasting few bits; and channel estimation and tracking for transmit beamforming in frequency-division duplex (FDD) mode.

In the case of spectrum sensing, each sensor is assumed to pass the received signal through a random wideband filter, measure the average power at the output of the filter, and send out a single bit to a fusion center (FC) depending on its measurement. Exploiting linearity with respect to the autocorrelation as well as important non-negativity properties in a novel linear programming (LP) formulation, it is shown that adequate power spectrum sensing is possible from few bits, even for dense spectra. The formulation can be viewed as generalizing classical nonparametric spectrum estimation to the case where the data is in the form of inequalities, rather than equalities. Taking into account fading and insufficient sample averaging considerations, a different convex maximum likelihood (ML) formulation is developed, outperforming the LP formulation when the power estimates prior to thresholding are noisy. Assuming availability of a downlink channel that the FC can use to send threshold information, active sensing strategies are developed which quickly narrow down the power spectrum estimate.

For the downlink channel tracking problem, the receiver is assumed to send back to the transmitter a coarsely quantized version of the received transmitter-beamformed pilot signal, instead of sending quantized channel information as in codebook-based beamforming. A novel channel tracking approach is proposed that exploits the quantization bits in a maximum a posteriori (MAP) estimation formulation, and closed-form expressions for the channel estimation mean-squared error and the corresponding signal-to-noise ratio are derived under certain conditions.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>i</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation and Context</td>
<td>2</td>
</tr>
<tr>
<td>1.1.1 Wideband Power Spectrum Sensing</td>
<td>2</td>
</tr>
<tr>
<td>1.1.2 Transmit Beamforming With Limited Feedback</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Thesis Outline and Contributions</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Notational Conventions</td>
<td>9</td>
</tr>
<tr>
<td>2 Frugal Sensing: Linear Programming Formulation</td>
<td>10</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Preliminaries</td>
<td>11</td>
</tr>
<tr>
<td>2.3 Network Sensing Model</td>
<td>13</td>
</tr>
<tr>
<td>2.4 Linear Programming Formulation</td>
<td>15</td>
</tr>
<tr>
<td>2.5 Simulations and Parameter Tuning</td>
<td>18</td>
</tr>
<tr>
<td>2.5.1 Threshold Selection</td>
<td>19</td>
</tr>
<tr>
<td>2.5.2 Filter Length and Type</td>
<td>21</td>
</tr>
<tr>
<td>2.6 Relevant Extensions</td>
<td>22</td>
</tr>
<tr>
<td>2.6.1 Model PS Estimation</td>
<td>23</td>
</tr>
</tbody>
</table>
3 Frugal Sensing: Maximum Likelihood Formulation

3.1 Introduction
3.2 Network Sensing Model
3.3 Maximum-Likelihood Formulation
3.4 Relevant Extensions
3.4.1 Model PS Estimation
3.4.2 Line Spectrum Estimation
3.4.3 Censoring
3.4.4 Multi-Bit Quantization
3.5 Active Sensing
3.6 Numerical Results
3.6.1 Model PS Estimation
3.6.2 Line Spectrum Estimation
3.6.3 Active Sensing

4 Channel Tracking for Transmit Beamforming with Frugal Feedback

4.1 Introduction
4.2 System Model
4.3 Analog-Amplitude Feedback
4.3.1 KF Approach
4.3.2 MMSE Approach
4.4 2-Bit Quantized Feedback
4.4.1 SOI-KF Approach
4.4.2 2-Bit MAP Approach
4.5 Performance Analysis
List of Tables

4.1 SNR loss comparison of different beamforming techniques . . . . . . . . . 77
B.1 List of acronyms used in the thesis . . . . . . . . . . . . . . . . . . . . . 110
List of Figures

2.1 Network sensing setup .................................................. 13
2.2 Sensor measurement chain ............................................ 14
2.3 Proposed frugal sensing approach in a sparse spectrum ............ 19
2.4 Proposed frugal sensing approach in a dense spectrum ............... 20
2.5 Trade-off between the NMSE and $\nu$ for signals with different $\eta$ .... 21
2.6 Tradeoff between the NMSE and the filter length $K$ .................... 22
2.7 Illustrative example for robust frugal sensing using (2.27) ........... 25
2.8 Illustrative example for the active sensing algorithm .................. 28
2.9 Active sensing ............................................................ 34
2.10 Performance of the active sensing and sensor polling algorithms .. . 35
3.1 The Gaussian distribution approximation of $\tilde{e}_m$ with different $L$ taps . . . 38
3.2 Log-likelihood function for a 2-tone signal with $\omega_1 = \pi/2$ and $\omega_2 = \pi$ . 45
3.3 ML PS estimation example ............................................. 49
3.4 Nonparametric PS: MSE of the LP and ML estimates and the CRB .......... 50
3.5 Model-based PS: MSE with large SER .................................. 52
3.6 Model-based PS: MSE with small SER .................................. 53
3.7 MSE with (a) Gaussian fading coefficients and (b) Gaussian errors .......... 54
3.8 Line Spectrum: Average CRB vs. threshold $t$ for different $L$ tones ....... 55
3.9 Line Spectrum: RMSE and CRB for a signal with 2 far-apart tones ....... 56
3.10 Line Spectrum: RMSE and CRB for a signal with 2 close tones ........... 57
3.11 Active sensing with Gaussian errors ................................... 58
4.1 Downlink frame structure and limited feedback beamforming model ....... 60
4.2 $\gamma_{\text{SOI-KF}}$ with $\alpha \in \{0.92, 0.96, 0.99\}$ and $\gamma_{\text{LA}}$ with $B \in \{10, 20, 50\}$ vs. $N$ ........... 74
4.3 Beamforming with $N = 10$ using Jake's model with $f_dT = 0.01$ ........... 76
Chapter 1

Introduction

Wireless communication networks increasingly rely on distributed sensing and estimation to improve communication, situational awareness, and network management. Spectrum sensing and channel estimation are two important examples of ‘background’ tasks needed for efficient wireless network operations.

Sensing tasks are often performed using special-purpose wireless sensor networks (WSNs – either stand-alone or overlaid on top of existing wireless communication infrastructure). WSNs comprise spatially distributed sensors that collect measurements about a certain phenomenon and communicate data to each other and to a fusion center (FC) responsible for detecting or estimating a signal of interest. Applications for WSNs include military surveillance, emergency services, environmental monitoring, smart agriculture, industrial monitoring, health monitoring, robotics, and spectrum sensing – see [3][5][25] for an overview of application areas. Wireless sensors must operate with limited (battery) power and transmission bandwidth (especially if the link is shared with ‘regular’ voice/data communication), and limited radio communication range. The latter means that multiple hops may be needed before the information reaches the FC. Hence, sending analog or finely quantized sample streams to the FC is a heavy burden in terms of communication overhead and sensor battery lifetime. It is therefore essential to develop bandwidth- and energy-efficient strategies for various network sensing and processing tasks.
Channel estimation and tracking is another important example where efficient quantization methods are necessary to support next-generation wireless networks and services, e.g., for transmit beamforming in frequency-division duplex (FDD) mode. In systems with many transmit antennas, the feedback overhead (in terms of power and rate) can be overwhelming, so the challenge is to limit the feedback to only a few bits that still provide sufficient information about the channel [50].

1.1 Motivation and Context

This thesis focuses on parameter estimation (and tracking) in wireless networks using very few quantization bits, in the seemingly data-starved ‘frugal’ regime. Two specific applications are considered: (1) crowdsourced power spectrum sensing using a network of low-end sensors, introduced in Section 1.1.1; and (2) channel estimation and tracking for transmit beamforming with limited feedback, introduced in Section 1.1.2.

1.1.1 Wideband Power Spectrum Sensing

The remarkable growth in mobile Internet and the explosion of wireless applications create an ever-increasing demand for the limited wireless spectrum. Exclusive licensing of bands to specific users or services is very inefficient, and lacks the agility needed to support new data-intensive applications. Twelve years ago, a U.S. Federal Communications Commission (FCC) Spectrum Policy Task Force report highlighted that the typical utilization of licensed bands is under 20% [1]. There is plenty of unused spectrum in most places, most of the time; the challenge is how to discover it in a timely fashion, and utilize it efficiently. Cognitive radio thus emerged as a promising cohabitation paradigm that facilitates the efficient utilization of the spectrum. Cognitive radios are fully programmable wireless devices that actively seek and exploit under-utilized spectral resources, while limiting interference to licensed primary users [4,77].

Spectrum sensing is arguably the most important component of cognitive radio, as it enables users to discover transmission opportunities, thus forming the basis for adaptive spectrum sharing. The goal of spectrum sensing is to detect spectral occupancy, and perhaps coarsely estimate power levels, under sensing constraints that typically preclude explicitly scanning the full band. This is essential in developing efficient power control.
schemes for secondary users in spectrum underlay settings [78]. Collaborative spectrum sensing using a network of spatially distributed sensors is important for measurement diversification – to increase reliability and alleviate shadow fading and the hidden terminal problem. If we could also control the signaling, power, and processing overhead so that the sensing task can be run as a lightweight job in background mode, that would open the door for crowdsourcing spectrum sensing to smart phones and other commodity wireless devices – creating a spectrum sensing web that spans across much of our living and working space.

A variety of spectrum sensing methods have been developed in recent years, ranging from narrowband energy detection to wideband sensing, mostly based on isolated hypothesis testing per narrowband channel ‘bin’, without taking into account dependence across frequency bins or exploiting any underlying parametrization. Reference [10] provides a good up-to-date review of spectrum sensing for cognitive radio. The premise of cognitive radio is that most of the band is idle, most of the time, i.e., measured spectra are typically sparse. Building upon this premise, compressive spectrum sensing has been introduced to exploit frequency-domain sparsity to obtain accurate spectrum estimates at sub-Nyquist sampling rates, without frequency sweeping [23, 72]. Cooperative spectrum sensing schemes that use compressive sensing have been considered in [14, 79], where the spectrum is estimated locally, then consensus on globally fused sensing outcomes is reached. However, the methods of [14] and [79] require sensors to perform complex local computations, and entail significant communication between sensors.

Most work on spectrum sensing focuses on detecting activity in the spectrum versus the power spectrum (PS), i.e., the Fourier transform of the signal, as opposed to the Fourier transform of its autocorrelation function. The PS is an expectation that reflects long-term spectral activity patterns; short-term effects such as fading are integrated out. PS sensing has been explored very recently in [6, 8, 47, 48, 64], where it was shown that neither Nyquist-rate sampling nor full-band scanning is necessary when the goal is to estimate only a finite set of correlation lags, which is then Fourier transformed to yield an estimate of the PS. This approach can decrease the sampling rate requirements by exploiting the ‘correlation parametrization’ (i.e., a low-order correlation model), without requiring spectrum sparsity. The key to this line of work is that power measurements
are linear in the autocorrelation function, hence a finite number of autocorrelation lags can be estimated by collecting enough power measurements to build an over-determined system of linear equations. In [64], the PS is estimated using sub-Nyquist rate sampling by exploiting the relationship between the autocorrelation function of the Nyquist-rate samples and that of the compressive measurements. The assumption that compressed measurements remain wide-sense stationary is relaxed in [8], where the under- and over-determined cases are considered. When over-determined, the PS is estimated using linear least-squares, without recourse to additional signal properties. When under-determined, the problem is regularized by minimizing the $\ell_1$ norm of the estimated PS, thus relying on sparsity in this case.

A bank of periodic modulators is considered in [6, 7], where each branch is sampled at a fraction of the Nyquist rate, and cross-correlations of the branch outputs are used to build a system of linear equations in the unknown input correlation for a fixed number of lags. This approach has been generalized to the case of cyclostationary signals in [47]. In [48], multi-coset sampling is employed producing multi-resolution power spectral estimates at arbitrarily low average sampling rates. A different approach exploiting spectrum sparsity has been proposed in [37], where $K$ wideband filters are used to detect occupancy in $N$ channels with $K < N$, assuming that the number of occupied channels is up to $O(K)$ (less than $K/2$). Note that [37] does not exploit the autocorrelation parametrization.

References [6–8, 14, 37, 47, 48, 64, 79] assume analog amplitude samples (i.e., ignore quantization issues), which is reasonable for lumped measurements taken with relatively accurate A/D converters at a high number of bits per sample. The situation is very different in a network sensing setting using scattered low-end sensors with limited communication capabilities, which is the scenario considered here. Suppose that each sensor can only down-convert, filter, and measure average power at the output of its filter. Depending on the computed power level, the sensor may send a binary signal to the FC, or broadcast it to its peers. Is it possible to form a satisfactory estimate of the ambient PS using just few such bits? This is a central question that is addressed in this thesis. PS sensing from few bits has never been considered in the past – yet is a natural extension of classical spectral estimation to the case where the data is in the form of power inequalities, rather than equalities.
The frugal sensing formulation may be reminiscent of one-bit compressed sensing \[20, 39, 63\], albeit there are significant differences between the two. Frugal sensing aims directly for the autocorrelation instead of the signal per se; it does not require (although it can exploit) sparsity in any domain – it instead relies on properties of autocorrelation sequences. Frugal sensing uses positive (power) thresholds, without the unit-sphere constraint of \[20, 39\], or the $\ell_1$-norm constraint of \[63\]; and it considers possible bit-flips due to pre-quantization measurement errors.

1.1.2 Transmit Beamforming With Limited Feedback

Transmit beamforming can enhance the performance of multiple-input multiple-output (MIMO) systems by exploiting channel state information (CSI) at the transmitter. In the FDD mode, where the downlink and uplink channels are not reciprocal, the receiver must feedback information about the downlink channel to the transmitter. In systems with many transmit antennas, the feedback overhead can be overwhelming; the challenge is to limit the feedback to only a few bits that still provide sufficient information about the channel.

Almost all work on transmit beamforming with limited feedback addresses this challenge by designing efficient beamformer weight vector quantization algorithms at the receiver. The focus is on designing a common beamformer codebook (known at the transmitter and receiver). At runtime, the receiver estimates the downlink channel, finds the best-matching beamforming vector in the codebook, and feeds back its index to the transmitter \[50\]. Codebook design can be based on maximizing the average signal-to-noise ratio (SNR) \[51\], maximizing the average mutual information \[46\], or minimizing the outage probability \[61\], and it can be viewed as a vector quantization problem, where the generalized Lloyd algorithm (GLA) can be used to construct the codebook \[75\]. This codebook-based framework assumes accurate CSI at the receiver, which in turn implies significant downlink pilot overhead. For large codebooks, which are necessary when the number of transmit-antennas is large, the feedback overhead can be significant, and the computational complexity of searching the codebook for the best beamformer can be prohibitive.

Another important issue is that most prior work assumes a Rayleigh block-fading model, according to which the channel remains constant over a block of symbols and
changes independently across different blocks. The block-fading assumption overlooks the channel temporal correlation, which can be exploited to decrease the feedback rate \cite{38,70}. In \cite{70} and \cite{38}, the temporal correlation of the channel is exploited by modeling the quantized CSI at the receiver as a finite-state Markov chain, and computing the transition probability of every codebook entry given the previous (one or more) codebook entries. In \cite{70}, variable-length Huffman source coding is applied to the transition probabilities of the Markov chain to compress the CSI feedback. This approach is not suitable for practical communication systems with limited feedback, which provision a fixed number of feedback bits per CSI slot, as in e.g., LTE \cite{2}. Considering this issue, a different fixed-length but lossy CSI compression algorithm is proposed in \cite{38}, where low-probability transitions between the Markov chain states are truncated. For large-size codebooks, computing the transition probabilities accurately for a large number of Markov states is an elusive task that requires very long training periods. Moreover, the transition probabilities are dependent on the specific channel model – new computations are necessary whenever the model varies significantly.

This thesis proposes a different approach for beamforming with limited feedback, that exploits the spatio-temporal channel correlation, and avoids the limitations of codebook-based feedback and Markov chain modeling. The transmitter is assumed to periodically transmit a beamformed pilot signal in the downlink, while the receiver quantizes the corresponding received signal (2-bit quantization is considered in this thesis), and sends the quantization bits to the transmitter through the uplink feedback channel. Therefore, instead of estimating the channel at the receiver and sending the quantized CSI to the transmitter as in codebook-based beamforming, the receiver feeds back a quantized (noisy) linear measurement of the channel. The challenge here is whether the transmitter can accurately estimate and track the channel using such few (periodic) feedback bits.

### 1.2 Thesis Outline and Contributions

Chapter \cite{2} considers the problem of wideband PS sensing from (few) power measurement bits received from scattered low-end sensors. Exploiting linearity with respect to autocorrelation and important non-negativity properties in a novel optimization-based
formulation, it is shown that the PS sensing problem can be reduced to linear programming (LP), and that adequate PS sensing is possible from few bits, even for dense spectra. The tradeoffs that emerge in the selection of key parameters, such as filter length and power threshold, and how these affect spectrum sensing performance and complexity are studied. The formulations are further extended to the case where the PS is modeled as a weighted sum of candidate spectral density primitives with unknown weights, similar to [67] and [13], but for coarsely quantized (1-bit) data. The special case of line spectra is also considered, wherein only few line frequencies and powers need to be estimated, instead of an entire function. Assuming availability of a ‘downlink’ channel that the FC can use to send threshold information, active sensing strategies are further developed that quickly narrow down and track the PS estimate, using ideas borrowed from cutting plane methods in optimization theory. Convergence of the proposed algorithms to the true finite-length autocorrelation, as more sensors report their measurement bits, is shown. To avoid the downlink threshold communication overhead, an alternative sensor polling algorithm is proposed, where the FC judiciously polls sensors to form its PS estimate. The results included in Chapter 2 have been reported in [53, 55–57].

The underlying assumption which enables using the LP formulation in Chapter 2 is that the soft power estimate at each sensor is accurate enough to avoid flipping the inequality after quantization. This assumption is relaxed in Chapter 3, where it is shown that the distribution of the error in the soft power estimates (prior to thresholding) due to frequency-selective fading and insufficient sample averaging can be approximated by a Gaussian distribution. The Gaussian distribution of the errors is then exploited in a maximum likelihood (ML) formulation that optionally includes a sparsity-inducing penalty term. It is shown that the ML estimate asymptotically achieves the Cramér-Rao bound (CRB), which is also derived here, as the number of reporting sensors grows. The ML formulation is then extended to the case where the PS is modeled as a weighted sum of candidate spectral density primitives and to the case of line spectra. In order to reduce the number of bits transmitted from the sensors to prolong battery life and minimize communication overhead, an ML/CRB-driven censoring scheme is proposed, where only sensors that provide the most useful information bits are permitted to send, while other sensors remain silent. The active sensing algorithms that are proposed in Chapter
Chapter 3 considers the problem of transmit beamforming with limited feedback in large-antenna FDD systems. Assuming that the channel can be modeled by an autoregressive (AR) model \cite{12}, and that the receiver feeds back the analog-amplitude (un-quantized or finely-quantized) received signal to the transmitter, Kalman filtering (KF) \cite{43} is used in \cite{68} to track the channel at the transmitter. However, sending the analog or finely-quantized received signal back to the transmitter is problematic in terms of uplink rate and transmit power. In this chapter, a 2-bit quantization scheme is considered that is based on the sign of innovation (SOI), demonstrating how the SOI-KF framework of \cite{66} can be extended and used for transmit beamforming with very limited feedback when the channel follows an AR model. Moreover, closed-form expressions for the channel estimation mean-squared error (MSE), and very tight closed-form approximations for the achievable average SNR, are derived under certain conditions. For general (non-AR or even unknown) channel models, a novel channel tracking approach is proposed that exploits the quantization bits in a maximum a posteriori (MAP) estimation formulation. Simulations confirm that by exploiting the spatio-temporal correlation of the channel, and with very limited feedback rate (i.e., 2-bits per channel dwell block), the performance achieved using the proposed approaches is close to that attainable with perfect CSI at the transmitter. Simulations also show that very large-size codebooks are required for codebook-based beamforming to achieve the same performance as the proposed approaches. The results advocate for using transmit beamforming for massive MIMO in FDD mode, whereas the focus of massive MIMO has so far been on time-division duplex (TDD) operation, because of the huge overhead associated with CSI feedback \cite{14}. The results included in Chapter 4 have been reported in \cite{58}.

A summary of the contributions and conclusions of the thesis and areas for future research are included in Chapter 5.
1.3 Notational Conventions

Boldface uppercase letters denote matrices, whereas boldface lowercase letters denote column vectors; $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote the complex conjugate, transpose and Hermitian (conjugate) transpose operators, respectively; Trace$(\cdot)$, $\| \cdot \|$, $| \cdot |$, ${\text{Re}}\{\cdot\}$ and $\text{Im}\{\cdot\}$ denote the trace, the Euclidean norm, the absolute value, the real, and the imaginary operators, respectively; MATLAB notations diag$(\mathbf{x})$ and toepplitz$(\mathbf{x})$ denote the diagonal matrix and the toeplitz matrix that are formed with vector $\mathbf{x}$, respectively; mod$(x,y)$ returns the modulus after division of $x$ by $y$; the operator $\odot$ corresponds to the Hadamard (elementwise) product of two matrices; $\mathbb{E}[]$ denotes the ensemble average; $\mathcal{CN}(\mathbf{a}, \mathbf{C})$ denotes the complex Gaussian distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{C}$; $\mathbf{I}$ denotes the identity matrix; the function sign$(x) = 1$ if $x \geq 0$ and -1 otherwise; and $Q(x) := \frac{1}{2\pi} \int_x^{\infty} e^{-u^2/2} du$ is the standard Gaussian tail integral.
Chapter 2

Frugal Sensing: Linear Programming Formulation

2.1 Introduction

This chapter addresses the challenge of wideband power spectrum (PS) sensing from few bits. A network sensing scenario is considered, where scattered low-end sensors randomly filter and measure the average signal power across a band of interest, and each sensor communicates a single bit (or coarsely quantized level) to a fusion center (FC), depending on whether its measurement is above a certain threshold. The focus is on the underdetermined case, where relatively few bits are available at the FC. To estimate the ambient PS using just few such bits, the linear relationship between the PS and the autocorrelation, and the PS non-negativity, are exploited in a linear programming (LP) formulation. The formulation can be viewed as generalizing classical nonparametric PS estimation to the case where the data is in the form of inequalities, rather than equalities. Simulations show that adequate PS sensing is possible from few bits, even for dense spectra. The tradeoffs that emerge in the selection of key parameters, such as filter length and power threshold, and how these affect spectrum sensing performance and complexity are studied.

The LP formulations are further extended to the case where the PS is modeled as a weighted sum of candidate spectral density primitives with unknown weights, and to the case of signals with line spectrum. Moreover, the sparse linear regression framework
is used to robustify the LP formulation against an outlier-contaminated model representing possible sensor-to-FC communication errors that may cause bit flips. Finally, assuming availability of a ‘downlink’ channel that the FC can use to send threshold information, active sensing strategies are proposed that quickly narrow down and track the PS estimate, using ideas borrowed from cutting plane methods in optimization theory. Convergence of the proposed algorithms to the true finite-length autocorrelation, as more sensors report their measurement bits, is shown. To avoid the downlink threshold communication overhead, an alternative sensor polling algorithm is proposed, where the FC judiciously polls sensors to form its PS estimate.

2.2 Preliminaries

Consider a discrete-time wide-sense stationary (WSS) signal \( x(n) \), and let \( r_x(k) := \mathbb{E}[x(n)x^*(n - k)] \) denote its autocorrelation sequence, where \( r_x(k) = r_x^*(-k) \), for \( k \geq 1 \), and \( r_x(0) \) is nonnegative, by definition. The PS of \( x(n) \), \( S_x(\omega) \), is the discrete-time Fourier transform (DTFT) of \( r_x(k) \), i.e., \( S_x(\omega) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-j\omega k} \), \( \omega \in [0, 2\pi) \), where \( S_x(\omega) \) is real and nonnegative. If only a finite \( K \)-lag autocorrelation sequence is available, then a windowed estimate of the PS can be obtained as \( \hat{S}_x(\omega) = \sum_{k=-K+1}^{K-1} r_x(k)e^{-j\omega k} \), \( \omega \in [0, 2\pi) \). Due to truncation to a finite number of lags, however, such an estimate is not guaranteed to be nonnegative at all frequencies.

Discretizing the frequency axis, an \( N_F \)-point estimate of the PS can be obtained as \( \hat{s}_x = \tilde{F}\tilde{r}_x \), where \( \hat{s}_x(f) = \hat{S}_x \left( \frac{2\pi f}{N_F} \right) \), \( f = 0, \ldots, N_F - 1 \), \( \tilde{r}_x := [r_x(1-K), \ldots, r_x(-1), r_x(0), r_x(1), \ldots, r_x(K-1)]^T \), and \( \tilde{F} \) is the \( N_F \times (2K-1) \) (phase-shifted) discrete Fourier transform (DFT) matrix:

\[
\tilde{F} = \begin{bmatrix}
1 & \cdots & 1 & \cdots & 1 \\
e^{-j\frac{2\pi}{N_F}(1-K)} & \cdots & 1 & \cdots & e^{-j\frac{2\pi}{N_F}(K-1)} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
e^{-j\frac{2\pi}{N_F}(N_F-1)(1-K)} & \cdots & 1 & \cdots & e^{-j\frac{2\pi}{N_F}(N_F-1)(K-1)}
\end{bmatrix}
\tag{2.1}
\]

Define the (real) autocorrelation vector

\[
r_x := [r_x(0), \text{Re}\{r_x(1)\}, \ldots, \text{Re}\{r_x(K-1)\}, \text{Im}\{r_x(1)\}, \ldots, \text{Im}\{r_x(K-1)\}]^T
\tag{2.2}
\]
and the matrix
\[
W := \begin{bmatrix}
0_{K-1} & \tilde{I}_{K-1} & -j \tilde{I}_{K-1} \\
1 & 0_{K-1}^T & 0 \\
0_{K-1} & I_{K-1} & j I_{K-1}
\end{bmatrix}
\] (2.3)

where \(0_{K-1}\) is the \((K - 1)\) zeros vector, \(I_{K-1}\) is the \((K - 1)\) identity matrix, and \(\tilde{I}_{K-1}\) is the flipped \((K - 1)\) identity matrix. Hence it is easy to see that \(\hat{s}_x = Fr_x\), where \(F := \hat{F}W\) (and \(r_x = W^{-1}r_x\)).

Define the \(K \times K\) Toeplitz-Hermitian autocorrelation matrix
\[
R_x = \begin{bmatrix}
r_x(0) & r_x(-1) & \cdots & r_x(1-K) \\
r_x(1) & r_x(0) & \ddots & \vdots \\
\vdots & \ddots & \ddots & r_x(-1) \\
r_x(K-1) & \cdots & r_x(1) & r_x(0)
\end{bmatrix}
\] (2.4)

The construction of \(R_x\) from \(r_x\) can be explicitly parameterized as follows. Let \(E_k\) denote the \(K \times K\) matrix with ones on the \(k\)-th lower diagonal and zeros elsewhere, \(k \in \{1, \ldots, K - 1\}\). Hence
\[
R_x = r_x(0)I + \sum_{k=1}^{K-1} \left[ (\text{Re}\{r_x(k)\} + j\text{Im}\{r_x(k)\}) E_k + (\text{Re}\{r_x(k)\} - j\text{Im}\{r_x(k)\}) E_k^T \right]
\]
\[
= r_x(0)I + \sum_{k=1}^{K-1} \left[ \text{Re}\{r_x(k)\} \tilde{E}_k + \text{Im}\{r_x(k)\} \tilde{E}_k \right]
\] (2.5)

where \(\tilde{E}_k := E_k + E_k^T\) and \(\tilde{E}_k := j(E_k - E_k^T)\).

Properties of the autocorrelation can be exploited to define an initial bounded feasible region for \(r_x\). First, an upper bound \(P_{\text{max}}\) for the total signal power can be obtained (due to the use of AGC at the front-end of the sensor processing chain) yielding the bounds \(0 \leq r_x(0) \leq P_{\text{max}}\). Another well-known property that can be exploited is that \(|r_x(k)| \leq r_x(0)\), for \(k = 1, \ldots, K - 1\). These inequalities yield the initial feasible region \(r_x \in \mathcal{P}\), where the bounded polyhedron
\[
\mathcal{P} := \left\{ r_x \in \mathbb{R}^{2K-1} : 0 \leq r_x(0) \leq P_{\text{max}}, -r_x(0) \leq \text{Re}\{r_x(k)\} \leq r_x(0), -r_x(0) \leq \text{Im}\{r_x(k)\} \leq r_x(0), k = 1, \ldots, K - 1 \right\}
\] (2.6)
2.3 Network Sensing Model

Consider $M$ scattered sensors measuring the ambient signal power and reporting to a FC, as shown in Fig. 2.1. In the presence of frequency-selective fading, the received signal at sensor $m \in \{1, \ldots, M\}$, sampled using a Nyquist-rate analog-to-digital converter (ADC), is the convolution of the primary WSS signal $x(n)$ with the $L$-tap linear finite impulse response (FIR) fading channel $\{h_m(\ell)\}_{\ell=0}^{L-1}$, expressed as

$$\tilde{y}_m(n) = \gamma_m \sum_{\ell=0}^{L-1} h_m(\ell)x(n-\ell) \quad (2.7)$$

where $\gamma_m$ is a sensor-specific constant that models path loss and frequency-flat shadowing and fading. In Appendix A.1, it is shown that frequency-selective fading can be mitigated by averaging the measurements over a long period of time (e.g., in a fast-fading environment). The Nyquist sampling requirement can be lifted by using an equivalent analog processing and integration chain - the details can be found in Appendix A.2. Sensor $m \in \{1, \ldots, M\}$ then uses automatic gain control (AGC) to adjust the scaling of its received signal, where the AGC output is expressed as $y_m(n) = \tilde{y}_m(n)/\gamma_m$. Note that sensors are not assumed to be synchronized; sensing time offsets and phase shifts are allowed. Since the PS is invariant with respect to timing offsets and phase
shifts, it is assumed in this chapter, without loss of generality, that all sensors sense the same signal after the AGC stage (i.e., \( y_m(n) = x(n) \) for \( m \in \{1, \ldots, M\} \)). Chapter 3 considers the case where frequency-selective fading realizations are not averaged over time (e.g., due to slow-fading environment).

Sensor \( m \in \{1, \ldots, M\} \) then passes \( y_m(n) \) through a wideband FIR filter with impulse response \( g_m(n) \) of length \( K \). In order to monitor a wide swath of spectrum with relatively few sensors, it is necessary to use broadband filters \( \{g_m(n)\}_{m=1}^{M} \), which should somehow provide, loosely speaking, independent yet complementary views of the underlying PS. A random complex binary pseudo-noise (PN) impulse response \( \{g_m(n)\} \) is considered, i.e.,

\[
g_m(n) = \begin{cases} 
\frac{1}{\sqrt{2K}}(\pm 1 \pm j) & \text{if } 0 \leq n \leq K - 1 \\
0 & \text{otherwise}
\end{cases}
\] (2.8)

The filter sequence (2.8) can be generated using a PN linear shift register, whose initial seed is unique for each sensor (e.g., its serial number), and is known to the FC. Using wideband PN filters is appealing because it ensures diversity; requires no coordination between sensors: a sensor may fail when its battery runs out, or new sensors may be added without re-programming the other ones; and also for its simplicity: convolution requires no multiplication. Using random PN filters can also be motivated from a random projections / incoherence point of view, as for the compression matrix applied to sparse signals [23].

The filter’s output sequence \( z_m(n) \) is the convolution of the signal \( x(n) \) with the impulse response \( g_m(n) \), expressed as \( z_m(n) = \sum_{k=0}^{K-1} g_m(k)x(n-k) \). Let \( \alpha_m := \mathbb{E}[|z_m(n)|^2] \) denote the average power of the WSS signal \( z_m(n) \). Each sensor estimates \( \alpha_m \) by sample

---

**Figure 2.2:** Sensor measurement chain.
averaging using \( N \) samples
\[
\hat{\alpha}_m = \frac{1}{N} \sum_{n=0}^{N-1} |z_m(n)|^2
\]  

with \( \lim_{N \to \infty} \hat{\alpha}_m = \alpha_m \) under appropriate mixing conditions [35, p. 171]. Finally, each sensor compares the estimated \( \hat{\alpha}_m \) to a sensor-specific threshold \( t_m \). If \( \hat{\alpha}_m \geq t_m \), then sensor \( m \) sends a single bit \( b_m = 1 \) to the FC, otherwise it sends \( b_m = -1 \). The power measurement bit of each sensor can thus be expressed as
\[
b_m = \text{sign}(\hat{\alpha}_m - t_m)
\]  

This sensor measurement chain is shown in Fig. 2.2. The objective is to estimate the PS of the signal \( x(n) \) at the FC from the measurement bits \( \{b_m\}_{m=1}^M \).

### 2.4 Linear Programming Formulation

Let \( q_m(k) \) denote the deterministic autocorrelation of \( g_m(n) \), defined as
\[
q_m(k) := \sum_{n=0}^{K-1} g_m(n) g_m^*(n+k)
\]  

Also, define the vectors
\[
g_m := [g_m^*(K-1), g_m^*(K-2), \ldots, g_m^*(0)]^T
\]
\[
q_m := [q_m(0), 2 \text{Re}\{q_m(1)\}, \ldots, 2 \text{Re}\{q_m(K-1)\}, 2 \text{Im}\{g_m(1)\}, \ldots, 2 \text{Im}\{g_m(K-1)\}]^T
\]  

Hence, it can be shown that
\[
\alpha_m = g_m^H R_x g_m = \sum_{k=1-K}^{K-1} r_x(k) q_m^*(k) = q_m^T r_x
\]  

where the autocorrelation vector \( r_x \) is defined in (2.2). It follows that, upon receipt of \( b_m = 1 \) (or \( b_m = -1 \)) from sensor \( m \), the FC learns that \( q_m^T r_x \geq t_m \) (resp. \( q_m^T r_x < t_m \)), assuming sufficient averaging such that sample averages converge to ensemble averages (i.e., assuming \( \hat{\alpha}_m \to \alpha_m \)). Note that since we only need to ensure that the inequality is not reversed (i.e., \( \text{sign}(\hat{\alpha}_m - t_m) = \text{sign}(\alpha_m - t_m) \)), sample averaging requirements are considerably relaxed relative to high-rate quantization.
The job of the FC is to estimate the ambient PS based on the information it received from the sensors. This can be accomplished by reconstructing the $K$-lag autocorrelation function $\hat{r}_x$, and then applying the DFT: $\hat{s}_x = \mathbf{F}\hat{r}_x$. Unlike classical spectral analysis, the data here is in the form of linear inequalities in the autocorrelation function. The setup is more heavily under-determined, and all available structural properties and prior information should be employed to obtain a meaningful estimate of the PS. For a valid autocorrelation vector $\mathbf{r}_x$ of any order, we know that the autocorrelation matrix $\mathbf{R}_x$ is positive semidefinite. This is an important structural property of autocorrelation sequences, which can be exploited to reduce under-determinacy and improve the estimation of $\mathbf{r}_x$. In the limit of $K \to \infty$, this also ensures that $\hat{S}_x(\omega) \geq 0$, $\forall \omega \in [0, 2\pi)$, but the windowed estimate $\hat{S}_x(\omega)$ that is obtained by taking the Fourier transform of a truncated $K$-lag autocorrelation is not guaranteed to be nonnegative at all frequencies. However, including the non-negativity constraint $\hat{s}_x = \mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ in the set of constraints when estimating $\mathbf{r}_x$ is essential in decreasing the under-determinacy of the estimation problem and obtaining a good estimate of $\mathbf{r}_x$. Towards this end, we propose including both $\mathbf{R}_x \succeq \mathbf{0}$ and $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ as explicit constraints in an optimization-based formulation.

The remaining issue is to find an appropriate cost function. A reasonable choice is to minimize the total signal power $r_x(0)$, consistent with the premise of cognitive radio that most of the spectrum is unused in most places, most of the time. Interestingly, by enforcing $\hat{s}_x = \mathbf{F}\mathbf{r}_x \geq \mathbf{0}$, and since $r_x(0) = \frac{1}{N_F} \sum_{f=0}^{N_F-1} \hat{s}_x(f)$, it follows that $||\hat{s}_x||_1 = \sum_{f=0}^{N_F-1} |\hat{s}_x(f)| = N_F r_x(0)$. It is well known that minimizing the $\ell_1$-norm induces sparsity \[1\], which means that minimizing the total signal power $r_x(0)$ implicitly encourages sparsity in the reconstructed PS.

Putting everything together, the autocorrelation vector $\mathbf{r}_x$ can be reconstructed at the FC by solving the following problem:

$$
\min_{\mathbf{r}_x \in P} \quad r_x(0) \\
\text{s.t. :} \quad b_m(\mathbf{q}_m^T \mathbf{r}_x - t_m) \geq 0, \ m = 1, \ldots, M, \tag{2.15}
$$

$$
\mathbf{R}_x \succeq \mathbf{0}, \quad \mathbf{F}\mathbf{r}_x \geq \mathbf{0}
$$

Note that the relation between $\mathbf{R}_x$ and $\mathbf{r}_x$ is a linear as expressed in (2.5). This implies that all the constraints in (2.15) are ordinary linear inequalities in $\mathbf{r}_x$ except for the constraint $\mathbf{R}_x \succeq \mathbf{0}$, which is a linear matrix inequality (LMI). Hence, problem (2.15) is
a semidefinite programming (SDP) that can be optimally solved using efficient interior point methods \cite{22}. The following proposition, however, asserts that the constraint $R_x \succeq 0$ is redundant; it is in fact implied by the constraint $F r_x \geq 0$.

**Proposition 1** For $N_F \geq 2K - 1$, $F r_x \geq 0$ implies that $R_x \succeq 0$. The converse is generally not true.

The proof can be found in Appendix A.3.

Proposition 1 implies that problem (2.15) is not affected by removing the constraint $R_x \succeq 0$. Thus, the SDP (2.15) can be expressed as the following linear program (LP):

\[
\begin{align*}
\min_{r_x \in \mathcal{P}} & \quad r_x(0) \\
\text{s.t.} & \quad b_m(q_m^T r_x - t_m) \geq 0, \quad m = 1, \ldots, M, \\
& \quad F r_x \geq 0
\end{align*}
\]

The significance of the reduction from an SDP to an LP is that the latter can be solved much more efficiently using specialized LP solvers.

Note that the final feasible region for $r_x$ when the FC receives all measurement bits $\{b_m\}_{m=1}^M$ (assuming that the true $r_x$ satisfies $F r_x \geq 0$) is defined by the polyhedron

\[
\mathcal{P}_M := \mathcal{P} \cap \{r_x \mid b_m(q_m^T r_x - t_m) \geq 0, \quad m = 1, \ldots, M\} \cap \{r_x \mid F r_x \geq 0\}
\]

The optimal solution of the LP (2.16) will always be on the boundary of $\mathcal{P}_M$ - in fact, without loss of optimality, it is a vertex of $\mathcal{P}_M$. Thus the boundary of $\mathcal{P}_M$ is associated with sparse feasible spectra. If the sought spectrum is known to be non-sparse, then it makes sense to steer away from the boundary of $\mathcal{P}_M$, and a good way to enforce this is to use the “center” of $\mathcal{P}_M$ as an estimate of $r_x$, instead of minimizing $r_x(0)$. In this case, the center of the maximum-volume inscribed ellipsoid, the analytic center (AC), or the Chebyshev center (CC) can be used as the center of a polyhedron $\mathcal{P}_M$ (see \cite{22} for the definition of each center). Obtaining any of these centers subject to $r_x \in \mathcal{P}_M$ is a convex optimization problem; however, only the CC is obtained using an LP \cite{22}.

The worst-case estimation error $\varepsilon_r := \max_{r_x \in \mathcal{P}_M} ||r_x - \hat{r}_x||$ can be upper bounded by the maximum euclidean distance between any two vertices of the polyhedron $\mathcal{P}_M$. This requires solving a *vertex enumeration* problem to obtain the set of vertices defining $\mathcal{P}_M$. 
See [9,26] for algorithms that solve the *vertex enumeration* problem in polynomial-time. Then, instead of exhaustively computing the euclidean distance between all vertices, which is clearly computationally prohibitive, we can compute the minimum volume ellipsoid (i.e., Löwner-John ellipsoid) that covers the set of vertices $V = \{v_1, \ldots, v_N\}$ of $P_M$ by solving the convex optimization problem [22] Sec. 8.4.1

$$\min_{A,b} \log \det A^{-1}$$

s.t. $||Av_i - b||_2 \leq 1, \ i = 1, \ldots, N, \ A > 0$ \hfill (2.18)

where the ellipsoid in this case is defined as $\rho = \{x \ | \ ||Ax - b||_2 \leq 1\}$. The maximum euclidean distance between any two vertices of $P_M$ is bounded by the maximum axis of $\rho$, which is computed as $2/\eta$ where $\eta$ as the minimum eigenvalue of $A$ obtained from (2.18) [22]. Thus, $\varepsilon_r \leq 2/\eta$. A looser bound can be obtained without *vertex enumeration* as follows. The maximum volume inscribed ellipsoid in a polyhedron $C = \{x \ | \ a_i^T x \leq c_i, \ i = 1, \ldots, L\}$ can be obtained by solving the convex optimization problem [22] Sec. 8.4.2

$$\min_{B,d} \log \det B^{-1}$$

s.t. $||Ba_i||_2 + a_i^T d \leq c_i, \ i = 1, \ldots, L, \ B > 0$ \hfill (2.19)

where the ellipsoid in this case is defined as $\rho = \{Bx + d \ | \ ||x||_2 \leq 1\}$. Denoting $\zeta$ as the maximum eigenvalue of $B$ obtained from (2.19), a (loose) upper bound on $\varepsilon_r$ can be obtained as $\varepsilon_r \leq 2(2K - 1)\zeta$ [22].

### 2.5 Simulations and Parameter Tuning

In this section, we provide simulation results and discuss the effect of some design parameters on the quality of the PS estimate. For simplicity, a single threshold $t_m = t$ is considered for all sensors. We begin with a simulation that illustrates what one can expect from the proposed approach. In Fig. 2.3 and Fig. 2.4 a scenario with $M = 100$ sensors was considered, and the estimated PS (dashed line) has been obtained by solving the LP (2.16). For Fig. 2.3 the true PS is sparse (solid line), filter length $K = 24$ was used, and the threshold $t$ was set such that 30 sensors report $b_m = 1$; whereas for Fig. 2.4 the true PS is dense, filter length $K = 10$ was used, and $t$ was set such that 50
sensors report $b_m = 1$. The plotted spectra have been normalized by the peak value of the true PS. The quality of the estimates in Figs. 2.3, 2.4 is very satisfactory considering that only 100 bits have been used as input data - corresponding roughly to three single precision IEEE floats, or about what it would take to transmit three accurate power measurements, or $r(0)$ and $r(1)$ (note that $r(1)$ is complex, requiring two floats).

In the rest of this section, we use the normalized mean square error (NMSE) to measure the quality of the PS estimate. The NMSE is defined as

$$\text{NMSE} := \mathbb{E}\left[ \frac{||s_x - \hat{s}_x||^2}{||s_x||^2} \right]$$

(2.20)

where the expectation is taken with respect to the random signal and the random impulse responses of the FIR filters, obtained via Monte-Carlo simulations. Note that using $\mathbb{E}[||s_x - \hat{s}_x||^2]/\mathbb{E}[||s_x||^2]$ instead of (2.20) to define NMSE made very little difference in our experiments - the results were almost identical.

2.5.1 Threshold Selection

This section shows that, from an estimation performance point of view, the threshold $t$ should be selected according to the sparsity level of the PS (assuming prior sparsity knowledge is available). Let $\eta$ denote the sparsity ratio, defined as the ratio of the
nonzero (or above a small quantity) entries to the total length of the PS vector $s_x$, and define $\nu$ as the ratio of the number of sensors with measurements above $t$ to the total number of reporting sensors (i.e., $\nu = \sum_{m=1}^{M} (b_m+1)$).

In Fig. 2.5, we plot the NMSE versus the ratio $\nu$, for signals with different sparsity ratios $\eta$. The sparse signal was fixed for each $\eta$, and 1000 Monte-Carlo simulations for each $\nu$ were used to obtain the corresponding NMSE (here the expectation was taken with respect to the random FIR filters only). The setup included $M = 60$ sensors and the filter length was set to $K = 8$. Two main points can be deduced from Fig. 2.5. First, we see that as the sparsity ratio $\eta$ increases, the NMSE is minimized at a higher ratio $\nu$. This means that the threshold $t$ should be tuned such that number of sensors reporting measurements above $t$ decreases as the PS becomes more sparse. Historical data can be used to get an expectation for $\eta$, and to identify the distribution of $\alpha_m$.

Exploiting such prior statistical information, the threshold $t$ can be selected such that $\nu$ minimizes the NMSE for the corresponding $\eta$. The second point that can be drawn from Fig. 2.5 is that the minimum NMSE increases as the PS becomes less sparse. This implies that the quality of the estimated PS using the proposed approach is relatively better for sparser signals.
Figure 2.5: Trade-off between the NMSE and $\nu$ for signals with different $\eta$.

### 2.5.2 Filter Length and Type

Next, we look at how the filter length $K$ affects the quality of the PS estimate, and also discuss two candidate classes of random filters. Note that the number of filter taps $K$ is also the number of estimated autocorrelation lags. Truncation of the autocorrelation sequence smears the estimated PS [71], and the smaller $K$ is, the more pronounced this smearing will be. This is the reason why $K = 24$ has been used in Fig. 2.3 where the spectrum is a sparse superposition of narrowband spectra, whereas $K = 10$ has been used in Fig. 2.4 which features two main lobes occupying more than half the bandwidth. On the other hand, $K$ is also the number of unknowns, and the larger $K$ is, relative to the number of inequality constraints in (2.16), the more under-determined the problem becomes, which counteracts the reduced smearing. The choice of $K$ thus determines the trade-off between smearing and inequalities-versus-unknowns considerations. In addition, the complexity of solving (2.16) is roughly $O(K^{3.5})$, which is another reason why $K$ should be kept moderate.

Fig. 2.6 illustrates this tradeoff, showing the NMSE as a function of $K$ for various $M$. In Fig. 2.6 two types of random impulse responses were used for the filters: (a)
complex binary antipodal $\pm 1 \pm j$-valued random PN, and (b) normalized white complex Gaussian random variables. Random sparse signals with $\eta = 0.25$ were generated and the reported NMSE for each $K$ is the result of averaging across more than 1000 Monte-Carlo simulations (with respect to the random signals and filters). Three scenarios were considered with $M = 50, 100$ and $200$ sensors, where $t$ was selected such that $12, 25$ and $50$ sensors report $b_m = 1$, respectively. Fig. 2.6 confirms our intuition about the trade-off in the choice of $K$. Fig. 2.6 also shows that the optimal $K$ is an increasing function of $M$, which can be understood by noting that as $M$ increases, the number of inequalities increases, hence one can afford more unknowns. Another point worth noting is that the performance of Gaussian filters (dotted lines) is almost identical to that of binary PN filters. However, binary PN filters are much simpler to implement via cheap linear shift registers, hence preferable to Gaussian filters.

2.6 Relevant Extensions

In this section we discuss some extensions and variations to the proposed frugal sensing scheme.
2.6.1 Model PS Estimation

Here we assume that the PS of the primary signal $x(n)$ can be expressed (approximated) by the model

$$S_x(\omega) = \sum_{\ell=1}^{L} \rho_\ell \Psi_\ell(\omega)$$

(2.21)

where $\{\Psi_\ell(\omega)\}_{\ell=1}^{L}$ are known functions and $\{\rho_\ell\}_{\ell=1}^{L}$ are unknown positive weights. For example, $\{\Psi_\ell(\omega)\}_{\ell=1}^{L}$ can correspond to (overlapping) raised cosine bases which can model transmit-spectra of multicarrier systems [13]. Note that the bandwidths of $\{\Psi_\ell(\omega)\}_{\ell=1}^{L}$ are allowed to be different. The linear model (2.21) can also be used assuming an overcomplete dictionary of bases functions $\{\Psi_\ell(\omega)\}_{\ell=1}^{L}$. Using the model (2.21), the PS can be reconstructed by estimating the weight vector $\rho := [\rho_1, \ldots, \rho_L]^T$.

From (2.21), the autocorrelation function can be expressed as

$$r_x(k) = \sum_{\ell=1}^{L} \rho_\ell \sum_{k=1}^{K-1} \psi_\ell(k)$$

(2.22)

where $\psi_\ell(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi_\ell(\omega) e^{j\omega k} d\omega$ is the inverse discrete-time Fourier transform (IDTFT) of $\Psi_\ell(\omega)$. Defining $v_{m,\ell} := \sum_{k=1}^{K-1} \psi_\ell(k) q_m^*(k)$ and $v_m := [v_{m,1}, \ldots, v_{m,\ell}]^T$, it can be shown that

$$\alpha_m = \sum_{\ell=1}^{L} \rho_\ell v_{m,\ell} = v_m^T \rho$$

(2.23)

The received power measurement bit (2.10) can be written in this case as $b_m = \text{sign}(v_m^T \rho - t_m)$, where the goal here is to estimate $\rho$ from $\{b_m\}_{m=1}^{M}$.

Note that $\rho$ can be upper bounded by $p_{\text{max}}$ (due to the use of AGC at the front-end of the sensor processing chain). This yields the box constraint $\rho \in B$, where $B := \{\rho \in \mathbb{R}^L \mid 0 \leq \rho \leq p_{\text{max}}\}$. Now the LP (2.16) can be modified to estimate $\rho$ as

$$\min_{\rho \in B} \sum_{\ell=1}^{L} \rho_\ell \quad \text{s.t.} \quad b_m(v_m^T \rho - t_m) \geq 0, \ m = 1, \ldots, M$$

(2.24)

where the objective function $\sum_{\ell=1}^{L} \rho_\ell$ induces sparsity in the reconstructed vector $\rho$. 
2.6.2 Line Spectrum Estimation

Consider that the primary signal is composed of $L$ tones (spectral components), $x(n) = \sum_{\ell=1}^{L} \gamma_{\ell} e^{j\omega_{\ell}n}$, where the complex number $\gamma_{\ell}$ represents the magnitude and phase of the $\ell$-th frequency component. Assuming that the phases are independent and uniformly distributed in $[0, 2\pi)$, the autocorrelation function can be expressed as $r_{x}(k) = \sum_{\ell=1}^{L} \rho_{\ell} e^{j\omega_{\ell}k}$, where $\rho_{\ell} = |\gamma_{\ell}|^2$. Given the measurement bits $\{b_{m}\}_{m=1}^{M}$, the goal is to estimate the frequencies and powers $\{\omega_{\ell}, \rho_{\ell}\}_{\ell=1}^{L}$.

In this case, the autocorrelation vector $r_{x}$ can be first estimated (non-parametrically) using the LP (2.16). Then by constructing an estimate for the $K \times K$ autocorrelation matrix, $\hat{R}_{x}$, from the estimated $r_{x}$, multiple signal classification (MUSIC - or any parametric frequency estimation method) can be used to estimate $\{\omega_{\ell}\}_{\ell=1}^{L}$ [71]. Assuming that the eigenvalues of $\hat{R}_{x}$ are sorted in decreasing order, the eigenvectors corresponding to the $L$ largest eigenvalues span the signal subspace, whereas the remaining $K - L$ eigenvectors span the orthogonal space where there is only noise. Defining $u(\omega) := [1, e^{j\omega}, \ldots, e^{j(K-1)\omega}]^T$, and $v_{i}$ as the eigenvector corresponding to the $i$-th strongest eigenvalue of $\hat{R}_{x}$, the frequency estimates $\{\hat{\omega}_{\ell}\}_{\ell=1}^{L}$ using MUSIC are the locations of the $L$ largest peaks of [71]:

$$\hat{S}_{\text{MUSIC}}(\omega) = \frac{1}{\sum_{i=L+1}^{K} |u(\omega)^{H}v_{i}|^2}$$

Finally, given the frequency estimates $\{\hat{\omega}_{\ell}\}_{\ell=1}^{L}$, the powers $\{\rho_{\ell}\}_{\ell=1}^{L}$ can be estimated via least squares [71]:

$$\{\hat{\rho}_{\ell}\}_{\ell=1}^{L} = \arg \min_{\{\rho_{\ell}\}_{\ell=1}^{L}} \sum_{k=0}^{K-1} \left| \hat{r}_{x}(k) - \sum_{\ell=1}^{L} \rho_{\ell} e^{j\hat{\omega}_{\ell}k} \right|^2$$

(2.26)

2.6.3 Robust Sensing: Outlier-Contaminated Model

For many reasons, such as sensor malfunctions and decoding errors in the sensor-to-FC communication links, the received power measurement bits at the FC can be flipped. This can result in an inconsistent constraint set $P_{M}$ in (2.16). Modeling such cases as outliers, the sparse regression framework [30,31] can be used to robustify (2.16) against the outlier-contaminated model.
A slack (outlier) variable $d_m \geq 0$, that represents the possible error in the measurement or reporting of $\alpha_m$, is added to the constraints of type $q_m^T r_x \geq t_m$ if $b_m = 1$ ($q_m^T r_x < t_m$ if $b_m = -1$), such that they become $q_m^T r_x + d_m \geq t_m$ (resp. $q_m^T r_x - d_m < t_m$). Then, a sparsity-inducing penalty $\|d\|_1 = \sum_{m=1}^{M} d_m$ is added to the cost function (where $d := [d_1, \ldots, d_M]^T$) to promote sparsity among the slack variables, in order to (approximately) minimize the number of inconsistent inequalities. Hence, problem (2.16) is modified to the following robust LP:

$$
\min_{r_x \in \mathcal{P}, d} \quad r_x(0) + \lambda \sum_{m=1}^{M} d_m \\
\text{s.t.} \quad q_m^T r_x + d_m \geq t_m, \quad \text{if } b_m = 1 \\
q_m^T r_x - d_m < t_m, \quad \text{if } b_m = -1 \\
F r_x \geq 0, \quad d_m \geq 0, \quad m = 1, \ldots, M
$$

(2.27)

where $\lambda \geq 0$ is a tuning parameter that controls the level of sparsity. Note that (2.27) is a LP in $r_x$ and $d$. It is worth mentioning that using the $\ell_1$-norm for robust estimation was introduced in [31], see also [30].

Figure 2.7: Illustrative example for robust frugal sensing using (2.27).

In Fig. 2.7, we consider a similar setup to that used for Fig. 2.3, assuming a sparse PS (solid line), $M = 100$, and $K = 24$. To model for inconsistencies and errors in
the reported measurement bits, an independent uniform random variable is added to each $\hat{\alpha}_m$. As a result, the FC received 20 wrong bits from the sensors (i.e., 20 reversed inequalities); 14 ‘-1’ bits are received as ‘1’, and 6 ‘1’ bits are received as ‘-1’. This resulted in an infeasible problem (2.16). The estimated PS that has been obtained by solving the robust LP (2.27) is plotted as the dotted line, where the tuning parameter $\lambda$ was set to 1. It is worth noting that the resulting sparse $d$ after solving (2.27) included only 16 nonzero entries (representing inconsistencies). If the true measurement bits are received by the FC (i.e., outlier-free model), the estimated PS obtained by solving (2.27) is given as the dashed line. Note that in this case problem (2.27) is equivalent to (2.16), since the added sparsity-inducing penalty $\lambda \sum_{m=1}^{M} d_m$ in the objective of (2.27) gives $d = 0$ (for $\lambda$ sufficiently large). As shown in the figure, the quality of the PS estimate using the robust LP (2.27) is very satisfactory, considering that 20% of the received measurement bits were flipped.

### 2.7 Active Sensing

The choice of $\{t_m\}_{m=1}^{M}$ is very important for achieving good estimates in (2.16). All sensing approaches discussed so far are passive in the sense that the thresholds $\{t_m\}_{m=1}^{M}$ are fixed and pre-assigned to sensors. What if the thresholds could be actively adapted online, based on the reports received from a subset of sensors up to a given point in time? This could yield a significant payoff in terms of sensing accuracy, provided that there is a way for the FC to communicate threshold information back to the sensors.

Consider a time-slotted bi-directional communication link between the $M$ sensors and the FC, comprising $M$ time slots. At the beginning of each time slot $m \in \{1, \ldots, M\}$, the FC sends the threshold $t_m$ to sensor $m$. Sensor $m$ then compares the measured $\hat{\alpha}_m$ with $t_m$, and responds with either $b_m = 1$ or $b_m = -1$ within the same slot. The final feasible region for $r_x$ when the FC receives all measurement bits $\{b_m\}_{m=1}^{M}$ is defined by the polyhedron $\mathcal{P}_M = \mathcal{P} \cap \{r_x \mid b_m(q_m^T r_x - t_m) \geq 0, \ m = 1, \ldots, M\}$.

The volume of the feasible region $\mathcal{P}_M$ gives a measure of ignorance or uncertainty about $r_x \in \mathcal{P}_M$; a small $\mathcal{P}_M$ implies that $r_x$ is localized to within a small neighborhood, whereas a large $\mathcal{P}_M$ means that there is still much uncertainty about $r_x$. In other words, a smaller feasible region $\mathcal{P}_M$ translates to higher accuracy in localizing $r_x$. Thus, our
objective here is to adaptively select the thresholds \( \{t_m\}_{m=1}^M \) to ensure that \( \mathcal{P}_M \) is as small as possible.

### 2.7.1 Basic Active Sensing Algorithm

Before introducing the proposed active sensing algorithm, we first discuss how to compute the Chebyshev center (CC) of a bounded polyhedron \( \mathcal{C} := \{x \mid \mathbf{a}_i^T x \leq c_i, \ i = 1, \ldots, L\} \), defined by a set of \( L \) linear inequalities. The CC is the center of the maximum ball that can be inscribed inside \( \mathcal{C} \), and it can be found by solving the LP [22, Sec. 8.5.1]

\[
\begin{align*}
\max_{R \geq 0, x} & \quad R \\
\text{s.t.} & \quad \mathbf{a}_i^T x + R||\mathbf{a}_i||_2 \leq c_i, \ i = 1, \ldots, L
\end{align*}
\]

The LP (2.28) finds the point inside \( \mathcal{C} \) that has the maximum distance to the closest point on the boundary hyperplanes defining \( \mathcal{C} \) (i.e., the exterior of \( \mathcal{C} \)).

Given the initial polyhedron \( \mathcal{P}_0 = \mathcal{P} \) from (2.6), its CC \( \mathbf{x}^{(0)} \), and \( \{\mathbf{q}_m\}_{m=1}^M \), the proposed active sensing algorithm can be explained as follows

For each time-slot/sensor \( m = 1, \ldots, M \), do

1. Set the threshold \( t_m = \mathbf{q}_m^T \mathbf{x}^{(m-1)} \), and send it to sensor \( m \) requesting its measurement bit \( b_m \).

2. Upon receiving \( b_m \), update the feasible polyhedron

\[
\mathcal{P}_m := \begin{cases} 
\mathcal{P}_{m-1} \cap \{x \mid \mathbf{q}_m^T x \geq t_m\} & \text{if } b_m = 1 \\
\mathcal{P}_{m-1} \cap \{x \mid \mathbf{q}_m^T x < t_m\} & \text{if } b_m = -1
\end{cases}
\]

3. Compute the CC \( \mathbf{x}^{(m)} \) of \( \mathcal{P}_m \).

The final autocorrelation estimate can be obtained as \( \hat{\mathbf{r}}_x = \mathbf{x}^{(M)} \), i.e., the CC of \( \mathcal{P}_M \).

Note that at the second step of the active sensing algorithm, the half-space \( \{x \mid b_m \mathbf{q}_m^T (x - \mathbf{x}^{(m-1)}) \leq 0\} \) is cut-off from the feasible region. The selection of the
threshold \( t_m = q_m^T x^{(m-1)} \) ensures that the CC of \( \mathcal{P}_{m-1} \) is a point in the trimmed half-space. Ideally, one would want half the volume of the feasible region to be cut-off after each received bit; however, this is an NP-hard problem, even if the filter of each sensor is a design parameter [27]. As an approximation, ensuring that the omitted half-space includes the CC guarantees that a large portion of \( \mathcal{P}_{m-1} \) is omitted from the feasible region, and that the new polyhedron \( \mathcal{P}_m \) is considerably smaller than \( \mathcal{P}_{m-1} \). It is worth noting that similar cutting-plane methods have been used in solving general convex and quasi-convex optimization problems [21, 28, 34].

An illustrative example for the active sensing algorithm in \( \mathbb{R}^2 \) is shown in Fig. 2.8 for \( M = 4 \). The grey-shaded region in the figure represents the union of the 4 planes inside \( \mathcal{P} \) that are cut-off from the feasible region after the 4 measurement bits are received, whereas the final feasible region \( \mathcal{P}_4 \) is unshaded. The figure shows that the unknown vector \( r_x \) is localized in a small region, implying a small error in the estimate \( \hat{r}_x \).

Figure 2.8: Illustrative example for the active sensing algorithm.

**Claim 1** The CC \( x^{(M)} \) converges linearly to the true autocorrelation vector \( r_x \) as \( M \to \infty \) using the proposed active sensing algorithm, under certain independence conditions (see Appendix A.4) on \( \{q_m\}_{m=1}^{M} \).

The proof can be found in Appendix A.4.
Non-negativity constraints

As mentioned in Section 2.2, the windowed estimate of the PS that is obtained by taking the Fourier transform of a finite $K$-lag autocorrelation is not guaranteed to be nonnegative at all frequencies. Therefore, including $F r_x \geq 0$ in the set of constraints when estimating $r_x$ may prevent the convergence of the CC $x^{(M)}$ to $r_x$ as $M \rightarrow \infty$. Instead, we consider including a relaxed linear non-negativity constraint. Define the Vandermonde vector $u(\omega) := [1, e^{j\omega}, \ldots, e^{j\omega(K-1)}]^T$. Any autocorrelation matrix $R_x$ is positive semidefinite. In unknown autocorrelation vector $r_x$, $R_x \succeq 0$ is a linear matrix inequality (LMI) instead of a ‘plain’ linear inequality. LMIs are convex, but entail higher computational cost than regular linear inequalities. Note however that

$$R_x \succeq 0 \implies u(\omega)^H R_x u(\omega) \geq 0$$

$$\implies \sum_{k=-K+1}^{K-1} (K - |k|) r_x(k) e^{-j\omega k} \geq 0$$

$\forall \omega \in [0,2\pi)$. Defining the diagonal matrix $P := \text{diag}([K,K-1,\ldots,1,K-1,\ldots,1])$, it is clear that $R_x \succeq 0$ implies (albeit is not equivalent to) $F P r_x \geq 0$. Including the relaxed linear non-negativity constraint $F P r_x \geq 0$ to the set of constraints reduces the feasible region and ensures convergence $x^{(M)} \rightarrow r_x$, unlike the constraint $F r_x \geq 0$. If the number of sensors $M$ is small, however, it is recommended to include the (more strict) constraint $F r_x \geq 0$ in order to decrease the under-determinacy of the estimation problem. The effects of including $F r_x \geq 0$ or $F P r_x \geq 0$ on the performance are further illustrated in Section 2.7.4. Whereas the semidefinite constraint $R_x \succeq 0$ can be included instead of the relaxed constraint $F P r_x \geq 0$, simulations have shown that the estimation performance using the linear constraint $F P r_x \geq 0$ is almost identical to using the LMI constraint $R_x \succeq 0$, at much lower complexity.

Pruning constraints

The number of linear inequalities defining $P_m$ increases at each iteration of the algorithm, and hence the computational effort to compute $x^{(m)}$ increases. For a polyhedron that is defined by $L$ linear inequalities, one approach is to keep only a fixed number $J \leq L$ of the most relevant inequality constraints while dropping the other $L - J$ less
relevant or redundant inequalities \cite{21,28}. With proper choice of \( J (> 10K) \), simulations have shown a negligible effect on the performance, at a dramatic decrease in the total computation time of the active sensing algorithm.

### 2.7.2 Low-Complexity Active Sensing

Instead of using the CC of the polyhedron \( P_{m-1} \) in computing \( t_m \), other options include the center of gravity, the center of the maximum volume inscribed ellipsoid, and the analytic center \cite{21}. The worst-case complexity of each method can be captured by the worst-case number of iterations (i.e., \( M \)) required to achieve an \( \epsilon \)-error estimate, \( \| \tilde{r}_x - r_x \|_2 \leq \epsilon \), in addition to the complexity of computing each center at each iteration. Using the center of gravity, the volume of the polyhedron is guaranteed to reduce by at least 37\% at each iteration of the algorithm, and the number of iterations required to achieve an \( \epsilon \)-error estimate is at most \( O(K \log(1/\epsilon)) \). However, computing the center of gravity of a polyhedron described by a set of linear inequalities is NP-hard \cite{21}. The number of iterations required using the center of the maximum volume inscribed ellipsoid is at most \( O(K^2 \log(1/\epsilon)) \) \cite{21}, and it can be computed by solving a convex problem \cite{22}, but its computation is prohibitive for large \( K \). The analytic center (AC) of a bounded polyhedron is the point that maximizes the product of distances to the defining hyperplanes, and is efficiently computed by minimizing the convex logarithmic barrier function \cite{22, Sec. 8.5.3}, while at most \( O(K^2/\epsilon^2) \) iterations are required to achieve error \( \leq \epsilon \) \cite{21}. Whereas no similar worst-case analysis has been developed for the CC (the CC can be strongly affected by scaling and affine transformations of coordinates), exhaustive simulations showed that the performance of the active sensing algorithm using the CC is at least as good as that which uses the AC, on average, with much smaller computation time to solve the LP \eqref{2.28}.

**Approximate AC:** In order to further decrease the computational complexity of the proposed active sensing algorithm, we consider efficiently computing an approximate AC, instead of solving the LP \eqref{2.28} to compute the CC, at each iteration. The AC of a bounded polyhedron \( C = \{ x \mid a_i^T x \leq c_i, \ i = 1, \ldots, L \} \) is obtained by minimizing the logarithmic barrier function \( \Omega(x) := - \sum_{i=1}^L \log(c_i - a_i^T x) \) \cite{22, Sec. 8.5.3}, which can be computed via Newton’s method \cite{22, Sec. 9.5}. In Newton’s method, multiple damped
Newton steps of type

\[ x = x - \beta \left( \nabla^2 \Omega(x) \right)^{-1} \nabla \Omega(x) \]  

(2.29)

are used to find the minimizer of \( \Omega(x) \), where

\[
\nabla \Omega(x) = \sum_{i=1}^{L} \frac{a_i}{c_i - a_i^T x}, \\
\nabla^2 \Omega(x) = \sum_{i=1}^{L} \frac{a_i a_i^T}{(c_i - a_i^T x)^2}
\]

(2.30)

and \( \beta \geq 0 \) is the step-size, which can be selected as \( \beta = 1 \) (among other options). The most computationally expensive operation in each Newton step is inverting the \((2K - 1) \times (2K - 1)\) Hessian matrix \( \nabla^2 \Omega(x) \). Therefore, an approximate AC can be computed using one (or few) damped Newton step(s) of type (2.29). Starting from \( x^{(0)} \), which can be any point inside the initial \( \mathcal{P} \), the cheap computation of the approximate AC can replace the CC computation in the third step of the proposed active sensing algorithm. The linear convergence \( x^{(M)} \rightarrow \mathbf{r}_x \) as \( M \rightarrow \infty \) using the approximate AC, and the tradeoff between the complexity and estimation performance as more Newton steps are used in each iteration, are shown in the simulations in Section 3.6.3. It is worth mentioning that linear convergence using approximate ACs, in the context of cutting plane methods for convex feasibility problems, was proven in [34] under certain conditions.

2.7.3 Sensor Polling with Pre-Assigned Thresholds

The proposed active sensing algorithm requires that the FC sends a computed threshold to each sensor at the beginning of its allocated time-slot. To avoid this threshold communication overhead, it may be more appealing to pre-assign the thresholds to the sensors. We propose that each sensor randomly selects \( t_m \) from a Gaussian distribution with the same mean as \( \alpha_m \), which can be considered a random variable with respect to the random signal and the PN filter of each sensor. Consider now that at the beginning of each time-slot the FC polls one sensor, which in turn responds with its feedback bit within the same time slot.

Let \( \mathcal{J} \) denote the set of sensors that have sent their feedback bits while \( \bar{\mathcal{J}} \) denotes the set of remaining sensors. The proposed sensor polling algorithm with pre-assigned thresholds can be described as follows:
Given the sets $\{q_m\}_{m=1}^M, \{t_m\}_{m=1}^M$, the initial polyhedron $P_0$, and its CC $x^{(0)}$.

Initialize $k = 1$. While $k \leq M$, do

1. For each $m \in \bar{J}$, find the distance between the hyperplane $\{x \mid q_m^T x - t_m = 0\}$ and the CC $x^{(k-1)}$:
   \[d_m = \frac{|q_m^T x^{(k-1)} - t_m|}{||q_m||}\] (2.31)

2. Select the sensor $m^* = \arg \min d_m$ to be polled requesting its measurement bit $b_{m^*}$.

3. Upon receiving $b_{m^*}$, delete $m^*$ from $\bar{J}$, add it to $J$, and update the polyhedron:
   \[P_k := \begin{cases} P_{k-1} \cap \{x \mid q_{m^*}^T x \geq t_{m^*}\} & \text{if } b_{m^*} = 1 \\ P_{k-1} \cap \{x \mid q_{m^*}^T x < t_{m^*}\} & \text{if } b_{m^*} = 0 \end{cases}\] (2.32)

4. Find the CC $x^{(k)}$ of $P_k$ by solving the LP (2.28).

5. Terminate if $x^{(k)} = x^{(k-1)}$ for the last $\tau$ consecutive iterations ($\tau$ is a design parameter). Else, increment $k$ and repeat.

Finally, an estimate of $\tilde{r}_x$ can be directly obtained as $\tilde{r}_x = x^{(k)}$, where $\bar{k}$ is the last $k$ at the algorithm’s termination.

Note that by polling sensor $m^*$ with the smallest distance between $x^{(k-1)}$ and the hyperplane $\{x \mid q_m^T x - t_m = 0\}$ at the $k$-th iteration, we try to ensure that the CC of $P_{k-1}$ is very close to the hyperplane which defines the trimmed half-space. The CC $x^{(k-1)}$ can be inside or outside the cut-off half-space. Similar to the active sensing algorithm, this ensures that a large portion of $P_{k-1}$ is omitted from the feasible region, and that the updated polyhedron $P_k$ is considerably smaller than $P_{k-1}$.

For sufficiently large $M$, the performance of the sensor polling algorithm is similar to the performance of the active sensing algorithm since $d_{m^*} \to 0$ as $M \to \infty$. In other words, if $M$ is sufficiently large, then at each iteration $k$ there exists a sensor $m \in \bar{J}$ with $d_m \approx 0$, and thus the sensor polling algorithm becomes almost identical to the active sensing algorithm. For small $M$, however, as more sensors are polled, it becomes more challenging to find sensors $m \in \bar{J}$ with small $d_m$. After some $k$ iterations, $d_{m^*}$ becomes relatively large such that the half-space information obtained by polling any of the remaining sensors $m \in \bar{J}$ is redundant, and thus the feasible region polyhedron will
not decrease (i.e., \( P_k = P_{k-1} \)). When this limit is reached, the CC does not change, and thus the sensor polling algorithm can be terminated prematurely. Note that we need to check that \( x(k) = x(k-1) \) for the last \( \tau \geq 1 \) iterations, since for some scenarios polling sensor \( m^* = \arg \min d_m \) does not change the Chabyshev center, whereas polling another sensor \( \bar{m} \in \bar{J} - \{ m^* \} \), where \( d_{\bar{m}} > d_{m^*} \), can yield a smaller polyhedron with a different CC. A small \( \tau = 5 \sim 10 \) is apparently sufficient. Due to the larger feasible region obtained at the termination of the sensor polling algorithm with limited \( M \), the estimation error with the sensor polling algorithm is generally larger than that obtained with the active sensing algorithm.

### 2.7.4 Numerical Results

To measure the quality of the estimated autocorrelation \( \hat{r}_x \) using the proposed active sensing algorithm, we use the normalized mean squared error (NMSE), defined as

\[
\text{NMSE}(m) := \mathbb{E} \left[ \frac{||r_x - \hat{r}_x^{(m)}||^2}{||r_x||^2} \right],
\]

where \( \hat{r}_x^{(m)} \) is the estimate of \( r_x \) when the FC receives the \( m \)-th bit, \( m = 1, \ldots, M \). The expectation is taken with respect to the random signals and the random impulse responses of the FIR filters, obtained via more than 100 Monte-Carlo simulations.

In Fig. 2.9, the primary signal is assumed to be a combination of 10 equispaced raised-cosine functions with roll-off factor = 0.5. The power coefficients of three raised-cosine functions were set to zero, while the remaining seven were drawn from a uniform distribution between 0.2 and 1 in each simulation run. A filter length \( K = 10 \) was used at all sensors. The NMSE is plotted as a function of the number of received bits \( m \) for the active sensing algorithm using the CC, the AC, and the approximate AC with 1, 5, and 10 Newton step(s). The figure shows that \( \text{NMSE}(m) \to 0 \) (i.e., \( \hat{r}_x^{(m)} \to r_x \)) as \( m \) increases for all considered algorithms, when using the relaxed linear non-negativity constraint \( \text{FP} r_x \geq 0 \), and confirms the linear rate of convergence. The figure also shows that the NMSE obtained when using the CC is better than using the AC, at much lower computation time. As expected, increasing the number of Newton steps of the low-complexity algorithm of Section 2.7.2 (from 1 to 5 to 10 Newton steps for each center computation) yields a better approximation to the AC and better performance, but the computational complexity is also increasing. When the (strict) non-negativity constraint \( \text{Fr}_x \geq 0 \) was included, the NMSE of the active sensing algorithm was better
than that obtained with the relaxed non-negativity constraint \( F^\top r_x \geq 0 \) for \( M < 100 \). But as more bits are received, the NMSE with the strict non-negativity constraint saturates at 0.0047, whereas with or even without the relaxed non-negativity constraint the NMSE continues decreasing to zero. This is because the strict constraint \( F^\top r_x \geq 0 \) is not valid in general (the FT of a truncated autocorrelation sequence is not necessarily non-negative), underscoring that including this strong constraint is only advisable for up to moderate \( M \); for higher \( M \) it should be omitted to enable convergence.

A different setting with filter length \( K = 12 \) is considered in Fig. 2.10. The figure shows that \( \text{NMSE}(k) \to 0 \) (i.e., \( \hat{r}_x^{(k)} \to r_x \)) as \( k \) increases with the active sensing algorithm, and confirms the linear rate of convergence. When the strict non-negativity constraint \( F^\top r_x \geq 0 \) was included to the initial polyhedron \( P_0 \), the NMSE of the active sensing algorithm was better than the default setting without the non-negativity constraint for \( M < 150 \). But as more bits are received, the NMSE with the non-negativity constraint saturates at 0.007, whereas the NMSE without the non-negativity constraint continues decreasing to zero. As mentioned earlier, this shows that including the non-negativity constraint improves the performance only up to moderate \( M \); for higher \( M \)
Figure 2.10: Performance of the active sensing and sensor polling algorithms.

it should be omitted to enable convergence. When the filter length was increased to $K = 16$, more sensors were needed to achieve the same NMSE level obtained with $K = 12$, due to the additional number of unknowns that need to be estimated. Note that, on the other hand, the resolution of the estimated PS increases with $K$. For the sensor polling algorithm, each $t_m$ was randomly drawn from a Gaussian with the same mean and variance as the random $\alpha_m$, obtained via Monte-Carlo simulations. Fig. 2.10 shows that, for the first 80 polled sensors, the NMSE of the sensor polling algorithm is similar to the NMSE of the active sensing algorithm; thereafter the NMSE of the sensor polling algorithm saturates at 0.017. The NMSE obtained using equal thresholds (with 100 out of 300 sensors reporting $b_m = 1$), and applying the $r_x$-estimation method (2.16), is also plotted in Fig. 2.10 yielding a larger NMSE compared to the active sensing and sensor polling algorithms. It is worth mentioning that almost identical results were obtained when only the most relevant $J = 150$ inequality constraints were considered when solving (2.25).
Chapter 3

Frugal Sensing: Maximum Likelihood Formulation

3.1 Introduction

The underlying assumption which enables using the LP formulation in Chapter 2 is that the power measurement prior to quantization at each sensor is accurate enough to avoid flipping the inequality. Frugal sensing is revisited in this chapter from a statistical estimation point of view. Taking into account frequency-selective fading and insufficient sample averaging considerations, it is shown that the distribution of the error in the soft power estimates (prior to thresholding) can be approximated by a Gaussian distribution. The Gaussian distribution of the errors is then exploited in a maximum likelihood (ML) formulation that optionally includes a sparsity-inducing penalty term. By deriving the Cramér-Rao bound (CRB), it is shown that the ML estimate asymptotically achieves the CRB as the number of reporting sensors grow.

The ML formulation is then extended to the case where the PS is modeled as a weighted sum of candidate spectral density primitives with unknown weights, the case of signals with line spectrum, and the active sensing (adaptive thresholding) setting. In order to reduce the number of bits transmitted from the sensors to prolong battery life and minimize communication overhead, an ML/CRB-driven censoring scheme is proposed, where only sensors that provide the most useful information bits are permitted to send, while other sensors remain silent. Simulations show that satisfactory wideband
PS estimates can be obtained with passive ML sensing from few bits, and much better performance can be attained using active sensing, even when a significant number of bits are flipped due to pre-quantization measurement errors at the sensors. Simulations also demonstrate the estimation performance improvement attained by exploiting the prior information on the signal in a parametric ML estimation formulation.

3.2 Network Sensing Model

The same network sensing setting and sensor measurement chain of Chapter 2 are considered here (see Fig. 2.1 and Fig. 2.2). After the AGC and ADC stages, the received signal at sensor \( m \in \{1, \ldots, M\} \) can be expressed as

\[
y_m(n) = \sum_{\ell=0}^{L-1} h_m(\ell)x(n - \ell)
\]  

where \( x(n) \) is the primary WSS signal and \( \{h_m(\ell)\}_{\ell=0}^{L-1} \) represent the \( L \)-tap linear finite impulse response (FIR) fading channel. It is assumed in this chapter that \( \{h_m(\ell)\}_{\ell=0}^{L-1} \) are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero-mean and variance \( 1/L \), and that the channel is time-invariant for the sensing epoch. Note that frequency-selective fading is not assumed to be mitigated by averaging the measurements over a long period of time in this chapter, unlike the case in Chapter 2.

Sensor \( m \in \{1, \ldots, M\} \) then passes \( y_m(n) \) through a wideband FIR filter with impulse response \( g_m(n) \), yielding the output sequence \( z_m(n) = \sum_{k=0}^{K-1} g_m(k)y_m(n - k) \). Let \( \alpha_m \) denote the average power of the filter’s output sequence \( z_m(n) \) (as in Chapter 2). Each sensor then computes \( \hat{\alpha}_m \) by sample averaging as (2.9) to estimate the average power of \( z_m(n) \), and compares it with the threshold \( t_m \) to obtain the measurement bit:

\[
b_m = \text{sign}(\hat{\alpha}_m - t_m).
\]

Define \( \tilde{z}_m(n) := \sum_{k=0}^{K-1} g_m(k)x(n - k) \) as the convolution of the primary signal \( x(n) \) and the random filter \( g_m(n) \) (i.e., ignoring fading) and let \( \tilde{\alpha}_m \) denote the average power of \( \tilde{z}_m(n) \). Hence, it can be shown that

\[
\hat{\alpha}_m = q_m^T r_x
\]
where the signal autocorrelation vector \( r_x \) is defined in (2.2) and the filter’s deterministic-autocorrelation vector \( q_m \) is defined in (2.13). Recall that the assumption in Chapter 2 is that \( \alpha_m = \tilde{\alpha}_m = q_m^T r_x \). Unlike the case in Chapter 2, \( z_m(n) \neq \tilde{z}_m(n) \) and \( \alpha_m \neq \tilde{\alpha}_m \) in this chapter due to frequency-selective fading.

Define the error due to the fading channel at sensor \( m \) as \( \tilde{e}_m := \alpha_m - \tilde{\alpha}_m \). Appendix A.5 shows that for large number of channel taps \( L \), the errors \( \{\tilde{e}_m\}_{m=1}^M \) can be approximated as i.i.d. zero-mean Gaussian random variables using the Lyapunov central limit theorem [18, pp. 371]. The corresponding approximate variances, \( \{\tilde{\sigma}_m^2\}_{m=1}^M \), are given in the appendix, and are shown to be inversely proportional to \( L \). Exhaustive simulations have indicated that approximating the distribution of \( \tilde{e}_m \) with a Gaussian distribution is a close approximation, even for relatively small \( L \). Fig. 3.1 compares the actual distribution of \( \tilde{e}_m \) obtained via Monte-Carlo simulations and the corresponding Gaussian distribution approximation, for different channel taps \( L \in \{10, 40, 100, 1000\} \), and filter length \( K = 20 \). The figure shows that the Gaussian distribution is a reasonable approximation for relatively small \( L = 10 \), and a very accurate one for relatively large \( L = 1000 \). The figure also shows the decrease in the variance of \( \tilde{e}_m \) as \( L \) increases.

Figure 3.1: The Gaussian distribution approximation of \( \tilde{e}_m \) with different \( L \) taps.
The estimation errors due to insufficient sample averaging, $\tilde{e}_m := \hat{\alpha}_m - \alpha_m$, can also be modeled as i.i.d. Gaussian random variables with zero-mean and variances $\{\bar{\sigma}_m^2\}_{m=1}^M$, by the central limit theorem. This means that $\hat{\alpha}_m$ can be modeled as $\hat{\alpha}_m = \tilde{\alpha}_m + \epsilon_m$, where $\epsilon_m := \tilde{e}_m + e_m$ is a Gaussian random variable with zero mean and variance $\sigma_m^2 = \tilde{\sigma}_m^2 + \bar{\sigma}_m^2$. Thus, the received power measurement bit of each sensor can be expressed as

$$b_m = \text{sign}(q_m^T r_x + \epsilon_m - t_m) \quad (3.3)$$

Define the vector $b := [b_1, \ldots, b_M]^T$. The objective, as in Chapter 2, is to estimate the PS of the signal $x(n)$ at the FC from the measurement bits $b$. In the next section, we consider the passive sensing case where the thresholds $\{t_m\}_{m=1}^M$ are fixed and pre-assigned to sensors, followed by the active sensing case where $\{t_m\}_{m=1}^M$ are adapted and communicated to sensors online in Section 3.5.

3.3 Maximum-Likelihood Formulation

The assumption that $\epsilon_m \approx 0$, enables using the linear inequality $b_m(q_m^T r_x - t_m) \geq 0$ as a (hard) constraint in the LP formulation (2.16) in Chapter 2. As the error variances $\{\sigma_m^2\}_{m=1}^M$ increase, the LP formulation (2.16) becomes inaccurate, and the constraints may become inconsistent as the number of flipped bits due to errors increase. In order to lift this limitation, the Gaussian distribution of $\{e_m\}_{m=1}^M$ is exploited in deriving a more flexible and powerful ML formulation.

An ML estimate of the $K$-lag autocorrelation $r_x$ can be obtained by exploiting the Gaussian distribution of $\{e_m\}_{m=1}^M$ as follows. The probability that $b_m = 1$ given $r_x$ can be expressed in terms of the $Q$-function as

$$p[b_m = 1|r_x] = p[q_m^T r_x + \epsilon_m \geq t_m] = Q \left( \frac{-q_m^T r_x - t_m}{\sigma_m} \right) \quad (3.4)$$

Define $\mathcal{M}_+ := \{m|b_m = 1\}$ and $\mathcal{M}_- := \{m|b_m = -1\}$. Since the errors $\{e_m\}_{m=1}^M$ are assumed i.i.d Gaussian random variables with zero-mean and variance $\{\sigma_m^2\}_{m=1}^M$, the
probability mass function (pmf) of $b$, given the autocorrelation $r_x$, is given as

$$p[b|r_x] = \prod_{m \in M_+} p(q_m^T r_x + e_m \geq t_m) \prod_{m \in M_-} p(q_m^T r_x + e_m < t_m)$$

$$= \prod_{m \in M_+} Q\left(-\frac{(q_m^T r_x - t_m)}{\sigma_m}\right) \prod_{m \in M_-} Q\left(\frac{q_m^T r_x - t_m}{\sigma_m}\right)$$

$$= \prod_{m=1}^M Q\left(-\frac{b_m(q_m^T r_x - t_m)}{\sigma_m}\right)$$

(3.5)

The log-likelihood function can be written as

$$\log p[b|r_x] = \sum_{m=1}^M \log Q\left(-\frac{b_m(q_m^T r_x - t_m)}{\sigma_m}\right)$$

(3.6)

Similar to the LP (2.16), the non-negativity constraint $F r_x \geq 0$ is essential in obtaining a good estimate of $r_x$. Furthermore, the PS sparsity can be exploited by minimizing $r_x(0)$, as discussed in Chapter 2. Therefore, an ML formulation, with a sparsity-inducing penalty term, is given as

$$\max_{r_x \in P} \sum_{m=1}^M \log Q\left(-\frac{b_m(q_m^T r_x - t_m)}{\sigma_m}\right) - \lambda r_x(0)$$

s.t. $F r_x \geq 0$

(3.7)

where $\lambda \geq 0$ is a tuning parameter that controls the sparsity of the PS estimate. Since the $Q$-function is log-concave [22, pp. 104], and $P \cap \{F r_x \geq 0\}$ is a convex set in $r_x$, problem (3.7) is convex and can be solved efficiently using interior-point algorithms [22]. Note that the maximizer of problem (3.7) always exists since $r_x$ is bounded in $P$.

Omitting the inequality constraint $F r_x \geq 0$ and the sparsity-inducing penalty ($\lambda = 0$) in (3.7), a sufficient condition for consistency of the ML estimate, i.e., for $\hat{r}_x$ to converge in probability to the true $r_x$ as $M \to \infty$, is that the second-moment matrix $R_q = E[q_m q_m^T]$ is full-rank [62] (see also [74]). This condition also ensures statistical identifiability of $r_x$ from the measurements. Clearly, for i.i.d. PN filter impulse responses $\{g_m(n)\}$ as in (2.8), the matrix $R_q$ is indeed diagonal and full-rank. Whereas including the sparsity-inducing penalty $\lambda r_x(0)$ and the constraint $F r_x \geq 0$ can yield better estimates for (relatively small) finite $M$, this can prevent convergence of $\hat{r}_x$ to
the true \( r_x \) as \( M \to \infty \), if \( \lambda \) is too large or if the true truncated \( K \)-lag autocorrelation does not satisfy \( \text{Fr}_x \geq 0 \).

**Cramér-Rao Bound**

The CRB gives a lower bound on the estimation error of any unbiased estimator and is asymptotically attained by the ML estimator \([43]\). Assuming that the true \( K \)-lag autocorrelation satisfies \( \text{Fr}_x \geq 0 \), and no prior PS sparsity information (i.e., \( \lambda = 0 \) in (3.7)), the CRB on the ML estimate \( \hat{r}_x \) can be derived as follows. First, we note that inequality constraints do not affect the CRB; only equality constraints yield a CRB that is lower than the unconstrained one \([33]\). Defining \( \mu_m := \log Q \left( \frac{-b_m (q_m^T r_x - t_m)}{\sigma_m} \right) \), the Fisher information matrix (FIM) with respect to \( r_x \) is computed as \([43]\)

\[
J := -E \left[ \nabla^2 \log p(b|r_x) \right] = -\sum_{m=1}^{M} E[\nabla^2 \mu_m] = Q^T D Q \tag{3.8}
\]

where \( Q = [q_1, \ldots, q_M]^T \) and

\[
D_{m,m} = \frac{e^{-(q_m^T r_x - t_m)^2}}{2\pi \sigma_m^2} \left[ \frac{1}{Q \left( \frac{q_m^T r_x - t_m}{\sigma_m} \right)} + \frac{1}{Q \left( \frac{t_m - q_m^T r_x}{\sigma_m} \right)} \right] \tag{3.9}
\]

is the \( m \)-th diagonal entry of the diagonal matrix \( D \) (the expectation in (3.8) is taken with respect to \( b \)). The FIM derivation is included in Appendix A.6. Finally, the CRB is obtained as \([1] \) \( \text{Trace}(J^{-1}) \). It is observed from (3.8) that the CRB depends on \( r_x \), \( \{q_m\} \), and \( \{t_m\} \). It can be shown that the thresholds that minimize the CRB, given \( r_x \) and \( \{q_m\} \), are obtained as \( t^*_m = q_m^T r_x \), for \( m = 1, \ldots, M \), i.e., \( t^*_m \) depends on \( r_x \) and \( q_m \) \([29]\).

If an upper bound on the number of nonzero frequency ‘bins’ (DFT points) in the PS is known \textit{a-priori}, it can be exploited by adding the cardinality constraint \( || \text{Fr}_x ||_0 \leq \kappa \), where the \( \ell_0 \)-(quasi)norm is the number of nonzero entries of \( \text{Fr}_x \). Note that this nonconvex \( \ell_0 \)-norm constraint can be relaxed by adding the sparsity-inducing penalty \( \lambda r_x(0) \) to the cost function as in (3.7). Computing the CRB with a cardinality constraint

\footnote{Assuming that \( M \geq (2K - 1) \) such that the FIM \( J \) is nonsingular. If \( J \) is singular, the pseudo-inverse can be used instead \([15]\), in the sense that it will yield a valid (albeit generally optimistic) lower bound.}
has been considered in [16] for a linear model and extended in [74] for a nonlinear model. It is shown in [16] and [74] that the CRB equals the unconstrained bound if the parameter to be estimated satisfies the cardinality constraint with strict inequality; whereas if the cardinality constraint is satisfied with equality, then the CRB coincides with the ‘clairvoyant’ one for when the nonzero locations are perfectly known. To compute the CRB if 
\[ \| \mathbf{F}_x \|_0 = \kappa \], let the rows of the \((2K - 1 - \kappa) \times (2K - 1)\) matrix \( \mathbf{\hat{F}} \) correspond to the \((2K - 1 - \kappa)\) rows of the \((2K - 1) \times (2K - 1)\) DFT matrix \( \mathbf{F} \) that satisfy \( \mathbf{\hat{F}} \mathbf{x} = \mathbf{0} \), and let the \((2K - 1) \times \kappa\) matrix \( \mathbf{U} \) satisfy \( \mathbf{\hat{F}} \mathbf{U} = \mathbf{0} \) and \( \mathbf{U}^T \mathbf{U} = \mathbf{I} \). Thus, the sparsity-constrained CRB can be obtained as [16]

\[
\text{Trace}(\mathbf{U}(\mathbf{U}^T \mathbf{JU})^{-1} \mathbf{U}^T) \tag{3.10}
\]

This gives the best achievable mean squared error obtained by estimators that have perfect knowledge of the support set of the PS \( \mathbf{s}_x \) to be estimated.

### 3.4 Relevant Extensions

#### 3.4.1 Model PS Estimation

Similar to Section 2.6.1, we consider that the PS of the primary signal \( x(n) \) can be expressed (approximated) by the model

\[
S_x(\omega) = \sum_{\ell=1}^{L} \rho_\ell \Psi_\ell(\omega) \tag{3.11}
\]

where \( \{ \Psi_\ell(\omega) \}_{\ell=1}^{L} \) are known functions and \( \{ \rho_\ell \}_{\ell=1}^{L} \) are unknown positive weights. In general, the PS model (3.11) can be used to model the PS of \( L \) primary transmitters, where \( \rho_\ell \) is the transmit-power of transmitter \( \ell \in \{1, \ldots, L\} \), and the function \( \Psi_\ell(\omega) \) characterizes the spectral mask and the carrier frequency [67]. In this case, the received signal at sensor \( m \in \{1, \ldots, M\} \), sampled using a Nyquist-rate ADC, can be expressed as

\[
y_m(n) = \sum_{\ell=1}^{L} h_m(\ell) \sqrt{\rho_\ell} x_\ell(n) \tag{3.12}
\]

where \( x_\ell(n) \) is the discrete-time WSS signal of the \( \ell \)-th transmitter (signals are assumed independent across transmitters), and \( \{ h_m(\ell) \}_{\ell=1}^{L} \) are time-invariant i.i.d. zero-mean
and unit-variance complex Gaussian random variables that represent the fading channel. This assumes that the long-term channel coefficient between transmitter $\ell$ and sensor $m$ is already known (e.g., using training sequences), and is removed from the received signal.

It can be shown that

$$\tilde{\alpha}_m = \mathbb{E}[|\tilde{z}_m(n)|^2] = \sum_{\ell=1}^{L} \rho_\ell v_{m,\ell}, \quad (3.13)$$

$$\alpha_m = \mathbb{E}[|z_m(n)|^2] = \sum_{\ell=1}^{L} |h_m(\ell)|^2 \rho_\ell v_{m,\ell} \quad (3.14)$$

where $v_{m,\ell} = \sum_{k=1}^{K-1} \psi_\ell(k)q_m^*(k)$ and $\psi_\ell(k)$ is the I-DTFT of $\Psi_\ell(\omega)$ (also the autocorrelation function of $x_\ell(n)$). Thus, the measurement error due to the fading channel

$$\tilde{e}_m = \alpha_m - \tilde{\alpha}_m = \sum_{\ell=1}^{L} \rho_\ell v_{m,\ell}(|h_m(\ell)|^2 - 1) \quad (3.15)$$

is a sum of independent random variables, which can be approximated as a zero-mean Gaussian random variable for large $L$ using the Lyapunov central limit theorem, as shown in Appendix A.5. Hence, defining $v_m := [v_{m,1}, \ldots, v_{m,L}]^T$, the power measurement bit (3.3) can be written in this case as a function of $\rho := [\rho_1, \ldots, \rho_L]^T$ as $b_m = \text{sign}(v_m^T \rho + e_m - t_m)$, where the goal here is to estimate $\rho$ from $\{b_m\}_{m=1}^{M}$. By exploiting the Gaussian distribution of $\{e_m\}_{m=1}^{M}$ in estimating $\rho$, the ML (3.7) is modified to

$$\max_{\rho \in \mathcal{B}} \sum_{m=1}^{M} \log Q \left( \frac{-b_m(v_m^T \rho - t_m)}{\sigma_m} \right) - \lambda \sum_{\ell=1}^{L} \rho_\ell \quad (3.16)$$

The CRB on the estimate of $\rho$ can be obtained in a manner similar to Section 3.3, by replacing $r_x$ with $\rho$ and $\{q_m\}$ with $\{v_m\}$ in the FIM expression (3.8). If prior information on the number of nonzero entries of the true $\rho$ is available, i.e., $||\rho||_0 = \kappa$, the constrained CRB is obtained using expression (3.10), where the $L \times \kappa$ matrix $U$ in (3.10) is now defined as the matrix of feasible directions consisting of the subset of columns of the identity matrix corresponding to the support set of $\rho$ [74].
3.4.2 Line Spectrum Estimation

As in Section 2.6.2, consider that the primary signal is composed of \( L \) tones (spectral components) \( x(n) = \sum_{\ell=1}^{L} \gamma_{\ell} e^{j\omega_{\ell} n} \), and that the autocorrelation function can be expressed as \( r_x(k) = \sum_{\ell=1}^{L} \rho_{\ell} e^{j\omega_{\ell} k} \), where \( \rho_{\ell} = |\gamma_{\ell}|^2 \). The goal is to estimate the frequencies and powers \( \{\omega_{\ell}, \rho_{\ell}\}_{\ell=1}^{L} \), given the measurement bits \( \{b_m\}_{m=1}^{M} \). Similar to Section 2.6.2, one approach is to first estimate \( r_x \) (non-parametrically) using the ML formulation (3.7), then MUSIC can be used to estimate \( \{\omega_{\ell}\}_{\ell=1}^{L} \), while \( \{\rho_{\ell}\}_{\ell=1}^{L} \) can be estimated using LS (2.26).

Another approach is to directly exploit the line spectrum structure of the signal in estimating \( \{\omega_{\ell}, \rho_{\ell}\}_{\ell=1}^{L} \) in a single step as follows. Note that for line spectrum, the power measurement \( \tilde{\alpha}_m \) can be explicitly expressed as a function of \( \{\omega_{\ell}, \rho_{\ell}\}_{\ell=1}^{L} \) as

\[
\tilde{\alpha}_m = \sum_{\ell=1}^{L} \sum_{k=1-K}^{K-1} \rho_{\ell} e^{jk\omega_{\ell} q_m^*(k)} (3.17)
\]

From (3.6), the log-likelihood function with line spectrum can be expressed as

\[
\Upsilon(\{\rho_{\ell}\}_{\ell=1}^{L}, \{\omega_{\ell}\}_{\ell=1}^{L}) := \sum_{m=1}^{M} \log Q \left( -b_m \left( \sum_{\ell=1}^{L} \sum_{k=1-K}^{K-1} \rho_{\ell} e^{jk\omega_{\ell} q_m^*(k)} - t_m \right) / \sigma_m \right)
\]

Therefore, the ML estimates are obtained by solving

\[
\{\hat{\omega}_{\ell}, \hat{\rho}_{\ell}\}_{\ell=1}^{L} = \arg \max \ \Upsilon(\{\rho_{\ell}\}_{\ell=1}^{L}, \{\omega_{\ell}\}_{\ell=1}^{L}) (3.19)
\]

Whereas problem (3.19) is convex in \( \{\rho_{\ell}\}_{\ell=1}^{L} \), it is nonconvex in \( \{\omega_{\ell}\}_{\ell=1}^{L} \).

In Fig. 3.2 a signal with \( L = 2 \) tones at angular frequencies \( \pi/2 \) and \( \pi \), and equal unit powers is considered. The network included \( M = 200 \) sensors, a single threshold \( t_m = t = 3 \) was used for all sensors, and the filter length was set to \( K = 20 \). The log-likelihood function \( \Upsilon \) is plotted as a function of \( \omega_1/2\pi \) and \( \omega_2/2\pi \), in 3-D, for a single realization of the random filters and error samples (with \( \sigma_m^2 = 1, \forall m \)). It is clear form the figure that \( \Upsilon \) is nonconvex in \( \omega_1 \) and \( \omega_2 \), and that the global maxima of \( \Upsilon \) is achieved at \( \omega_1 \approx \pi/2 \) and \( \omega_2 \approx \pi \) (and vice versa due to symmetry).

We propose a coordinate descent grid search (CDGS) algorithm that approximately solves the nonconvex problem (3.19). Dividing the search range for each \( \omega_{\ell} \in [0, 2\pi) \)
Figure 3.2: Log-likelihood function for a 2-tone signal with $\omega_1 = \pi/2$ and $\omega_2 = \pi$.

into $N_F$ (uniform) grid points, and assuming initial estimates $\{\hat{\omega}_\ell, \hat{\rho}_\ell\}_{\ell=1}^L$, the proposed CDGS algorithm is explained as follows:

1. For $\ell = 1, \ldots, L$, perform exhaustive grid search for $\omega_\ell$ that maximizes the log-likelihood function, fixing $\{\hat{\omega}_i\}_{i=1, i \neq \ell}^L$ and $\{\hat{\rho}_\ell\}_{\ell=1}^L$, i.e., solve (3.19) in the variable $\omega_\ell$ only using grid search, and set this value as $\hat{\omega}_\ell$.

2. Solve the convex problem (3.19) in the variable powers $\{\hat{\rho}_\ell\}_{\ell=1}^L$ only while fixing the frequency estimates $\{\hat{\omega}_\ell\}_{\ell=1}^L$ (e.g., using interior-point methods [22]).

3. Terminate on convergence; otherwise, repeat.

The proposed CDGS terminated after 2-to-4 iterations for all the scenarios that we have considered. Note that the complexity of solving (3.19) using CDGS grows with $L$ (i.e., $2L$ variables), whereas the complexity of the LP (2.16) and ML (3.7) grows with
\( K \) (i.e., \((2K - 1)\) variables). In order to increase the convergence speed of the CDGS algorithm when \( L \) is large, the initial estimates \( \{\hat{\omega}_\ell, \hat{\rho}_\ell\}^L_{\ell=1} \) can be obtained using the “nonparametric ML + MUSIC” technique.

The CRB on the estimates of \( \{\omega_\ell, \rho_\ell\}^L_{\ell=1} \), can be derived as follows. Define \\
\( \mu_m := \log Q\left( -\frac{b_m(\tilde{\alpha}_m - t_m)}{\sigma_m} \right) \) and the length-(2\( L \)) gradient vector \( \nabla \mu_m \), where the \( \ell \)-th entry of \( \nabla \mu_m \) is the partial derivative \( \frac{\partial \mu_m}{\partial \omega_\ell} \), while the \((L + \ell)\)-th entry is \( \frac{\partial \mu_m}{\partial \rho_\ell} \), for \( \ell = 1, \ldots, L \). Similar to (3.8) and the derivation in Appendix A.6, the FIM (with respect to \( \{\omega_\ell, \rho_\ell\}^L_{\ell=1} \)) can be computed as

\[
J = -E[\nabla^2 \Upsilon] = -\sum_{m=1}^M E[\nabla^2 \mu_m] = UDU^T \tag{3.20}
\]

where the entries of the diagonal matrix \( D \) are defined in (3.9) and the \( \ell \)-th row and \( m \)-th column entry of the \((2L \times M)\) matrix \( U \) is obtained as

\[
U_{\ell,m} = \begin{cases} \sum_{k=1}^{K-1} e^{jk\omega_\ell} q^*_m(k), & \ell = 1, \ldots, L \\ \sum_{k=1}^{K-1} e^{jk\omega_{L+\ell}} q^*_m(k), & \ell = L + 1, \ldots, 2L \end{cases} \tag{3.21}
\]

Finally, the CRB on the ML estimates \( \{\hat{\omega}_\ell, \hat{\rho}_\ell\}^L_{\ell=1} \) is obtained as \( \text{Trace}(J^{-1}) \).

### 3.4.3 Censoring

It is important to reduce the number of bits transmitted from the sensors so as to prolong battery lifetimes and to minimize the communication overhead [59]. With censoring, only sensors that provide the most useful information bits are permitted to send, while other less-informative sensors remain silent. For the Gaussian tail function, we know that \( \log Q(-|u|) \approx 0 \), if \( |u| > 4 \). Hence, the log-likelihood function (3.6) is almost unaffected by sensor \( m \) if the value of \( |\tilde{\alpha}_m - t_m|/\sigma_m \) is too large. It can also be shown in (3.9) that \( D_{m,m} \approx 0 \), and that the FIM (3.8) is almost unaffected by sensor \( m \), if the value of \( |\tilde{\alpha}_m - t_m|/\sigma_m \) is too large. This means that sensor \( m \) is almost useless in the estimation problem if \( |\tilde{\alpha}_m - t_m|/\sigma_m \) is too large. Thus, censoring can be employed such that sensor \( m \) sends \( b_m \) only if \( |\tilde{\alpha}_m - t_m| \leq \zeta_m \), where \( \zeta_m > 0 \) is a censoring threshold that is known at the sensor. More insights on the choice of \( \zeta_m \) are discussed in Section 3.6.1.
3.4.4 Multi-Bit Quantization

So far, only a single-bit quantization has been considered. Consider now that $\hat{\alpha}_m$ is quantized to multiple bits using the quantization function $Q(\hat{\alpha}_m)$. Without loss of generality, the quantization $Q(\hat{\alpha}_m)$ implies that $\hat{\alpha}_m$ is bounded as $t^L_m \leq \hat{\alpha}_m \leq t^U_m$.

Hence, the pmf of $Q(\hat{\alpha}_m)$ given $r_x$ can be expressed as

$$p[Q(\hat{\alpha}_m)|r_x] = p[t^L_m \leq q^T_m r_x + e_m \leq t^U_m] = Q\left(\frac{t^L_m - q^T_m r_x}{\sigma_m}\right) - Q\left(\frac{q^T_m r_x - t^U_m}{\sigma_m}\right)$$

In this case, the ML formulation (3.7) can be modified to

$$\max_{r_x \in \mathcal{P}} \sum_{m=1}^{M} \log \left[ Q\left(\frac{t^L_m - q^T_m r_x}{\sigma_m}\right) - Q\left(\frac{q^T_m r_x - t^U_m}{\sigma_m}\right) \right] - \lambda r_x(0)$$

s.t. $Fr_x \geq 0$

Problem (3.23) is also convex in $r_x$ as shown in [65].

3.5 Active Sensing

The same active sensing setting of Section 2.7 is considered here, where the objective is to adaptively design the thresholds online to ensure fast convergence to the autocorrelation $r_x$. Unlike the situation in Section 2.7, the feasible region does not decrease after receiving a measurement bit due to the uncertainty induced by Gaussian measurement errors. Similar to the error-free case, however, it is still desirable to select the threshold $t_m$ to be as close as possible to $q^T_m r_x$. This is also motivated by the fact that the CRB is minimized with $t^*_m = q^T_m r_x$, for $m = 1, \ldots, M$ [29].

Therefore, after the FC receives the bits $b_1, \ldots, b_m$, the proposed active sensing algorithm sets $t_{m+1} = q^T_{m+1} r^{(m)}_x$ for sensor $m + 1$, where $r^{(m)}_x$ is the ML estimate of $r_x$ given the first $m$ measurement bits, which is obtained by solving

$$r^{(m)}_x = \arg \max_{r_x \in \mathcal{P}, Fr_x \geq 0} \sum_{i=1}^{m} \log Q\left(\frac{-b_i(q^T_i r_x - t_i)}{\sigma_i}\right)$$

The initial $r^{(0)}_x$ can be chosen any point inside $\mathcal{P}$. Solving (3.24) exactly to compute $t_m$ for each sensor, however, is computationally expensive.
Similar to the proposed low complexity algorithm of Section 2.7.2, an approximate solution to (3.24) can be obtained using Newton’s method. Defining \( \Gamma_m(x) := - \sum_{i=1}^{m} \log Q \left( \frac{-b_i(q_i^T x - t_i)}{\sigma_i} \right) \), it can be shown that

\[
\nabla \Gamma_m(x) = - \sum_{i=1}^{m} \frac{b_i e^{\frac{(q_i^T x - t_i)^2}{2\sigma_i^2}}}{2\pi\sigma_i^2 Q \left( \frac{-b_i(q_i^T x - t_i)}{\sigma_i} \right)} q_i
\]

(3.25)

\[
\nabla^2 \Gamma_m(x) = \sum_{i=1}^{m} \left[ \frac{e^{\frac{(q_i^T x - t_i)^2}{2\sigma_i^2}}}{2\pi\sigma_i^2 \left[ Q \left( \frac{-b_i(q_i^T x - t_i)}{\sigma_i} \right) \right]^2} + \frac{b_i e^{\frac{(q_i^T x - t_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi\sigma_i^2 Q \left( \frac{-b_i(q_i^T x - t_i)}{\sigma_i} \right)}} \right] q_i q_i^T
\]

(3.26)

Thus, starting from any point \( x^{(0)} \) inside the initial \( \mathcal{P} \), an approximate ML estimate can be computed for the \( m \)-th sensor/time-slot using a single (or few) Newton step(s), similar to (2.29), as

\[
r_x^{(m)} = r_x^{(m-1)} - \left( \nabla^2 \Gamma_m(r_x^{(m-1)}) \right)^{-1} \nabla \Gamma_m(r_x^{(m-1)})
\]

(3.27)

Convergence of \( r_x^{(M)} \rightarrow r_x \) as \( M \rightarrow \infty \) using a single Newton step approximate ML is shown in the simulations in Section 3.6.3. It is worth mentioning that this low-complexity (approximate) ML algorithm can be used in the passive sensing setting by incrementally updating the PS estimate as new measurement bits are received, without having to wait for all sensors to report their measurements prior to estimation. This is particularly important for online sensing applications.

It is also worth noting that the same estimation performance can be achieved in a different decentralized scenario as follows. Consider that each of the spatially-distributed sensors aims at estimating \( r_x \) collaboratively with no FC. Consider a time-slotted structure with \( M \) time slots, where sensor \( m \in \{1, \ldots, M\} \) broadcasts \( b_m \) in time slot \( m \) such that all other sensors listen to and acquire \( b_m \). Assuming that \( \{q_m\}_{m=1}^{M} \) is known at all sensors, \( t_m \) can be computed at all sensors as \( t_m = q_m^T r_x^{(m-1)} \) where \( r_x^{(m-1)} \) is computed as (3.24) (or (3.27)) using the priorly broadcasted bits \( b_1, \ldots, b_{m-1} \). In this case, all sensors can obtain a common estimate \( \hat{r}_x \), with more sophisticated processing at each sensor (to solve (3.24) or (3.27) at each time slot). Clearly, the estimated \( \hat{r}_x \) in this setting is the same as the case of centralized active sensing with FC discussed above.
3.6 Numerical Results

We begin with a simulation in Fig. 3.3 that illustrates what one can expect from the proposed nonparametric ML estimation using (3.7) and the model-based ML estimation using (3.16). A scenario with \( M = 150 \) sensors was considered and a single threshold \( t_m = t \) was selected such that \( \tilde{\alpha}_m \geq t \) for 50 sensors. The error \( e_m = \hat{\alpha}_m - \tilde{\alpha}_m \) caused the flipping of 10 bits, i.e., \( \text{sign}(\tilde{\alpha}_m - t_m) \neq \text{sign}(\hat{\alpha}_m - t_m) \) for 10 sensors. The true PS of the primary signal, which is comprised of 8 equispaced raised-cosine functions with 0.5
roll-off factor and different power coefficients, is plotted with a solid line in Fig. 3.3(a) whereas the Fourier transform (FT) of the true truncated ($K = 10$)-lag autocorrelation is plotted with a solid line in Fig. 3.3(b). The estimated model-based PS obtained using (3.16) (with $\lambda = 50$) by exploiting the information of the raised-cosine model is plotted with a dashed line in Fig. 3.3(a), whereas the estimated PS obtained by estimating the ($K = 10$)-lag autocorrelation nonparametrically using (3.16) (with $\lambda = 0$) is plotted with a dashed line in Fig. 3.3(b). The quality of the estimates in Figs. 3.3(b) and 3.3(a) is very satisfactory, considering that 10 of the 150 bits that are used as input data have been flipped. As expected, much better PS estimates can be obtained by exploiting available information on the PS model.

![Figure 3.4](image)

**Figure 3.4:** Nonparametric PS: MSE of the LP and ML estimates and the CRB.

In Fig. 3.4, the mean squared error (MSE) of the estimated ($K = 10$)-lag autocorrelation using the LP (2.16) and the ML (3.7) (with $\lambda = 0$), and the CRB, are plotted versus the number of sensors $M$. The true primary signal is comprised of 10 equi-spaced raised-cosine functions with 0.5 roll-off factor, where the corresponding power coefficients are randomly chosen from a uniform distribution, then normalized (i.e., non-sparse spectrum). A single threshold $t_m = t$ was selected such that $\alpha_m \geq t$ for 50% of the sensors, and the error variance $\sigma_m = \sigma$ was selected such that the measurement bits
reported by 17\% of the sensors were flipped on average. The expectation of the MSE is taken with respect to the random impulse responses of the FIR filters, the random Gaussian error samples, and the random raised-cosine power coefficients, obtained via 1000 Monte-Carlo simulation runs. The figure shows the decrease of the CRB and the MSE of the ML estimate as $M$ increases, as expected, whereas the LP (2.16) fails to provide a meaningful estimate, due to the flipped bits. The figure also shows that the MSE of the ML estimate is asymptotically converging to the CRB as $M$ grows.

### 3.6.1 Model PS Estimation

To test the performance of the proposed model-based PS estimation techniques in Section 3.4.1, we assume that the primary signal PS model is the combination of $L = 20$ identical raised cosine functions, each with roll-off factor $= 0.5$ and bandwidth $= 3\pi/L$. The corresponding center frequencies are $\{\omega_\ell = 2\pi(\ell - 1)/L\}_{\ell=1}^L$, implying equispaced and overlapping functions. A sparse vector $\mathbf{\rho}$ with 5 uniformly distributed nonzero entries out of 20 is randomly generated and normalized by $\sum_{\ell=1}^L \rho_\ell$ in each simulation run. We use the MSE, defined as $\text{MSE}_{\rho} := E[||\mathbf{\rho} - \hat{\mathbf{\rho}}||^2]$, to measure the performance of the proposed estimation techniques, where the expectation is taken with respect to the random impulse responses of the FIR filters, the random Gaussian error samples, and the random weights vector $\mathbf{\rho}$, obtained via 1000 Monte-Carlo simulation runs. Typically, the MSE should be computed with respect to the estimated PS; however, the symmetry in the considered PS model allows using $\text{MSE}_{\rho}$ instead. A single threshold $t_m = t = 0.052$ and a filter length $K = 25$ were used at all sensors. This $t$ is numerically computed as the minimizer of the expected CRB (3.10) across different $\mathbf{\rho}$ and filter realizations. The sparsity tuning parameter in (3.16) was fixed to $\lambda = 100$ in the simulations. For the same error variance, $\sigma_m = \sigma$, $\forall m$, we define the signal-to-error ratio (SER) as $\text{SER} := \sum_{\ell=1}^L \rho_\ell \psi_\ell(0)/\sigma^2$. For brevity, we name the estimate of $\mathbf{\rho}$ obtained using (2.24) as the LP estimate, whereas the estimate obtained using (3.16) is named the ML estimate. The MSE of the LP estimate, the MSE of the ML estimate, and the (oracle) CRB computed using (3.10), are plotted versus the number of sensors $M$ in Figs. 3.5 and 3.6, for SER $= 5 \times 10^4$ and SER $= 500$, respectively.

In Fig. 3.5, where a relatively large SER is considered, the random errors in this case cause the flipping of 2\% of the sensor measurement bits, on average. In other words,
sign($\tilde{\alpha}_m - t_m$) $\neq$ sign($\hat{\alpha}_m - t_m$) for 0.02$M$ sensors, on average. The figure shows the decrease of the CRB and the MSE of the ML estimate as $M$ increases, as expected. The big gap between the MSE of the ML estimate and the CRB is justified since the CRB is computed with perfect knowledge of the support of the true $\rho$. The figure also shows that the MSE of the LP estimate is decreasing and stays close to the MSE of the ML estimate for $M \leq 300$, then it increases when $M$ increases to 500. The reason is that as $M$ increases, the constraints in (2.24) become more stringent, and the flipped bits due to errors drive the solution far away from the true one. The performance of the censoring scheme proposed at the end of Section 3.3 is also considered in Fig. 3.5. The dashed line in the figure shows the MSE of the ML estimate when the censoring threshold $\zeta$ was selected such that only the best 80 out of $M$ sensors are active in each simulation run (i.e., $|\tilde{\alpha}_m - t| \leq \zeta$ for 80 sensors), on average. As shown in the figure, the performance with censoring is very close to the case when all sensors are reporting. Censoring in this case is very efficient for large $M$ (e.g., performance with 80/500 reporting sensors is almost the same as with all 500 sensors reporting).

In Fig. 3.6, where a relatively small SER is considered, the random errors cause the flipping of 16% of the sensor measurement bits, on average. The figure shows the decrease of the CRB and the MSE of the ML estimate as $M$ increases, albeit at larger
values than their counterparts in Fig. 3.5 due to the higher number of flipped bits due to the measurement errors. It is satisfying to see that good ML estimates can be obtained, despite the flipping of so many measurement bits. We also note that the gap between the MSE of the ML estimate and the CRB was reduced as the SER decreased from Fig. 3.5 to Fig. 3.6. The performance of the LP estimate in this figure is severely limited by the relatively large number of flipped bits, which are not accounted for in (2.16). The MSE of the ML estimate when employing the same fixed censoring strategy as in Fig. 3.5, represented by the dashed line in Fig. 3.6, is increasing with $M$. This is because the number of flipped bits among the 80 measurement bits that are reported in each simulation run, on average, increases as $M$ increases, which seriously degrades performance. This is because censoring chooses to transmit bits whose measurements are close to the respective thresholds, as the ‘most informative’ bits, by design. These are also the bits that are most likely to be flipped due to measurement errors. As an alternative, we considered an adaptive censoring scheme, where the censoring threshold $\zeta$ is selected such that 67% of the sensors (i.e., $0.67M$) are reporting in each simulation run, on average. The MSE of the ML estimate with this adaptive censoring scheme is shown with the dotted line in Fig. 3.6. Although the adaptive censoring performance is...
improving with $M$, we can see that the performance is significantly worse than the case when all sensors are reporting. We conclude that when the SER is relatively small, it is better that all sensors report to combat the increasing number of flipped measurement bits due to errors, whereas censoring is more efficient when the SER is relatively large.

Figure 3.7: MSE with (a) Gaussian fading coefficients and (b) Gaussian errors.

In Fig. 3.7 we consider the same signal model and sensor settings as Figs. 3.5 and 3.6 and compare between the ML estimate obtained using (3.16) (with $\lambda = 0$) and the benchmark estimate that is obtained by solving the (box-constrained) least squares: $\min_{\rho \in B} ||\alpha - V\rho||_2^2$, where $\alpha := [\hat{\alpha}_1, \ldots, \hat{\alpha}_M]^T$. The benchmark estimate is obtained assuming that each sensor sends the analog power measurement $\hat{\alpha}_m$. The MSE of the ML and benchmark estimates are plotted for two scenarios: (a) the case of i.i.d. randomly generated fading channel coefficients $\{h_m(\ell)\}_{\ell=1}^L$ from a zero-mean and unit-variance complex Gaussian distribution such that $\hat{\alpha}_m = \sum_{\ell=1}^L |h_m(\ell)|^2 \rho_{\ell} v_{m,\ell}$; and (b) the approximation case where $\hat{\alpha}_m = \sum_{\ell=1}^L \rho_{\ell} v_{m,\ell} + e_m$ and $e_m$ is randomly generated from a zero-mean Gaussian distribution using the same variance of the error in scenario (a). The average SER across different filter and $\rho$ realizations was found to be 50, and the errors caused the flipping of 31.5% and 33.7% of the measurement bits,
on average, for scenario (a) and scenario (b), respectively. The figure shows that the Gaussian approximation of the distribution of the errors \( \{e_m\}_{m=1}^{M} \) is a good one. More interestingly, the figure also shows that the performance with 1-bit measurements is close to the benchmark performance with analog (i.e., finely quantized) measurements. For example, the same performance can be achieved using 200 1-bit-sensors or 150 analog-sensors. Assuming that each analog signal is quantized to 8 bits, this means that 1200 bits are required to achieve the same performance that can be obtained using only 200 bits with 1-bit-sensors.

3.6.2 Line Spectrum Estimation

![Figure 3.8: Line Spectrum: Average CRB vs. threshold \( t \) for different \( L \) tones.](image)

Assuming that the threshold is the same for all sensors \( t_m = t, \forall m \), a good choice for \( t \) is the minimizer of the CRB. In Fig. 3.8 we plot the expected CRB as a function of \( t \) for a signal with \( L = 1, 2, 4, \) and 8 tones, where the expectation is taken with respect to the random impulse responses of the FIR filters obtained via 100 Monte-Carlo simulation runs. The power of each tone was fixed to unity and the generated tones were equispaced with spacing \( \Delta \omega = 0.2\pi \). The setup included \( M = 200 \) sensors, the filter length was set to \( K = 25 \), and \( \sigma_m^2 = 1, \forall m \). The figure shows that the number of sensors above \( t \) should increase as \( L \) increases to minimize the average CRB. For
example, \( t \) should be selected such that 44 sensors send \( b_m = 1 \) for \( L = 1 \) (i.e., 22% of the sensors), whereas \( t \) should be selected such that 94 sensors send \( b_m = 1 \) for \( L = 8 \) (i.e., 47% of the sensors), on average. The figure also shows that the minimum average CRB is increasing (by more than double) as the number of existing tones \( L \) doubles, as expected.

![Figure 3.9: Line Spectrum: RMSE and CRB for a signal with 2 far-apart tones.](image)

We use the root-mean-square error (RMSE) to compare the performance of the three proposed estimation techniques in Figs. 3.9 and 3.10, where the expectation is taken with respect to the random impulse responses of the FIR filters and the random error samples, obtained via 1000 Monte-Carlo simulation runs. A single threshold \( t_m = t = 2.3 \) and a filter length \( K = 25 \) was used for all sensors, and the error variance was assumed to be \( \sigma_m^2 = 1, \forall m \). This \( t = 2.3 \) is the minimizer of the expected CRB for 2 tones across different filter realizations, as shown in Fig. 3.8. A signal consisting of two far-apart tones at angular frequencies \( \omega_1 = 0.4\pi \) and \( \omega_2 = 0.8\pi \) was considered in Fig. 3.9, whereas two very close tones at angular frequencies \( \omega_1 = 0.4\pi \) and \( \omega_2 = 0.46\pi \) were considered in Fig. 3.10. The power of each tone was set to unity (not estimated). The \( 2\pi \) range of each \( \omega_\ell \) was divided into \( N_F = 1000 \) search points for the grid search step of the CDGS. In both figures, the RMSE and CRB of the estimates \( \hat{\omega}_1 \) and \( \hat{\omega}_2 \) are
plotted versus the number of sensors $M$.

![Figure 3.10: Line Spectrum: RMSE and CRB for a signal with 2 close tones.](image)

Figures 3.9 and 3.10 show that the RMSE of the proposed estimation techniques and the CRB decrease as $M$ increases, as expected. In Fig. 3.9 the RMSE of the ML estimates obtained by solving (3.19) using CDGS meets the CRB for $M \geq 150$, and outperforms the 2-step-estimation technique for all considered $M$. In Fig. 3.10 where $|\omega_1 - \omega_2| = 0.06\pi < 0.08\pi = 2\pi/K$, the RMSE of the proposed estimation techniques and the CRB are larger than their counterparts in Fig. 3.9 as expected. The figure also shows that the nonparametric ML estimate of the autocorrelation followed by MUSIC outperforms the direct parametric ML + CDGS estimation when the number of sensors are relatively small, due to the superior high-resolution frequency detection performance of MUSIC, whereas the ML estimates obtained by solving (3.19) using CDGS meets the CRB and outperforms the 2-step-estimation technique for $M \geq 200$.

### 3.6.3 Active Sensing

We now switch to testing the performance of the proposed active sensing algorithm. To measure the quality of the estimated autocorrelation $\hat{r}_x$, we use the normalized
mean squared error (NMSE), defined as $\text{NMSE}(m) := \mathbb{E}\left[\frac{\|\mathbf{r}_x - \hat{\mathbf{r}}_x^{(m)}\|^2}{\|\mathbf{r}_x\|^2}\right]$, where $\hat{\mathbf{r}}_x^{(m)}$ is the estimate of $\mathbf{r}_x$ when the FC receives the $m$-th bit, $m = 1, \ldots, M$. The primary signal is assumed to be a combination of 10 equispaced raised-cosine functions with roll-off factor $= 0.5$. The power coefficients were drawn from a uniform distribution between 0.2 and 1 in each simulation run. A filter length $K = 10$ was used at all sensors. In Fig. 3.11, the NMSE is plotted as a function of $m$ for the active sensing algorithm with the exact ML estimate (3.24), the active sensing algorithm with the approximate ML estimate obtained using the single-step Newton (3.27), and the case of passive sensing with a single threshold $t_m = t$ for all sensors selected such that 50% of the sensors are reporting $b_m = 1$. The generated Gaussian errors caused the flipping of 68 bits from the total 300 received bits, on average. The figure shows the much faster convergence of active sensing using the exact and approximate ML estimates as opposed to the passive sensing case. The figure also shows that the performance with the exact ML is slightly better than the single-step Newton approximation; however, there is a huge complexity and computation-time reduction when the approximate ML is used instead.
Chapter 4

Channel Tracking for Transmit Beamforming with Frugal Feedback

4.1 Introduction

This chapter addresses the challenge of downlink channel tracking using few (periodic) feedback bits for transmit beamforming, exploiting the spatio-temporal correlation of the channel. The huge feedback overhead is the main limitation in the employment of transmit beamforming with large-antenna systems in the frequency-division duplex (FDD) mode. Instead of estimating the channel at the receiver then sending the quantized channel state information (CSI) to the transmitter as in the codebook-based beamforming framework, the transmitter is assumed to periodically transmit a beamformed pilot signal in the downlink, while the receiver simply sends back a quantized version of the received pilot signal.

A bound on the performance with quantization is first obtained assuming complex analog-amplitude feedback of the received pilot signal. Then, using a 2-bit quantization scheme and assuming that the channel can be modeled by an autoregressive (AR) model, a Kalman filtering (KF) approach that is based on the sign of innovations (SOI)
is considered, followed by a novel channel tracking approach that exploits the quantization bits in a maximum a posteriori (MAP) estimation formulation for general (non-AR or unknown) channel models. In the AR channel case, closed-form expressions for the resulting channel estimation mean-squared error (MSE) and very tight approximations for the corresponding SNR are derived, assuming circular single-antenna beamforming for the pilots. Careful simulations show that by exploiting the spatio-temporal correlation of the channel, the performance achieved using the proposed frugal feedback approaches is close to that attainable with perfect CSI at the transmitter. Simulations also show that very large-size codebooks are required for codebook-based beamforming to achieve the same performance as the proposed approaches.

4.2 System Model

Figure 4.1: Downlink frame structure and limited feedback beamforming model.

Consider a downlink transmit beamforming setting comprising a transmitter with $N$ antennas and a receiver with a single receive antenna. Extensions to account for
multiple receive antennas and multiple receivers are discussed at the end of Section 4.5.

We consider a time-slotted downlink frame structure, where the duration of each slot is \( T \) seconds. We assume that at the beginning of each time slot \( n \), the transmitter sends a unit-power pilot symbol \( s(n) \) that is known at the receiver (i.e., downlink pilot rate is \( 1/T \) symbols/s), followed by data transmission for the remainder of the slot duration. The pilot symbol \( s(n) \) is \textit{beamformed} with a unit-norm \( N \times 1 \) beamforming vector \( w(n) \) (i.e., \( w(n) \) corresponds to the weights applied to the \( N \) transmit-antenna elements when transmitting \( s(n) \)), whereas the data symbols are \textit{beamformed} with a different unit-norm \( N \times 1 \) beamforming vector \( \tilde{w}(n) \).

We assume that the complex \( N \times 1 \) vector that models the frequency-flat channel between the \( N \) transmit-antennas and the receive antenna at time slot \( n \), denoted by \( h(n) \), is complex Gaussian distributed with zero mean and covariance matrix \( C_h \), i.e., \( h(n) \sim \mathcal{CN}(0, C_h) \), for all \( n \). The covariance \( C_h \) describes the spatial correlation of the channel, and is assumed to be known at the transmitter and the receiver. The channel vector \( h(n) \) is assumed to be fixed within time slot \( n \), and the random process \( \{h(n)\} \) is assumed to be stationary, ergodic, and \textit{temporally correlated}. A simple model for \( \{h(n)\} \), which allows specifying the temporal correlation of the channel, is the first-order AR model:

\[
    h(n) = \sqrt{\alpha} h(n - 1) + \sqrt{1 - \alpha} u(n) \tag{4.1}
\]

where \( u(n) \sim \mathcal{CN}(0, C_h) \), \( h(n - 1) \) is statistically independent of \( u(n) \) for all \( n \), and \( \alpha \leq 1 \) controls the degree of temporal correlation of the channel, \( E[h(n)h^H(n - k)] = \alpha^{k/2} C_h \). The AR model (4.1) has been widely considered in the literature to model the temporal progression of the channel (see, for example, \[19\] \[49\] \[68\]). Extending (4.1) to higher orders is straightforward \[43\] Ch. 13]. The channel is not restricted to the model (4.1) in this work, but (4.1) is considered for its analytical tractability. Note that unlike the common assumption in the literature on limited feedback (cf. \[50\] and references therein), the channel is \textit{not} assumed perfectly known at the receiver.

The received signal that corresponds to the transmitted pilot \( s(n) \) can be expressed as

\[
    \bar{y}(n) = w^H(n) h(n) s(n) + \tilde{v}(n) \tag{4.2}
\]

where the random variable \( \tilde{v}(n) \sim \mathcal{CN}(0, \sigma_v^2) \) models the additive white Gaussian noise (AWGN), and \( \{\tilde{v}(n)\} \) are independent and identically distributed (i.i.d.). Multiplying
the received signal $\tilde{y}(n)$ by $s^*(n)$ (i.e., de-scrambling) at the receiver yields

$$y(n) := s^*(n)\tilde{y}(n) = w^H(n)h(n) + v(n)$$

(4.3)

where $v(n) \sim \mathcal{C}\mathcal{N}(0, \sigma_v^2)$ and $\{v(n)\}$ are i.i.d.

The receiver then passes $y(n)$ through a quantizer, and the output quantization bits are sent to the transmitter through an uplink feedback channel. The challenge at the transmitter is to estimate and track the channel $h(n)$ using such few (periodic) feedback bits. The transmitter then uses the channel estimate $\hat{h}(n)$ to design the beamforming vector that is used for data transmission in time slot $n$ as $\hat{w}(n) = \frac{h(n)}{||h(n)||}$. Assuming that the data symbols are temporally white with zero-mean and unit-variance, and that the AWGN is zero-mean and unit-variance, the average receive-SNR can be expressed as

$$\gamma = \mathbb{E}[|\hat{w}^H(n)h(n)|^2].$$

The time-slotted downlink frame structure and the proposed limited feedback beamforming system are illustrated in Fig. 4.1.

In Section 4.3, we first consider the case where the receiver feeds back the complex analog-amplitude (or finely-quantized) signal $y(n)$ to the transmitter at each time slot, yielding a bound on the performance with quantization. The more practical case with very limited feedback, where the receiver feeds back only 2 bits to the transmitter at each time slot, is then considered in Section 4.4.

### 4.3 Analog-Amplitude Feedback

Here we assume that the receiver will send the complex analog-amplitude (or finely-quantized) signal $y(n)$ to the transmitter through an uplink feedback channel. Assuming an AR channel model, we first consider a KF approach for estimating and tracking $h(n)$, followed by a minimum mean-square error (MMSE) approach that can be applied for any channel model.

#### 4.3.1 KF Approach

Assuming an AR channel evolution model as (4.1), in addition to the linear observation model of $y(n)$ as (4.3), the transmitter can apply the KF iterations to estimate and track

---

1 Feedback delay is not considered in this work. The effect of feedback delay on the throughput has been considered in [38].
\( h(n) \) from \( \{y(k)\}_{k=1}^n \) [43, Ch. 13]. KF has been considered for tracking a time-correlated channel in [19, 49, 68].

Define the vector of observations \( y_n := [y(n), y(n-1), \ldots, y(1)]^T \) and the innovation

\[
\hat{y}(n) := y(n) - w^H(n)\hat{h}(n)
\]

where \( \hat{h}(n) := E[h(n)|y_{n-1}] \) is the predicted channel vector, which equals \( \sqrt{\alpha}\hat{h}(n-1) \) for the considered AR model. Exploiting that the posterior distribution \( p(h(n)|y_n) \) is Gaussian for the linear Gaussian state and observation models considered, the MMSE estimate of \( h(n) \) can be recursively obtained by the KF equations [43, Ch. 13]:

\[
\hat{h}_{KF}(n) = E[h(n)|y_n] = E[h(n)|y_{n-1}] + E[h(n)|\hat{y}(n)]
\]

\[
= \sqrt{\alpha}\hat{h}_{KF}(n-1) + \frac{\bar{M}(n)w(n)}{w^H(n)\bar{M}(n)w(n) + \sigma_v^2}\hat{y}(n)
\]

where the prediction error covariance matrix (ECM) is

\[
\bar{M}(n) = E[(h(n) - \hat{h}(n))(h(n) - \hat{h}(n))^H] = \alpha\bar{M}_{KF}(n-1) + (1-\alpha)C_h
\]

and the estimation ECM is

\[
\bar{M}_{KF}(n) = E[(h(n) - \hat{h}(n))(h(n) - \hat{h}(n))^H] = \bar{M}(n) - \frac{\bar{M}(n)w(n)w^H(n)\bar{M}(n)}{w^H(n)\bar{M}(n)w(n) + \sigma_v^2}
\]

For a general (non-AR) channel model, one approach is to approximate the actual channel evolution by the AR model (4.1), using \( \alpha \) that gives the best performance (e.g., \( \alpha \) that minimizes the average estimation error or maximizes the average achieved SNR). The performance of this approach is illustrated in Section 4.6. We next consider a different channel tracking approach that does not require a specific channel evolution model.

### 4.3.2 MMSE Approach

Here we consider a simple and general approach that does not assume a model for \( h(n) \). When estimating \( h(n) \) using the current and prior observations \( \{y(k)\}_{k=1}^n \), more weight should be given to recent observations, while older observations should be given less weight. Motivated by the exponentially-weighted recursive least-squares (RLS) algorithm [69, Ch. 30], we consider approximating the set of observations \( \{y(k) = \)
$w^H(k)h(k) + v(k)\}_{k=1}^n$ with the set $\{y(k) = w^H(k)h(n) + \lambda \frac{k-n}{2} v(k)\}_{k=1}^n$, where $0 \leq \lambda \leq 1$. The role of the forgetting factor $\lambda$ is to (exponentially) increase the noise variance of the older observations, implying more uncertainty in the approximate equality of the linear measurement $y(k) = w^H(k)h(n)$ as $n-k$ increases.

Define the beamforming matrix $W_n := [w(n), w(n-1), \ldots, w(1)]^H$ and the diagonal noise covariance matrix $C_v = \sigma_v^2 \text{diag}([1, \lambda^{-1}, \ldots, \lambda^{-n}])$. Hence, the MMSE estimate of $h(n)$, assuming the linear Gaussian observations $\{y(k) = w^H(k)h(n) + \lambda^{-n-k}/2 v(k)\}_{k=1}^n$, can be obtained as \[ \hat{h}_{\text{MMSE}}(n) = C_h W_n^H (W_n C_h W_n^H + C_v)^{-1} y_n \] (4.8)

The matrix $C_h W_n^H (W_n C_h W_n^H + C_v)^{-1}$ can be pre-computed for each $n$ in order to reduce the run-time computational complexity. Note that, because of the exponential decay, only finite-size matrices $W_n$ and $C_v$ are needed to compute $\hat{h}_{\text{MMSE}}(n)$ using (4.8), as $n \to \infty$. The main challenge in this MMSE approach is to find the value of $\lambda$ that gives the best performance for each channel model. Performance comparisons between the KF approach and the MMSE approach are considered in Section 4.6 for different channel models.

It is worth mentioning that if $h(n)$ is assumed deterministic instead of random, the exponentially-weighted RLS algorithm can be applied to estimate and track $h(n)$ from $\{y(k)\}_{k=1}^n$ [69, Ch. 30]. It is also worth mentioning that if second order statistics are available, i.e., $E[h(n)h^H(n-k)]$ for all $k$, then Wiener filtering (WF) can be applied [43, Ch. 12]. Assuming, for example, that $E[h(n)h^H(n-k)] = \rho_k C_h$ (where $\rho_0 = 1$ and $\rho_k$ is known for $k \geq 1$), the WF channel estimate can be obtained as:

\[ \hat{h}_{\text{WF}}(n) = C_h \tilde{W}_n^H (\tilde{W}_n C_h \tilde{W}_n^H \odot \Gamma + \tilde{C}_v)^{-1} y_n \] (4.9)

where $\tilde{C}_v := \sigma_v^2 I$, $\tilde{W}_n := [w(n), \rho_1 w(n-1), \ldots, \rho_n w(1)]^H$ and $\Gamma := \text{toeplitz}([\rho_0, \rho_1, \ldots, \rho_n])$.

### 4.4 2-Bit Quantized Feedback

Sending the complex analog-amplitude (or finely-quantized) signal $y(n)$ via the uplink feedback channel entails a large overhead in terms of the uplink resources (rate,
transmit-power). Instead, consider the following 2-bit quantization scheme at the receiver. It is easy to see that the KF channel tracking approach in (4.5) depends on the innovation \( \hat{y}(n) \) defined in (4.4), i.e., the difference between the current observation and the predicted observation based on past observations. Thus, we consider one-bit quantization for the real part of \( \hat{y}(n) \), and one-bit quantization for the imaginary part \( \hat{y}(n) \). This can be expressed as

\[
\begin{align*}
    b_r(n) &= \text{sign} [\text{Re}\{y(n)\} - d_r(n)] \\
    b_i(n) &= \text{sign} [\text{Im}\{y(n)\} - d_i(n)]
\end{align*}
\]

where \( d_r(n) := \text{Re}\{w^H(n)\tilde{h}(n)\} \), \( d_i(n) := \text{Im}\{w^H(n)\tilde{h}(n)\} \), and \( \tilde{h}(n) \) is the predicted channel given the past observations. In order to compute \( d_r(n) \) and \( d_i(n) \) that are required to perform the 2-bit quantization in (4.10) and (4.11), the receiver has to know the beamforming vector \( w(n) \), and must compute \( \tilde{h}(n) \) in the same way as the transmitter, as will be discussed later.

After the quantization, the receiver sends the two bits \( b_r(n) \) and \( b_i(n) \) to the transmitter via the uplink feedback channel. The feedback channel is assumed free of errors, which is a typical assumption in the literature on limited feedback [50]. Note that with such 2-bit quantization, the downlink pilot rate is only \( 1/T \) symbols/s, and the uplink feedback rate is only \( 2/T \) bits/s. The challenge here is whether the transmitter can accurately estimate and track the complex \( N \)-dimensional channel \( h(n) \), using only the periodically received pairs of feedback bits \( b_r(n) \) and \( b_i(n) \). To address this challenge, we first consider a SOI-KF approach (based on [66]) that is suitable for the AR channel model, followed by a novel MAP approach that is applicable for general channel models.

### 4.4.1 SOI-KF Approach

Here we assume the AR channel model in (4.1), and the binary observation model given by (4.10) and (4.11), where \( \tilde{h}(n) = \sqrt{\alpha}\tilde{h}(n-1) \) for the AR model. To estimate and track \( h(n) \) at the transmitter using \( \{b_r(k)\}_{k=1}^n \) and \( \{b_i(k)\}_{k=1}^n \), we extend the SOI-KF framework from the real vector space considered in [66] to the complex vector space. To facilitate operating in the more convenient real domain, consider the following
the following KF-like recursive equations:

\[
\begin{align*}
\mathbf{b}_n & := [b_r(1), \ldots, b_r(n), b_i(1), \ldots, b_i(n)]^T, \\
\mathbf{w}_r(n) & := [\text{Re}\{\mathbf{w}(n)\}]^T, \text{Im}\{\mathbf{w}(n)\}]^T, \\
\mathbf{w}_i(n) & := [-\text{Im}\{\mathbf{w}(n)\}]^T, \text{Re}\{\mathbf{w}(n)\}]^T, \\
\mathbf{h}(n) & := [\text{Re}\{\mathbf{h}(n)\}]^T, \text{Im}\{\mathbf{h}(n)\}]^T, \\
\mathbf{C}_h & := \mathbb{E}[\mathbf{h}(n)\mathbf{h}(n)^H]
\end{align*}
\]

such that \(\text{Re}\{\mathbf{w}^H(n)\mathbf{h}(n)\} = \mathbf{w}_r^T(n)\tilde{\mathbf{h}}(n)\) and \(\text{Im}\{\mathbf{w}^H(n)\mathbf{h}(n)\} = \mathbf{w}_i^T(n)\tilde{\mathbf{h}}(n)\).

The distribution \(p(\mathbf{h}(n)|\mathbf{b}_n)\) is not necessarily Gaussian because the binary observation model is not linear, and hence the exact MMSE estimator, i.e., \(\mathbb{E}[\mathbf{h}(n)|\mathbf{b}_n]\), requires solving multiple nested numerical integrations to compute the posterior distribution \(p(\mathbf{h}(n)|\mathbf{b}_n)\) \cite{66}. Assuming that \(p(\mathbf{h}(n)|\mathbf{b}_{n-1}) = \mathcal{N}(\sqrt{\alpha}\mathbf{h}(n-1), \tilde{\mathbf{M}}(n))\), and utilizing the results of \cite{66}, the MMSE estimate \(\hat{\mathbf{h}}_{\text{SOI-KF}}(n) := \mathbb{E}[\mathbf{h}(n)|\mathbf{b}_n]\) can be obtained using the following KF-like recursive equations:

\[
\hat{\mathbf{h}}_{\text{SOI-KF}}(n) = \sqrt{\alpha} \hat{\mathbf{h}}_{\text{SOI-KF}}(n-1) + \frac{\sqrt{2/\pi}\tilde{\mathbf{M}}(n)\mathbf{w}_r(n)b_r(n)}{\sqrt{\mathbf{w}_r^T(n)\tilde{\mathbf{M}}(n)\mathbf{w}_r(n) + \sigma_v^2/2}} + \frac{\sqrt{2/\pi}\tilde{\mathbf{M}}(n)\mathbf{w}_i(n)b_i(n)}{\sqrt{\mathbf{w}_i^T(n)\tilde{\mathbf{M}}(n)\mathbf{w}_i(n) + \sigma_v^2/2}}
\]  

(4.12)

where

\[
\begin{align*}
\tilde{\mathbf{M}}(n) & = a\mathbf{M}_{\text{SOI-KF}}(n-1) + (1-a)\mathbf{C}_h \\
\mathbf{M}(n) & = \mathbf{M}(n) - \frac{(2/\pi)\mathbf{M}(n)\mathbf{w}_r(n)\mathbf{w}_r^T(n)\tilde{\mathbf{M}}(n)}{\mathbf{w}_r^T(n)\tilde{\mathbf{M}}(n)\mathbf{w}_r(n) + \sigma_v^2/2} \\
\mathbf{M}_{\text{SOI-KF}}(n) & = \mathbf{M}(n) - \frac{(2/\pi)\mathbf{M}(n)\mathbf{w}_i(n)\mathbf{w}_i^T(n)\tilde{\mathbf{M}}(n)}{\mathbf{w}_i^T(n)\tilde{\mathbf{M}}(n)\mathbf{w}_i(n) + \sigma_v^2/2}
\end{align*}
\]  

(4.13), (4.14), (4.15)

There are two underlying assumptions in the SOI-KF approach: (1) the actual channel model follows an AR model; and (2) the distribution \(p(\mathbf{h}(n)|\mathbf{b}_{n-1})\) is Gaussian. Relaxing both assumptions, we next develop a MAP estimation and tracking approach that does not assume a specific channel evolution model, and which can yield superior performance relative to the SOI-KF approach, as we will show in the simulations.
4.4.2 2-Bit MAP Approach

We consider the same exponential weighting idea that is used in Section 4.3.2, where the set of measurements \( \{ y(k) = w^H(k)h(k) + v(k) \}_{k=1}^n \) is approximated and replaced with the set \( \{ y(k) = w^H(k)h(n) + \lambda h(k) - n^2 v(k) \}_{k=1}^n \) for \( 0 \leq \lambda \leq 1 \). Using this assumption, we formulate a MAP estimation problem for \( h(n) \), given the \( 2^n \) measurement bits \( \{ b_r(k) \}_{k=1}^n \) and \( \{ b_i(k) \}_{k=1}^n \) [43, Ch. 11]. Note that without assuming a specific channel model, the predicted channel can be taken to be the same as its most recent estimate, i.e., \( \tilde{h}(n) = \hat{h}(n-1) \).

The probability that \( b_r(k) = 1 \) (and similarly for the probability that \( b_i(k) = 1 \)) at time slot \( n \) given \( h(n) \) can be expressed in terms of the \( Q \)-function as

\[
p \left[ b_r(k) = 1 | \bar{h}(n) \right] = \frac{Q \left( \frac{d_r(k) - w^T_r(k)\bar{h}(n)}{\sigma_n(k)} \right)}{Q \left( \frac{d_r(k) - w^T_r(k)\bar{h}(n)}{\sigma_n(k)} \right) + 1 - \frac{1}{2} \bar{h}(n)^T \bar{C}_h^{-1} \bar{h}(n) \}
\]

(4.16)

where \( \sigma_n(k) := \lambda^{k-1} \sigma_v / \sqrt{2} \). Since the noise samples \( \{ v(k) \}_{k=1}^n \) are independent, the probability mass function (pmf) of \( b_n \), given \( \bar{h}(n) \), is given as

\[
p \left[ b_n | \bar{h}(n) \right] = \prod_{k=1}^n Q \left( \frac{-b_r(k) \left( w^T_r(k)\bar{h}(n) - d_r(k) \right)}{\sigma_n(k)} \right) Q \left( \frac{-b_i(k) \left( w^T_i(k)\bar{h}(n) - d_i(k) \right)}{\sigma_n(k)} \right)
\]

(4.17)

Now the MAP estimate can be obtained as

\[
\hat{h}_{\text{MAP}}(n) = \arg \max_{\bar{h}(n)} p \left[ b_n | \bar{h}(n) \right] p \left[ \bar{h}(n) \right]
\]

\[
= \arg \max_{\bar{h}(n)} \sum_{k=1}^n \left[ \log Q \left( \frac{-b_r(k) \left( w^T_r(k)\bar{h}(n) - d_r(k) \right)}{\sigma_n(k)} \right) \right.

+ \left. \log Q \left( \frac{-b_i(k) \left( w^T_i(k)\bar{h}(n) - d_i(k) \right)}{\sigma_n(k)} \right) \right] - \frac{1}{2} \bar{h}(n)^T \bar{C}_h^{-1} \bar{h}(n)
\]

(4.18)

Since the \( Q \)-function is log-concave [22 pp. 104], problem (4.18) is convex and can be solved efficiently using Newton’s method [22, Sec. 9.5].

In Newton’s method, defining the function \( \Phi_n(x) \) as the negative of the objective function in (4.18) (defined explicitly in (4.20)), and starting from a feasible initial point
are used to find the minimizer of the convex function $\Phi_n(x)$ (where $\beta \geq 0$ is the step-size). Closed form expressions for the gradient vector $\nabla \Phi_n(x)$ and the Hessian matrix $\nabla^2 \Phi_n(x)$ are derived in (4.21) and (4.22), respectively.

$$\begin{align*}
\Phi_n(x) := & - \sum_{k=1}^{n} \left[ \log Q \left( \frac{-b_r(k)(w_r^T(k)x - d_r(k))}{\sigma_n(k)} \right) \
& + \log Q \left( \frac{-b_i(k)(w_i^T(k)x - d_i(k))}{\sigma_n(k)} \right) \right] + \frac{1}{2} x^T C_h^{-1} x
\end{align*}$$

(4.20)

$$\begin{align*}
\nabla \Phi_n(x) = & - \sum_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma_n^2(k)}} \left[ b_r(k) e^{\frac{-(w_r^T(k)x - d_r(k))^2}{2\sigma_n^2(k)}} w_r(k) \
& + b_i(k) e^{\frac{-(w_i^T(k)x - d_i(k))^2}{2\sigma_n^2(k)}} w_i(k) \right] + C_h^{-1} x
\end{align*}$$

(4.21)

$$\begin{align*}
\nabla^2 \Phi_n(x) = & \sum_{k=1}^{n} \left[ e^{\frac{-(w_r^T(k)x - d_r(k))^2}{\sigma_n^2(k)}} \frac{1}{2} \frac{-(w_r^T(k)x - d_r(k))^2}{\sigma_n^2(k)} w_r(k) w_r^T(k) \right] \
& + \sum_{k=1}^{n} \left[ e^{\frac{-(w_i^T(k)x - d_i(k))^2}{\sigma_n^2(k)}} \frac{1}{2} \frac{-(w_i^T(k)x - d_i(k))^2}{\sigma_n^2(k)} w_i(k) w_i^T(k) \right] + C_h^{-1}
\end{align*}$$

(4.22)
In order to reduce the complexity of solving (4.18) exactly, we consider applying only a single iteration of Newton’s method (with unit-step $\beta = 1$) to obtain $\hat{h}(n)$, using $\hat{h}(n - 1)$ as the initial point. The proposed low-complexity approximate MAP (AMAP) estimate can be expressed as

$$\hat{h}_{\text{AMAP}}(n) = \hat{h}_{\text{AMAP}}(n - 1) - \left(\nabla^2 \Phi_n(\hat{h}_{\text{AMAP}}(n - 1))\right)^{-1} \nabla \Phi_n(\hat{h}_{\text{AMAP}}(n - 1))$$

(4.23)

Intuitively, when the channel is tracked well, the actual channel $\bar{h}(n)$ at time $n$ is very close to the estimated channel $\hat{h}_{\text{AMAP}}(n - 1)$ at time $n - 1$, hence a single Newton step is sufficient to obtain a close approximation of the exact MAP estimate (4.18). For the rest of this chapter, references to the 2-bit MAP approach will mean the AMAP in (4.23), not the exact MAP in (4.18).

The complexity of computing $\hat{h}_{\text{AMAP}}(n)$ using (4.23) is determined by computing and inverting the $2N \times 2N$ Hessian matrix $\nabla^2 \Phi_n(\hat{h}_{\text{AMAP}}(n - 1))$. Note that because of the exponential increase of $\sigma_n(k)$ as $n - k$ increases, the number of measurement bits that are required to compute $\nabla^2 \Phi_n(\hat{h}_{\text{AMAP}}(n - 1))$ and $\nabla \Phi_n(\hat{h}_{\text{AMAP}}(n - 1))$ (and the corresponding terms in the summation), as $n \to \infty$, are finite. The 2-bit MAP approach is computationally more complex than the SOI-KF approach; however, the performance of the 2-bit MAP approach can be better than that of the SOI-KF approach, as shown in Section 4.6. It is also worth mentioning that, in terms of applications, the proposed 2-bit MAP approach is not restricted to channel tracking – it can be used for general estimation and tracking problems with (very) limited feedback.

### 4.5 Performance Analysis

It is clear that the performance of the considered channel tracking schemes depends on the actual channel model and the choice of beamforming vectors \{w(n)\}. In this section we restrict attention to the analytically tractable AR channel model (4.1).

A greedy beamforming design strategy for the KF approach is to use the beamforming vector $w(n)$ that minimizes $\text{Trace}(M_{\text{KF}}(n))$ at time $n$. This has been considered
From (4.7), the optimization problem can be expressed as

\[ w(n) = \arg \max_w \text{ s.t. } ||w||^2 = 1 \]

\[ \frac{\text{Trace} \left( \hat{M}(n)ww^H \hat{M}(n) \right)}{w^H \hat{M}(n)w + \sigma_v^2} \]

(4.24)

The objective function in (4.24) can be expressed as a Rayleigh quotient as

\[ \text{Trace} \left( \hat{M}(n)ww^H \hat{M}(n) \right) = \frac{w^H \hat{M}^2(n)w}{w^H (\hat{M}(n) + \sigma_v^2 I)w} = \frac{z^H B^{-1/2} \hat{M}^2(n) B^{-1/2} z}{z^H z} \]

where \( B = \hat{M}(n) + \sigma_v^2 I \) and \( z = B^{1/2} w \). The optimal \( z \) that maximizes the Rayleigh quotient \( \frac{z^H E z}{z^H z} \), where \( E := B^{-1/2} \hat{M}^2(n) B^{-1/2} \), is the eigenvector that corresponds to the maximum eigenvalue of \( E \), denoted \( z^* \). Then the optimal beamforming vector solution to (4.24) is obtained as \( w(n) = \frac{B^{-1/2} z^*}{||B^{-1/2} z^*||} \).

Note that there are no guarantees that this greedy beamforming approach yields the best overall estimation/tracking performance for more than one time slot. In fact, we show in the next section via simulations that a different simple beamforming scheme can outperform this greedy beamforming approach, when the channel is spatially correlated (i.e., \( C_h \) is not a diagonal matrix). If \( C_h = \sigma_h^2 I \), and the initial ECM \( M_{KF}(0) = \nu I \), \( \nu \geq 0 \), it is easy to see that the greedy optimization (4.24) selects a single antenna for each \( n \), with different antennas selected in a round-robin fashion, i.e., the \( i \)-th entry of \( w(n) \) is 1 if \( \text{mod}(n, N) + 1 = i \) and 0 otherwise. In the sequel, we will refer to this beamforming scheme as single-antenna beamforming.

The following proposition gives a closed-form expression for the channel estimation MSE with the KF and SOI-KF approaches (for sufficiently large \( n \)), using single-antenna beamforming, and assuming \( C_h = \sigma_h^2 I \).

**Proposition 2** Consider the AR channel model (4.1), the linear observation model (4.3), the single-antenna beamforming scheme, and assume that \( C_h = \sigma_h^2 I \) (and that the
distribution \( p(h(n)|b_{n-1}) \) is Gaussian for the SOI-KF approach. Then,

\[
\varepsilon_{\text{KF}} := \lim_{n \to \infty} \text{Trace} \left( M_{\text{KF}}(n) \right)
= N\sigma_h^2 - \left( \sigma_h^2 - \left( \sqrt{c_1^2 + c_2 - c_1} \right) \right) \frac{1 - \alpha^N}{1 - \alpha}
\]

(4.25)

\[
\varepsilon_{\text{SOI-KF}} := \lim_{n \to \infty} \text{Trace} \left( M_{\text{SOI-KF}}(n) \right)
= N\sigma_h^2 - \left( \sigma_h^2 - \frac{\sqrt{c_4^2 + c_5 c_4}}{c_3} \right) \frac{1 - \alpha^N}{1 - \alpha}
\]

(4.26)

where \( c_1 = (1 - \alpha^N)(\sigma_h^2 + \sigma_v^2)/2\alpha^N \), \( c_2 = (1 - \alpha^N)\sigma_h^2 \sigma_v^2/\alpha^N \), \( c_3 = \alpha^N - \alpha^{2N} (1 - \frac{2}{\pi}) \), \( c_4 = \frac{\sigma_h^2}{2\pi} (1 - \alpha^N) + \frac{\sigma_v^2}{2\alpha} (1 - \alpha^N) (1 - 2\alpha^N (1 - \frac{2}{\pi})) \), and \( c_5 = \sigma_v^2 \sigma_h^2 (1 - \alpha^N) + \sigma_h^4 (1 - \alpha^N)^2 (1 - \frac{2}{\pi}) \).

The proof is provided in the Appendix. Note that analogous closed-form results are not available for general KF or SOI-KF; what allows these results here is our specific choice of pilot beamforming strategy (single-antenna beamforming), which, as we will show in the simulations, also happens to be the best among several alternatives that we tried.

Using the same assumptions as Proposition 2 and the relations \( e(n) := h(n) - \hat{h}_{\text{KF}}(n) \), where \( \mathbb{E}[e(n)] = 0 \), \( \mathbb{E}[||e(n)||^2] = \text{Trace} \left( M_{\text{KF}}(n) \right) \), \( \mathbb{E}[\hat{h}_{\text{KF}}^H(n)e(n)] = 0 \) (orthogonality principle), and \( \mathbb{E}[||h(n)||^2] = N\sigma_h^2 = \mathbb{E}[||\hat{h}_{\text{KF}}(n)||^2] + \mathbb{E}[||e(n)||^2] \), a lower bound on the average achieved SNR with the KF approach for large \( n \) can be obtained as

\[
\gamma_{\text{KF}} := \lim_{n \to \infty} \mathbb{E} \left[ \frac{||\hat{h}_{\text{KF}}^H(n)h(n)||^2}{||\hat{h}_{\text{KF}}(n)||^2} \right]
= \lim_{n \to \infty} \mathbb{E} \left[ ||\hat{h}_{\text{KF}}(n)|| + \frac{\hat{h}_{\text{KF}}^H(n)e(n)}{||\hat{h}_{\text{KF}}(n)||} \right]
= \lim_{n \to \infty} \mathbb{E} [||\hat{h}_{\text{KF}}(n)||^2] + \lim_{n \to \infty} \mathbb{E} \left[ \frac{||\hat{h}_{\text{KF}}(n)\hat{h}_{\text{KF}}(n)||^2}{||\hat{h}_{\text{KF}}(n)||^2} \right]
\geq N\sigma_h^2 - \varepsilon_{\text{KF}}
\]

(4.27)
since \( \psi(n) := E \left[ \frac{\hat{h}_{KF}^H(n)e(n)}{|\hat{h}_{KF}(n)|^2} \right] \geq 0 \). Denoting the \( i \)-th entry of \( \hat{h}_{KF}(n) \) as \( a_i \) for brevity,

\[
\lim_{n \to \infty} \psi(n) = \lim_{n \to \infty} E \left[ \frac{\sum_{i=1}^{N} a_i^2 e_i(n)}{\sum_{i=1}^{N} |a_i|^2} \right] = \lim_{n \to \infty} E \left[ \frac{\sum_{i=1}^{N} |a_i|^2 |e_i(n)|^2}{\sum_{i=1}^{N} |a_i|^2} \right] \approx \varepsilon_{KF}/N \tag{4.28}
\]

where the last approximation step in (4.28) is obtained assuming that \( \hat{h}_{KF}^H(n) \) and \( e(n) \) are independent (they are uncorrelated but not necessarily independent). Hence \( \gamma_{KF} \) can be closely approximated as

\[
\gamma_{KF} \approx N\sigma_h^2 - \left( \frac{N - 1}{N} \right) \varepsilon_{KF} \tag{4.29}
\]

Similarly, a lower bound on the average achieved SNR with the SOI-KF approach at large \( n \) can be obtained as

\[
\gamma_{SOI-KF} := \lim_{n \to \infty} E \left[ \frac{|\hat{h}_{SOI-KF}^H(n)h(n)|^2}{||\hat{h}_{SOI-KF}(n)||^2} \right] \geq N\sigma_h^2 - \varepsilon_{SOI-KF} \tag{4.30}
\]

and a close approximation is obtained as

\[
\gamma_{SOI-KF} \approx N\sigma_h^2 - \left( \frac{N - 1}{N} \right) \varepsilon_{SOI-KF} \tag{4.31}
\]

The approximations (4.29) and (4.31), are evaluated in Section 4.6.

It is easy to verify in Proposition 2 that if \( \alpha \to 1 \) (i.e., the channel is time-invariant), then \( \varepsilon_{KF}, \varepsilon_{SOI-KF} \to 0 \) and \( \gamma_{KF}, \gamma_{SOI-KF} \to N\sigma_h^2 \). In other words if the channel is time-invariant, then the estimation error will go to zero, and the average SNR will reach the case with perfect CSI at the transmitter, as \( n \to \infty \). It is also easy to check that \( \varepsilon_{KF} \) and \( \varepsilon_{SOI-KF} \) are increasing functions in \( N, \sigma_h^2, \) and \( \sigma_v^2 \). An empirical observation made in our simulations is worth mentioning: we noticed that Trace \((M_{KF}(n))\) converges to the limit in (4.25) for \( n \geq 2N \), while Trace \((M_{SOI-KF}(n))\) converges to the limit in (4.26) for \( n \geq 4N \).

A generalization to single-antenna beamforming is the case where the beamforming vector \( w(n) \) is selected as one of the columns of a \( N \times N \) unitary matrix \( U \) in a round-robin fashion, i.e., \( w(n) \) is the \( i \)-th column of \( U \) if \( \text{mod}(n,N) + 1 = i \). We will refer to this scheme as unitary beamforming, and note that single-antenna beamforming is a special case of unitary beamforming with \( U = I \). Based on extensive numerical tests,
we conjecture that the closed-form expressions for $\varepsilon_{KF}$ and $\varepsilon_{SOI-KF}$ in (4.25) and (4.26), respectively, are also applicable for the general case of unitary beamforming, using any unitary matrix $U$. Moreover, we conjecture the optimality of the unitary beamforming scheme in terms of minimizing $\varepsilon_{KF}$ and $\varepsilon_{SOI-KF}$ (and maximizing $\gamma_{KF}$ and $\gamma_{SOI-KF}$), if $C_h = \sigma_h^2 I$.

Intuitively, the beamforming vectors that are used for learning/tracking the channel should provide complementary views of the entire channel vector $h(n)$. For example, the $N \times N$ matrix $[w(n), w(n-1), \ldots, w(n-N+1)]$ should be full-rank if $C_h = \sigma_h^2 I$. Thus, the beamforming vectors $\{w(n)\}$ that are used for pilots $\{s(n)\}$ for channel tracking should be different than the beamforming vectors $\{\tilde{w}(n)\}$ that are used for data transmission. Choosing $w(n) = \tilde{w}(n) = \frac{h(n)}{|h(n)|}$, which is the case considered in [68], yields poor performance. This point is further elaborated in Section 4.6.

4.5.1 Comparing with Codebook-Based Beamforming

As mentioned earlier, the state-of-the-art in transmit beamforming with limited feedback is focused on designing a common beamformer codebook (known at the transmitter and the receiver). The setup assumes that the receiver will accurately estimate the downlink channel, search the codebook, and feed back the index of the best beamformer in the codebook to the transmitter [50]. In [75], it is stated that for beamforming over i.i.d. Rayleigh fading channels with beamformer codebook of size $2^B$ designed by the GLA, the achieved average SNR $\gamma_{LA}$ can be closely approximated as

$$\gamma_{LA} = N \sigma_h^2 - (N-1)\sigma_v^2 2^{-B/(N-1)}$$  \hspace{1cm} (4.32)

Note that expression (4.32) is obtained ignoring the temporal correlation of the channel and assuming perfect CSI at the receiver (unlike the case for $\gamma_{SOI-KF}$).

Figure 4.2 plots the lower bound on $\gamma_{SOI-KF}$ from (4.30) and $\gamma_{LA}$ from (4.32) as $N$ increases, assuming $C_h = I$, $\sigma_v^2 = 0.001$, $\alpha \in \{0.92, 0.96, 0.99\}$, and $B \in \{10, 20, 50\}$. The figure shows the increase of $\gamma_{SOI-KF}$ as $N$ increases and as $\alpha$ increases (i.e., channel becomes more correlated across time). The figure also shows that a large number of feedback bits $B$ (i.e., large codebook) is required for codebook-based beamforming to achieve the same performance as the SOI-KF approach, which is obtained using only 2 feedback bits per channel dwell time block of length $T$. The number of bits $B$ required
for $\gamma_{\text{LA}}$ to achieve $\gamma_{\text{SOI-KF}}$ increases as $N$ or $\alpha$ increases. For example, the figure shows that $\gamma_{\text{SOI-KF}}$ (with $\alpha = 0.99$) outperforms $\gamma_{\text{LA}}$ with $B = 10$ feedback bits for $N \geq 3$, and outperforms $\gamma_{\text{LA}}$ with $B = 50$ feedback bits for $N \geq 22$.

Exploiting the channel temporal correlation to reduce the feedback rate, references [70] and [38] propose modeling the quantized CSI at the receiver using a finite-state Markov chain. As shown in Fig. 4.2 at least $B = 50$ bits are needed to achieve the same SNR performance that is achieved with only 2 feedback bits using the SOI-KF approach when $N = 20$ and $\alpha = 0.99$, for example. This means that at least $2^{50}$ Markov states need to be modeled and $2^{100}$ transition probabilities must be computed in order to apply the compression techniques in [38] and [70], which is clearly computationally prohibitive.

Before moving to the numerical results, two practice-oriented remarks are in order.

- **Variable-length quantization.** To further decrease the feedback rate to 1 bit per $T$, the receiver can send only the bits that correspond to the real measurements, $\{b_r(n)\}$, in even time slots, while the bits that correspond to the imaginary measurements, $\{b_i(n)\}$, are sent in odd time slots (or vice versa). On the other hand,
the estimation performance can be improved by increasing the feedback quantization bits (at the cost of higher feedback rate) using the iteratively quantized Kalman filter approach introduced in [60], where the quantization bits are iteratively formed using the sign of the difference between the observation $y(n)$ and its estimate based on past observations along with previous bits of the current observation.

- **Multiple receive antennas.** Extending this work to a setting with more than one receive antennas (or multiple receivers) is straightforward if the receive antennas are uncorrelated. A separate estimation/tracking problem can be set up for the channel vector that corresponds to each receive antenna.

### 4.6 Numerical Results

To test the performance of the proposed beamforming and feedback techniques, we consider the widely used Jake’s channel model [40] in figures 4.3, 4.4, 4.5, and 4.6. According to Jake’s model, the spatio-temporal correlation matrix can be expressed as $E[h(n)h^H(n-k)] = \rho_k C_h$, for $k \geq 0$, where $\rho_k := J_0(2\pi f_d T k)$, $J_0$ is the 0-th-order Bessel function, and $f_d$ denotes the Doppler frequency. The unitary beamforming scheme that is described in Section 4.5 is used for all figures. The SNR loss, defined as the ratio of the average SNR achieved with perfect CSI at the transmitter (i.e., $E[||h(n)||^2]$) to the average SNR achieved with the estimated channel (i.e., $E[\frac{||h^H(n)h(n)||^2}{||h(n)||^2}]$), is used to measure and compare the performance of the proposed techniques.

The setup for Fig. 4.3 considers a transmitter with $N = 10$ antennas, Doppler frequency $f_d = 10$ Hz, time slot duration $T = 1$ ms (same performance for any values of $f_d$ and $T$ that satisfy $f_d T = 0.01$), spatial correlation matrix $C_h = \sigma_h^2$ Toeplitz $([0.5^0, 0.5^1, \ldots, 0.5^9])$, where $\sigma_h^2 = 0.1$, and observation noise variance $\sigma_v^2 = 0.01$. The figure illustrates the trade-off between the SNR loss of the KF and SOI-KF approaches and $\alpha$, the trade-off between the SNR loss of the MMSE and 2-bit MAP approaches and $\lambda$, the SNR loss using the WF (4.9) (which requires additional knowledge of $\rho_k = J_0(2\pi f_d T k)$ for all $k$) as a baseline. The SNR loss plots are obtained via 1000 Monte-Carlo simulation runs, where each run includes 400 time slots.
Interestingly, Fig. 4.3 shows that the difference between the average receive-SNR achieved using the proposed 2-bit MAP approach with only 2 feedback bits every $T$ seconds (at the optimal $\lambda^* = 0.83$), and the *Genie* receive-SNR achieved with perfect CSI at the transmitter, is less than 1 dB. The figure also shows that the average receive-SNR achieved using the proposed 2-bit MAP approach (at $\lambda^* = 0.83$) is 0.2 dB larger than that achieved using the SOI-KF approach (at $\alpha^* = 0.94$), and is only 0.6 dB less than that achieved using WF (4.9). In other words, the cost of quantizing the received signal $y(n)$ into 2 feedback bits, as compared to the analog-amplitude $y(n)$ feedback, is less than 0.6 dB. Note that in the case of analog-amplitude feedback, it is assumed that $y(n)$ is perfectly known at the transmitter (in addition to the knowledge of $\{\rho_k\}$); accounting for additional uplink (or quantization) errors in the analog feedback case will further decrease the 0.6 dB difference. Another observation from the figure is that the MMSE approach (at $\lambda^* = 0.83$) and the KF approach (at $\alpha^* = 0.94$) are very close in performance.

Table 4.1 uses the same setup as Fig. 4.3 and reports the SNR loss (in dB) with different beamforming schemes at $\lambda^* = 0.83$ and $\alpha^* = 0.94$. The considered beamforming
Table 4.1: SNR loss comparison of different beamforming techniques.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WF</td>
<td>0.39</td>
<td>0.39</td>
<td>0.60</td>
<td>0.84</td>
<td>3.29</td>
</tr>
<tr>
<td>KF</td>
<td>0.54</td>
<td>0.54</td>
<td>0.83</td>
<td>0.96</td>
<td>4.24</td>
</tr>
<tr>
<td>MMSE</td>
<td>0.57</td>
<td>0.58</td>
<td>0.92</td>
<td>1.07</td>
<td>4.79</td>
</tr>
<tr>
<td>SOI-KF</td>
<td>1.14</td>
<td>1.10</td>
<td>1.40</td>
<td>1.42</td>
<td>4.45</td>
</tr>
<tr>
<td>2-Bit MAP</td>
<td>0.94</td>
<td>0.92</td>
<td>1.19</td>
<td>1.27</td>
<td>4.04</td>
</tr>
</tbody>
</table>

schemes, which correspond to the columns of the table, are (in order): (i) the unitary beamforming scheme described in Section 4.5; (ii) the single-antenna beamforming scheme described in Section 4.5; (iii) a random beamforming scheme where $w(n)$ is a normalized Gaussian random vector for each $n$; (iv) the greedy beamforming scheme where $w(n)$ is obtained by solving (4.24); and (v) the case where $w(n)$ corresponds to the most recent channel estimate using the KF approach (i.e., $w(n) = \hat{h}_{KF}(n-1)/||\hat{h}_{KF}(n-1)||$).

The table shows that the performance of the unitary beamforming is almost identical to that of the single-antenna beamforming (small difference within the sample averaging error), which is superior to other considered beamforming schemes. The table also verifies that the greedy beamforming scheme using (4.24) is not optimal, and that using $w(n) = \hat{h}_{KF}(n-1)/||\hat{h}_{KF}(n-1)||$ yields poor performance, as discussed in Section 4.5.

In Fig. 4.4, a large system with $N = 100$ antennas is considered, with Doppler frequency $f_d = 5$ Hz, spatial correlation matrix $C_h = \sigma_h^2$ Toeplitz $([0.9^0, 0.9^1, \ldots, 0.9^{99}])$, where $\sigma_h^2 = 0.01$, and observation noise variance $\sigma_v^2 = 0.001$. Similar to Fig. 4.3, Fig. 4.4 illustrates the trade-off between the SNR loss and the parameters $\lambda$ and $\alpha$, and confirms that the proposed 2-bit MAP approach with only 2 feedback bits every $T$ seconds is applicable even with large $N$. At the optimal $\lambda^* = 0.91$, the SNR achieved with 2-bit MAP approach is 1.7 dB less than the case with perfect CSI at the transmitter, 0.6 dB less than WF with analog-signal feedback, and 0.2 dB higher than the SOI-KF approach (at the optimal $\alpha^* = 0.97$). The results shown in this figure help pave the way for using massive MIMO systems in FDD mode [44], by exploiting the high spatio-temporal channel correlation.

Figure 4.5 considers the same setup and network parameters as Fig. 4.3. The SNR loss that corresponds to the different considered estimation/tracking techniques is plotted versus the Doppler frequency, using the numerically optimized $\lambda$ and $\alpha$. The
SNR loss is increasing with $f_d$ as expected. The figure shows that the SNR loss due to the 2-bit quantization (i.e., 2-bit MAP and SOI-KF approaches), as compared to the case with analog-signal feedback (i.e., KF, MMSE, and WF approaches), is small for small $f_d$, and increases as $f_d$ increases. The figure also shows that the 2-bit MAP approach outperforms the SOI-KF approach for the considered $f_d$ range, and that the MMSE and KF approaches are very close in performance.

In Fig. 4.6, the average achieved SNR using the numerically optimized $\lambda$ and $\alpha$ is plotted as a function of $N$, considering a setup with $f_d = 10$ Hz, $C_h = \sigma_h^2 I$, $\sigma_h^2 = 0.1$, and $\sigma_v^2 = 0.01$. The figure shows that the average SNR is increasing with $N$ as expected, and that the gap between the average SNR achieved with 2-bit quantization (using the 2-bit MAP and SOI-KF approaches) and the average SNR achieved with analog-signal feedback (using the KF, MMSE, and WF approaches), is increasing as $N$ increases. The figure also shows that the 2-bit MAP approach outperforms the SOI-KF approach for the considered range of $N$, and that the MMSE and KF approaches are very close in performance. Using the average SNR expression (4.32) achieved using GLA for the codebook-based beamforming framework (assuming perfect CSI at the receiver), it can
be shown that at least $B = 40$ bits are required to achieve the same performance as the 2-bit MAP approach when $N = 16$ (1.33 dB), and at least $B = 45$ bits are required when $N = 36$ (3.33 dB). Computing the transition probabilities for the finite-state Markov chain model, as considered in [70] and [38], is clearly prohibitive in these cases.

Figure 4.7 considers the AR channel model (4.1), with $N = 10$, $C_h = \sigma_h^2 I$, $\sigma_h^2 = 0.1$, and $\sigma_v^2 = 0.01$. The SNR loss for the considered techniques is plotted versus $\alpha$, where the numerically optimized $\lambda$ is used for the MMSE and 2-bit MAP approaches. The figure also plots the analytical approximations for the KF and SOI-KF approaches using (4.29) and (4.31), respectively. Note that for the AR model (4.1), the performance of the KF (4.5) and the WF (4.9) are identical for large $n$ [43]. The figure shows the decrease of the SNR loss as $\alpha$ increases as expected. The figure also shows that the SOI-KF approach outperforms the 2-bit MAP approach for the considered AR channel model, and that the performances of the MMSE and KF approaches are very close. Moreover, the figure shows that the approximations derived in (4.29) and (4.31) are very tight, particularly for large $\alpha$. Considering the average SNR achieved using GLA for the codebook-based beamforming, it can be shown using (4.32) that at least $B = 12$
bits and $B = 25$ bits are required to achieve the same performance of the SOI-KF approach when $\alpha = 0.95$ and $\alpha = 0.99$, respectively.
Figure 4.7: Average SNR vs. $\alpha$ using the AR model (4.1) with $N = 10$. 
Chapter 5

Conclusions and Future Research

This thesis considered two applications for parameter estimation (and tracking) in wireless networks from few bits, namely, power spectrum sensing using a network of low-end sensors, and channel tracking for transmit beamforming with limited feedback. In this final chapter, the main contributions of this thesis are summarized and possible directions for future research are pointed out.

5.1 Thesis Conclusions

In chapters 2 and 3 a network sensing scenario was considered, where scattered low-end sensors pass the received signal through a random wideband filter, measure average power at the output of the filter, and send out a bit or coarsely quantized power level to a FC. In Chapter 2, the FC obtains an estimate of the power spectrum by solving an under-determined linear program comprising inequality constraints derived from the sensor data, plus prior information in the form of the cost function and non-negativity constraints. The formulation here can be viewed as generalizing classical nonparametric power spectrum estimation to the case where the data is in the form of inequalities, rather than equalities. It was shown that adequate power spectrum sensing is possible from relatively few bits, even for dense spectra. The selection of some key design parameters was considered, and important trade-offs were revealed and illustrated in pertinent simulations. It was demonstrated that judicious choice of the filter length
is needed to balance smearing effects against inequalities-versus-unknowns considerations, and the detection threshold at the sensors should be tuned such that number of sensors reporting measurements above it decreases as the power spectrum becomes more sparse. Assuming availability of a ‘downlink’ channel that the FC can use to send threshold information, active sensing strategies were developed that adaptively select the thresholds online using ideas borrowed from cutting plane methods, yielding significantly faster convergence to the true finite-length autocorrelation compared to passive sensing. The results underscore the importance of judicious threshold design / adaptation in the context of distributed power spectrum sensing.

Frugal sensing was revisited from a statistical estimation point of view in Chapter 3, taking into account the effects of fading and insufficient sample averaging on the soft power measurements prior to quantization. The distribution of the corresponding error is shown to be approximately Gaussian, which was exploited by formulating ML estimation as a convex optimization problem that yields consistent estimates, and optionally includes a sparsity-inducing penalty term and non-negativity constraints for better estimation performance in the small sample-size regime. Simulations have shown that satisfactory PS estimates can be obtained with ML sensing from few bits, even when relatively many bits are flipped due to fading-induced measurement errors. Extensions to parametric estimation were considered for the case of line spectra and the case where the emitter spectral shapes are known. Furthermore, a censoring scheme was proposed where only sensors that provide the most useful information bits are permitted to send, while other less-informative sensors remain silent.

A new approach for channel estimation and tracking for transmit beamforming with (very) limited feedback was proposed in Chapter 4. Instead of putting the burden of channel estimation and codebook search on the receiver, the bulk of the work is shifted to the transmitter. Using separate beamforming weight vectors for pilot and payload transmission, the transmitter sends a single pilot symbol per channel dwell time block, while the receiver simply sends back a coarsely quantized 2-bit version of the received pilot signal (or the corresponding innovation, in the case of AR modeling). For channel tracking, a novel 2-bit MAP algorithm was proposed, as a ‘universal’ complement to an extended version of the SOI-KF framework, which is advocated when the channel can be modeled as an AR process. In the AR case, closed-form expressions for the
resulting channel estimation MSE and very tight approximations for the corresponding
SNR were derived, assuming circular single-antenna beamforming for the pilots. Careful
simulations confirmed that by exploiting the spatio-temporal correlation of the channel,
the performance achieved using the proposed frugal feedback approaches is close to that
attainable with perfect CSI at the transmitter. Simulations also showed that very
large-size codebooks are required for codebook-based beamforming to achieve the same
performance as the proposed approaches. The results of this chapter help pave the
way for using transmit beamforming for massive MIMO in FDD instead of TDD mode,
which was almost impossible with the prior state-of-art.

The frugal estimation (and tracking) techniques proposed in this thesis are not re-
stricted to the power spectrum sensing and transmit beamforming applications; they
can be used with a wide range of applications. Consider the general quantization-
based observation / measurement model $b = \text{sign}(y - t)$, where $b = [b_1, \ldots, b_N]^T$,
$t = [t_1, \ldots, t_N]^T$, $y = [y_1, \ldots, y_N]^T$, $y_n = a_n^T x + v_n$, $v_n \sim \mathcal{N}(0, \sigma_v^2)$, and \{v_n\}_{n=1}^N are
i.i.d. The objective is to estimate the vector $x \in \mathcal{F}$ from $b$ (i.e., estimation from in-
equalities), where $\mathcal{F}$ is the convex feasible set for $x$. Note that the estimation task using
this (nonlinear) quantization model is very different from traditional statistical signal
processing techniques where it is assumed that $y$ is accessible at the central processing
unit to estimate $x$ \cite{43} (i.e., estimation from equalities). Consider the following cases:

- If $x$ is deterministic and the errors \{v_n\} are negligible, then an estimate $\hat{x}$ can be
  obtained using an LP formulation, similar to (2.16) in Chapter 2.

- If $x$ is deterministic (or random with known prior distribution $x \sim p(x)$) and the
  errors \{v_n\} cause bit-flips, then an estimate $\hat{x}$ can be obtained using an ML (resp.
  MAP) formulation, similar to (3.7) in Chapter 3.

- If $x$ is changing slowly with time, then $\hat{x}$ can be tracked using an ML (or MAP)
  tracking formulation with an exponential forgetting factor, similar to (4.18) in
  Chapter 4.

In addition to the power spectrum sensing and transmit beamforming applications con-
sidered in this thesis, similar one bit quantization models (but different estimation
formulations) have been considered for compressed sensing \cite{20,39,63,80}, distributed
estimation in WSNs [29, 59, 60], and consumer preference measurement using conjoint analysis [74].

5.2 Future Work

Frugal sensing has focused on wide-sense stationary signals, but digital communication signals are in fact cyclostationary. It is therefore appealing to extend our approach to cover cyclostationary signals. Interestingly, preliminary results have confirmed that the frugal sensing setup considered in this thesis is also applicable with cyclostationary signals. If applied to a cyclostationary signal, the LP and ML formulations in chapters 2 and 3, respectively, will estimate the averaged cyclic PS.

Another interesting extension is to consider parametric frugal sensing for signals that can be modeled by autoregressive (AR) models, moving-average (MA) models, or autoregressive-moving-average (ARMA) models [71]. Preliminary results for such signal models indicate that the PS estimation problem can be formulated as a nonconvex quadratically constrained quadratic problem (QCQP – NP-hard in general), where the semidefinite relaxation (SDR) technique can be used to find an approximate solution [52]. An alternative approach to the SDR in this case is to linearly approximate the nonconvex quadratic constraints (i.e., using first-order Taylor series), then solve a sequence of convex second-order cone programs (SOCPs) until convergence, similar to the approach presented in [73]. Preliminary simulations have shown that for MA signal models, the performance of the parametric estimation of the PS, obtained using the SDR or the sequential-SOCP techniques, is better than that of the nonparametric estimation techniques presented in this thesis.

One interesting research direction that can be pursued in the future is to use the scattered low-end sensors in forming a power spectral map in space and frequency, i.e., spectrum cartography. Sensors can cooperate to estimate the distribution of power across spatial locations and frequencies, enabling the identification of the (un)used frequency bands at arbitrary locations, thus facilitating spatial frequency reuse. Today’s smart phones and tablets are ideal platforms for crowdsourc ing spectrum sensing, and they can be used to create a spectrum sensing web that spans across much of our living and working space. Each sensor can measure its local (in space and time) power.
spectral density, or some summary statistic, and encode the information concisely into a message that it posts using a service model like Twitter. Secondary users wishing to communicate in a given area can elect to “follow” the tweets posted by sensors that are geographically close to themselves, and each secondary user creates an estimate of the spatio-temporal spectral occupancy in its immediate vicinity from the tweets it receives. This paradigm brings up challenging research questions. A tweet is (at most) a mere 140 characters (1120 bits); GHz-order sensing practically implies compression ratios in the order of 1 bit/MHz. Effective power spectrum compression, fusion, spatial interpolation, and spatio-temporal prediction methods are in turn needed, together with agile measurement-based access protocols that take advantage of such an advanced distributed sensing fabric.
References


[70] C. Simon and G. Leus, “Feedback reduction for spatial multiplexing with linear pre-


Proc. of IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), vol. 4,

[73] L. N. Tran, M. Hanif, and M. Juntti, “A conic quadratic programming approach to
physical layer multicasting for large-scale antenna arrays,” IEEE Signal Processing

analysis through maximum likelihood estimation,” IEEE Transactions on Signal

[75] P. Xia and G. B. Giannakis, “Design and Analysis of Transmit-Beamforming based

[76] F. Zhang, Matrix Theory: Basic Results and Techniques. New York: Springer-
Verlag, 1999.


[78] Q. Zhao and A. Swami, “A survey of dynamic spectrum access: signal processing
and networking perspectives,” Proc. of IEEE Int. Conf. Acoustics, Speech and

[79] F. Zeng, C. Li, and Z. Tian, “Distributed compressive spectrum sensing in cooperative
multihop cognitive networks,” IEEE J. Select. Topics Signal Process., vol. 5,
Appendix A

Technical Details and Proofs

A.1 Fading Considerations

First note that if the discrete signal $y_m(n)$ is received in presence of frequency-flat fading, then the difference in the received power spectrum across sensors can be compensated for using AGC. Consider now a more general frequency-selective fading scenario. The received signal $y_m(n)$ is the convolution of the transmitted discrete-time WSS signal $x(n)$ with the linear (possibly time-varying) finite-impulse response fading channel \( \{h_m(n; \ell)\}_{\ell=0}^{L-1} \), expressed as $y_m(n) = \sum_{\ell=0}^{L-1} h_m(n; \ell)x(n - \ell)$. Assuming that $x(n)$ is independent of $\{h_m(n; \ell)\}$, the received autocorrelation is thus given as

$$
\mathbb{E}[y_m(n)y_m^*(n-k)] = \mathbb{E} \left[ \sum_{\ell_1=0}^{L-1} h_m(n; \ell_1)x(n - \ell_1) \sum_{\ell_2=0}^{L-1} h_m^*(n - k; \ell_2)x^*(n - k - \ell_2) \right] \\
= \sum_{\ell_1=0}^{L-1} \sum_{\ell_2=0}^{L-1} \mathbb{E}[h_m(n; \ell_1)h_m^*(n - k; \ell_2)]r_x(k + \ell_2 - \ell_1). \tag{A.1}
$$

Next, we consider two scenarios for the fading channel.

Scenario 1: $\{h_m(n; \ell)\}$ is random, time-invariant, and the correlation between two filter taps is only a function of the ordinal distance between them. This implies that

$$
\mathbb{E}[h_m(n; \ell_1)h_m^*(n - k; \ell_2)] = \mathbb{E}[h_m(\ell_1)h_m^*(\ell_2)] = r_{h_m}(\ell_1 - \ell_2).
$$

96
Then, from (A.1):

$$E[y_m(n)y_m^*(n-k)] = \sum_{\ell_1=0}^{L-1} \sum_{\ell_2=0}^{L-1} r_{hm}(\ell_1 - \ell_2)r_x(k + \ell_2 - \ell_1)$$

$$= \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{hm}(\ell)r_x(k - \ell)$$

$$= r_{ym}(k)$$

and thus $y_m(n)$ is WSS, and the received power spectrum is expressed as

$$S_{ym}(\omega) = \sum_{k=-\infty}^{\infty} r_{ym}(k)e^{-j\omega k}$$

$$= \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{hm}(\ell)e^{-j\omega \ell} \sum_{k=-\infty}^{\infty} r_x(k)e^{-j\omega \ell}$$

$$= S_{hm}(\omega)S_x(\omega)$$

where $S_{hm}(\omega) := \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{hm}(\ell)e^{-j\omega \ell}$. Note that since the channel frequency response is given as $H_m(\omega) = \sum_{\ell=0}^{L-1} h_m(\ell)e^{-j\omega \ell}$, then

$$E[|H_m(\omega)|^2] = \sum_{\ell_1=0}^{L-1} \sum_{\ell_2=0}^{L-1} E[h_m(\ell_1)h_m^*(\ell_2)]e^{-j\omega(\ell_1-\ell_2)}$$

$$= \sum_{\ell=-L+1}^{L-1} (L - |\ell|)r_{hm}(\ell)e^{-j\omega \ell}$$

$$= S_{hm}(\omega)$$

Assuming that $E[|H_m(\omega)|^2]$ is the same across all sensors, and that sensors acquire sufficient samples with different channel realizations such that the sample average converges to the expectation, then all sensors will be reporting consistent power spectrum measurements. This effectively assumes that the channel remains constant over a relatively long period of time, then jumps to a new realization, dwells there for another measurement epoch, and so on. This is a reasonable model if each sensor only spends a small part of its time to sense the spectrum, while it does other things most of the
time. Every time it returns to the spectrum sensing task, it will encounter a new channel realization, not only because of drift but also due to acquiring a new carrier/phase lock. If the reported measurements reflect averaging over many such epochs, then the proposed model is well-motivated.

Scenario 2: The Wide Sense Stationary Uncorrelated Scattering (WSSUS) channel model [32, Sec. 3.3], first introduced by Bello [17], where $h_m(n; \ell)$ is WSS with respect to the time variable $n$ and uncorrelated across the lag variable $\ell$. This implies that $E[h_m(n; \ell_1)h_m^*(n - k; \ell_2)] = r_{h_m}(k; \ell_1)\delta(\ell_1 - \ell_2)$. Hence, substituting in (A.1) yields:

$$E[y_m(n)y_m^*(n - k)] = \sum_{\ell=0}^{L-1} E[h_m(n; \ell)h_m^*(n - k; \ell)]r_x(k) = \phi_m(k)r_x(k) = r_y(k)$$

where $\phi_m(k) := \sum_{\ell=0}^{L-1} E[h_m(n; \ell)h_m^*(n - k; \ell)]$. For slowly varying channels, $h_m(n; \ell) \approx h_m(n - k; \ell)$ for the (small) range of autocorrelation lags considered here, which implies that $\phi_m(k) \approx \sum_{\ell=0}^{L-1} E[|h_m(n; \ell)|^2]$ is approximately constant (not a function of $k$). Hence, all sensors will be reporting consistent power spectrum measurements, assuming that sensors acquire sufficient samples such that the sample average converges to the expectation.
A.2 Analog Sensor Measurement Chain

Figure A.1: Sensor measurement chain: analog processing.

Assume that the complex-valued analog signal \( x(t) \) is bandlimited with two-sided bandwidth \( 1/T \) (i.e., Nyquist rate = \( 1/T \)). Let \( \tilde{g}_m(t) \) be the impulse response of the analog filter of duration \( KT \) that corresponds to the FIR filter \( g_m(n) \), satisfying \( \tilde{g}_m(t) = g_m(n) \) for \( nT < t \leq (n+1)T \), where \( n = 0, \ldots, K - 1 \), and \( \tilde{g}_m(t) = 0 \) for \( t > KT \) and \( t < 0 \). Let the discrete-time signal \( x(n) \) be the output samples from passing \( x(t) \) through an integrate and dump device operating at Nyquist rate:

\[
x(n) = \int_{(n-1)T}^{nT} x(t)dt.
\]

Passing the signal \( x(t) \) through the filter \( \tilde{g}_m(t) \) yields

\[
\tilde{z}_m(t) = \int_0^{KT} \tilde{g}_m(\tau)x(t-\tau)d\tau
\]

\[
= \int_0^{T} \tilde{g}_m(\tau)x(t-\tau)d\tau + \int_T^{2T} \tilde{g}_m(\tau)x(t-\tau)d\tau + \ldots + \int_{(K-1)T}^{KT} \tilde{g}_m(\tau)x(t-\tau)d\tau
\]

\[
= g_m(0) \int_0^{T} x(t-\tau)d\tau + g_m(1) \int_T^{2T} x(t-\tau)d\tau + \ldots + g_m(K-1) \int_{(K-1)T}^{KT} x(t-\tau)d\tau
\]

\[
= \sum_{\ell=0}^{K-1} g_m(\ell) \int_{\ell T}^{(\ell+1)T} x(t-\tau)d\tau
\]
Now, consider the Nyquist-rate samples of $\tilde{z}_m(t)$ at $t = nT$,

$$
\tilde{z}_m(nT) = \sum_{\ell=0}^{K-1} g_m(\ell) \int_{\ell T}^{(\ell+1)T} x(nT - \tau)d\tau
= \sum_{\ell=0}^{K-1} g_m(\ell) \int_{(n-\ell)T}^{(n-\ell-1)T} x(\tilde{\tau})d\tilde{\tau}
= \sum_{\ell=0}^{K-1} g_m(\ell)x(n - \ell)
$$

which is the discrete-time convolution of $x(n)$ and $g_m(n)$. This shows that

$$
\frac{1}{T} \int_0^{NT} |\tilde{z}_m(t)|^2 dt \approx \sum_{n=0}^{N} |z_m(n)|^2
$$

The modified analog measurement chain is depicted in Fig. A.1.

### A.3 Proof of Proposition 1

We show that enforcing nonnegativity of the discretized $N_F$-point power spectrum estimate, i.e., $s_x = \tilde{F}\tilde{r}_x \geq 0$, where $s_x(f) = S_x \left( \frac{2\pi f}{N_F} \right)$, $f \in \{0, \ldots, N_F-1\}$, implies a positive semidefinite $K \times K$ autocorrelation matrix $R_x$. We consider $N_F \geq (2K - 1)$ and assume that $N_F$ is odd (extending the proof to even $N_F$ follows along the same lines). Define the $N_F \times 1$ vector $\bar{r}_x$ as the zero-padded extension of $\tilde{r}_x$, $\bar{r}_x := [0 \ldots 0 \tilde{r}_x^T 0 \ldots 0]$. Also, define $n_F := \frac{N_F-1}{2}$ and let $\bar{F}$ be the square $N_F \times N_F$ phase-shifted DFT matrix:

$$
\bar{F} = \begin{bmatrix}
1 & \cdots & 1 & \cdots & 1 \\
e^{-j \frac{2\pi}{N_F} (-n_F)} & \cdots & e^{-j \frac{2\pi}{N_F} (n_F)} & \cdots & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
e^{-j \frac{2\pi}{N_F} (N_F-1)(-n_F)} & \cdots & 1 & \cdots & e^{-j \frac{2\pi}{N_F} (N_F-1)(n_F)}
\end{bmatrix}.
$$

It is easy to verify that $\bar{F}\bar{r}_x = \bar{F}\tilde{r}_x = s_x$. Let matrix $W$ be the original (non-phase-shifted) $N_F$-point DFT matrix, vector $v$ be the first column of $\bar{F}$, and define the diagonal matrix $D := \text{diag}(v)$ with elements of $v$ on the main diagonal, such that $\bar{F} = DW$ (and $W = D^H\bar{F}$).
Let \( \tilde{r}_x^{(j)} \) be the \( j \)-th circular shift of \( \tilde{r}_x \) obtained by removing the last \( j \) entries of \( \tilde{r}_x \) and putting them as the first \( j \) entries (with \( \tilde{r}_x^{(0)} = \tilde{r}_x \)). A negative \( j \) signifies a shift in the reverse direction. Define the \( N_F \times N_F \) circulant matrix \( \mathbf{R}_c := [\tilde{r}_x^{(-N_F)}, \ldots, \tilde{r}_x^{(0)}, \ldots, \tilde{r}_x^{(N_F)}] \). For example, for \( K = 2 \) and \( N_F = 5 \),

\[
\mathbf{R}_c = \begin{bmatrix}
  r(0) & r(-1) & 0 & 0 & r(1) \\
  r(1) & r(0) & r(-1) & 0 & 0 \\
  0 & r(1) & r(0) & r(-1) & 0 \\
  0 & 0 & r(1) & r(0) & r(-1) \\
  r(-1) & 0 & 0 & r(1) & r(0)
\end{bmatrix}.
\]

Circulant matrices are diagonalized by a DFT: \( \mathbf{R}_c = \frac{1}{N_F} \mathbf{W}^H \mathbf{A} \mathbf{W} \), where \( \mathbf{A} = \text{diag} \left( \mathbf{W} \tilde{r}_x^{(-N_F)} \right) \) holds the eigenvalues of \( \mathbf{R}_c \) [76, p. 107]. Note that \( \mathbf{W} \tilde{r}_x^{(-N_F)} = \tilde{\mathbf{R}}_x^{(0)} \).

Since we enforce \( \tilde{\mathbf{F}} \tilde{r}_x = \mathbf{s}_x \geq 0 \), this directly implies that \( \mathbf{R}_c \) is positive semidefinite. Next, it is easy to see that the \( K \times K \) autocorrelation matrix \( \mathbf{R}_x \) can be obtained by deleting the last \( N_F - K \) rows and the last \( N_F - K \) columns of \( \mathbf{R}_c \), i.e., \( \mathbf{R}_x \) is the \( K \)-th order leading principal submatrix of \( \mathbf{R}_c \). Sylvester’s criterion states that a matrix is positive semidefinite if and only if the determinant of every principal submatrix is nonnegative [76, p. 160]. This implies that if \( \mathbf{R}_c \succeq 0 \), then the principal submatrix \( \mathbf{R}_x \succeq 0 \). Hence, we showed that enforcing \( \tilde{\mathbf{F}} \tilde{r}_x \geq 0 \) implies that \( \mathbf{R}_x \succeq 0 \). The converse is not true since \( \mathbf{R}_x \succeq 0 \) does not necessarily imply that \( \mathbf{R}_c \succeq 0 \).

### A.4 Proof of Claim 1

Let \( R^{(m)} \) denote the radius of the largest ball centered at \( x^{(m)} \) that lies inside \( \mathcal{P}_m \), which is the solution to the LP (2.28). The convergence of the sequence \( \{R^{(m)}\}_{m=1}^{\infty} \) to zero has been established by Theorem 1 in [28]; we present the proof here for completeness. It is easy to see that \( R^{(m)} \geq 0 \) and \( R^{(m)} \geq R^{(m+1)}, \forall m \). It is also easy to see that \( \|x^{(k)} - x^{(m)}\| \geq R^{(k)}, \forall k > m \). Since \( \{R^{(m)}\}_{m=1}^{\infty} \) is a bounded monotone sequence, then \( \lim_{m \to \infty} R^{(m)} = \bar{R} \geq 0 \). Since any sequence \( \{x^{(m)}\}_{m=1}^{\infty} \) in compact set has a subsequence that converges to a point in the set, and every convergent sequence is a Cauchy sequence, this means that for every \( \epsilon > 0 \), there exists \( N, m, \) and \( k \), such that \( \|x^{(k)} - x^{(m)}\| < \epsilon \), for \( m, k \geq N \). Now, suppose \( \bar{R} > 0 \). This means that \( \|x^{(k)} - x^{(m)}\| \geq R^{(k)} \geq \bar{R} \),
and for $0 < \epsilon < \bar{R}$, there does not exist $N$ such that $||x^{(k)} - x^{(m)}|| < \epsilon$ for $m, k \geq N$, which contradicts that $\{x^{(m)}\}_{m=1}^{\infty}$ must have a Cauchy subsequence. Therefore, $\bar{R} = 0$.

It is also shown in [28], Theorem 7, that the sequence $\{R^{(m)}\}_{m=1}^{M}$ has a linear rate of convergence. The remaining issue is to show that the convergence of $R^{(M)}$ to zero as $M \to \infty$ implies that $x^{(M)}$ converges to the single point $r_x$.

Consider the set of linear inequalities defining the bounded polyhedron $\mathcal{P}_M = \{x \in \mathbb{R}^n| a^T_m x \leq c_m, m = 1, \ldots, M\}$, where $n := 2K - 1$, $a_m := -b_m q_m$, and $c_m := -b_m t_m$ (ignoring the initial $\mathcal{P}$ without loss of generality). An inequality $a^T_m x \leq c_m$ is redundant if it can be omitted without changing $\mathcal{P}_M$, whereas if $\mathcal{P}_M$ changes by removing an inequality $a^T_m x \leq c_m$, then we denote this inequality as active. Define the set $S_M$ as the set of indices that correspond to active inequalities defining $\mathcal{P}_M$, such that $m \in S_M$ implies that $a^T_m x \leq c_m$ is active, and $\mathcal{P}_M = \{x \mid a^T_m x \leq c_m, m \in S_M\}$. Note that $n + 1 \leq |S_M| \leq M$, since an $n$-dimensional polyhedron is bounded by the intersection of at least $n + 1$ half-spaces. Assuming that any $n$ vectors from the set $\{a_m\}_{m \in S_M}$ are linearly independent, then $\mathcal{P}_M$ is non-degenerate, i.e., $\mathcal{P}_M$ cannot be contained in a single $n$-dimensional (or $< n$) hyperplane. This means that at the limit $R^{(M)} \to 0$, $\mathcal{P}_M$ converges to a single point, and the inequality constraints become equalities $a^T_m x = c_m$, for $m \in S_M$. Since the set of constraints are consistent, any $n$ vectors of $\{a_m\}_{m \in S_M}$ are linearly independent, $|S_M| \geq n + 1$, and $r_x \in \mathcal{P}_M$, then the unique solution to $a^T_m x = c_m$, for $m \in S_M$, is $x = r_x$. The linear independence condition for $\{a_m\}_{m \in S_M}$ is guaranteed with high probability if the random vectors $\{q_m\}_{m=1}^{M}$ are chosen from a discrete distribution with large $K$ (probability increases with $K$), and with probability-one if $\{q_m\}_{m=1}^{M}$ are chosen from a continues distribution.
A.5 Sensor Measurement Error due to Fading

Here we show that for large number of channel taps \( L \), the errors due to the fading channel \( \{ \tilde{e}_m \}_{m=1}^M \) can be approximated as i.i.d. zero-mean Gaussian random variables. Let \( S_{z_m}(\omega) \) denote the PS of the WSS signal \( z(n) \), \( H_m(\omega) := \sum_{\ell=0}^{L-1} h_m(\ell) e^{-j\omega\ell} \) denote the frequency response of the channel, and \( G_m(\omega) := \sum_{\ell=0}^{K-1} g_m(\ell) e^{-j\omega\ell} \) denote the frequency response of the filter, such that \( S_{z_m}(\omega) = |H_m(\omega)|^2 |G_m(\omega)|^2 S_x(\omega) \). Thus

\[
\alpha_m = E[|z_m(n)|^2] = r_{z_m}(0)
= \frac{1}{2\pi} \int_0^{2\pi} S_{z_m}(\omega) d\omega
\approx \frac{1}{L} \sum_{k=0}^{L-1} \left| \tilde{H}_m \left( \frac{2\pi k}{L} \right) \right|^2 \left| \tilde{G}_m \left( \frac{2\pi k}{L} \right) \right|^2 \tilde{S}_x \left( \frac{2\pi k}{L} \right)
= \frac{1}{L} \sum_{k=0}^{L-1} \left| \tilde{H}_m(k) \right|^2 \left| \tilde{G}_m(k) \right|^2 \tilde{S}_x(k) \tag{A.2}
\]

where \( \tilde{H}_m(k) \) and \( \tilde{G}_m(k) \) correspond to the \( L \)-point DFT of \( h_m(n) \) and \( g_m(n) \), respectively, and \( \tilde{S}_x(k) = S_x \left( \frac{2\pi k}{L} \right) \). The approximation is accurate for large \( L \), \( r_x(\ell) \approx 0 \) for large \( \ell \) (implying slowly varying \( S_x(\omega) \) with \( \omega \)), and \( \tilde{S}_x(k) > 0 \) for large number of samples. Similarly, it is easy to see that

\[
\tilde{\alpha}_m \approx \frac{1}{L} \sum_{k=0}^{L-1} \left| \tilde{G}_m(k) \right|^2 \tilde{S}_x(k) \tag{A.3}
\]

Therefore, the sensor measurement error due to fading is approximated as

\[
\tilde{e}_m = \alpha_m - \tilde{\alpha}_m
\approx \frac{1}{L} \sum_{k=0}^{L-1} \left| \tilde{G}_m(k) \right|^2 \tilde{S}_x(k) \left( \left| \tilde{H}_m(k) \right|^2 - 1 \right)
= \frac{1}{L} \sum_{k=0}^{L-1} a_m(k) \delta_m(k) \tag{A.4}
\]

where \( a_m(k) := |\tilde{G}_m(k)|^2 \tilde{S}_x(k) \geq 0 \) and \( \delta_m(k) := |\tilde{H}_m(k)|^2 - 1 \). Note that the random variable \( |\tilde{H}_m(k)|^2 \) corresponds to a sum of squares of 2 independent zero-mean
and $1/\sqrt{2}$-variance Gaussian random variables, which yields a unit-mean and unit-variance exponentially distributed random variable (i.e., distribution parameter $\lambda = 1$). Since $\{h_m(\ell)\}_{\ell=0}^{L-1}$ are assumed i.i.d. Gaussian random variables with zero-mean and variance $1/L$, then the $L$-point DFT, $\{\tilde{H}_m(k)\}_{k=0}^{L-1}$, are i.i.d. Gaussian random variables with zero-mean and unit-variance. Thus, the error $\tilde{e}_m$ in (A.4) is approximated as a (weighted) sum of $L$ independent random variables $\{\delta_m(\ell)\}_{\ell=0}^{L-1}$. Hence $E[\delta_m(k)] = E[|\tilde{H}_m(k)|^2] - 1 = 0$ and $E[\delta_m^2(k)] = E[|\tilde{H}_m(k)|^4] - 2E[|\tilde{H}_m(k)|^2] + 1 = 1$, which implies that $E[\tilde{e}_m] = 0$ and $\text{var}(\tilde{e}_m) = \frac{1}{L^2} \sum_{k=0}^{L-1} a_m^2(k) \approx \frac{1}{L} E[a_m^2(k)]$. Since $E[a_m^2(k)]$ is independent of $L$, this means that $\text{var}(\tilde{e}_m)$ is increasing with $1/L$ (and is independent of $K$).

**Lyapunov central limit theorem.** Define $X_k = \frac{1}{L} a_m(k) \delta_m(k)$ and $s_T^2 := \sum_{k=0}^{L-1} \text{var}(X_k) = \frac{1}{L^2} \sum_{k=0}^{L-1} a_m^2(k)$. The Lyapunov central limit theorem [18, pp. 371] states that if $\{X_1, X_2, \ldots\}$ is a sequence of independent random variables, each with zero-mean and finite variance, and if for some $\zeta > 0$, the Lyapunov condition $\lim_{L \to \infty} \frac{1}{s_T^2} \sum_{k=1}^{L} E[|X_k|^{2+\zeta}] = 0$ is satisfied, then $\frac{1}{s_T} \sum_{k=1}^{L} X_k \overset{d}{\to} \mathcal{N}(0, 1)$. To verify that the Lyapunov condition is satisfied in our case, set $\zeta = 1$. It is easy to see that $E[\delta_m^3(k)] = 2$ and $E[X_k^3] = 2a_m^3(k)$. Without loss of generality, we assume $a_m(k) > 0, \forall k$ (ignoring zero values). Define $a_{\text{max}} := \max_k a_m(k)$ and $a_{\text{min}} := \min_k a_m(k) > 0$. Hence, it is easy to see that $\frac{\sum_{k=1}^{L} a_m^2(k)}{(\sum_{k=0}^{L-1} a_m^2(k))^{3/2}} \leq \frac{La_{\text{max}}^3}{(L a_{\text{min}}^2)^{3/2}}$. Since $\lim_{L \to \infty} \frac{a_{\text{max}}^2}{L^{1/2} a_{\text{min}}^2} = 0$, then we have shown that the Lyapunov condition is satisfied. Therefore, according to the Lyapunov central limit theorem, $\left(\frac{L}{\sqrt{\sum_{k=0}^{L-1} a_m^2(k)}}\right) \tilde{e}_m$ converges in distribution to a zero-mean and unit-variance Gaussian distribution as $L$ increases.
A.6 Fisher Information Matrix Derivation

The gradient and Hessian of \( \mu_m := \log Q \left( \frac{-b_m(q^T_m r_x - t_m)}{\sigma_m} \right) \) (with respect to \( r_x \)) are derived as

\[
\nabla \mu_m = \frac{b_m e^{-\frac{(q^T_m r_x - t_m)^2}{2\sigma_m^2}} q_m}{\sqrt{2\pi \sigma_m^2} Q \left( \frac{-b_m(q^T_m r_x - t_m)}{\sigma_m} \right)} \tag{A.5}
\]

\[
\nabla^2 \mu_m = -\left[ \frac{e^{-\frac{(q^T_m r_x - t_m)^2}{2\sigma_m^2}}}{2\pi \sigma_m^2} \left[ Q \left( \frac{-b_m(q^T_m r_x - t_m)}{\sigma_m} \right) \right]^2 + \frac{b_m(q^T_m r_x - t_m) e^{-\frac{(q^T_m x - t_m)^2}{2\sigma_m^2}}}{\sqrt{2\pi \sigma_m^2} Q \left( \frac{-b_m(q^T_m r_x - t_m)}{\sigma_m} \right)} \right] q_m q^T_m \tag{A.6}
\]

The PMF of \( b_m \) can be expressed as

\[
b_m = \begin{cases} 
1 & \text{with probability } Q \left( \frac{-q^T_m r_x - t_m}{\sigma_m} \right) \\
-1 & \text{with probability } Q \left( \frac{q^T_m r_x - t_m}{\sigma_m} \right) 
\end{cases} \tag{A.7}
\]

Hence, taking the expectation with respect to \( b_m \), it is easy to see that \( \mathbb{E} [\nabla \mu_m] = 0 \) (satisfying the regularity condition on the log-likelihood [43]), and

\[
\mathbb{E} [\nabla^2 \mu_m] = -\frac{e^{-\frac{(q^T_m r_x - t_m)^2}{2\sigma_m^2}}}{2\pi \sigma_m^2} \mathbb{E} \left[ \frac{1}{Q \left( \frac{-b_m(q^T_m r_x - t_m)}{\sigma_m} \right)} \right]^2 q_m q^T_m \tag{A.8}
\]

Substituting with \( \mathbb{E} [\nabla^2 \mu_m] \) yields the FIM expression [338].
A.7 Proof of Proposition 2

We first focus on the KF approach. It is easy to see from (4.6) and (4.7) that $\tilde{M}(n)$ and $M_{KF}(n)$ are diagonal matrices for sufficiently large $n$ when single-antenna beamforming is used. Let $\{X_0, X_1, \ldots, X_{N-1}\}$ denote the sorted (ascendingly) diagonal entries of $M_{KF}(n)$, and $\{\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{N-1}\}$ denote the sorted (ascendingly) diagonal entries of $\tilde{M}(n)$, for large $n$. Since the channel entries are i.i.d. ($C_h = \sigma_h^2 I$), it is easy to see that the values of $\{X_0, X_1, \ldots, X_{N-1}\}$ (and $\{\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{N-1}\}$) are the same for any sufficiently large $n$ (i.e., $n \to \infty$) because the KF estimator will be based on present and infinite past observations - only the location of $X_i$ (and $\tilde{X}_i$) in the diagonal of $M_{KF}(n)$ (resp. $\tilde{M}(n)$) differs for different $n$.

From (4.6), we have the relation $\tilde{X}_k = \alpha X_k + (1-\alpha)\sigma_h^2$, for $k = 0, \ldots, N-1$. Assume that antenna $i$ is used to send $s(n)$ at time $n$ (i.e., the $i$-th entry of $w(n)$ equals 1). Prior to time $n$, antenna $i$ was last accessed at time $n-N$ with the single-antenna beamforming, and thus the $i$-th diagonal entry of $\tilde{M}(n)$ is the largest entry $\tilde{X}_{N-1}$. From (4.7), only the $i$-th diagonal entry of $\tilde{M}(n)$ is affected by the recursion in (4.7), yielding the smallest diagonal entry $X_0$ of $M_{KF}(n)$, whereas the rest of the diagonal entries of $\tilde{M}(n)$ are duplicated in $M_{KF}(n)$. These relations can be expressed as

$$X_k = \alpha X_{k-1} + (1-\alpha)\sigma_h^2 = \alpha^k X_0 + (1-\alpha)\sigma_h^2 \sum_{i=0}^{k-1} \alpha^i$$  \hspace{1cm} (A.9)

for $k = 1, \ldots, N-1$, whereas from (4.7),

$$X_0 = \tilde{X}_{N-1} - \frac{\tilde{X}_{N-1}^2}{X_{N-1} + \sigma_v^2} = \frac{\tilde{X}_{N-1}\sigma_v^2}{X_{N-1} + \sigma_v^2}$$  \hspace{1cm} (A.10)

From (A.9)

$$\tilde{X}_{N-1} = \alpha X_{N-1} + (1-\alpha)\sigma_h^2 = \alpha^N X_0 + (1-\alpha^N)\sigma_h^2$$  \hspace{1cm} (A.11)

Substituting with $\tilde{X}_{N-1}$ from (A.11) in (A.10), we obtain the quadratic equation in $X_0$:

$$X_0^2 + (\sigma_v^2 + \sigma_h^2)\left(\frac{1-\alpha^N}{\alpha^N}\right) X_0 - \sigma_v^2\sigma_h^2\left(\frac{1-\alpha^N}{\alpha^N}\right) = 0$$  \hspace{1cm} (A.12)
The only positive solution for (A.12) is \( X_0 = -c_1 + \sqrt{c_1^2 + c_2} \), where \( c_1 \) and \( c_2 \) are defined in Proposition 2. Finally, using (A.9),

\[
\varepsilon_{\text{KF}} = \sum_{k=0}^{N-1} X_k = X_0 \sum_{k=0}^{N-1} \alpha^k + N\sigma_h^2 - \sigma_h^2 \sum_{k=0}^{N-1} \alpha^k
\]

\[
= N\sigma_h^2 + (X_0 - \sigma_h^2) \frac{1 - \alpha^N}{1 - \alpha}
\]

which proves (4.13).

The proof of (4.26) for the SOI-KF approach follows along the same lines. Note that the \( 2N \times 2N \) matrix \( M_{\text{SOI-KF}}(n) \) is diagonal for sufficiently large \( n \), where the upper-left \( N \times N \) sub-matrix (which corresponds to the real part) is identical to the lower-right \( N \times N \) sub-matrix (which corresponds to the imaginary part). Focusing only on the upper-left sub-matrix, and defining \( \{Y_0, Y_1, \ldots, Y_{N-1}\} \) and \( \{\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_{N-1}\} \) as the sorted diagonal entries of the upper-left sub-matrix of \( M_{\text{SOI-KF}}(n) \) and \( \hat{M}(n) \), respectively, an expression for \( Y_0 \) in this case can be obtained from (4.15) as

\[
Y_0 = \hat{Y}_{N-1} - \frac{2\hat{Y}_{N-1}^2}{\hat{Y}_{N-1} + \sigma_v^2/2}
\]

(A.14)

Substituting with \( \hat{Y}_{N-1} = \alpha^N Y_0 + (1 - \alpha^N)\sigma_h^2/2 \) in (A.14), we obtain a quadratic equation in \( Y_0 \), which is solved to obtain the only positive solution \( Y_0 = \frac{-c_4 + \sqrt{c_4^2 + c_5^2}}{2c_3} \), where \( c_3, c_4 \) and \( c_5 \) are defined in Proposition 2. Then,

\[
\varepsilon_{\text{SOI-KF}} = 2 \sum_{k=0}^{N-1} Y_k = N\sigma_h^2 + (2Y_0 - \sigma_h^2) \frac{1 - \alpha^N}{1 - \alpha}
\]

(A.15)

which proves (4.26).
Appendix B

Acronyms

This appendix gives a table of acronyms and their meanings in alphabetical order.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>analytic center</td>
</tr>
<tr>
<td>ADC</td>
<td>analog-to-digital converter</td>
</tr>
<tr>
<td>AGC</td>
<td>automatic gain control</td>
</tr>
<tr>
<td>AMAP</td>
<td>approximate maximum a posteriori</td>
</tr>
<tr>
<td>AR</td>
<td>autoregressive</td>
</tr>
<tr>
<td>ARMA</td>
<td>autoregressive-moving-average</td>
</tr>
<tr>
<td>AWGN</td>
<td>additive white Gaussian noise</td>
</tr>
<tr>
<td>CC</td>
<td>Chebyshev center</td>
</tr>
<tr>
<td>CDGS</td>
<td>coordinate descent grid search</td>
</tr>
<tr>
<td>CRB</td>
<td>Cramer-Rao bound</td>
</tr>
<tr>
<td>CSI</td>
<td>channel state information</td>
</tr>
<tr>
<td>DFT</td>
<td>discrete Fourier transform</td>
</tr>
<tr>
<td>DTFT</td>
<td>discrete-time Fourier transform</td>
</tr>
<tr>
<td>ECM</td>
<td>error covariance matrix</td>
</tr>
<tr>
<td>FC</td>
<td>fusion center</td>
</tr>
<tr>
<td>FDD</td>
<td>frequency-division duplex</td>
</tr>
<tr>
<td>FIM</td>
<td>Fisher information matrix</td>
</tr>
<tr>
<td>FIR</td>
<td>finite impulse response</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>GLA</td>
<td>generalized Lloyd algorithm</td>
</tr>
<tr>
<td>I-DTFT</td>
<td>inverse discrete-time Fourier transform</td>
</tr>
<tr>
<td>i.i.d</td>
<td>independent and identically distributed</td>
</tr>
<tr>
<td>KF</td>
<td>Kalman filtering</td>
</tr>
<tr>
<td>LMI</td>
<td>linear matrix inequality</td>
</tr>
<tr>
<td>LP</td>
<td>linear programming</td>
</tr>
<tr>
<td>LTE</td>
<td>long term evolution</td>
</tr>
<tr>
<td>MA</td>
<td>moving-average</td>
</tr>
<tr>
<td>MAP</td>
<td>maximum a posteriori</td>
</tr>
<tr>
<td>MIMO</td>
<td>multiple-input multiple-output</td>
</tr>
<tr>
<td>ML</td>
<td>maximum likelihood</td>
</tr>
<tr>
<td>MMSE</td>
<td>minimum mean-square error</td>
</tr>
<tr>
<td>MSE</td>
<td>mean square error</td>
</tr>
<tr>
<td>MUSIC</td>
<td>multiple signal classification</td>
</tr>
<tr>
<td>NMSE</td>
<td>normalized mean square error</td>
</tr>
<tr>
<td>pmf</td>
<td>probability mass function</td>
</tr>
<tr>
<td>PN</td>
<td>pseudo-noise</td>
</tr>
<tr>
<td>PS</td>
<td>power spectrum</td>
</tr>
<tr>
<td>QCQP</td>
<td>quadratically constrained quadratic problem</td>
</tr>
<tr>
<td>RLS</td>
<td>recursive least-squares</td>
</tr>
<tr>
<td>RMSE</td>
<td>root-mean-square error</td>
</tr>
<tr>
<td>SDP</td>
<td>semidefinite program</td>
</tr>
<tr>
<td>SDR</td>
<td>semidefinite relaxation</td>
</tr>
<tr>
<td>SER</td>
<td>signal-to-error ratio</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SOCP</td>
<td>second-order cone programming</td>
</tr>
<tr>
<td>SOI</td>
<td>sign of innovations</td>
</tr>
<tr>
<td>TDD</td>
<td>time-division duplex</td>
</tr>
<tr>
<td>VQ</td>
<td>vector quantization</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>WF</td>
<td>Wiener filtering</td>
</tr>
<tr>
<td>WSN</td>
<td>wireless sensor network</td>
</tr>
<tr>
<td>WSS</td>
<td>wide-sense stationary</td>
</tr>
</tbody>
</table>

Table B.1: List of acronyms used in the thesis.