Parametric Frugal Sensing of Power Spectra for Moving Average Models

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

Aritra Konar

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

Professor Nicholas D. Sidiropoulos, Advisor

MAY, 2015

Acknowledgements

I would like to express my sincere gratitude to several people who helped me in producing this thesis. First and foremost, I wish to thank my advisor, Professor Nikos Sidiropoulos for his invaluable advice and deep patience, without which this work would not have been possible. His immense knowledge, fatherly guidance and dedication to research has served as an immense inspiration to me and has motivated me to strive to excel. I feel extremely fortunate and honored to have had the chance to work alongside him.

I am also thankful to Professor Georgios Giannakis and Professor Shuzhong Zhang for agreeing to serve on my defense committee and am grateful for their support and valuable feedback.

I would also like to thank my (former and current) labmates Omar Mehanna, Balasubramanian Gopalakrishnan, Kejun Huang, Xiao Fu, John Tranter and Artem Mosesov for our stimulating discussions which produced valuable ideas and insights that significantly improved the work of this thesis, and in the process also helped me to become more independent in my research.

A word of thanks also goes out to my friends Akshay Soni, Tanvi Sharma, Gaurav Chopra and Swayambhoo Jain for helping me get adjusted to graduate life in the beginning and for their support.

Finally, I would like to extend my deepest gratitude to my family: my parents, sister and brother-in-law, and my nieces, whose unconditional love and unwavering support provides me the strength to keep me going on my journey.

Dedication

To my nieces Parthivee and Sudiksha.

Abstract

Wideband spectrum sensing is a fundamental component of cognitive radio and other applications. A novel *frugal sensing* scheme was recently proposed as a means of crowdsourcing the task of spectrum sensing. Using a network of scattered low-end sensors transmitting randomly filtered power measurement bits to a fusion center, a non-parametric approach to spectral estimation was adopted to estimate the ambient power spectrum. Here, a parametric spectral estimation approach is considered within the context of frugal sensing. Assuming a Moving-Average (MA) representation for the signal of interest, the problem of estimating admissible MA parameters, and thus the MA power spectrum, from single bit quantized data is formulated. This turns out being a non-convex quadratically constrained quadratic program (QCQP), which is NP–Hard in general. Approximate solutions can be obtained via semi-definite relaxation (SDR) followed by randomization; but this rarely produces a feasible solution for this particular kind of QCQP. A new Sequential Parametric Convex Approximation (SPCA) method is proposed for this purpose, which can be initialized from an infeasible starting point, and yet still produce a feasible point for the QCQP, when one exists, with high probability. Simulations not only reveal the superior performance of the parametric techniques over the globally optimum solutions obtained from the non-parametric formulation, but also the better performance of the SPCA algorithm over the SDR technique.

Contents

List of Figures

Chapter 1

Introduction

Wideband spectrum sensing is one of the core components of cognitive radio, since it forms the basis for adaptive spectrum sharing [1]. In cognitive radio, discovering transmission opportunities requires the unlicensed secondary users to scan a wide band of frequencies. Spectrum sensing aims to detect spectral occupancy in a particular frequency band, preferably without scanning the entire band. In order to overcome issues related to reliability, fading, and the hidden terminal problem, collaborative spectrum sensing schemes are essential. These involve a network of scattered sensors taking judicious measurements across space, time and frequency. Apart from cognitive radio, collaborative spectrum sensing schemes are also employed in other diverse fields such as industrial monitoring, smart agriculture, weather forecasting, military surveillance, disaster response, and health monitoring applications [2].

Wideband spectrum sensing at very high sampling rates requires expensive analog-to-digital converters (ADCs) which cannot easily fit in portable devices, and have high power consumption. In cognitive radio settings, however, it is commonly assumed that the frequency band is under-utilized, i.e., most of the band is idle during most of the time, which in turn implies frequency domain sparsity. Exploiting this prior information, compressive spectrum sensing techniques have been developed, which provide accurate spectrum estimates at sub-Nyquist rates, without using frequency sweeping [3], [4]. Cooperative protocols for distributed compressive spectrum sensing were introduced in [5], [6], which involve estimating the spectrum locally, followed by achieving consensus on globally fused sensing outcomes. The main drawback is that these methods are

computationally intensive in terms of the calculations performed at each sensor node and also require significant message-passing between sensors.

Although the aim of most work on spectrum sensing (e.g., [1]– [6]) has been on the reconstruction of the signal's *Fourier spectrum*, in cognitive radio and many other applications, it is sufficient to only consider reconstructing the *power spectrum* [7]. Recent work on Power spectrum (PS) sensing, (e.g. [7–11]) establishes that spectrum sparsity is not essential for relaxing sampling rate requirements. Instead, by exploiting the fact that the power measurements are linear in the autocorrelation function, the utilization of a low-order correlation model enables estimation of a finite number of autocorrelation lags from a system of over-determined linear equations, resulting in a reduction in sampling rate. If spectral information, in the form of spectral masks and carrier frequencies are available beforehand, it is possible to obtain a further reduction in the sampling rate [12].

The PS estimation methods described in [7]– [12] assume analog amplitude samples, which is reasonable if ADCs that operate at a high sampling rate are employed. However, this is undesirable in distributed spectrum sensing scenarios, since transmitting streams of finely quantized bits to a fusion center creates a large communications overhead and has an adverse effect on battery lifetime. In [13], a network sensing scenario was introduced, where the case of each sensor transmitting a single, randomly filtered power measurement bit to the FC was considered. A non-parametric model for the spectrum was assumed, and the problem of estimating the power spectrum from the received bits was formulated as a Linear Programming (LP) problem, generalizing classical non-parametric PS estimation to the case where the data is in the form of inequalities, instead of equalities.

If we know *a priori* that the signal admits a representation in terms of a parametric model of a certain order, then it is well-known [14] that classical parametric PS estimation methods are more accurate as compared to the non-parametric methods, provided the modeling assumptions are valid. A parametric model provides a more parsimonious representation of the PS, since it requires estimation of fewer parameters as compared to a non-parametric model. This is the main motivation for our present work, where we assume that the signal of interest admits a Moving Average (MA) representation.

From an application point of view, an MA parametrization is well-suited for sensing digital communication signals, which are pulse-shaped using finite impulse response (FIR) filters, and transmitted over wireless channels which are also modeled as FIR filters[1] .

The problem of MA parameter estimation has been extensively studied in signal processing, statistics, econometrics, and other areas. One approach is to estimate the autocorrelation sequence first, from which the MA parameters can then be determined by solving a system of non-linear equations by the method of moments (MOM) [15]. However, it is well known that MOM does not yield good estimates in general. Another technique utilizes the autocorrelation estimates to obtain the MA parameters by employing a spectral factorization step. This method fails in the event that the estimated autocorrelation sequence is not positive semidefinite, in which case the estimated spectrum will take negative values at certain frequencies, and will not admit factorization. This drawback was overcome by the MA autocorrelation estimation method proposed in [16], which can be used to determine the optimal MA autocorrelation approximation of an invalid autocorrelation sequence, in the least square sense. An approximate maximum-likelihood approach, based on the likelihood function of the data, is described in [17, p. 281], which results in a non-linear optimization problem. An iterative algorithm for the MA parameter estimates is presented, but the method can only converge to a local minimum of the likelihood function, and convergence is not guaranteed. Other methods, which involve approximating the MA model by an autoregressive (AR) model of sufficiently large order (e.g., Durbin's method [18] and the inverse covariance method [19]), are known to produce highly biased estimates for MA models with zeros close to the unit circle. It is to be noted that none of the aforementioned methods work directly with the underlying MA representation of the signal in their formulations, which is in contrast to our formulation that explicitly uses the MA structure from the outset. We obtain an admissible MA parameter vector and take the magnitude square of the discrete time Fourier transform (DTFT) to obtain the spectral estimate.

Considering the same network sensing scenario as in [13], we develop parametric

[1] When the channel is random and different from sensor to sensor, but with the same second-order statistics, and each sensor averages out the power measurements over multiple fading realizations, it has been shown in [13] that all sensors will report consistent measurements, as if all channels were equal to a spectral factor.

methods for PS estimation of an MA model from 1 bit quantized power sensing data. Parametrizing the MA autocorrelation sequence in terms of the MA parameters, it is shown that the problem of estimating admissible MA parameters from 1 bit data can be expressed as a non-convex Quadratically Constrained Quadratic Program (QCQP). This is NP–Hard [20] in general, and hence cannot be solved to global optimality in polynomial time. Instead, we resort to two approximation algorithms in order to obtain polynomial time sub-optimal solutions.

First, a semi-definite relaxation (SDR) approach is considered. SDR is conceptually simple and widely used together with *randomization* procedures to obtain high quality approximate solutions for a large ensemble of NP–Hard problems that arise in engineering practice (e.g., [21–24]). However, it is known that SDR followed by randomization may fail to obtain a feasible point when the constraints are stringent [25]. The second method is a sequential parametric convex approximation approach (SPCA) that requires solving a sequence of second-order cone programs (SOCPs), initialized from a starting point that is designed to be feasible for only one subset of constraints of the non-convex QCQP. Although several SPCA algorithms exist in the current literature ([26–29]), they require initialization from a starting point that lies in the feasible set of the non-convex problem. However, finding a feasible point is hard in general for non-convex QCQPs. This is the main motivation for pursuing the development of an SPCA algorithm whose initial starting point is infeasible. Using a linear restriction to approximate the non-convex constraints, adding slack variables to the convex ones to ensure feasibility, and imposing an ℓ_1-norm penalty on the slacks to minimize constraint violations, the modified problem is formulated as a SOCP. Upon iteratively solving the SOCP, using the solution of the current iteration as the point about which the linear restriction is computed for the next iteration, a feasible point for the non-convex QCQP can be obtained in a few iterations, for a large percentage of problem instances. We also consider a two-step approach that combines the non-parametric LP formulation in [13] with a second step that imposes MA structure to the autocorrelation estimate provided by [13]. A comprehensive comparison of these methods is then carried out in different scenarios and various aspects of their performance are evaluated.

Relative to [30], this journal version adds the more effective SPCA/SOCP approach, proofs and derivations, and comprehensive simulations providing interesting insights

and new findings. The particular SPCA approach that we advocate here for MA power spectrum sensing from 1-bit data is closely related to our recent work on feasible point pursuit for a general class of non-convex QCQPs [31] – in fact the MA power spectrum sensing application paved the way to [31]. There are two important differences between the specific method proposed here and the generic one in [31]. The first is that we tighten the linear restriction used to lower bound the non-convex constraints, which results in a better approximation. The second is that we use a more intelligent initialization strategy that ensures satisfaction of all non-convex constraints. These two custom modifications make a difference in terms of the quality of the estimated power spectra.

The rest of the thesis is organized as follows. We begin with some preliminaries in Chapter 2, followed by a description of the signal model in Chapter 3. The formulations of the proposed parametric PS estimation schemes are presented in Chapter 4. Comprehensive simulations are provided in Chapter 5, along with a discussion of the results, insights, and how these guide the choice of design parameters. Conclusions are drawn in Chapter 6. Technical derivations are relegated to the Appendices.

Throughout the rest of the article, we adopt the following notations. The superscript * denotes conjugation, whereas inline $*$ denotes convolution, as is customary. The superscript ^{H} is used to denote the Hermitian (conjugate) transpose of a vector/matrix, while ^{T} denotes plain transposition. Capital boldface is used for matrices, while vectors are denoted by small boldface. Scalar terms are represented in the normal face. The circularly symmetric complex Gaussian distribution is denoted by $\mathcal{CN}(.,.)$. $\mathbb{E}(.)$ and $\nabla(f(.))$ denote the expectation operator and the gradient of the function $f(.)$ respectively. The $n \times n$ Identity matrix is denoted by $\mathbf{I}_n$.

Chapter 2

Preliminaries

Consider a discrete-time wide-sense stationary (WSS) signal $x(n)$ and let $r(k) = \mathbb{E}[x(n)x^*(n-k)]$ denote its autocorrelation sequence, where $r(k) = r^*(-k), \forall k$. Assuming that $x(n)$ admits an MA representation, we can characterize it as being generated by passing complex, circularly symmetric, uncorrelated, zero mean White Gaussian Noise (WGN) of unit variance through a linear shift invariant FIR filter with an impulse response $\mathbf{h} = [h(0), h(1), ..., h(q)]^T$, where q is the order of the MA process. Hence, $x(n)$ can be expressed as

$$x(n) = \sum_{k=0}^{q} h(k)w(n-k) \tag{2.1}$$

where $w(n) \sim \mathcal{CN}(0,1)$. We can parametrize $r(k)$ in terms of the MA parameters $\mathbf{h}$ as

$$r(k) = \begin{cases} h(k) * h^*(-k) & : |k| \leq q \\ 0 & : |k| > q \end{cases} \tag{2.2}$$

$$= \begin{cases} \sum_{i=0}^{q-|k|} h^*(i)h(i+|k|) & : |k| \leq q \\ 0 & : |k| > q \end{cases} \tag{2.3}$$

$$= \begin{cases} \mathbf{h}^H \mathbf{\Theta}_k^{q+1} \mathbf{h} & : |k| \leq q \\ 0 & : |k| > q \end{cases} \tag{2.4}$$

where $\mathbf{\Theta}_k^{q+1}$ is the elementary $(q+1)\times(q+1)$ Toeplitz matrix with ones on the $(k+1)^{th}$ sub-diagonal and zeros everywhere else (note that $\mathbf{\Theta}_0 = \mathbf{I}_{q+1}$) . Given a finite length

data sequence $\mathbf{x} = [x(n), x(n-1), \ldots, x(n-K+1)]^T$ of length K, we define the $K \times K$ Toeplitz-Hermitian autocorrelation matrix $\mathbf{R}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ as

$$\mathbf{R}_x = \begin{bmatrix} r(0) & r(1) & \cdots & r(K-1) \\ r^*(1) & r(0) & \cdots & r(K-2) \\ \vdots & \vdots & \ddots & \vdots \\ r^*(K-1) & r^*(K-2) & \cdots & r(0) \end{bmatrix} \tag{2.5}$$

The PS of $x(n)$, according to the parametric model, is given by $S_x(e^{j\omega}) = |H(e^{j\omega})|^2$, where $H(e^{j\omega})$ is the discrete-time Fourier Transform (DTFT) of $\mathbf{h}$ and is given by $H(e^{j\omega}) = \sum_{n=0}^{q} h(n)e^{-j\omega n}$.

Chapter 3

System Model

A network sensing scenario is considered (as in [13]), where M scattered sensors measure the ambient signal power and report to a fusion center (FC). For simplicity of exposition, let us assume that all sensors sense a signal that is common up to a sensor-specific constant that models the effects of path loss, frequency-flat fading and shadowing. Frequency-selective fading that varies from sensor to sensor can be accommodated, provided sensors average out their measurements over many fading states, and the 'expected fading spectrum' is the same across sensors (not the fading realizations); see [13]. The received signal at each sensor $m \in \{1, \ldots, M\}$ is downconverted to baseband, and then automatic gain control (AGC) is used to scale each to a common reference signal $x(t)$. Since the power spectrum is invariant to sensing timing offsets and phase shifts, it is not necessary to compensate for these effects. After the AGC stage, the signal is sampled using a Nyquist rate ADC to yield the WSS sequence $x(n)$. As shown in [13], the Nyquist rate sampling requirement can be relaxed by using an equivalent analog processing and integration chain. Then, $x(n)$ is passed through a wideband FIR filter with complex binary pseudo-noise (PN) impulse response $g_m(n)$ of length K, where each filter coefficient is drawn uniformly from the set of 4 possible QPSK symbols, i.e.,

$$g_m(n) = \begin{cases} \sim \frac{1}{\sqrt{2K}} \mathcal{U}(\{1+j, -1+j, 1-j, -1-j\}) & : 0 \leq n \leq K-1 \\ 0 & : otherwise \end{cases} \tag{3.1}$$

where $\mathcal{U}(S)$ denotes the uniform probability mass function over the finite set S. The filter output is given by $z_m(n) = \sum_{k=0}^{K-1} g_m^*(k) x(n-k)$, where $\mathbf{g}_m = [g_m(0), \ldots, g_m(K-$

$1)]^T$ are the filter tap weights and $\mathbf{x} = [x(n), x(n-1), \ldots, x(n-K+1)]^T$ are the tap inputs. The use of PN random filters promotes diversity, simplifies the convolution operation, there being no need for multiplications, and eliminates the need for coordination between sensors. Denote the average power of the WSS signal $z_m(n)$ by ρ_m, i.e., $\rho_m = \mathbb{E}[|z_m(n)|^2]$. Each sensor obtains soft estimates of ρ_m by averaging over N samples

$$\hat{\rho}_m = \frac{1}{N} \sum_{n=0}^{N-1} |z_m(n)|^2 \tag{3.2}$$

Finally, each sensor compares its estimate $\hat{\rho}_m$ to a single threshold t. If $\hat{\rho}_m \geq t$, then a $'1'$ is transmitted, otherwise a $'0'$ is transmitted to the FC. We define the sets $\mathcal{M}_a := \{m : \hat{\rho}_m \geq t\}$ and $\mathcal{M}_b := \{m : \hat{\rho}_m < t\}$, with $M_a = |\mathcal{M}_a|$ and $M_b = |\mathcal{M}_b|$ such that $M_a + M_b = M$. Since $z_m(n) = \mathbf{g}_m^H \mathbf{x}$, it can be shown that $\rho_m = \mathbb{E}[|\mathbf{g}_m^H \mathbf{x}|^2] = \mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m$ where $\mathbf{R}_x$ is the $K \times K$ Toeplitz-Hermitian autocorrelation matrix of $\mathbf{x}$ given in (2.5). Thus, on receiving a $'1'$ (or a $'0'$) from a sensor, the FC learns that $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m \geq t$ (or $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m < t$), assuming sufficient averaging to ensure that sample averages converge to ensemble averages. As an alternative to using PN random filters for diversity, one can also consider exploiting the diversity due to the random fading channels. One may naturally wonder why we employ PN filters instead of exploiting the inherent frequency selectivity that is different from sensor to sensor? The answer is two-fold.

- First, we do not know the fading channel realization for each sensor, and the task of estimating it is much more challenging than power spectrum estimation *per se*. Since we do not know the random fading channel realization for each sensor, we cannot plug it in the linear inequality constraint associated with the corresponding binary measurement. To circumvent this issue, we average out these random fading effects, and rely on pseudo-random filters seeded using the serial number of each sensor (which is known to the FC) to collect diversity.

- Second, since our goal is to estimate the power spectrum instead of the Fourier spectrum, averaging is in fact required to get rid of short-term effects such as fading.

The job of the FC is to estimate the ambient PS of the signal $x(n)$ from these inequalities, which are linear in the autocorrelation sequence $r(k)$. In the next section, it is shown

that exploiting the underlying MA parametrization of $x(n)$, the inequalities can be explicitly expressed in terms of MA model parameters $\mathbf{h}$.

Chapter 4

Problem Formulation

Define p as our postulated model order of the MA process i.e., we believe $x(n)$ is generated by passing complex, circularly symmetric, uncorrelated, zero mean WGN of unit variance through the linear shift invariant FIR filter whose frequency response is given by $H(e^{j\omega}) = \sum_{n=0}^{p} h(n)e^{-j\omega n}$, with impulse response $\mathbf{h} = [h(0), \ldots, h(p)]^T$. Note that it is not necessary to have $p = q$, where q is the true model order. Instead, p may represent an upper bound on the true model order. $\mathbf{R}_x$, as defined in (2.5), can also be expressed as

$$\mathbf{R}_x = r(0)\mathbf{\Theta}_0^K + \sum_{k=1}^{\min(K-1,p)} (r(k)\mathbf{\Theta}_k^K + r^*(k)\mathbf{\Theta}_{-k}^K) \tag{4.1}$$

where $\mathbf{\Theta}_k^K$ is the $K \times K$ elementary Toeplitz matrix with ones on the k^{th} diagonal and zeros elsewhere, and $\mathbf{r_x} = [r^*(K-1), \ldots, r^*(1), r(0), r(1), \ldots, r(K-1)]^T$ is the autocorrelation sequence. The upper limit on the summation term stems from the fact that $r(k) = 0, \forall\ |k| > p$, and thus, depending on whether we set K to be larger than p or not, we get the corresponding number of terms in the sum. Using (4.1), each

$\rho_m = \mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m$ can be expressed as

$$\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m = \mathbf{g}_m^H \Bigg(r(0)\boldsymbol{\Theta}_0^K + \sum_{k=1}^{\min(K-1,p)} (r(k)\boldsymbol{\Theta}_k^K + r^*(k)\boldsymbol{\Theta}_{-k}^K) \Bigg) \mathbf{g}_m \tag{4.2}$$

$$= \underbrace{\mathbf{g}_m^H \boldsymbol{\Theta}_0^K \mathbf{g}_m}_{c_{m,0}} r(0) + \sum_{k=1}^{min(K-1,p)} \Bigg(\underbrace{\mathbf{g}_m^H \boldsymbol{\Theta}_k^K \mathbf{g}_m}_{c_{m,k}} r(k) + \underbrace{\mathbf{g}_m^H \boldsymbol{\Theta}_{-k}^K \mathbf{g}_m}_{c_{m,-k}} r^*(k) \Bigg) \tag{4.3}$$

$$= c_{m,0} r(0) + \sum_{k=1}^{\min(K-1,p)} (c_{m,k} r(k) + c_{m,-k} r^*(k)) \tag{4.4}$$

where in (4.3), $c_{m,k}$ represents the k^{th} lag of the deterministic autocorrelation sequence of the m^{th} broadband filter with impulse response $\mathbf{g}_m$. Using the expression for $r(k)$ given by (2.4), we obtain

$$\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m = \mathbf{h}^H \Bigg(c_{m,0} \boldsymbol{\Theta}_0^{p+1} + \sum_{k=1}^{min(K-1,p)} (c_{m,k} \boldsymbol{\Theta}_k^{p+1} + c_{m,-k} \boldsymbol{\Theta}_{-k}^{p+1}) \Bigg) \mathbf{h} \tag{4.5}$$

$$= \mathbf{h}^H \mathbf{C}_m \mathbf{h} \tag{4.6}$$

where by construction, each $\mathbf{C}_m$ matrix is also Toeplitz, Hermitian and positive semi-definite. We refer the reader to Appendix A for a proof. Hence, the linear inequalities $\mathbf{g}_m^H \mathbf{R}_x \mathbf{g}_m \gtrless t$ can now be expressed as the quadratic inequalities in the MA parameters $\mathbf{h}^H \mathbf{C}_m \mathbf{h} \gtrless t$.

In order to estimate the PS from these inequalities, an admissible set of MA parameters $\mathbf{h}$ are estimated first, and then the spectrum is computed as $S(e^{j\omega}) = |H(e^{j\omega})|^2$ where $H(e^{j\omega})$ is the DTFT of $\mathbf{h}$. The total signal power $r(0) = \mathbb{E}[|x(n)|^2] = \mathbf{h}^H \mathbf{h}$ is chosen as a cost function to minimize, since in the cognitive radio setting, it is implicitly assumed that most of the spectrum is idle at most times. Overall, we obtain the

following formulation.

$$\begin{aligned} \underset{\mathbf{h}\in\mathbb{C}^{p+1}}{\text{minimize}} \quad & \mathbf{h}^H\mathbf{h} && \text{(4.7a)} \\ \text{subject to} \quad & \mathbf{h}^H\mathbf{C}_m\mathbf{h} \geq t,\ m \in \mathcal{M}_a && \text{(4.7b)} \\ & \mathbf{h}^H\mathbf{C}_m\mathbf{h} < t,\ m \in \mathcal{M}_b && \text{(4.7c)} \end{aligned}$$

Remark 1 *Note that identifiability of* $\mathbf{h}$ *cannot be guaranteed, since the phase of its Fourier transform cannot be estimated even from exact analog power measurements - let alone quantized ones. Our ultimate goal, however, is to estimate the power spectrum (i.e., the magnitude squared of the DTFT of h) which is unaffected by this spectral factorization ambiguity. Furthermore, our choice of cost function does not discriminate amongst the spectral factors (by Parseval's Theorem they all have the same cost), so the lack of identifiability of* $\mathbf{h}$ *does not hurt our ultimate goal of power spectrum estimation.*

Problem (4.7) is a QCQP, where the ellipsoid exterior constraints (4.7b) are non-convex, and is NP–Hard in general. We now present two formulations which produce approximate solutions for this problem.

4.1 SDR Formulation

The non-convex QCQP (4.7) can be recast as

$$\begin{aligned} \underset{\mathbf{H}\in\mathbb{C}^{(p+1)\times(p+1)}}{\text{minimize}} \quad & \text{Trace}(\mathbf{H}) && \text{(4.8a)} \\ \text{subject to} \quad & \text{Trace}(\mathbf{C}_m\mathbf{H}) \geq t,\ m \in \mathcal{M}_a && \text{(4.8b)} \\ & \text{Trace}(\mathbf{C}_m\mathbf{H}) < t,\ m \in \mathcal{M}_b && \text{(4.8c)} \\ & \mathbf{H} \succeq \mathbf{0}, && \text{(4.8d)} \\ & \text{rank}(\mathbf{H}) = 1 && \text{(4.8e)} \end{aligned}$$

where we have defined $\mathbf{h}\mathbf{h}^H = \mathbf{H}$ and utilized the fact that $\mathbf{h}^H\mathbf{C}_m\mathbf{h} = \text{Trace}(\mathbf{C}_m\mathbf{H})$ and similarly, $\mathbf{h}^H\mathbf{h} = \text{Trace}(\mathbf{h}\mathbf{h}^H)$. $\mathbf{H}$ is a $(p+1)\times(p+1)$ complex, rank one, symmetric positive semidefinite (PSD) matrix. Our reformulation results in an equivalent problem,

with linear objective and linear trace constraints, while the set of symmetric PSD matrices is convex. The non-convexity has been isolated in the form of the rank constraint on $\mathbf{H}$. The technique of semidefinite relaxation (e.g., see [32]) now entails dropping the rank constraints to obtain the following relaxed problem.

$$\begin{aligned}
\underset{\mathbf{H}\in\mathbb{C}^{(p+1)\times(p+1)}}{\text{minimize}} \quad & \text{Trace}(\mathbf{H}) && \text{(4.9a)}\\
\text{subject to} \quad & \text{Trace}(\mathbf{C}_m\mathbf{H}) \geq t,\ m \in \mathcal{M}_a && \text{(4.9b)}\\
& \text{Trace}(\mathbf{C}_m\mathbf{H}) < t,\ m \in \mathcal{M}_b && \text{(4.9c)}\\
& \mathbf{H} \succeq \mathbf{0} && \text{(4.9d)}
\end{aligned}$$

which is a Semidefinite Programming (SDP) problem. Using modern interior-point algorithms, SDP problems can be solved efficiently to global optimality at a complexity cost that is at most $O(M + (p+1)^2)^{3.5}$ [33] and is usually much less. The cone programming solver SeDuMi [34] can be used to solve (4.9) efficiently. However, solving the relaxed problem does not solve the actual NP–Hard problem in general. Even so, the process of rank relaxation to obtain the SDP problem can be justified since doing so yields the Lagrange bi-dual of the original non-convex QCQP, which, in a certain sense, is the closest convex problem to the original NP–Hard problem [35]. Using rank relaxation, various post-processing procedures have been developed for obtaining approximate solutions for the original problem from the optimal solution of the relaxed problem. This may be done via *randomization* techniques, which are computationally inexpensive compared to solving the relaxed SDP problem.

Randomization Algorithm. In general, solving the relaxed SDP problem does not result in a rank 1 solution. If it does, then the principal component of the solution $\mathbf{H}_{\text{opt}}$, will be the optimal solution to the original problem. Otherwise, we use the following *randomization* approach to convert the globally optimal solution $\mathbf{H}_{\text{opt}}$ for problem (4.9) into an approximate solution that is feasible for the original problem (4.7).

1. Consider the case of determining the rank 1 approximation of $\mathbf{H}_{\text{opt}}$. Let $r = \text{rank}(\mathbf{H}_{\text{opt}})$ and define $\mathbf{H}_{\text{opt}} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathbf{H}} = \sum_{i=1}^{r} \lambda_i \mathbf{q}_i \mathbf{q}_i^H$ as the eigen-decomposition of $\mathbf{H}_{\text{opt}}$, where $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_r \geq 0$ are the eigen-values and $\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_r \in \mathbb{C}^{p+1}$ are the corresponding eigen-vectors. Since the best rank 1 approximation of $\mathbf{H}_{\text{opt}}$

(in the Frobenius norm sense) is given by $\lambda_1 \mathbf{q}_1 \mathbf{q}_1^H$, we can define $\mathbf{h}_{\text{pc}} = \sqrt{\lambda_1}\mathbf{q}_1$ as a possible candidate solution. Unless $r = 1$, $\mathbf{h}_{\text{pc}}$ will not be feasible for the problem (4.7). In the event that $\mathbf{H}_{\text{opt}}$ is approximately rank 1, we propose scaling $\mathbf{h}_{\text{pc}}$ to satisfy both set of constraints of (4.7). We first scale $\mathbf{h}_{\text{pc}}$ up a factor $\alpha > 1$ until it satisfies all the constraints in the set (4.7b), which can be determined as

$$\alpha = \sqrt{\frac{t}{\min\limits_{m \in \mathcal{M}_a} (\mathbf{h}_{\text{pc}})^H \mathbf{C}_m (\mathbf{h}_{\text{pc}})}} \tag{4.10}$$

Thus, we obtain a new candidate vector $\tilde{\mathbf{h}}_{\text{A}} = \alpha \mathbf{h}_{\text{pc}}$. If the candidate vector also satisfies all the constraints in the set (4.7c), then it is kept as a candidate solution. However, if it violates one or more constraints in (4.7c), it is discarded.

2. In general, $\mathbf{H}_{\text{opt}}$ will not even be approximately rank 1. Hence, we employ the technique of *Gaussian randomizations* to generate approximate solutions to the original problem (4.7). Using $\mathbf{H}_{\text{opt}}$, we generate a series of candidate Gaussian random vectors $\{\mathbf{h}_l^c\}_{l=0}^{L}$ from which the 'best' solution is chosen, where L denotes the number of randomization rounds. We calculate the eigen-decomposition of $\mathbf{H}_{\text{opt}}$ and in each randomization round, generate $\mathbf{h}_l^c = \mathbf{Q}\Lambda^{1/2}\mathbf{e}_l$, where $\mathbf{e}_l \sim \mathcal{CN}(0, \mathbf{I}_{p+1})$. A feasible vector $(\tilde{\mathbf{h}}_{\text{B}})_l$ that satisfies both sets of constraints can be found by scaling $\mathbf{h}_l^c$ to first satisfy the set of constraints (4.7b) and then checking to see if the constraints (4.7c) are satisfied too for the given choice of scaling. The scaled vector $(\tilde{\mathbf{h}}_{\text{B}})_l$ is given by $(\tilde{\mathbf{h}}_{\text{B}})_l = \beta_l \mathbf{h}_l^c$, where β_l is defined as

$$\beta_l = \sqrt{\frac{t}{\min\limits_{m \in \mathcal{M}_a} (\mathbf{h}_l^c)^H \mathbf{C}_m (\mathbf{h}_l^c)}} \tag{4.11}$$

If $(\tilde{\mathbf{h}}_{\text{B}})_l$ violates one or more constraints in the set (4.7c), then it is discarded and a new randomization round begins. Finally, amongst the feasible candidate vectors, the one which has the smallest objective value given by (4.7a), is chosen as the suboptimal solution $\tilde{\mathbf{h}}_{\text{B}}$.

3. In many cases, it is not even possible to obtain any suboptimal solution that is feasible for the original problem (4.7) by the previously outlined approaches. In such an instance, we propose to drop the convex constraint set (4.7c) and scale the

candidate vectors to satisfy the non-convex constraints (4.7b) only. For example, we scale the principal component $\mathbf{h}_{\text{pc}}$ of $\mathbf{H}_{\text{opt}}$ by α as defined in (4.10). This scaling ensures that $\tilde{\mathbf{h}}_{\text{C}} = \alpha \mathbf{h}_{\text{pc}}$ satisfies the constraints (4.7b), but we do not check to see if any of the convex constraints (4.7c) are violated. Similarly, for Gaussian randomization, after drawing a random vector $\mathbf{h}_c^l$, the constraint set (4.7c) is dropped and the vector scaled by β_l to be only feasible for the set (4.7b). Thereafter, the scaled vector that minimizes the objective is chosen as the solution $\tilde{\mathbf{h}}_{\text{D}}$. Our basic intuition for doing so is that in the frugal sensing setting the set (4.7b) is more informative, as it corresponds to activity detection events, and the choice of cost function in (4.7a) places an upper bound on each quadratic term in the set (4.7b), by virtue of the Rayleigh-Ritz criterion which upper bounds the Rayleigh quotient by its principal eigenvalue, thus controlling the violation of the dropped constraints on average. Hence, we expect this approach to also yield a good quality, albeit infeasible estimate.

Overall, after solving an instance of the problem (4.9) to obtain $\mathbf{H}_{\text{opt}}$, we proceed as follows. First, it is checked whether $\mathbf{H}_{\text{opt}}$ is rank 1 or not. If it is, then its principal component is the globally optimal solution to the problem (4.7). Otherwise, we check if a suboptimal solution $\tilde{\mathbf{h}}_{\text{A}}$ or $\tilde{\mathbf{h}}_{\text{B}}$ that satisfies both sets of constraints exists. If both $\tilde{\mathbf{h}}_{\text{A}}$ and $\tilde{\mathbf{h}}_{\text{B}}$ exist, then the one having the smaller objective value is chosen as the solution. In the event that both sets of constraints cannot be satisfied, then the set (4.7c) is dropped, and a pair of candidate solutions $\tilde{\mathbf{h}}_{\text{C}}$ and $\tilde{\mathbf{h}}_{\text{D}}$ is obtained. Again, the one having the smaller objective value is chosen as the solution.

4.2 Sequential Parametric Convex Approximation

We will refer to the non-convex QCQP (4.7) as (problem) W. In order to tackle the non-convexity, we employ the following affine approximation of the left hand side of the non-convex constraints (4.7b). Since $\mathbf{C}_m \succeq 0$, $f_m(\mathbf{h}) = \mathbf{h}^H \mathbf{C}_m \mathbf{h} \geq 0$, $\forall\ \mathbf{h}$, on expanding

$f_m(\mathbf{h}-\mathbf{p}) \geq 0$ and rearranging terms, we obtain the following inequality

$$f_m(\mathbf{h}) = \mathbf{h}^H \mathbf{C}_m \mathbf{h} \tag{4.12a}$$
$$\geq 2\mathcal{R}\mathbf{e}\{\mathbf{p}^H \mathbf{C}_m \mathbf{h}\} - \mathbf{p}^H \mathbf{C}_m \mathbf{p} \tag{4.12b}$$
$$= \mathcal{R}\mathbf{e}\{\mathbf{a}_m^H \mathbf{h}\} - b_m \tag{4.12c}$$
$$= F_m(\mathbf{h}, \mathbf{p}) \tag{4.12d}$$

where the right hand side of (4.12b) represents the first order power series expansion of $f_m(\mathbf{h})$ about the point $\mathbf{p}$ and in (4.12c), we have defined the constants $\mathbf{a}_m = 2\mathbf{C}_m\mathbf{p}$ and $b_m = \mathbf{p}^H \mathbf{C}_m \mathbf{p}$. Hence, for a chosen $\mathbf{p}$, the function $F_m(\mathbf{h}, \mathbf{p})$ is affine in $\mathbf{h}$, and is a lower bound to $f_m(\mathbf{h})$.

We now propose to solve the original problem W via a sequence of convex problems. At each step k, we replace the left hand side $f_m(\mathbf{h})$ of each non-convex constraint $m \in \mathcal{M}_a$ by the affine lower bound $F_m(\mathbf{h}, \mathbf{p}_k)$ at an appropriately chosen point $\mathbf{p}_k$. The convex constraints are left unchanged. Thus, at step k, we obtain a convex problem of the form

$$(U_k) \quad \underset{\mathbf{h}\in\mathbb{C}^{p+1}}{\text{minimize}} \quad \mathbf{h}^H \mathbf{h} \tag{4.13a}$$
$$\text{subject to} \quad F_m(\mathbf{h}, \mathbf{p}_k) \geq t, \; m \in \mathcal{M}_a \tag{4.13b}$$
$$f_m(\mathbf{h}) < t, \; m \in \mathcal{M}_b \tag{4.13c}$$

where the vector $\mathbf{p}_k$ is a fixed parameter depending on the solution of the problem instance U_{k-1}. Let $\mathbf{h}_k$ be the optimal solution to U_k (assuming it is feasible). Since for every $k \geq 0$ and $m \in \mathcal{M}_a$ we have $f_m(\mathbf{h}_k) \geq F_m(\mathbf{h}_k, \mathbf{p}_k) \geq t$, it follows that $\mathbf{h}_k$ is feasible for the original problem W. We update the parameter vector $\mathbf{p}$ at each iteration k by setting $\mathbf{p}_{k+1} = \mathbf{h}_k$. For this choice of update, $\mathbf{h}_k$ is also feasible for U_{k+1}. This is because $F_m(\mathbf{h}_k, \mathbf{p}_{k+1}) = F_m(\mathbf{h}_k, \mathbf{h}_k) = f_m(\mathbf{h}_k) \geq t$. The inequality holds since $\mathbf{h}_k$ is optimal for U_k. Since $\mathbf{h}_k$ is also feasible for U_{k+1}, it follows that $\|\mathbf{h}_k\|_2^2 \geq \|\mathbf{h}_{k+1}\|_2^2$. Hence, the method generates a monotonically non-increasing cost sequence. The overall procedure is as follows.

Algorithm 1

Input: An arbitrary point $\mathbf{p}_0$ that is feasible for the set $\mathcal{M}_a$ of the original problem (W). Let $k = 0, 1, \cdots$

Output: An approximate solution of (W)

1. Solve the problem (U_k) to obtain an approximate solution $\mathbf{h}_k$ of (W).
2. Set $\mathbf{p}_{k+1} = \mathbf{h}_k$, $k = k + 1$
3. Repeat until stopping criterion is satisfied

SOCP Formulation with Slack Variables. At each problem instance U_k, we are effectively restricting the feasible set of the original problem W, since using the affine lower bounds in place of the actual non-convex constraints forms a convex subset of $\mathcal{M}_a$. Owing to this restriction, and the fact that the starting point $\mathbf{p}_0$ is only feasible for the $\mathcal{M}_a$ part of the constraints, problem U_k may be infeasible from the very outset. In order to ensure feasibility at every step, we propose adding non-negative slack variables $\mathbf{s} = [s_1, \cdots, s_{M_b}]^T \succeq \mathbf{0}$ to each of the convex constraints in the set $\mathcal{M}_b$, and augment the objective with a penalty $\|\mathbf{s}\|_1 = \sum_{i=1}^{M_b} s_i$ in order to enforce sparsity in $\mathbf{s}$, such that the number of violated inequalities is approximately minimized. We also modify the first-order approximation of the non-convex constraints to ensure a better approximation of the constraint set $\mathcal{M}_a$ at each step. Towards this end, we translate the hyperplane $F_m(\mathbf{h}, \mathbf{p})$ until it becomes tangent to the hyper-ellipse defined by $\mathbf{h}^H \mathbf{C}_m \mathbf{h} = t$. This is done by scaling $\mathbf{p}$ by α_m until it touches the hyper-ellipse $\mathbf{h}^H \mathbf{C}_m \mathbf{h} = t$, i.e., $(\alpha_m \mathbf{p})^H \mathbf{C}_m (\alpha_m \mathbf{p}) = t$, from which we obtain $\alpha_m = \sqrt{\frac{t}{\mathbf{p}^H \mathbf{C}_m \mathbf{p}}}$. After linearizing $f_m(\mathbf{h})$ about $\alpha_m \mathbf{p}$, we obtain a hyperplane $F_m(\mathbf{h}, \alpha_m \mathbf{p}) = \mathcal{R}\mathbf{e}\{\tilde{\mathbf{a}}_m^H \mathbf{h}\} - \tilde{b}_m$ that is tangent to the hyper-ellipse $\mathbf{h}^H \mathbf{C}_m \mathbf{h} = t$ at the point $\alpha_m \mathbf{p}$, where $\tilde{\mathbf{a}}_m = 2\alpha_m \mathbf{C}_m \mathbf{p}$ and $\tilde{b}_m = \alpha_m^2 \mathbf{p}^H \mathbf{C}_m \mathbf{p}$. As a result of tightening the hyperplanes, note that the following inequalities hold.

$$f_m(\mathbf{h}) \geq F_m(\mathbf{h}, \alpha_m \mathbf{p}) \geq F_m(\mathbf{h}, \mathbf{p}), \forall\ m \in \mathcal{M}_a \tag{4.14}$$

This procedure is applied to each of the non-convex constraints. Overall, we can formulate the problem as follows

$$
\begin{aligned}
(V_k) \quad & \underset{\mathbf{h}\in\mathbb{C}^{p+1},\mathbf{s}\in\mathbb{R}^{M_b}}{\text{minimize}} && \mathbf{h}^H\mathbf{h} + \lambda\sum_{i=1}^{M_b} s_i && \text{(4.15a)}\\
& \text{subject to} && F_m(\mathbf{h}, \alpha_m\mathbf{p}_k) \geq t,\ m \in \mathcal{M}_a && \text{(4.15b)}\\
& && \mathbf{h}^H\mathbf{C}_m\mathbf{h} < t + s_m,\ m \in \mathcal{M}_b && \text{(4.15c)}\\
& && \mathbf{s} \succeq \mathbf{0} && \text{(4.15d)}
\end{aligned}
$$

where λ is a positive weighting factor. The parameter update equation remains unchanged. When solving the sequence of convex problems V_k, the corresponding cost sequence will be non-increasing, i.e., if $\mathbf{h}_k$, $\mathbf{s}_k$ are the optimal solution of V_k, then $\|\mathbf{h}_k\|_2^2 + \|\mathbf{s}_k\|_1$ is monotonically non-increasing in k. This is because $\mathbf{h}_k, \mathbf{s}_k$ are also feasible for V_{k+1}. To see this, note that for the linearized constraints, we have $F_m(\mathbf{h}_k, \alpha_m\mathbf{p}_{k+1}) = F_m(\mathbf{h}_k, \alpha_m\mathbf{h}_k) \geq F_m(\mathbf{h}_k, \mathbf{h}_k) = f_m(\mathbf{h}_k) \geq t$, while it is obvious that the convex constraints (4.15c) and (4.15d) are also satisfied by $\mathbf{h}_k, \mathbf{s}_k$. Hence, the optimal cost of V_{k+1} cannot be larger than that of V_k. Moreover, extensive simulations have revealed that not only is the overall objective monotonically non-increasing, but the two parts of the objective are also non-increasing, i.e. $\|\mathbf{h}_{k+1}\|_2^2 \leq \|\mathbf{h}_k\|_2^2$ and $\|\mathbf{s}_{k+1}\|_1 \leq \|\mathbf{s}_k\|_1$. Simulations also strongly suggest that the slacks $\mathbf{s}_k \longrightarrow \mathbf{0}$ (i.e., a feasible point for the non-convex QCQP W is found) in a finite number of iterations, in numerous spectrum sensing scenarios where W is feasible by construction. When the slacks become zero, it can be further shown using arguments similar to those in [28], that continued iterations produce a sequence of feasible vectors $\mathbf{h}_k$ with monotonically non-increasing cost, which eventually converges to a Karush-Kuhn-Tucker (KKT) point of W, under the assumption that the linear independence constraint qualification (LICQ) [36, p. 320] is satisfied.

We now reformulate this problem as an SOCP. Since the constraints (4.15b) and (4.15d) are linear in $\mathbf{h}$ and $\mathbf{s}$ respectively, no reformulation is needed. In order to represent the constraints (4.15c) in the second order cone (SOC) form, we proceed as follows

$$\mathbf{h}^H\mathbf{C}_m\mathbf{h} < t + s_m \tag{4.16a}$$

$$\Leftrightarrow \mathbf{h}^H\mathbf{C}_m\mathbf{h} - s_m < t \tag{4.16b}$$

$$\Leftrightarrow \begin{bmatrix}\mathbf{h}^H & s_m\end{bmatrix}\begin{bmatrix}\mathbf{L}_m\mathbf{L}_m^H & \mathbf{0} \\ \mathbf{0}^T & 0\end{bmatrix}\begin{bmatrix}\mathbf{h} \\ s_m\end{bmatrix} + \begin{bmatrix}\mathbf{0}^T & -1\end{bmatrix}\begin{bmatrix}\mathbf{h} \\ s_m\end{bmatrix} - t < 0 \tag{4.16c}$$

$$\Leftrightarrow \mathbf{x}^H\mathbf{A}_m^H\mathbf{A}_m\mathbf{x} + \mathbf{b}^H\mathbf{x} + c < \mathbf{0} \tag{4.16d}$$

$$\Leftrightarrow \left\|\begin{matrix}(1 + \mathbf{b}^H\mathbf{x} + c)/2 \\ \mathbf{A}_m\mathbf{x}\end{matrix}\right\|_2 < (1 - \mathbf{b}^H\mathbf{x} - c)/2 \tag{4.16e}$$

where in (4.16c), we have taken $\mathbf{C}_m = \mathbf{L}_m\mathbf{L}_m^H = (\mathbf{U}_m\mathbf{\Lambda}_m^{\frac{1}{2}}\mathbf{U}_m^H)^2$ as the square-root decomposition of $\mathbf{C}_m$, and $\mathbf{0}$ is the $(p+1)\times 1$ vector of zeros. In (4.16d), we have defined the new variables $\mathbf{x} = \begin{bmatrix}\mathbf{h}^H & s_m\end{bmatrix}^H$, $\mathbf{A}_m = \begin{bmatrix}\mathbf{L}_m^H & \mathbf{0}\end{bmatrix}, \mathbf{b} = \begin{bmatrix}\mathbf{0}^T & -1\end{bmatrix}^H$ and $c = -t$, and in (4.16e), we have converted the quadratic constraint of the previous step into a SOC constraint. Since the objective of V_k is quadratic in $\mathbf{h}$, we introduce an upper bound $\mathbf{h}^H\mathbf{h} \leq \beta$ and minimize the upper bound instead. Putting everything together, we have the overall formulation for the k^{th} problem instance

$$(\tilde{V_k}) \quad \underset{\mathbf{h}\in\mathbb{C}^{p+1},\mathbf{s}\in\mathbb{R}^{M_b},\beta\in\mathbb{R}}{\text{minimize}} \quad \beta + \lambda\sum_{i=1}^{B} s_i \tag{4.17a}$$

$$\text{subject to} \quad \tilde{\mathbf{a}}_m^H\mathbf{h} - \tilde{b}_m \geq t,\ m \in \mathcal{M}_a \tag{4.17b}$$

$$\left\|\begin{matrix}(1 + \mathbf{b}^H\mathbf{x} + c)/2 \\ \mathbf{A}_m\mathbf{x}\end{matrix}\right\|_2 < (1 - \mathbf{b}^H\mathbf{x} - c)/2,$$

$$\mathbf{x} = \begin{bmatrix}\mathbf{h}^H & s_m\end{bmatrix}^H,\ \mathbf{C}_m = \mathbf{L}_m\mathbf{L}_m^H,$$

$$\mathbf{A}_m = \begin{bmatrix}\mathbf{L}_m^H & \mathbf{0}\end{bmatrix},\ \mathbf{b} = \begin{bmatrix}\mathbf{0}^T & -1\end{bmatrix}^H,\ c = -t,\ m \in \mathcal{M}_b$$

$$\mathbf{s} \succeq \mathbf{0} \tag{4.17c}$$

$$\|\mathbf{h}\|_2 \leq \sqrt{\beta} \tag{4.17d}$$

which is an SOCP that can be solved using standard packages, such as SeDuMi. Feasibility of each problem instance is guaranteed because of the added slack variables. Hence, in each iteration, we solve an SOCP with a worst case per iteration complexity of $\mathcal{O}(M_b + p + 1)^3$ [37]. The overall procedure is as follows.

Algorithm 2

Input: A randomly generated point $\mathbf{p_0}$ that satisfies the set of constraints $\mathcal{M}_a$ for the original problem (W). Let $k = 0, 1, \cdots$

Output: An approximate solution of (W)

1. Solve the problem $\tilde{V_k}$ to obtain an approximate solution $\mathbf{h}_k$ of (W).
2. Set $\mathbf{p}_{k+1} = \mathbf{h}_k$, $k = k + 1$
3. Repeat until stopping criterion is satisfied (See text)

The initial starting point $\mathbf{p}_0$ is obtained by first generating 1000 random directions $\{\mathbf{p}_r\}_{r=1}^{1000}$ on the unit norm ball, and then scaling each vector by a factor γ_r until it satisfies the constraints in the set $\mathcal{M}_a$. We choose the scaling $\gamma_r = \sqrt{\frac{t}{\min\limits_{m \in \mathcal{M}_a} \mathbf{p}_r^H \mathbf{C}_m \mathbf{p}_r}}$, which is the smallest required to make $\gamma_r \mathbf{p}_r$ feasible for the constraint set $\mathcal{M}_a$. Finally, we set $\mathbf{p}_0$ to be the scaled vector which has the smallest norm. The algorithm is terminated when a feasible point for the original problem (4.7) is found. If feasibility is not achieved in 30 iterations, then the algorithm is re-initialized from a different starting point. If, after 5 such re-initializations, feasibility is still not achieved, then the algorithm is terminated and the solution returned by the last SOCP is chosen. Throughout our experiments, we use $\lambda = 10$.

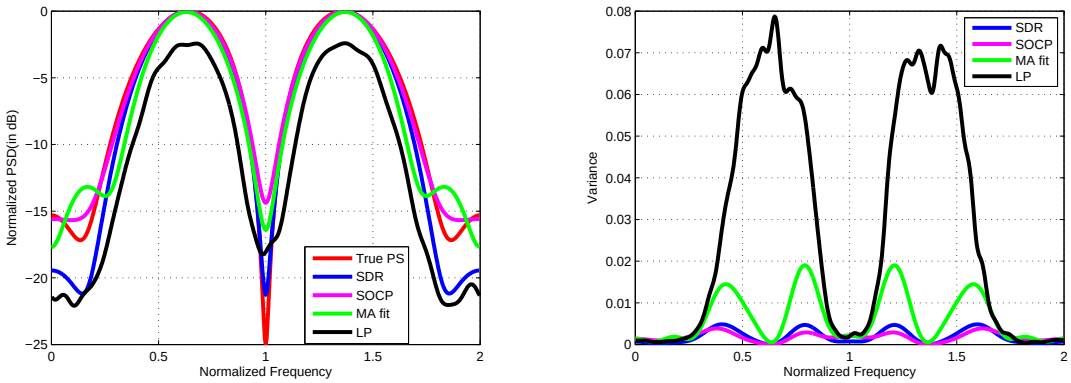
Chapter 5

Numerical Results

In this section, we present pertinent simulation results to gain insight regarding the effects of various parameters on the quality of spectral estimates. In addition to the previously described parametric and non-parametric approaches, we also include another technique for comparison, which involves fitting an MA model using the autocorrelation estimate obtained from the LP in [13]. The details of this method are discussed in Appendix B. First, we present a simulation which provides a comparison of the parametric methods presented here and the non-parametric LP formulation. A scenario was considered with $M = 100$ sensors, setting filter length $K = 24$ and the threshold t such that $\mathcal{M}_a = 30$. Fig. 5.1 shows the PS estimation results for a signal generated by a real MA(5) process while Fig. 5.2 shows the same for a complex MA(4) process under the same setting. It is assumed that the true order of the process is known *a priori* in each case, i.e., we have $p = q$. The spectra were normalized by their peak values and the plots were obtained over 500 Monte-Carlo trials. A few remarks are now in order:

- The quality of the PS estimates, obtained from only 100 input bits, is very satisfactory. The proposed parametric methods not only provide superior spectral estimates on average as compared to the non-parametric methods, but they also exhibit lower variance.

- Even though the SDR and the iterative SOCP algorithms produce approximate solutions for the non-convex QCQP, their performance is superior as compared to non-parametric LP followed by MA model fitting, for which both steps can be



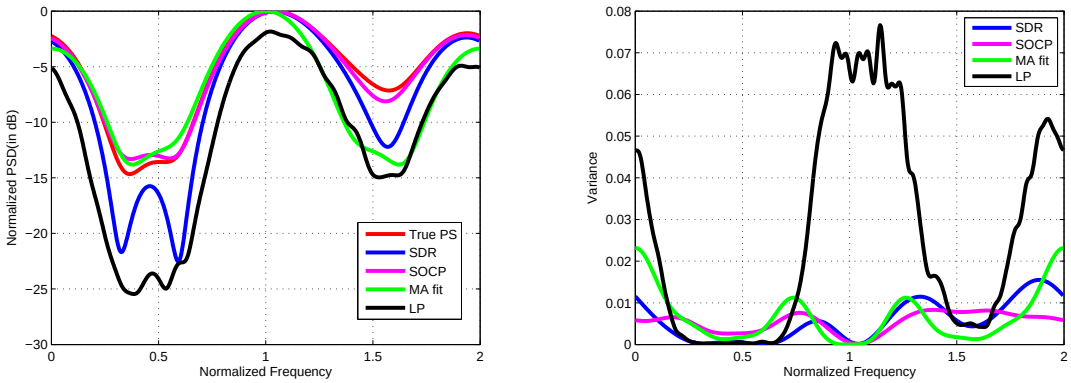
(A) (B)

Figure 5.1: Real MA(5) Model: (A) Mean Normalized Spectra (B) Variance of Normalized Spectra.

solved to global optimality. This is because the former methods utilize the MA representation from the outset, while the latter two-step method exploits the MA parametrization only in the second step.

- SDR followed by randomization failed to find a feasible solution which satisfied both sets of constraints for the original problem 99.8% and 100% of the trials, for the real and complex model cases respectively. In such instances, the solutions obtained by dropping the lower set of constraints still provide good quality spectral estimates, as explained previously.

- Solving the sequence of SOCPs, a feasible solution for the original problem was obtained in 100% of the trials, requiring an average of only 2.3 iterations (for the real case) and 3.9 iterations (for the complex case) to achieve feasibility. The choice to terminate the algorithm once feasibility is attained was based on the observation that additional iterations, which resulted in a feasible solution with a lower cost, did not, in general, bring about an improvement in the spectral estimate. Re-initializations of the algorithm from a different starting point was required in



(A) (B)

Figure 5.2: Complex MA(4) Model: (A) Mean Normalized Spectra (B) Variance of Normalized Spectra.

0.4% and 0.6% of the trials for the real and the complex case, respectively. [1]

- Although the worst-case computational complexity of solving each SOCP is theoretically much lower than that of the SDR, and few SOCP iterations are needed, in practice it was observed that the iterative SOCP algorithm was slower than SDR.

To better illustrate the performance of these methods, we use the Normalized Mean Square Error (NMSE) to define the quality of the PS estimate as

$$\mathbf{NMSE} = \mathbb{E}\left[\frac{\|S_x - \hat{S}_x\|^2}{\|S_x\|^2}\right] \tag{5.1}$$

where S_x is the true spectrum, $\hat{S}_x$ is the estimated spectrum, with both spectra being normalized by their peak values, and the expectation is taken with respect to the random signal and the random impulse response of the broadband FIR filters. We present

[1] The iterative SOCP method may also be initialized using SDR followed by the proposed randomization algorithm in Section IV A, resulting in a marginal improvement in performance, in terms of spectral NMSE. However, the number of re-initializations and average number of iterations taken to reach feasibility were similar to the random initialization strategy described in Section IV B. Since the SDR initialization does not result in a substantial all-round performance improvement, and has a theoretically worse computational complexity as compared to the SOCP method itself, we feel that its use is not justified in this case.

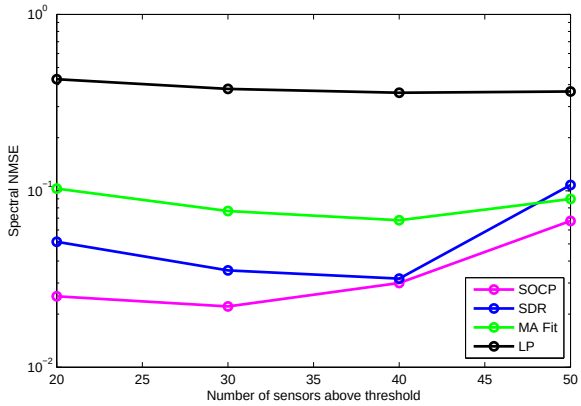


Figure 5.3: NMSE vs M_a for MA(9) models

another simulation in Fig. 5.3 where 50 real MA models of order 9 were randomly generated, in a scenario with $M = 100$, $K = 34$, and the spectral NMSE was plotted as a function of the number of sensors reporting above threshold M_a [2] . For each value of M_a, the spectral NMSE was computed for each model over 100 Monte Carlo trials, with the final NMSE value obtained by averaging across the models. The superior performance of the parametric methods is once again noted. The non-parametric method is significantly worse - off in comparison. While fitting an MA model to the autocorrelation lags brings about an improvement, the SDR and iterative SOCP algorithms still perform better, except when M_a is large. Some statistics regarding the solutions obtained from the SDR followed by randomization and the iterative SOCP algorithms are summarized in Tables 5.1 and 5.2 respectively.

Table 5.1 reports the percentage of simulation runs where SDR resulted in rank 1 solutions, feasible solutions after scaling the principal component / randomization candidates, and solutions after dropping the convex constraint set and scaling the principal component / randomization candidates, as the number of sensors reporting above threshold was varied. It is observed that the percentage of feasible solutions obtained from SDR, (either directly from rank 1 solutions or after randomization and scaling) is small to begin with, and as M_a increases, i.e., the number of non-convex constraints

[2] This experiment (and also others) were repeated for the case where each MA model realization was excited by streams of symbols drawn independently from BPSK and QPSK constellations, instead of complex i.i.d. WGN. The results obtained were very similar to the ones we present in this manuscript.

increases, it decreases even further. In the majority of the cases, SDR failed to return a feasible solution and we had to resort to dropping the convex constraints in order to obtain a working estimate.

In contrast, using the iterative SOCP algorithm always yielded a feasible solution (in 100% of the trials, including re-initializations for each value of M_a). Table 5.2 shows that the average number of iterations required by the algorithm to achieve feasibility is small, and the percentage of simulation runs which required re-initializing the algorithm from a different starting point is also very low.

M_a	20	30	40	50
SDR Rank-1 solution	1.20%	0.24%	0.04%	0.00%
Feasible sol. after SDR	0.98%	0.58%	0.20%	0.08%
Infeasible sol. after SDR and dropping convex constraints	97.82%	99.18%	99.76%	99.92%

Table 5.1: Results using the SDR approach.

M_a	20	30	40	50
Feasible solution	100%	100%	100%	100%
Avg. itrs. for feasibility	2.47	2.63	2.72	2.88
Re - initializations	0.00%	0.02%	0.04%	0.02%

Table 5.2: Results using the iterative SOCP approach.

5.1 Threshold Selection

The choice of threshold t plays an important role in determining the quality of the PS estimate in terms of NMSE. In order to determine the optimal choice of threshold, for which minimum NMSE is attained, we considered the following experiment. Setting $K = 30, M = 80$, the threshold t was changed in order to vary M_a, for different MA model orders. For each model order, 50 MA models were generated and for a given model, for each value of M_a, the spectral NMSE calculated was averaged over 100

Monte Carlo trials, with the final NMSE value obtained by averaging across all models. Prior knowledge of the true model order was assumed.

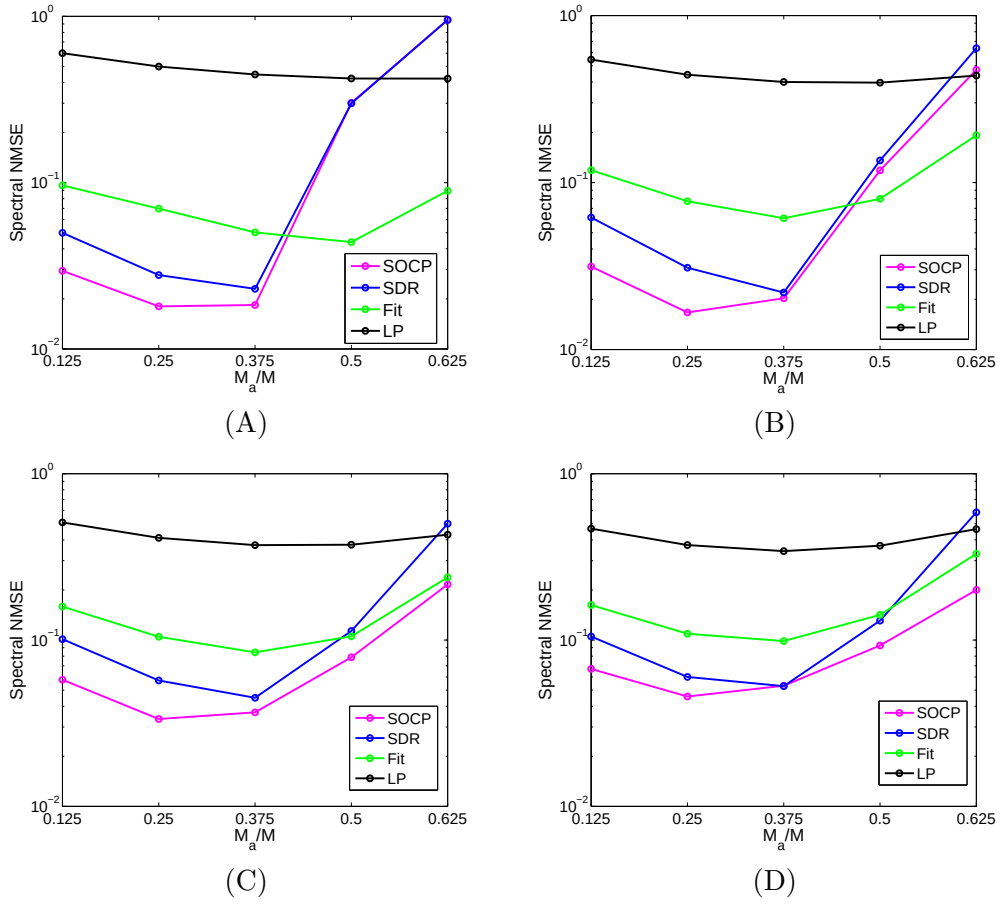


Figure 5.4: NMSE vs M_a/M for different model orders: (A) MA(3) models (B) MA(6) models, (C) MA(9) models and (D) MA(12) models

The results are plotted in Fig. 5.4. It was noted that the SDR and iterative SOCP approaches achieve the lowest NMSE when the threshold is selected such that approximately $25 - 35\%$ of the sensors report above threshold, for the range of model orders depicted in the plot. However, the minimum NMSE value for the parametric model-based algorithms increases with model order. This happens because the number of parameters to be estimated from a fixed set of data (80 bits in this case) increases with

increase in model order, which in turn results in an increase in the minimum NMSE.

5.2 Number of Sensors

Increasing the number of sensors makes the problem less underdetermined, i.e., estimating a fixed set of parameters using a larger number of bits should improve the estimation performance, since the number of inequality constraints increases. To investigate this, a scenario was considered where 50 MA models of order 12 were randomly generated, with $K = 30$, $p = 12$ and the spectral NMSE was computed for each model as a function of the number of sensors over 100 Monte-Carlo trials. In each trial, the threshold t was fixed such that 50 sensors reported above threshold. The results were averaged out across all model realizations and are depicted in Figure 5.5. It is observed that all methods exhibit improved performance as the number of sensors is increased. However, the parametric methods achieve lower NMSE values as compared to the other methods. In particular, as M increases (i.e., $\frac{M_a}{M}$ decreases), the iterative SOCP method's performance improves, while in contrast, SDR does not improve beyond a certain performance floor. Thus, as M increases, so that $\frac{M_a}{M}$ decreases with M, very low spectral NMSE can be achieved using the proposed iterative SOCP method.

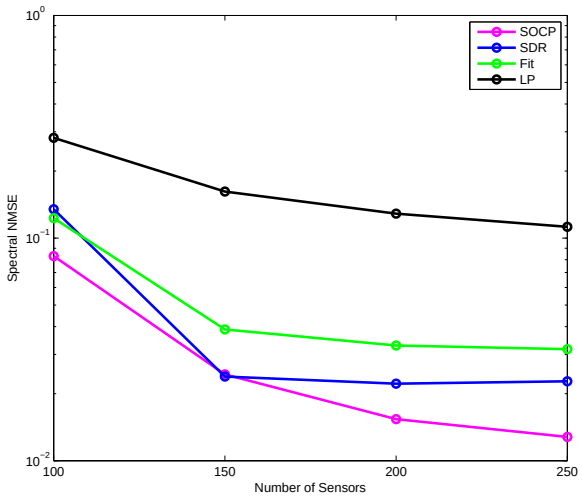


Figure 5.5: NMSE vs M for MA(9) models

5.3 Broadband Filter Length K

Setting the broadband filter length K determines the length of the $2K-1$ autocorrelation lag window. For an MA(q) process, the length of the autocorrelation lag window is $2q+1$, since $r(k) = 0, \forall\ |k| > q$. Hence, K should be set atleast equal to $q+1$, in order to capture the correct parametrization of the underlying MA model. From [13], it is known that for the LP formulation, setting K to be too low results in a smeared spectral estimate, while setting it too large also degrades the spectral estimate, since it involves estimating more unknowns from a fixed number of linear inequalities. We considered an experiment where $M = 100$, $M_a = 20$, with 50 randomly generated MA models and $p = q$. The spectral NMSE was plotted as a function of the filter length K in Fig. 5.6, from which it can be seen that for $K = 2$, the estimation results are very poor, since the underlying MA model is incorrectly parametrized. For $K = 7$, the number of autocorrelation lags estimated ($2K - 1 = 13$) is exactly equal to the true number of autocorrelation lags $2q + 1 = 13$. In this case, the LP formulation exactly matches the right extent of the underlying finite autocorrelation sequence, albeit it still does not use a model for it. Subsequently using the MA model to fit the finite autocorrelation obtained from LP results in an even better approximation of the true PS. For this choice of K, LP and LP followed by MA both outperform SDR and iterative SOCP, even though it is assumed that the true model order q is known to both of the latter. The reason this happens is that the LP problem and the SQP (see Appendix B) can be solved to global optimality while the SDR and iterative SOCP approaches only generate approximate solutions for the non-convex QCQP. However as K is increased further, the non-parametric model ceases to be a good approximation to the MA model. Hence, it's performance deteriorates relative to the parametric techniques. Also, the problem setup becomes more under-determined as the number of unknowns increases. The SDR and iterative SOCP do not suffer from this drawback since in their problem setup, the number of unknowns is independent of K. It is to be noted that for these algorithms, the NMSE decreases very slowly once K is increased to roughly $2-3$ times the true model order q. Even for high values of K, the minimum NMSE achieved with the SDR and iterative SOCP approaches is larger than that achieved with the LP problem followed by the MA model fit for $K = q + 1$.

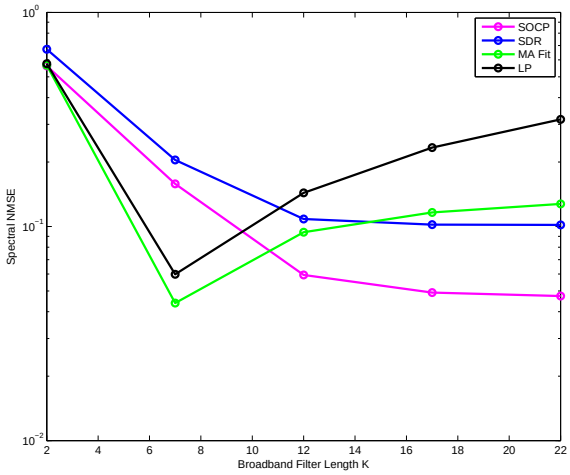


Figure 5.6: NMSE vs K for MA(6) models

5.4 Estimated Model Order p

An important observation from Fig. 5.6 is that SDR and iterative SOCP methods are robust to model order over-estimation. In reality, it is not easy to change the length of the random sensing filters to match the order of the operational model, which is unknown. Assuming that an upper bound on the true model order is known is much more realistic. It is known that underestimating the model order results in a highly smoothed spectral estimate, while overestimation introduces spurious details into the spectrum. Several model order selection criteria are available in the literature. We refer the interested reader to [14, Appendix C] for an overview of these techniques. For most of these rules, with a large data record, the probability of underestimating the model order approaches zero, while the probability of overestimating the model order remains non-zero even when the size of the data record tends to infinity.

Keeping this fact in mind, we consider the case where we overestimate the model order by a certain factor of the true order q. In Fig. 5.7, a scenario was considered with $M = 100$, $K = 24$, $M_a = 40$ where 50 minimum phase real MA(3) models were randomly generated and the spectral NMSE was plotted as a function of the assumed model order p. The NMSE was computed for each model over 100 Monte Carlo trials and the final values were averaged across the models. The performance of the LP remains constant,

since it does not make any model order postulations. The performance of the parametric methods is very satisfactory, even when the model order is overestimated by a significant amount. Moreover, the iterative SOCP method appears to more robust to model order overestimation, as compared to the SDR and the MA model fitting methods. Certain statistics of the solutions obtained from the SDR and SOCP algorithms are reported in Tables 5.3 and 5.4 respectively. Note that $p+1$ represents the number of problem unknowns. Hence, by increasing p, the problem dimensions are also increased. It is observed that the percentage of trials where SDR finds a feasible solution decreases as p increases, since in higher dimensions, not only is a rank 1 solution very rare, but finding a feasible solution after randomization and scaling is either extremely difficult or very expensive in terms of the number of randomization rounds. Thus, in the near total majority of these cases, the lower set of constraints is dropped in order to obtain a working estimate. For iterative SOCP, a feasible solution was obtained in more than 98% of the trials, at an average of approximately only $3-5$ iterations. Although the parametric approaches are robust to model order overestimation, one should refrain from setting p too high, since the computational cost of solving the optimization problems increases.

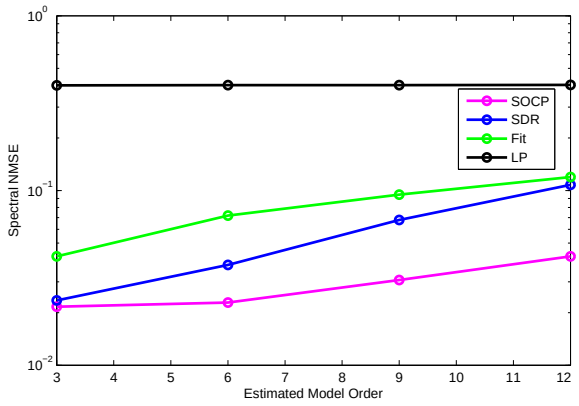


Figure 5.7: NMSE vs p for MA(3) models

p	3	6	9	12
SDR Rank-1 solution	6.14%	0.00%	0.00%	0.00%
Feasible sol. after SDR	9.22%	0.06%	0.00%	0.00%
Infeasible sol. after SDR and dropping convex constraints	84.64%	99.94%	99.98%	100.00%

Table 5.3: Results using the SDR approach.

p	3	6	9	12
Feasible solution	98.14%	99.78%	100%	100%
Avg. itrs. for feasibility	2.46	2.66	3.11	3.49
Re - initializations	3.64%	0.66%	0.00%	0.00%

Table 5.4: Results using the iterative SOCP approach.

Chapter 6

Conclusion

Considering a distributed spectrum sensing scenario, where a network of low end, scattered sensors transmitting randomly filtered power measurement bits to a fusion center is employed, it was shown that under the assumption that the signal of interest admits an MA representation, the problem of estimating admissible MA parameters, and hence the power spectrum, from 1 bit quantized data can be expressed as a non-convex QCQP.

Although this problem is NP–Hard, high quality approximate solutions were obtained by using the rank relaxation technique of SDR to solve a SDP problem, and a SPCA-based technique, formulated as a sequence of SOCPs, initialized from a point that is only known to be feasible for the non-convex constraints. The methods were compared with the non-parametric LP formulation in [13] and a two-step LP followed by MA model fitting approach, and their performance was assessed in various scenarios, through simulations. The averaged spectral estimates obtained from the parametric approaches exhibited lower NMSE and variance as compared to their non-parametric counterparts. Also, in the majority of the cases, SDR failed to find a feasible solution for the non-convex QCQP, in contrast to the iterative SOCP technique. In order to obtain the best performance from the parametric approaches, in terms of spectral NMSE, the choice of threshold was revealed to be such that roughly $25 - 35\%$ of the total number of sensors reported above threshold, for a wide range of MA model orders.

It was further demonstrated that when the true model order is known, and if it is possible to select the sensing filter length to match the resulting truncated autocorrelation model to the extent of the true autocorrelation sequence, then solving the LP

followed by MA model fitting is superior compared to the parametric methods. However, when the true model order is unknown and only an upper bound is available, it was shown that the parametric methods outperform the non-parametric LP and the two-step LP-MA approach. In particular, the iterative SOCP algorithm features notable robustness to model order overestimation.

Overall, we can conclude by saying that when the signal of interest admits an MA parametrization, then employing the parametric formulations yields better spectral estimates as compared to the non-parametric techniques, even though the former can only be solved approximately, as opposed to the latter. Amongst the parametric approximation techniques, the iterative SOCP algorithm conclusively exhibited superior performance over the SDR technique.

Whereas we focused on single-bit quantization, multi-bit sensor reports can be incorporated in two ways. The first is by equipping each sensor with multiple PN filters, instead of a single one. This can be implemented by using a single linear PN shift register at each sensor, and taking different shifts of the PN sequence as the impulse responses of the filters. Due to the shifts, the different impulse responses will be (approximately) uncorrelated, and hence, so will be the filter measurements. In essence, each sensor would perform the job of several sensors and would report multiple bits to the fusion center, from which it follows that we can use the proposed estimation methods without modification. Alternatively, each sensor could use a single broadband filter yet transmit multiple bits to the fusion center by using more quantization levels. However, in order to mitigate the effects of frequency selective fading, each sensor would be required to acquire a larger number of samples in order to average out its measurements across multiple fading states to within the higher accuracy required by multi-bit quantization. In our case, using a single threshold allows us to relax the sample averaging requirements since we only need sufficient averaging to ensure that the sign of the inequality corresponding to each power measurement is not reversed. Hence, there is a tradeoff in the number of samples to be acquired and the number of quantization levels employed - the finer the quantization, the higher sample averaging is needed.

References

[1] Erik Axell, Geert Leus, Erik G Larsson, and H Vincent Poor. Spectrum sensing for cognitive radio: State-of-the-art and recent advances. *IEEE Signal Process. Mag.*, 29(3):101–116, May 2012.

[2] Eric J Msechu and Georgios B Giannakis. Sensor-centric data reduction for estimation with WSNs via censoring and quantization. *IEEE Trans. Signal Process.*, 60(1):400–414, Jan. 2012.

[3] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Process. Mag.*, 25(2):21–30, Mar. 2008.

[4] Z. Tian and G. B. Giannakis. Compressed sensing for wideband cognitive radios. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, volume 4, pages 1349–1352. IEEE, Honolulu, HI, USA, Apr. 2007.

[5] J. A. Bazerque and G. B. Giannakis. Distributed spectrum sensing for cognitive radio networks by exploiting sparsity. *IEEE Trans. Signal Process.*, 58(3):1847–1862, Mar. 2010.

[6] Z. Fanzi, C. Li, and Z. Tian. Distributed compressive spectrum sensing in cooperative multihop cognitive networks. *IEEE J. Sel. Topics Signal Process.*, 5(1):37–48, Feb. 2011.

[7] D. D. Arianandа and G. Leus. Compressive wideband power spectrum estimation. *IEEE Trans. Signal Process.*, 60(9):4775–4789, Sept. 2012.

[8] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus. Compressive wide-band spectrum sensing. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pages 2337–2340. IEEE, Taipei, Taiwan, Apr. 2009.

[9] G. Leus and D. D. Arianandа. Power spectrum blind sampling. *IEEE Signal Process. Lett.*, 18(8):443–446, Aug. 2011.

[10] M. Lexa, M. Davies, J. Thompson, and J. Nikolic. Compressive power spectral density estimation. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pages 3884–3887. IEEE, Prague, Czech Republic, May 2011.

[11] V. Havary-Nassab, S. Hassan, and S. Valaee. Compressive detection for wide-band spectrum sensing. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pages 3094–3097. IEEE, Dallas, TX, USA, Mar. 2010.

[12] D. Romero and G. Leus. Wideband spectrum sensing from compressed measurements using spectral prior information. *IEEE Trans. Signal Process.*, 61(24):6232–6246, Dec 2013.

[13] O. Mehanna and N. D. Sidiropoulos. Frugal sensing: wideband power spectrum sensing from few bits. *IEEE Trans. Signal Process.*, 61(10):2693–2703, May 2013.

[14] Petre Stoica and Randolph L Moses. *Spectral Analysis of Signals*. Prentice Hall Upper Saddle River, NJ, 2005.

[15] George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. *Time series analysis: forecasting and control*. John Wiley & Sons, 2013.

[16] Petre Stoica, Tomas McKelvey, and Jorge Mari. MA estimation in polynomial time. *IEEE Trans. Signal Process.*, 48(7):1999–2012, July 2000.

[17] SM Kay. *Modern spectral estimation: theory and application*. Prentice Hall Englewood Cliffs, NJ, 1988.

[18] James Durbin. Efficient estimation of parameters in moving-average models. *Biometrika*, 142:306–316, 1959.

[19] Christopher Chatfield. Inverse autocorrelations. *J. R. Statist. Soc. A*, 32:363–377, 1979.

[20] Alexandre d'Aspremont and Stephen Boyd. Relaxations and randomized methods for nonconvex qcqps. *EE392o Class Notes, Stanford University*, 2003.

[21] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, volume 2. MPS - SIAM Series on Optimization. Cambridge UK: Cambridge University Press, 2001.

[22] Shuzhong Zhang. Quadratic maximization and semidefinite relaxation. *Mathematical Programming*, 87(3):453–465, 2000.

[23] Zhi-Quan Luo, Nicholas D Sidiropoulos, Paul Tseng, and Shuzhong Zhang. Approximation bounds for quadratic optimization with homogeneous quadratic constraints. *SIAM J. Optim.*, 18(1):1–28, 2007.

[24] Ami Wiesel, Yonina C Eldar, and Shlomo Shamai. Semidefinite relaxation for detection of 16-QAM signaling in MIMO channels. *IEEE Signal Process. Lett.*, 12(9):653–656, Sep. 2005.

[25] Khoa T Phan, Sergiy A Vorobyov, Nicholas D Sidiropoulos, and Chintha Tellambura. Spectrum sharing in wireless networks via qos-aware secondary multicast beamforming. *IEEE Trans. Signal Process.*, 57(6):2323–2335, June 2009.

[26] G Scutari, F Facchinei, L Lampariello, and P Song. Parallel and distributed methods for nonconvex optimization. In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. IEEE, Florence, Italy, May 2014.

[27] L-N Tran, Muhammad Fainan Hanif, and Markku Juntti. A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays. *IEEE Signal Process. Lett.*, 21(1):114–117, Jan. 2014.

[28] Amir Beck, Aharon Ben-Tal, and Luba Tetruashvili. A sequential parametric convex approximation method with applications to nonconvex truss topology design problems. *J. Global Optim.*, 47(1):29–51, 2010.

[29] Barry R Marks and Gordon P Wright. A general inner approximation algorithm for nonconvex mathematical programs. *Oper. Res.*, 26(4):681–683, 1978.

[30] A. Konar and N. D. Sidiropoulos. Parametric Frugal Sensing of Moving Average Power Spectra. In *Proc. IEEE ICASSP 2015 (submitted)*.

[31] Omar Mehanna, Kejun Huang, Balasubramanian Gopalakrishnan, Aritra Konar, and Nicholas D. Sidiropoulos. Feasible Point Pursuit and Successive Approximation of Nonconvex QCQPs. *IEEE Signal Process. Lett.*, 22(7):804–808, July 2015.

[32] Zhi-Quan Luo, Wing-kin Ma, AM-C So, Yinyu Ye, and Shuzhong Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Process. Mag.*, 27(3):20–34, May 2010.

[33] Yurii Nesterov, Arkadii Nemirovskii, and Yinyu Ye. *Interior-point Polynomial Algorithms in Convex Programming*, volume 13. SIAM, 1994.

[34] Jos F Sturm. Using SeDuMi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

[35] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, volume 27. Springer, 2000.

[36] Jorge Nocedal and Stephen J Wright. *Numerical Optimization*. Springer, 2006.

[37] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear algebra and its applications*, 284(1):193–228, 1998.

[38] Shao-Po Wu, Stephen Boyd, and Lieven Vandenberghe. FIR filter design via semidefinite programming and spectral factorization. In *Proc. IEEE Conf. Dec. and Control*, volume 1, pages 271–276. IEEE, 1996.

Appendix A

Properties of $\mathbf{C}_m$ Matrices

From (4.5), we have $\mathbf{C}_m$ given by

$$\mathbf{C}_m = c_{m,0}\mathbf{\Theta}_0^{p+1} + \sum_{k=1}^{\min(K-1,p)} (c_{m,k}\mathbf{\Theta}_k^{p+1} + c_{m,-k}\mathbf{\Theta}_{-k}^{p+1}) \tag{A.1}$$

where $c_{m,k} = \mathbf{g}_m^H \mathbf{\Theta}_k^K \mathbf{g}_m$ is the deterministic autocorrelation of the m^{th} broadband filter with impulse response $\mathbf{g}_m$, $\mathbf{\Theta}_k^K$ and $\mathbf{\Theta}_k^{p+1}$ are $K \times K$ and $(p+1) \times (p+1)$ elementary Toeplitz matrices respectively, with ones on the k^{th} sub-diagonal and zeros elsewhere. From (A.1), it is evident that $\mathbf{C}_m$ possesses Toeplitz structure. Since we also have

$$\begin{aligned} c_{m,k}^* &= (\mathbf{g}_m^H \mathbf{\Theta}_k^K \mathbf{g}_m)^* \\ &= \mathbf{g}_m^H \mathbf{\Theta}_{-k}^K \mathbf{g}_m = c_{m,-k} \end{aligned}$$

$\mathbf{C}_m$ is also Hermitian. In order to show positive semi-definiteness, we consider the following cases, when $K \geq p+1$ and $K < p+1$.

A.1 $K \geq p+1$

In this case, from (A.1), we have

$$\begin{aligned}\mathbf{C}_m &= c_{m,0}\mathbf{\Theta}_0^{p+1} \\ &\quad + \sum_{k=1}^{p}(c_{m,k}\mathbf{\Theta}_k^{p+1} + c_{m,-k}\mathbf{\Theta}_{-k}^{p+1})\end{aligned} \tag{A.2}$$

Note that $\mathbf{C}_m$ corresponds to the $(p+1)^{th}$ principal submatrix of the $K \times K$ deterministic autocorrelation matrix of the m^{th} broadband filter. Since the autocorrelation matrix is positive semi-definite, it follows that all it's principal submatrices are also positive semi-definite. Hence, $\mathbf{C}_m$ is also positive semi-definite.

A.2 $K < p+1$

From (A.1), we have

$$\begin{aligned}\mathbf{C}_m &= c_{m,0}\mathbf{\Theta}_0^{p+1} \\ &\quad + \sum_{k=1}^{K-1}(c_{m,k}\mathbf{\Theta}_k^{p+1} + c_{m,-k}\mathbf{\Theta}_{-k}^{p+1}) \\ &\quad + \sum_{k=K}^{p}(0\mathbf{\Theta}_k^{p+1} + 0\mathbf{\Theta}_{-k}^{p+1})\end{aligned} \tag{A.3}$$

where the last term indicates that we substitute zeros on the sub - diagonals $|k| \geq K$. In this case, $\mathbf{C}_m$ can be interpreted as being the $(p+1)\times(p+1)$ deterministic autocorrelation matrix of the filter with impulse response $\mathbf{g'}_m = [\mathbf{g}_m^H, \underbrace{0, \ldots, 0}_{p+1-K \text{ zeros}}]^H$, which is also positive semi-definite.

Appendix B

MA Model Fitting

In this section, a different parametric MA spectral estimation method is considered , which uses the estimate of the autocorrelation vector of the signal of interest, obtained from the solution of the LP in [13], as a starting point. Here, we make use of the signal's MA representation in a second step, in contrast with the previously defined methods which utilize the MA parametrization from the very outset. Given the $2K-1$ lag autocorrelation sequence $\hat{\mathbf{r}}_{\mathbf{x}}$ from the LP, we denote the non-parametric spectral estimate as $\hat{S}_x(e^{j\omega}) = \sum_{k=-(K-1)}^{K-1} \hat{r}_x(k)e^{-j\omega k}$. Let $S_x(e^{j\omega})$ be the PSD of an MA process of order p; where p is our estimate of the order of the process. Thus, we have, $S_x(e^{j\omega}) = \sum_{k=-p}^{p} r_x(k)e^{-j\omega k}$ where $\mathbf{r}_{\mathbf{x}}$ constitutes the $2p+1$ lag autocorrelation sequence of the MA process. We seek to determine the $S_x(e^{j\omega})$ that is closest to $\hat{S}_x(e^{j\omega})$ in the following least squares sense

$$\text{minimize} \quad \frac{1}{2\pi}\int_{-\pi}^{\pi}[\hat{S}_x(e^{j\omega}) - S_x(e^{j\omega})]^2 \mathrm{d}\omega \tag{B.1}$$

As shown in [14, p. 131], the spectral least squares criterion can be rewritten as a weighted least squares autocorrelation fitting criterion with an infinite number of spectral non-negativity constraints. Exploiting the trace parametrization of an MA autocorrelation sequence to express the non-negativity constraints, the problem is reformulated as a semidefinite quadratic program (SQP) in terms of the MA autocorrelation sequence $\{r_x(k)\}_{k=0}^{p}$, which can be efficiently solved using interior point methods in MATLAB

(e.g., using SeDuMi) at a complexity of $\mathcal{O}(p^4)$ flops. Taking the DTFT of the sequence $\{r_x(k)\}_{k=0}^{p}$ gives us the spectral estimate. If a minimum phase estimate of $\mathbf{h}$, $\mathbf{h}_{min}$ is desired, then using $\{r_x(k)\}_{k=0}^{p}$, a spectral factorization step, as described in [38] is carried out. The spectral estimate may then also be obtained by taking the magnitude square of the DTFT of $\mathbf{h}_{min}$. Note that these two techniques of obtaining the spectral estimate from $\{r_x(k)\}_{k=0}^{p}$ are equivalent due to the trace parametrization of $\{r_x(k)\}_{k=0}^{p}$ in terms of $\mathbf{h}$.