Maximum Likelihood Passive and Active Sensing of Wideband Power Spectra From Few Bits

Omar Mehanna, *Student Member, IEEE*, and Nicholas D. Sidiropoulos, *Fellow, IEEE*

***Abstract*—Wideband power spectrum sensing is essential for cognitive radio and many other applications. Aiming to crowdsource spectrum sensing operations, a novel *frugal sensing* framework was recently proposed, employing a network of low duty-cycle sensors (e.g., running in background mode on consumer devices) reporting randomly filtered broadband power measurement bits to a fusion center, which in turn estimates the ambient power spectrum. Frugal sensing is revisited here from a statistical estimation point of view. Taking into account fading and insufficient sample averaging considerations, maximum likelihood (ML) formulations are developed which outperform the original minimum power and interior point solutions when the soft power estimates prior to thresholding are noisy. Assuming availability of a *downlink* channel that the fusion center can use to send threshold information, *active sensing* strategies are developed that quickly narrow down and track the power spectrum estimate, using ideas borrowed from cutting plane methods to develop active ML solutions. Simulations show that satisfactory wideband power spectrum estimates can be obtained with passive ML sensing from few bits, and much better performance can be attained using active sensing. Various other aspects, such as known emitter spectral shapes and different types of non-negativity constraints, are also considered.**

***Index Terms*—Cognitive radio, collaborative sensing, spectral analysis, spectrum sensing.**

I. Introduction

SPECTRUM sensing is the most important component of cognitive radio as it enables users to discover transmission opportunities, thus forming the basis for adaptive spectrum sharing [3]. Spectrum estimation is also essential in developing efficient power control schemes for secondary users in spectrum underlay settings [4]. Collaborative spectrum sensing, involving many sensors taking relatively sparse measurements across space, time, and frequency, is crucial to increase reliability and alleviate the harmful effects of fading and the hidden terminal problem. In addition to cognitive radio, collaborative sensing plays a fundamental role in various applications such as smart agriculture, industrial monitoring, weather forecasting, military surveillance, disaster response, and health monitoring applications [5].

Wideband spectrum sensing requires very high sampling rates and thus high power consumption and high-cost analog-to-digital converters (ADCs) [3]. Exploiting frequency-domain sparsity, compressive spectrum sensing can obtain accurate spectrum estimates at sub-Nyquist sampling rates, without frequency sweeping [6], [7]. Cooperative spectrum sensing schemes that use compressive sensing have been considered in [8], [9], where the spectrum is estimated locally, then consensus on globally fused sensing outcomes is reached. However, the methods of [8] and [9] require sensors to preform complex local computations, and entail significant communication between sensors.

Whereas most work on spectrum sensing (e.g., [3], [5]–[9]) has focused on reconstructing the signal's *Fourier spectrum* (i.e., the Fourier transform of the signal itself), in cognitive radio and many other applications only the *power spectrum* (PS) is needed (i.e., the Fourier transform of the signal's autocorrelation)—there is no need to reconstruct or demodulate the original signal itself [10]. Power spectrum estimation methods have been developed in [10]–[14], where it is shown that sampling rate requirements can be considerably relaxed by exploiting a low-order correlation model, without even requiring spectrum sparsity. In [15], it is shown that further reduction in sampling rate is possible by exploiting available prior information on the spectrum, such as spectral masks and carrier frequencies. The main idea in 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 building an over-determined system of linear equations using sufficient power measurements.

The PS estimation methods considered in [10]–[15] assume analog amplitude samples (i.e., ignore quantization issues). In distributed spectrum sensing scenarios, however, sending analog or finely quantized sample streams to a fusion center (FC) is a heavy burden in terms of communication overhead and battery lifetime. This was relaxed in [16], where a network sensing scenario was considered, with each sensor reporting a single randomly filtered power measurement bit to the FC, which estimates the ambient PS from the collected bits. A Linear Programming (LP) formulation was introduced in [16], generalizing classical nonparametric PS estimation to the case where the data is in the form of inequalities, rather than equalities, and exploiting the autocorrelation parametrization and pertinent non-negativity properties.

Frugal sensing [16]—PS estimation from relatively few 1-bit measurements—seems reminiscent of *1-bit compressed*

Manuscript received April 16, 2014; revised October 30, 2014; accepted December 19, 2014. Date of publication January 12, 2015; date of current version February 11, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Akbar Sayeed. Supported in part by NSF ECCS-1231504, NSF AST-1247885. Conference versions [1], [2] of parts of this work have been presented at the Thirty-Eighth International Conference on Acoustics, Speech, and Signal Processing, Vancouver, Canada, May 26–31, 2013, and the Twenty-First European Signal Processing Conference, Marrakech, Morroco, September 9–13, 2013.

The authors are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: meha0006@umn.edu; nikos@umn.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2391073

sensing (CS) [17]–[20], albeit there are significant differences between the two. Frugal sensing aims directly for the autocorrelation; 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, which can be intelligently selected to ensure asymptotic consistency, and robustness to bit flips, as we will show in this paper, whereas errors were not considered in [17]–[20].

Contributions: The underlying assumption which enables using the LP formulation in [16] is that the power measurement prior to quantization at each sensor is accurate enough to avoid flipping the reported power measurement bit. We relax this assumption here, and show 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. Interestingly, with the exception of the autocorrelation-specific constraints, this ML formulation turns out being similar to the formulation considered in [21] for quantized CS measurements, and in [22] for a seemingly very different problem—consumer preference measurement using so-called *conjoint analysis*. Despite these similarities, the use of autocorrelation-specific constraints differentiates our work here from [21], [22], and is indeed instrumental in obtaining meaningful PS estimates with a small number of bits, as shown in simulations and comparisons to the Cramér-Rao bound (CRB) that is also derived here. In order to reduce the number of bits transmitted from the sensors to prolong battery life and minimize communication overhead, we also propose a ML/CRB-driven censoring scheme, where only sensors that provide the most useful information bits are permitted to send, while other sensors remain silent. We further extend our formulations to the case where the PS is modeled as a weighted sum of candidate spectral density primitives with unknown weights, similar to [15] and [23], but for coarsely quantized (1-bit) data.

All approaches discussed up to this point are passive, in the sense that the FC simply listens to the received bits and forms a PS estimate—there is no FC-to-sensor communication. Assuming availability of a 'downlink' channel that the FC can use to send threshold information, we further develop *active sensing* strategies that quickly narrow down and track the PS estimate, using ideas borrowed from cutting plane methods in optimization theory. This line of work is the *second major contribution of this paper*. Convergence of the proposed algorithms to the true finite-length autocorrelation, as more sensors report their measurement bits, is shown. It is worth mentioning that a different adaptive thresholding algorithm has been developed in [24] for the 1-bit CS setting; however, the Bayesian framework considered in [24] is computationally intractable and assumes a signal with separable distribution that is known a-priori, which is not a valid assumption in general.

Relative to the conference submissions [1] and [2], this journal version i) covers nonparametric PS estimation in addition to the case where the PS follows a certain model (only model-based PS estimation was considered in [1]); ii) proves that the effect of frequency-selective fading can be approximated by adding a Gaussian random error variable to the fading-free soft power measurement prior to quantization; iii) proposes low-complexity active sensing strategies that take into account possible bit-flips due to errors in the soft power measurements before quantization; and iv) fleshes-out numerical results and comparisons.

The rest of the paper is organized as follows. The network sensing model and the formulation of [16] are presented in Section II. The ML formulation for PS estimation with a passive FC is provided in Section III, followed by the proposed active sensing algorithms in Section IV. Simulations and discussions on the various design trade-offs are presented in Section V, and conclusions are drawn in Section VI. Technical derivations and proofs are deferred to the Appendices.

II. Formulation and Background

We begin by describing the network sensing scenario, the sensor measurement chain, and the baseline frugal sensing (LP) formulation considered in [16]. The sensor measurement chain in this paper is the same as in [16]. Section III uses the same (passive) network sensing scenario as [16], while a new active network sensing paradigm is considered in Section IV.

A. Network Sensing Model and Sensor Measurement Chain

Consider a network comprised of M scattered sensors that measure the ambient signal power and report to a FC. In the presence of frequency-selective fading, the received signal at sensor $m \in \{1,\dots,M\}$, sampled using a Nyquist-rate ADC, is the convolution of the primary wide-sense stationary (WSS) signal $x(n)$ with the T-tap linear finite impulse response (FIR) fading channel $\{h_m(\ell)\}_{\ell=0}^{T-1}$, expressed as

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

where γ_m models the associated sensor-specific propagation loss. A very simple sensor measurement chain is considered, where sensor $m \in \{1,\dots,M\}$ first uses automatic gain control (AGC) to adjust the scaling of its received signal such that $y_m(n) = \tilde{y}_m(n)/\gamma_m$, then passes $y_m(n)$ through a wideband random FIR filter of length-K. The length-K impulse response sequence of the FIR filter at sensor m, $g_m(n)$, is pseudo-randomly and independently generated, where each filter coefficient is uniformly drawn from the set of 4 possible QPSK symbols, i.e.,

$$g_m(n)\begin{cases}\sim \mathcal{U}(\{1+j, 1-j, -1+j, -1-j\}) & n\in[0,K-1]\\ =0 & \text{otherwise}\end{cases} \tag{2}$$

where $\mathcal{U}(S)$ stands for the uniform probability mass function over the finite set S. The filter sequence $g_m(n)$ can be generated using a pseudo-noise (PN) linear shift register, whose initial seed is unique for each sensor (e.g., its serial number), and is known to the FC. Using random PN filters is appealing because it ensures diversity, and also for its simplicity: convolution requires no multiplication, and there is no need for coordination between sensors on who-covers-what. PN filters can also be motivated from a random projections/incoherence point of view, similar to compressed sensing [6].

The output sequence of the random FIR filter is expressed as $z_m(n) = \sum_{k=0}^{K-1} g_m(k) y_m(n-k)$. Let α_m denote the average power of the WSS signal $z_m(n)$, i.e., $\alpha_m := \mathbb{E}[|z_m(n)|^2]$. Note that the expectation is obtained given a single realization

of $g_m(n)$ for each sensor. Each sensor estimates α_m by sample averaging using N samples

$$\hat{\alpha}_m = \frac{1}{N}\sum_{n=0}^{N-1} |z_m(n)|^2. \tag{3}$$

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 bit $b_m = 1$ to the FC, otherwise it sends $b_m = -1$. The objective is to estimate the PS of the signal $x(n)$ at the FC from the measurement bits $\{b_m\}_{m=1}^M$. It is worth noting that the Nyquist-rate sampling requirement can be lifted by using an equivalent analog processing and integration chain, as shown in [16], and that the sensors are not required to be synchronized since the PS of the signal $x(n)$ is invariant with respect to timing offsets and phase shifts.

B. Frugal Sensing LP Formulation [16]

Here we summarize the main technique proposed in [16] for PS estimation from the measurement bits $\{b_m\}_{m=1}^M$. 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 impulse response of the random FIR filter $g_m(n)$ (i.e., ignoring fading), and define $\tilde{\alpha}_m := \mathbb{E}[|\tilde{z}_m(n)|^2]$. Let $r_x(k) := \mathbb{E}[x(n)x^*(n-k)]$ denote the autocorrelation sequence of $x(n)$, and define the *deterministic* autocorrelation of $g_m(n)$ as

$$q_m(k) := \sum_{n=0}^{K-1} g_m(n)g_m^*(n+k). \tag{4}$$

Hence, it can be shown that

$$\tilde{\alpha}_m = \sum_{k=1-K}^{K-1} r_x(k)q_m^*(k) = \mathbf{q}_m^T\mathbf{r}_x \tag{5}$$

where

$$\begin{aligned}\mathbf{r}_x &:= \big[r_x(0), \mathrm{Re}\{r_x(1)\},\ldots,\mathrm{Re}\{r_x(K-1)\},\\ &\qquad \mathrm{Im}\{r_x(1)\},\ldots,\mathrm{Im}\{r_x(K-1)\}\big]^T\\ \mathbf{q}_m &:= \big[q_m(0), 2\,\mathrm{Re}\{q_m(1)\},\ldots,2\,\mathrm{Re}\{q_m(K-1)\},\\ &\qquad 2\,\mathrm{Im}\{q_m(1)\},\ldots,2\,\mathrm{Im}\{q_m(K-1)\}\big]^T\end{aligned}$$

with $\mathrm{Re}\{\cdot\}$ and $\mathrm{Im}\{\cdot\}$ denoting the real and imaginary parts, respectively. In [16], it is assumed that the random frequency-selective fading channels to the different sensors have the same second-order statistics, and the effect of frequency-selective fading is mitigated by averaging the measurements at each sensor over a long period of time (many channel fading states). In this case, the averaged channel effect is common to all sensors, and can be absorbed in $\mathbf{r}_x$, cf. the Appendix of [16]. In this case, we can take $\hat{\alpha}_m = \tilde{\alpha}_m = \mathbf{q}_m^T\mathbf{r}_x$.

A windowed *nonparametric* estimate of the PS can be obtained from the K-lag autocorrelation as $\hat{S}_x(\omega) = \sum_{k=-K+1}^{K-1} r_x(k)e^{-j\omega k}$, $\omega \in [0, 2\pi)$. A discrete N_F-point estimate is thus obtained as $\hat{\mathbf{s}}_x = \mathbf{F}\mathbf{r}_x$, where $\hat{s}_x(f) = \hat{S}_x\left(\frac{2\pi f}{N_F}\right)$, $f = 0,\ldots,N_F-1$, $\mathbf{F} = \tilde{\mathbf{F}}\mathbf{W}$, $\tilde{\mathbf{F}}$ is the $N_F \times (2K-1)$ (phase-shifted) discrete Fourier transform (DFT) matrix, and

$$\mathbf{W} := \begin{bmatrix} \mathbf{0}_{K-1} & \tilde{\mathbf{I}}_{K-1} & -j\,\tilde{\mathbf{I}}_{K-1} \\ 1 & \mathbf{0}_{K-1}^T & \mathbf{0}_{K-1}^T \\ \mathbf{0}_{K-1} & \mathbf{I}_{K-1} & j\,\mathbf{I}_{K-1} \end{bmatrix} \tag{6}$$

where $\mathbf{0}_{K-1}$ is a $(K-1)$ zeros vector, $\mathbf{I}_{K-1}$ is the $(K-1)$ identity matrix, and $\tilde{\mathbf{I}}_{K-1}$ is the *flipped* $(K-1)$ identity matrix. Thus, without any prior knowledge on the PS, the goal is to estimate the autocorrelation vector $\mathbf{r}_x$ at the FC from the measurement bits $\{b_m\}_{m=1}^M$ received from the M sensors.

Properties of the autocorrelation can be exploited to define an initial bounded feasible region for $\mathbf{r}_x$. First, an upper bound $P_{\max}$ 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 $0 \leq r_x(0) \leq P_{\max}$. 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 $\mathbf{r}_x \in \mathcal{P}$, where the *bounded* polyhedron

$$\begin{aligned}\mathcal{P} := \{\mathbf{r}_x \in \mathbb{R}^{2K-1} \mid &-r_x(0) \leq \mathrm{Re}\{r_x(k)\} \leq r_x(0),\\ &-r_x(0) \leq \mathrm{Im}\{r_x(k)\} \leq r_x(0),\ 0 \leq r_x(0) \leq P_{\max},\\ &k = 1,\ldots,K-1\}.\end{aligned} \tag{7}$$

Note that the choice of K determines the trade-off between the level of smearing in the estimated PS [26] and the level of under-determinacy (i.e., number of unknowns) in the estimation problem, as shown in [16].

When the FC receives the few bits $\{b_m\}_{m=1}^M$, the setup is still 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 non-negative at all frequencies. However, including the non-negativity constraint $\hat{\mathbf{s}}_x = \mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ 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 $\mathbf{r}_x$. Interestingly, it has been shown in [16] that the *semidefinite* constraint $\mathbf{R}_x \succeq 0$ is already implied by the *linear* PS non-negativity constraint $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$.

To further decrease the under-determinacy of the setup, the expected sparsity of the PS can be exploited by minimizing the sparsity-inducing ℓ_1-norm of $\hat{\mathbf{s}}_x$. Enforcing $\hat{\mathbf{s}}_x \geq \mathbf{0}$ implies that $\|\hat{\mathbf{s}}_x\|_1 = \sum_{f=0}^{N_F-1} \hat{s}_x(f) = N_F\hat{r}_x(0)$, i.e., minimizing the total signal power implicitly encourages sparsity in the reconstructed PS. Assuming that $\hat{\alpha}_m = \tilde{\alpha}_m$ such that $b_m = 1$ implies $\mathbf{q}_m^T\mathbf{r}_x \geq t_m$ and $b_m = -1$ implies $\mathbf{q}_m^T\mathbf{r}_x < t_m$, [16] proposed the following LP formulation

$$\begin{aligned}\min_{\mathbf{r}_x\in\mathcal{P}} \quad & r_x(0)\\ \text{s.t.} \quad & \mathbf{F}\mathbf{r}_x \geq \mathbf{0}, \quad b_m(\mathbf{q}_m^T\mathbf{r}_x - t_m) \geq 0, \quad m = 1,\ldots,M.\end{aligned} \tag{8}$$

The LP (8) can be efficiently solved using specialized algorithms.

The underlying assumption which enables using the LP formulation (8) is that the soft power estimate at each sensor is accurate enough i.e., $\hat{\alpha}_m \approx \tilde{\alpha}_m$. Unlike [16], this assumption is relaxed in this work where we explicitly model the distribution

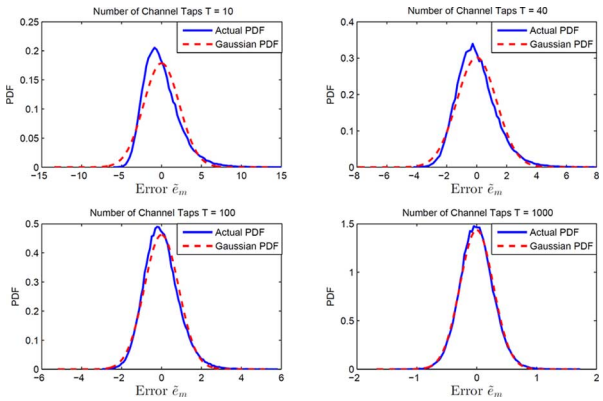


Fig. 1. The Gaussian distribution approximation of $\tilde{e}_m$ with different channel taps $T \in \{10,40,100,1000\}$.

of the error in the soft power estimates (prior to thresholding) due to frequency-selective fading and insufficient sample averaging by a Gaussian distribution, and exploit this error model in a more robust ML formulation. In the next section, we first 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 IV.

III. Passive ML Sensing

Here we relax the assumption that $\hat{\alpha}_m \approx \tilde{\alpha}_m$ which is used to derive the LP formulation (8) in [16]. Assume that the T taps of the FIR fading channel $\{h_m(\ell)\}_{\ell=0}^{T-1}$ are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero-mean and variance $1/T$, and that the channel is time-invariant for the sensing epoch. Define the error due to the fading channel at sensor m as $\tilde{e}_m := \alpha_m - \tilde{\alpha}_m$.

1) Claim 1: As $T \to \infty$, the errors $\{\tilde{e}_m\}_{m=1}^M$ can be approximated as i.i.d. zero-mean Gaussian random variables.

The proof is sketched in Appendix A, and uses the Lyapunov central limit theorem [25, pp. 371]. Exhaustive simulations have indicated that approximating the distribution of $\tilde{e}_m$ with a Gaussian distribution is a close approximation, even for relatively small T. Fig. 1 compares the actual distribution of $\tilde{e}_m$ obtained via Monte-Carlo simulations and the corresponding Gaussian distribution approximation, for different channel taps $T \in \{10,40,100,1000\}$, and filter length $K = 20$. The figure shows that the Gaussian distribution is a reasonable approximation for relatively small $T = 10$, and a very accurate one for relatively large $T = 1000$. The figure also shows the decrease in the variance of $\tilde{e}_m$ as T increases.

The estimation errors due to insufficient sample averaging, $\bar{e}_m := \hat{\alpha}_m - \alpha_m$, 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 $\hat{\alpha}_m = \tilde{\alpha}_m + e_m$, where $e_m := \tilde{e}_m + \bar{e}_m$ is a Gaussian random variable with zero mean and variance $\sigma_m^2 = \tilde{\sigma}_m^2 + \bar{\sigma}_m^2$. Thus, the power measurement bit of each sensor can be expressed as

$$b_m = \text{sign}(\underbrace{\mathbf{q}_m^T \mathbf{r}_x + e_m}_{\hat{\alpha}_m} - t_m) \tag{9}$$

where $\text{sign}(u) = 1$ if $u \geq 0$ and -1 otherwise, for $u \in \mathbb{R}$.

The assumption that $\text{sign}(\hat{\alpha}_m - t_m) = \text{sign}(\tilde{\alpha}_m - t_m)$, $\forall m$, allows using linear inequality constraints in (8). As the error variances $\{\sigma_m^2\}_{m=1}^M$ increase, however, formulation (8) becomes inaccurate, and the constraints may become inconsistent as the number of flipped bits due to errors increase. In order to lift this limitation, we propose to exploit the Gaussian distribution of $\{e_m\}_{m=1}^M$ to derive a more flexible and robust ML formulation. The nonparametric PS estimation case is first considered in the next subsection, followed by a parametric PS case.

A. Nonparametric ML Formulation

We may obtain an ML estimate of the K-lag autocorrelation $\mathbf{r}_x$ by exploiting the Gaussian distribution of $\{e_m\}_{m=1}^M$ as follows. 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 of the bits $b_1, \ldots, b_M$, given the autocorrelation $\mathbf{r}_x$, is

$$\begin{aligned} &f(b_1, \ldots, b_M|\mathbf{r}_x) \\ &= \prod_{m\in\mathcal{M}_+} \Pr(\tilde{\alpha}_m + e_m \geq t_m) \prod_{m\in\mathcal{M}_-} \Pr(\tilde{\alpha}_m + e_m < t_m) \\ &= \prod_{m\in\mathcal{M}_+} \Phi\left(\frac{\mathbf{q}_m^T\mathbf{r}_x - t_m}{\sigma_m}\right) \prod_{m\in\mathcal{M}_-} \Phi\left(-\frac{\mathbf{q}_m^T\mathbf{r}_x - t_m}{\sigma_m}\right) \end{aligned} \tag{10}$$

where $\Phi(u) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt$ is the cumulative distribution function (CDF) of the Gaussian distribution. The log-likelihood function can be written as

$$\log f(b_1, \ldots, b_M|\mathbf{r}_x) = \sum_{m=1}^{M} \log \Phi\left(\frac{b_m(\mathbf{q}_m^T\mathbf{r}_x - t_m)}{\sigma_m}\right). \tag{11}$$

Similar to the LP (8), the non-negativity constraint $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ is essential in obtaining a good estimate of $\mathbf{r}_x$, and the sparsity of $\hat{\mathbf{s}}_x$ can be exploited by minimizing $r_x(0)$. Therefore, an ML formulation, with a sparsity-inducing penalty term, is given as

$$\begin{aligned} \max_{\mathbf{r}_x\in\mathcal{P}} \quad & \sum_{m=1}^{M} \log \Phi\left(\frac{b_m(\mathbf{q}_m^T\mathbf{r}_x - t_m)}{\sigma_m}\right) - \lambda\, r_x(0) \\ \text{s.t.} \quad & \mathbf{F}\mathbf{r}_x \geq \mathbf{0} \end{aligned} \tag{12}$$

where $\lambda \geq 0$ is a tuning parameter that controls the sparsity of the PS estimate. Since the Gaussian CDF $\Phi(\cdot)$ is log-concave [27, pp. 104], problem (12) is convex and can be solved efficiently using interior-point algorithms [27]. Note that the maximizer of problem (12) always exists since $\mathbf{r}_x$ is bounded in $\mathcal{P}$. It is worth mentioning that the autocorrelation-specific constraints $\mathbf{r}_x \in \mathcal{P}$ and $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ considered in (12), which are not present in the ML formulations in [21], [22], can yield significantly better-quality estimates with fewer M.

Omitting the inequality constraint $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ and the sparsity-inducing penalty ($\lambda = 0$) in (12), a sufficient condition for consistency of the ML estimate, i.e., for $\hat{\mathbf{r}}_x$ to converge in probability to the true $\mathbf{r}_x$ as $M \to \infty$, is that the second-moment matrix $\mathbf{R}_q := \mathbb{E}[\mathbf{q}_m\mathbf{q}_m^T]$ is full-rank [28] (see also [22]). This condition also ensures statistical identifiability of $\mathbf{r}_x$ from the measurements. Clearly, for i.i.d. PN filter impulse responses $\{g_m(n)\}$ as in (2), the matrix $\mathbf{R}_q$ is indeed diagonal and full-

rank. Whereas including the sparsity-inducing penalty $\lambda r_x(0)$ and the constraint $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ can yield better estimates for (relatively small) finite M, this can prevent convergence of $\hat{\mathbf{r}}_x$ to the true $\mathbf{r}_x$ as $M \to \infty$, if λ is too large or if the true K-lag autocorrelation does not satisfy $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$.

Cramér-Rao Bound: Assuming that the true K-lag autocorrelation satisfies $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$, and no prior PS sparsity information, the CRB on the estimated $\hat{\mathbf{r}}_x$ 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 [29]. Defining $\mu_m := \log \Phi\left(\frac{b_m(\tilde{\alpha}_m - t_m)}{\sigma_m}\right)$, it can be shown that [22]

$$\mathbb{E}[\nabla^2 \mu_m] = \frac{e^{-\frac{(\tilde{\alpha}_m - t_m)^2}{\sigma_m^2}}}{2\pi\sigma_m^2}\left[\frac{1}{\Phi\left(\frac{\tilde{\alpha}_m - t_m}{\sigma_m}\right)} + \frac{1}{\Phi\left(\frac{t_m - \tilde{\alpha}_m}{\sigma_m}\right)}\right]\mathbf{q}_m\mathbf{q}_m^T \quad (13)$$

where the expectation is with respect to $\{b_m\}_{m=1}^M$. Hence, the Fisher information matrix (FIM) is computed as

$$\begin{aligned}\mathbf{J} &:= \mathbb{E}[\nabla^2 \log f(b_1, \ldots, b_M | \mathbf{r}_x)] \\ &= \sum_{m=1}^M \mathbb{E}[\nabla^2 \mu_m] = \mathbf{Q}^T \mathbf{D} \mathbf{Q}\end{aligned} \quad (14)$$

where $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_M]^T$ and

$$D_{m,m} = \frac{e^{-\frac{(\tilde{\alpha}_m - t_m)^2}{\sigma_m^2}}}{2\pi\sigma_m^2}\left[\frac{1}{\Phi\left(\frac{\tilde{\alpha}_m - t_m}{\sigma_m}\right)} + \frac{1}{\Phi\left(\frac{t_m - \tilde{\alpha}_m}{\sigma_m}\right)}\right] \quad (15)$$

is the m-th entry of the diagonal matrix $\mathbf{D}$. Finally, the CRB on $\hat{\mathbf{r}}_x$ is obtained as[1] $\mathrm{Trace}(\mathbf{J}^{-1})$.

If an upper bound on the number of nonzero frequency 'bins' (FFT points) in the PS is known *a-priori*, it can be exploited by adding the cardinality constraint $\|\mathbf{F}\mathbf{r}_x\|_0 \leq \kappa$, where the ℓ_0-(quasi)norm is the number of nonzero entries of $\mathbf{F}\mathbf{r}_x$. Note that this nonconvex ℓ_0-norm constraint can be relaxed by adding the sparsity-inducing penalty $\lambda r_x(0)$ to the cost function as in (12). Computing the CRB with a cardinality constraint has been considered in [31] for a linear model and extended in [22] for a nonlinear model. It is shown in [31] and [22] 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 $\|\mathbf{F}\mathbf{r}_x\|_0 = \kappa$, let the rows of the $(2K-1-\kappa) \times (2K-1)$ matrix $\breve{\mathbf{F}}$ correspond to the $(2K-1-\kappa)$ rows of the $(2K-1) \times (2K-1)$ DFT matrix $\mathbf{F}$ that satisfy $\breve{\mathbf{F}}\mathbf{r}_x = \mathbf{0}$, and let the $(2K-1) \times \kappa$ matrix $\mathbf{U}$ satisfy $\breve{\mathbf{F}}\mathbf{U} = \mathbf{0}$ and $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Thus, the sparsity-constrained CRB can be obtained as [31]

$$\mathrm{Trace}(\mathbf{U}(\mathbf{U}^T\mathbf{J}\mathbf{U})^{-1}\mathbf{U}^T). \quad (16)$$

[1]Assuming that $M \geq (2K-1)$ such that the FIM $\mathbf{J}$ is nonsingular. If $\mathbf{J}$ is singular, the pseudo-inverse can be used instead [30], in the sense that it will yield a valid (albeit generally optimistic) lower bound.

This gives the best achievable mean squared error obtained by estimators that have perfect knowledge of the support set of the PS to be estimated.

B. Dictionary Model PS Estimation

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

$$S_x(\omega) = \sum_{\ell=1}^L \rho_\ell \Psi_\ell(\omega) \quad (17)$$

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 [23]. Using the model (17), the PS can be reconstructed by estimating the weight vector $\boldsymbol{\rho} := [\rho_1, \ldots, \rho_L]^T$. If the center frequencies are unknown, the linear model (17) can also be used assuming an overcomplete dictionary of bases functions $\{\Psi_\ell(\omega)\}_{\ell=1}^L$. Note that the bandwidths of $\{\Psi_\ell(\omega)\}_{\ell=1}^L$ are allowed be different.

In general, the PS model (17) can be used to model the PS of L primary transmitters, where ρ_ℓ 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 [15]. In this case, the received signal at sensor $m \in \{1, \ldots, M\}$, sampled using a Nyquist-rate ADC, can be expressed as

$$y_m(n) = \sum_{\ell=1}^L h_m(\ell)\sqrt{\rho_\ell} x_\ell(n) \quad (18)$$

where $x_\ell(n)$ is the discrete-time WSS signal of the ℓ-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 ℓ and sensor m is already known (e.g., using training sequences), and is removed from the received signal.

From (17), the autocorrelation of the primary signal $x(n) = \sum_{\ell=1}^L \sqrt{\rho_\ell} x_\ell(n)$ can be expressed as

$$r_x(k) = \sum_{\ell=1}^L \rho_\ell \sum_{k=1-K}^{K-1} \psi_\ell(k) \quad (19)$$

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 (I-DTFT) of $\Psi_\ell(\omega)$ (also the k-th lag autocorrelation of $x_\ell(n)$). Defining $v_{m,\ell} := \sum_{k=1-K}^{K-1} \psi_\ell(k) q_m^*(k)$, it can be shown that

$$\tilde{\alpha}_m = \mathbb{E}[|\tilde{z}_m(n)|^2] = \sum_{\ell=1}^L \rho_\ell v_{m,\ell}, \quad (20)$$

$$\alpha_m = \mathbb{E}[|z_m(n)|^2] = \sum_{\ell=1}^L |h_m(\ell)|^2 \rho_\ell v_{m,\ell}. \quad (21)$$

Thus, the measurement error due to the fading channel

$$\tilde{e}_m = \alpha_m - \tilde{\alpha}_m = \sum_{\ell=1}^L \rho_\ell v_{m,\ell}(|h_m(\ell)|^2 - 1) \quad (22)$$

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. Hence, defining $\mathbf{v}_m := [v_{m,1}, \ldots, v_{m,L}]^T$, the received power measurement bit (9) can be written in this case as $b_m = \text{sign}(\mathbf{v}_m^T\boldsymbol{\rho} + e_m - t_m)$, where the goal here is to estimate $\boldsymbol{\rho}$ from $\{b_m\}_{m=1}^M$.

Note that $\boldsymbol{\rho}$ can be upper bounded by $\mathbf{p}_{\max}$ (due to the use of AGC at the front-end of the sensor processing chain). This yields the box constraint $\boldsymbol{\rho} \in \mathcal{B}$, where $\mathcal{B} := \{\boldsymbol{\rho} \in \mathbb{R}^L | \mathbf{0} \leq \boldsymbol{\rho} \leq \mathbf{p}_{\max}\}$. With accurate power measurements, the LP (8) can be modified to estimate $\boldsymbol{\rho}$ as

$$\min_{\rho\in\mathcal{B}} \sum_{\ell=1}^{L} \rho_\ell \quad \text{s.t. } b_m(\mathbf{v}_m^T\boldsymbol{\rho} - t_m) \geq 0,\ m = 1, \ldots, M. \quad (23)$$

Similarly, by exploiting the Gaussian distribution of $\{e_m\}_{m=1}^M$, the ML (12) is modified to

$$\max_{\rho\in\mathcal{B}} \sum_{m=1}^{M} \log\Phi\left(\frac{b_m(\mathbf{v}_m^T\boldsymbol{\rho} - t_m)}{\sigma_m}\right) - \lambda\sum_{\ell=1}^{L}\rho_\ell. \quad (24)$$

The CRB on the estimate of $\boldsymbol{\rho}$ can be obtained in a manner similar to Section III-B, by replacing the matrix $\mathbf{Q}$ with $\mathbf{V} := [\mathbf{v}_1, \ldots, \mathbf{v}_M]^T$ in the FIM expression (14). If prior information on the number of nonzero entries of the true $\boldsymbol{\rho}$ is available, i.e., $\|\boldsymbol{\rho}\|_0 = \kappa$, the constrained CRB is obtained using expression (16), where the $L \times \kappa$ matrix $\mathbf{U}$ in (16) 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 $\boldsymbol{\rho}$ [22].

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 [5]. 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 CDF, we know that $\Phi(u)$ is an increasing function in u and that $\log\Phi(|u|) \approx 0$, if $|u| > 4$. Hence, the log-likelihood function (11) 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 (15) that $D_{m,m} \approx 0$, and that the FIM (14) 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 $|\hat{\alpha}_m - t_m| \leq \zeta_m$, where $\zeta_m > 0$ is a censoring threshold that is known at the sensor. More insights on the choice of ζ_m are discussed in Section V-A.

IV. Active Sensing

The choice of $\{t_m\}_{m=1}^M$ is very important for achieving good estimates in (8) and (12). All approaches discussed in Section III are passive in the sense that the thresholds $\{t_m\}_{m=1}^M$ are fixed and pre-assigned to sensors. In [16], a common threshold $t_m = t, \forall m$, was chosen for all sensors, and empirically tuned to ensure that a certain percentage of sensors generate positive reports. Another appealing option is to use t that minimizes the average CRB across different signals and filter realizations. But what if the threshold 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.

A. Active Sensing With Accurate Power Measurements

We first consider the case of accurate power measurements, i.e., $\hat{\alpha}_m = \tilde{\alpha}_m$. The final feasible region for $\mathbf{r}_x$ when the FC receives all measurement bits $\{b_m\}_{m=1}^M$ is defined by the polyhedron

$$\mathcal{P}_M = \mathcal{P} \cap \{\mathbf{r}_x \mid b_m(\mathbf{q}_m^T\mathbf{r}_x - t_m) \geq 0,\ m = 1, \ldots, M\}. \quad (25)$$

The volume of the feasible region $\mathcal{P}_M$ gives a measure of ignorance or uncertainty about $\mathbf{r}_x \in \mathcal{P}_M$; a small $\mathcal{P}_M$ implies that $\mathbf{r}_x$ is localized to within a small neighborhood, whereas a large $\mathcal{P}_M$ means that there is still much uncertainty about $\mathbf{r}_x$. In other words, a smaller feasible region $\mathcal{P}_M$ translates to higher accuracy in localizing $\mathbf{r}_x$. 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.

Before introducing the proposed active sensing algorithm, we first discuss how to compute the *Chebyshev center* (CC) of a bounded polyhedron $\mathcal{C} := \{\mathbf{x} \mid \mathbf{a}_i^T\mathbf{x} \leq c_i,\ i = 1, \ldots, L\}$, 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 [27, Sec. 8.5.1]

$$\begin{aligned} \max_{R\geq 0, \mathbf{x}} \quad & R \\ \text{s.t.:} \quad & \mathbf{a}_i^T\mathbf{x} + R\|\mathbf{a}_i\|_2 \leq c_i,\ i = 1, \ldots, L. \end{aligned} \quad (26)$$

The LP (26) 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 (7), its CC $\mathbf{r}_{\text{CC}}^{(0)}$, 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 $t_m = \mathbf{q}_m^T\mathbf{r}_{\text{CC}}^{(m-1)}$, and send it to sensor m requesting its measurement bit b_m.

2) Upon receiving b_m, update the feasible polyhedron

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

3) Compute the CC $\mathbf{r}_{\text{CC}}^{(m)}$ of $\mathcal{P}_m$.

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

Note that at the second step of the active sensing algorithm, the half-space $\{\mathbf{r}_x \mid b_m\mathbf{q}_m^T(\mathbf{r}_x - \mathbf{r}_{\text{CC}}^{(m-1)}) \leq 0\}$ is cut-off from the feasible region. The selection of the threshold $t_m = \mathbf{q}_m^T\mathbf{r}_{\text{CC}}^{(m-1)}$ ensures that the CC of $\mathcal{P}_{m-1}$ is a point in the trimmed half-space. Ideally, one would want half the volume

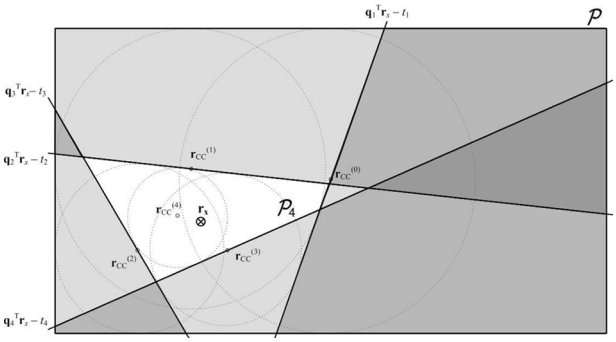


Fig. 2. Illustrative example for the active sensing algorithm.

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 [32]. 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 [33]–[35].

An illustrative example for the active sensing algorithm in $\mathbb{R}^2$ is shown in Fig. 2 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 $\mathbf{r}_x$ is localized in a small region, implying a small error in the estimate $\hat{\mathbf{r}}_x$.

Claim 2: The CC $\mathbf{r}_{\rm CC}^{(M)}$ converges linearly to the true autocorrelation vector $\mathbf{r}_x$ as $M \to \infty$ using the proposed active sensing algorithm, under certain independence conditions (see Appendix B) on $\{\mathbf{q}_m\}_{m=1}^M$.

The proof can be found in Appendix B.

Non-Negativity Constraints: As stated in Section III, 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 non-negative at all frequencies. Therefore, including $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ in the set of constraints when estimating $\mathbf{r}_x$ may prevent the convergence of the CC $\mathbf{r}_{\rm CC}^{(M)}$ to $\mathbf{r}_x$ as $M \to \infty$. Instead, we consider including a *relaxed* linear non-negativity constraint. Define the Vandermonde vector $\mathbf{u}(\omega) := [1, e^{j\omega}, \ldots, e^{j\omega(K-1)}]^T$. Any autocorrelation matrix $\mathbf{R}_x$ is positive semidefinite. In unknown autocorrelation vector $\mathbf{r}_x$, $\mathbf{R}_x \succeq 0$ 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

$$\begin{aligned}\mathbf{R}_x \succeq 0 \Longrightarrow & \mathbf{u}(\omega)^H \mathbf{R}_x \mathbf{u}(\omega) \geq 0 \\ \Longrightarrow & \sum_{k=-K+1}^{K-1} (K-|k|) r_x(k) e^{-j\omega k} \geq 0.\end{aligned}$$

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

Pruning Constraints: The number of linear inequalities defining $\mathcal{P}_m$ increases at each iteration of the algorithm, and hence the computational effort to compute $\mathbf{r}_{\rm CC}^{(m)}$ 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 [33], [34]. 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.

B. Low-Complexity Active Sensing Algorithm

Instead of using the CC of the polyhedron $\mathcal{P}_{m-1}$ in computing t_m, other options include the center of gravity, the center of the maximum volume inscribed ellipsoid, and the analytic center [34]. The worst-case complexity of each method can be captured by the worst-case number of iterations (i.e., M) required to achieve an ϵ-error estimate, $\|\hat{\mathbf{r}}_x - \mathbf{r}_x\|^2 \leq \epsilon$, 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 ϵ-error estimate is at most $O\left(K \log(1/\epsilon)\right)$. However, computing the center of gravity of a polyhedron described by a set of linear inequalities is NP-hard [34]. The number of iterations required using the center of the maximum volume inscribed ellipsoid is at most $O(K^2 \log(1/\epsilon))$ [34], and it can be computed by solving a convex problem [27], 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 [27, Sec. 8.5.3], while at most $O\left(K^2/\epsilon^2\right)$ iterations are required to achieve error $\leq \epsilon$ [34]. 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 (26).

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 (26) to compute the CC, at each iteration. The AC of a bounded polyhedron $\mathcal{C} = \{\mathbf{x} \mid \mathbf{a}_i^T \mathbf{x} \leq c_i,\ i = 1, \ldots, L\}$ is obtained by minimizing the logarithmic barrier function $\Omega(\mathbf{x}) := -\sum_{i=1}^L \log(c_i - \mathbf{a}_i^T \mathbf{x})$ [27, Sec. 8.5.3], which can be computed via Newton's method [27, Sec. 9.5]. In Newton's method, multiple damped Newton steps of type

$$\mathbf{x} = \mathbf{x} - \beta \left(\nabla^2 \Omega(\mathbf{x})\right)^{-1} \nabla \Omega(\mathbf{x}) \tag{27}$$

are used to find the minimizer of $\Omega(\mathbf{x})$, where

$$\nabla\Omega(\mathbf{x}) = \sum_{i=1}^{L} \frac{\mathbf{a}_i}{c_i - \mathbf{a}_i^T\mathbf{x}}, \quad \nabla^2\Omega(\mathbf{x}) = \sum_{i=1}^{L} \frac{\mathbf{a}_i\mathbf{a}_i^T}{(c_i - \mathbf{a}_i^T\mathbf{x})^2} \quad (28)$$

and $\beta \geq 0$ is the step-size. The most computationally expensive operation in each Newton step is inverting the $(2K-1)\times(2K-1)$ Hessian matrix $\nabla^2\Omega(\mathbf{x})$. Therefore, an *approximate* AC can be computed using one (or few) damped Newton step(s) of type (27). Starting from $\mathbf{r}_{\rm AC}^{(0)}$, 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 in Section IV-A. The linear convergence $\mathbf{r}_{\rm AC}^{(M)} \rightarrow \mathbf{r}_x$ 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 V-B. It is worth mentioning that linear convergence using approximate ACs, in the context of cutting plane methods for convex feasibility problems, was proven in [35] under certain conditions.

C. Active Gaussian ML Sensing

We now develop active ML sensing for Gaussian errors, i.e., $e_m = \hat{\alpha}_m - \tilde{\alpha}_m$ is a zero-mean Gaussian random variable with variance σ_m^2. Unlike the situation in the previous scenario, the feasible region does not decrease after receiving a measurement bit due to the uncertainty induced by 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 $\tilde{\alpha}_m = \mathbf{q}_m^T\mathbf{r}_x$. Therefore, after the FC receives the bits $b_1,\ldots,b_m$, it sets $t_{m+1} = \mathbf{q}_{m+1}^T\mathbf{r}_{\rm ML}^{(m)}$ for sensor $m+1$, where $\mathbf{r}_{\rm ML}^{(m)}$ is the ML estimate of $\mathbf{r}_x$ given the first m measurement bits, which is obtained by solving

$$\mathbf{r}_{\rm ML}^{(m)} = \arg \max_{\mathbf{r}_x\in\mathcal{P}, \mathbf{F}\mathbf{P}\mathbf{r}_x\geq\mathbf{0}} \sum_{i=1}^{m} \log\Phi\left(\frac{b_i(\mathbf{q}_i^T\mathbf{r}_x - t_i)}{\sigma_i}\right) \quad (29)$$

and $\mathbf{r}_{\rm ML}^{(0)}$ can be chosen any point inside $\mathcal{P}$. However, solving (29) exactly to compute t_m for each sensor is computationally expensive.

Similar to the low complexity algorithm of Section IV-B, an approximate solution to (29) can be obtained using Newton's method. Defining $\Gamma_m(\mathbf{x}) := -\sum_{i=1}^{m}\log\Phi\left(\frac{b_i(\mathbf{q}_i^T\mathbf{x}-t_i)}{\sigma_i}\right)$, it can be shown that

$$\nabla\Gamma_m(\mathbf{x}) = \sum_{i=1}^{m} \frac{b_i\, e^{-\frac{(\mathbf{q}_i^T\mathbf{x}-t_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi\sigma_i^2}\,\Phi\left(\frac{b_i(\mathbf{q}_i^T\mathbf{x}-t_i)}{\sigma_i}\right)}\,\mathbf{q}_i \quad (30)$$

$$\nabla^2\Gamma_m(\mathbf{x}) = -\sum_{i=1}^{m}\left[\frac{e^{-\frac{(\mathbf{q}_i^T\mathbf{x}-t_i)^2}{\sigma_i^2}}}{2\pi\sigma_i^2\left[\Phi\left(\frac{b_i(\mathbf{q}_i^T\mathbf{x}-t_i)}{\sigma_i}\right)\right]^2} + \frac{\frac{b_i(\mathbf{q}_i^T\mathbf{x}-t_i)}{\sigma_i}\, e^{-\frac{(\mathbf{q}_i^T\mathbf{x}-t_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi}\sigma_i^2\,\Phi\left(\frac{b_i(\mathbf{q}_i^T\mathbf{x}-t_i)}{\sigma_i}\right)}\right]\mathbf{q}_i\mathbf{q}_i^T. \quad (31)$$

Thus, starting from any point $\mathbf{r}_{\rm AML}^{(0)}$ 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 (27), as

$$\mathbf{r}_{\rm AML}^{(m)} = \mathbf{r}_{\rm AML}^{(m-1)} - \beta_m\left(\nabla^2\Gamma_m(\mathbf{r}_{\rm AML}^{(m-1)})\right)^{-1}\nabla\Gamma_m(\mathbf{r}_{\rm AML}^{(m-1)}). \quad (32)$$

Convergence of $\mathbf{r}_{\rm AML}^{(M)} \rightarrow \mathbf{r}_x$ as $M \rightarrow \infty$ using a single Newton step approximate ML is shown in the simulations in Section V-B. Finally, it is worth noting 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.

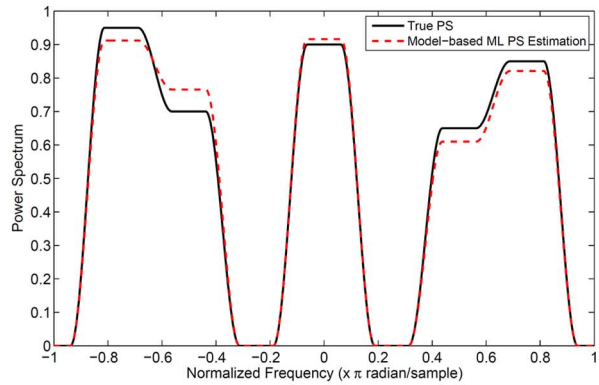
V. Numerical Results

In the next subsection, we first consider the passive sensing case with fixed thresholds $t_m = t$ for all m, followed by the active sensing case with adaptive thresholds.

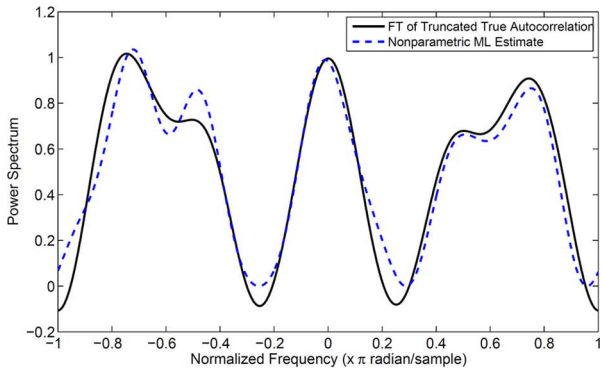
A. Passive Sensing: Fixed Thresholds

We begin with a simulation in Fig. 3 that illustrates what one can expect from the proposed nonparametric ML estimation using (12) and the model-based ML estimation using (24). 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(a), whereas the Fourier transform (FT) of the true truncated $(K = 10)$-lag autocorrelation is plotted with a solid line in Fig. 3(b). The estimated model-based PS obtained using (24) (with $\lambda = 50$) by exploiting the information of the raised-cosine model is plotted with a dashed line in Fig. 3(a), whereas the estimated PS obtained by estimating the $(K = 10)$-lag autocorrelation nonparametrically using (24) (with $\lambda = 0$) is plotted with a dashed line in Fig. 3(b). The quality of the estimates in Figs. 3(b) and 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.

In Fig. 4, the mean squared error (MSE) of the estimated $(K = 10)$-lag autocorrelation using the LP (8) and the ML (12) (with $\lambda = 0$), and the CRB, are plotted versus the number of sensors M. The true primary signal is comprised of 10 equispaced 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 (8) fails to provide a meaningful estimate, due to the flipped bits. The figure also shows that the



(a)



(b)

Fig. 3. ML PS estimation example. (a) Model-Based ML estimation. (b) Nonparametric ML estimation.

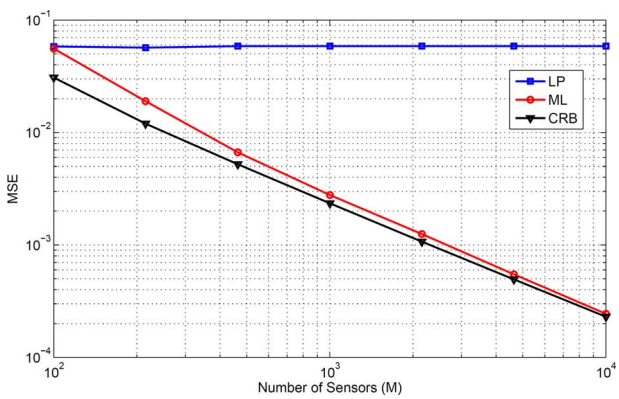


Fig. 4. Nonparametric PS: MSE of the LP estimate, MSE of the ML estimate, and the CRB versus M.

MSE of the ML estimate is asymptotically converging to the CRB as M grows.

To test the performance of the proposed model-based PS estimation techniques in Section III-C, we assume that the primary signal PS model is the combination of $L = 20$ identical raised cosine functions, each with $\text{roll-off factor} = 0.5$ and $\text{bandwidth} = 3\pi/L$. The corresponding center frequencies are $\{\omega_\ell = 2\pi(\ell-1)/L\}_{\ell=1}^{L}$, implying equispaced and overlapping functions. A sparse vector $\boldsymbol{\rho}$ 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 $\text{MSE}_\rho := E[\|\boldsymbol{\rho}-\hat{\boldsymbol{\rho}}\|^2]$, to measure the performance of the proposed estimation techniques, where the expectation is taken

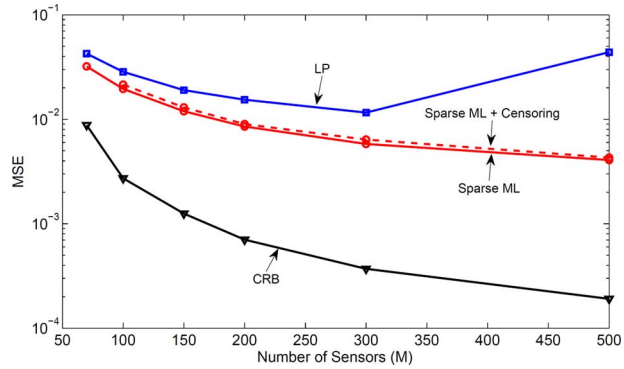


Fig. 5. Model-based PS: MSE of the LP estimate, MSE of the ML estimate, and the (oracle) CRB versus M for large SER.

with respect to the random impulse responses of the FIR filters, the random Gaussian error samples, and the random weights vector $\boldsymbol{\rho}$, 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 MSE_ρ 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 (16) across different $\boldsymbol{\rho}$ and filter realizations. The sparsity tuning parameter in (24) 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 $\boldsymbol{\rho}$ obtained using (23) as the *LP estimate*, whereas the estimate obtained using (24) is named the *ML estimate*. The MSE of the LP estimate, the MSE of the ML estimate, and the (oracle) CRB computed using (16), are plotted versus the number of sensors M in Figs. 5 and 6, for $\text{SER} = 5\times 10^4$ and $\text{SER} = 500$, respectively.

In Fig. 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, $\text{sign}(\hat{\alpha}_m - t_m) \neq \text{sign}(\hat{\alpha}_m - t_m)$ for $0.02M$ 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 $\boldsymbol{\rho}$. 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 (23) 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 III-B is also considered in Fig. 5. The dashed line in the figure shows the MSE of the ML estimate when the censoring threshold ζ was selected such that only the *best* 80 out of M sensors are active in each simulation run (i.e., $|\hat{\alpha}_m - t| \leq \zeta$ 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. 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

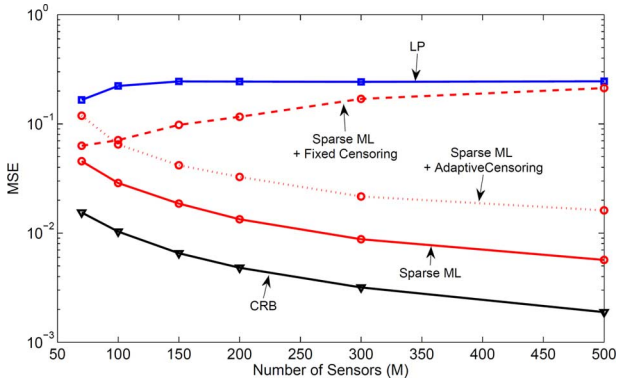


Fig. 6. Model-based PS: MSE of the LP estimate, MSE of the ML estimate, and the (oracle) CRB, versus M for small SER.

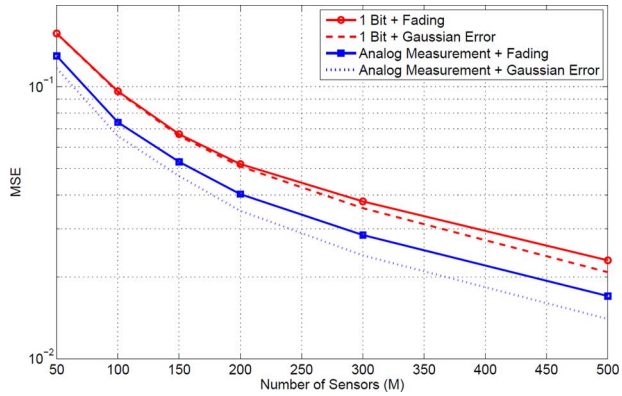


Fig. 7. MSE of the ML estimate and MSE of benchmark estimate for (a) complex Gaussian fading coefficients and (b) Gaussian errors.

values than their counterparts in Fig. 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. 5 to Fig. 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 (8). The MSE of the ML estimate when employing the same *fixed censoring* strategy as in Fig. 5, represented by the dashed line in Fig. 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 ζ 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. 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.

In Fig. 7, we consider the same signal model and sensor settings as Figs. 5 and 6, and compare between the ML estimate obtained using (24) (with $\lambda = 0$) and the *benchmark estimate* that is obtained by solving the (box-constrained) least squares: $\min_{\rho\in\mathcal{B}} \|\boldsymbol{\alpha} - \mathbf{V}\boldsymbol{\rho}\|_2^2$, where $\boldsymbol{\alpha} := [\hat{\alpha}_1, \ldots, \hat{\alpha}_M]^T$. 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 $\boldsymbol{\rho}$ 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.

B. Active Sensing

We now switch to testing the performance of the proposed active sensing algorithms. To measure the quality of the estimated autocorrelation $\hat{\mathbf{r}}_x$, we use the normalized mean squared error (NMSE), defined as $\mathrm{NMSE}(m) := \mathbb{E}\left[\frac{\|\mathbf{r}_x - \hat{\mathbf{r}}_x^{(m)}\|^2}{\|\mathbf{r}_x\|^2}\right]$, 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 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. 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. Active sensing algorithms for the case of accurate power measurements (i.e., $\mathrm{sign}(\tilde{\alpha}_m - t_m) = \mathrm{sign}(\hat{\alpha}_m - t_m)\ \forall m$) are considered in Fig. 8, whereas active Gaussian ML sensing algorithms are considered in Fig. 9.

In Fig. 8, 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 $\mathrm{NMSE}(m) \to 0$ (i.e., $\hat{\mathbf{r}}_x^{(m)} \to \mathbf{r}_x$) as m increases for all considered algorithms, when using the relaxed linear non-negativity constraint $\mathbf{F}\mathbf{P}\mathbf{r}_x \geq \mathbf{0}$, 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 IV-B (from 1 to 5 to 10 Newton steps

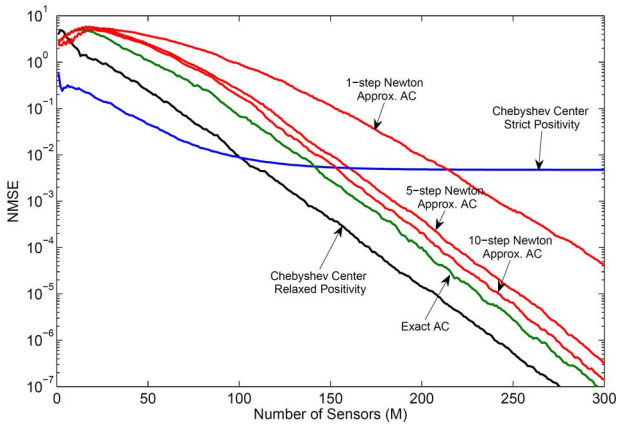


Fig. 8. Active sensing with accurate power measurements.

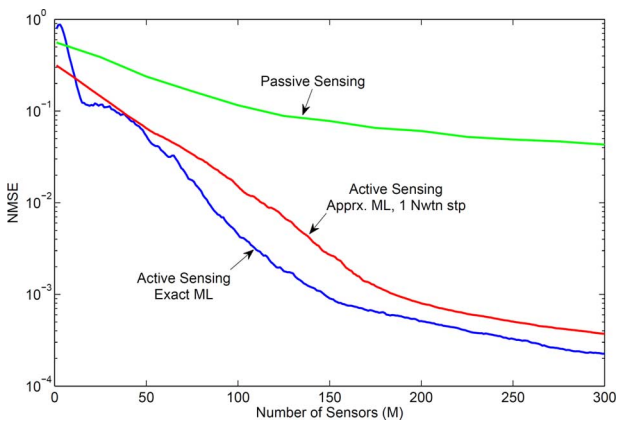


Fig. 9. Active sensing with Gaussian errors.

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 $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ was included, the NMSE of the active sensing algorithm was better than that obtained with the relaxed non-negativity constraint $\mathbf{F}\mathbf{P}\mathbf{r}_x \geq \mathbf{0}$ 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 $\mathbf{F}\mathbf{r}_x \geq \mathbf{0}$ is not valid in general (the FT of a truncated autocorrelation sequence is not necessarily non-negative), underscoring that including this strong constaint is only advisable for up to moderate M; for higher M it should be omitted to enable convergence.

In Fig. 9, the NMSE is plotted as a function of m for the active sensing algorithm with the exact ML estimate (29), the active sensing algorithm with the approximate ML estimate obtained using the single-step Newton (32), 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.

VI. Conclusions

Frugal sensing was revisited from a statistical estimation point of view, taking into account the effects of fading and insufficient sample averaging on the soft power measurements prior to quantization. We showed that the distribution of the corresponding error is approximately Gaussian, and exploited this result 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 passive ML sensing from few bits, even when relatively many bits are flipped due to fading-induced measurement errors. Assuming availability of a 'downlink' channel that the FC can use to send threshold information, we developed active sensing strategies that adaptively select the thresholds online, yielding significantly faster convergence to the true finite-length autocorrelation compared to passive sensing.

Our methods can be extended to handle the case where each sensor sends multiple bits to the fusion center. One way to do this is to take different pieces of the sensor's PN sequence to create multiple measurement filter impulse responses taking uncorrelated measurements. This way, each multi-bit sensor will do the job of multiple single-bit sensors. Alternatively, multi-bit quantization of the output of a single PN filter can be used, but note that our experiments in Fig. 7 suggest that for moderately large M the gap between 1-bit and analog measurements is relatively small.

Appendix

A. Sensor Measurement Error due to Fading

Here we show that for large number of channel taps T, the errors $\{\tilde{e}_m\}_{m=1}^M$ due to the fading channel can be approximated as i.i.d. zero-mean Gaussian random variables. Let $S_{z_m}(\omega)$ denote the PS of the WSS signal $z_m(n)$, $H_m(\omega) := \sum_{\ell=0}^{T-1} h_m(\ell)e^{-j\omega\ell}$ 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

$$\begin{aligned}
\alpha_m &= \mathbb{E}[|z_m(n)|^2] = r_{z_m}(0) \\
&= \frac{1}{2\pi}\int_0^{2\pi} S_{z_m}(\omega)d\omega \\
&\approx \frac{1}{T}\sum_{k=0}^{T-1} S_{z_m}\left(\frac{2\pi k}{T}\right) \\
&= \frac{1}{T}\sum_{k=0}^{T-1} \left|H_m\left(\frac{2\pi k}{T}\right)\right|^2 \left|G_m\left(\frac{2\pi k}{T}\right)\right|^2 S_x\left(\frac{2\pi k}{T}\right) \\
&= \frac{1}{T}\sum_{k=0}^{T-1} \left|\tilde{H}_m(k)\right|^2 \left|\tilde{G}_m(k)\right|^2 \tilde{S}_x(k) \qquad (33)
\end{aligned}$$

where $\tilde{H}_m(k)$ and $\tilde{G}_m(k)$ correspond to the T-point DFT of $h_m(n)$ and $g_m(n)$, respectively, and $\tilde{S}_x(k) = S_x\left(\frac{2\pi k}{T}\right)$. Note that the expectation used to compute α_m in (33) is obtained given a single realization of $g_m(n)$ and $h_m(n)$, i.e., the expectation is taken with respect to the random primary signal $x(n)$ only. The approximation is accurate for large T, $r_x(\ell) \approx 0$ for large ℓ (implying slowly varying $S_x(\omega)$ with ω), and $\tilde{S}_x(k) > 0$ for large number of samples. Similarly, it is easy to see that

$$\tilde{\alpha}_m \approx \frac{1}{T}\sum_{k=0}^{T-1}\left|\tilde{G}_m(k)\right|^2 \tilde{S}_x(k). \tag{34}$$

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

$$\begin{aligned}\tilde{e}_m &= \alpha_m - \tilde{\alpha}_m \\ &\approx \frac{1}{T}\sum_{k=0}^{T-1}\left|\tilde{G}_m(k)\right|^2 \tilde{S}_x(k)\left(\left|\tilde{H}_m(k)\right|^2 - 1\right) \\ &= \frac{1}{T}\sum_{k=0}^{T-1} a_m(k)\delta_m(k)\end{aligned} \tag{35}$$

where $a_m(k) := |\tilde{G}_m(k)|^2\tilde{S}_x(k) \geq 0$ and $\delta_m(k) := |\tilde{H}_m(k)|^2 - 1$. Since $\{h_m(\ell)\}_{\ell=0}^{T-1}$ are assumed i.i.d. Gaussian random variables with zero-mean and variance $1/T$, then the T-point DFT, $\{\tilde{H}_m(k)\}_{k=0}^{T-1}$, are i.i.d. Gaussian random variables with zero-mean and unit-variance. The random variable $|\tilde{H}_m(k)|^2$ corresponds to a sum of squares of 2 independent zero-mean and $1/2$-variance Gaussian random variables, which yields a unit-mean and unit-variance exponentially distributed random variable (i.e., distribution parameter $\lambda = 1$). Thus, the error $\tilde{e}_m$ in (35) is approximated as a (weighted) sum of T independent random variables $\{\delta_m(k)\}_{k=0}^{T-1}$. Hence $\mathbb{E}[\delta_m(k)] = \mathbb{E}[|\tilde{H}_m(k)|^2] - 1 = 0$ and $\mathbb{E}[\delta_m^2(k)] = \mathbb{E}[|\tilde{H}_m(k)|^4] - 2\,\mathbb{E}[|\tilde{H}_m(k)|^2] + 1 = 1$, which implies that $\mathbb{E}[\tilde{e}_m] = 0$ and $\mathrm{var}(\tilde{e}_m) = \frac{1}{T^2}\sum_{k=0}^{T-1} a_m^2(k) \approx \frac{1}{T}\frac{1}{2\pi}\int_0^{2\pi}(|G_m(\omega)|^2 S_x(\omega))^2 d\omega$. Since the integral is independent of T, this means that $\mathrm{var}(\tilde{e}_m) \sim 1/T$ (and is independent of K).

Lyapunov Central Limit Theorem: Define $X_k = \frac{1}{T}a_m(k)\delta_m(k)$ and $s_T^2 := \sum_{k=0}^{T-1}\mathrm{var}(X_k) = \frac{1}{T^2}\sum_{k=0}^{T-1} a_m^2(k)$. The Lyapunov central limit theorem [25, 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 $\lim_{T\to\infty}\frac{1}{s_T^{2+\zeta}}\sum_{k=1}^{T}\mathbb{E}[|X_k|^{2+\zeta}] = 0$ is satisfied, then $\frac{1}{s_T}\sum_{k=1}^{T} X_k \xrightarrow{d} \mathcal{N}(0,1)$. To verify that the Lyapunov condition is satisfied in our case, set $\zeta = 1$. It is easy to see that $\mathbb{E}[\delta_m^3(k)] = 2$ and $\mathbb{E}[X_k^3] = 2\frac{a_m^3(k)}{T^3}$. Without loss of generality, we assume $a_m(k) > 0$, $\forall k$ (ignoring zero values). Define $a_{\max} := \max_k a_m(k)$ and $a_{\min} := \min_k a_m(k) > 0$. Hence, it is easy to see that $\frac{\sum_{k=1}^{T} a_m^3(k)}{\left(\sum_{k=0}^{T-1} a_m^2(k)\right)^{3/2}} \leq \frac{Ta_{\max}^3}{(Ta_{\min}^2)^{3/2}}$. Since $\lim_{T\to\infty}\frac{a_{\max}^3}{T^{1/2}a_{\min}^3} = 0$, then we have shown that the Lyapunov condition is satisfied. Therefore, according to the *Lyapunov central limit theorem*, $\left(T/\sqrt{\sum_{k=0}^{T-1} a_m^2(k)}\right)\tilde{e}_m$ converges in distribution to a zero-mean and unit-variance Gaussian distribution as T increases.

B. Proof of Claim 2

Let $R^{(m)}$ denote the radius of the largest ball centered at $\mathbf{x}^{(m)}$ that lies inside $\mathcal{P}_m$, which is the solution to the LP (26). The convergence of the sequence $\{R^{(m)}\}_{m=1}^{\infty}$ to zero has been established by Theorem 1 in [33]; we sketch the proof here to relate associated non-degeneracy conditions to our context. 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 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\| \geq R^{(k)}, \forall k > m$. 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 $\{\mathbf{x}^{(m)}\}_{m=1}^{\infty}$ in a 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 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\| < \epsilon$, for $m, k \geq N$. Now, suppose $\bar{R} > 0$. This means that $\|\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\| \geq R^{(k)} \geq \bar{R}$, and for $0 < \epsilon < \bar{R}$, there does not exist N such that $\|\mathbf{x}^{(k)} - \mathbf{x}^{(m)}\| < \epsilon$ for $m, k \geq N$, which contradicts that $\{\mathbf{x}^{(m)}\}_{m=1}^{\infty}$ must have a Cauchy subsequence. Therefore, $\bar{R} = 0$. It is also shown in [33], 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 $\mathbf{x}^{(M)}$ converges to the single point $\mathbf{r}_x$.

Consider the set of linear inequalities defining the bounded polyhedron $\mathcal{P}_M = \left\{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{a}_m^T\mathbf{x} \leq c_m,\ m = 1, \ldots, M\right\}$, where $n := 2K - 1$, $\mathbf{a}_m := -b_m\mathbf{q}_m$, and $c_m := -b_m t_m$ (ignoring the initial $\mathcal{P}$ without loss of generality). An inequality $\mathbf{a}_m^T\mathbf{x} \leq c_m$ is *redundant* if it can be omitted without changing $\mathcal{P}_M$, whereas if $\mathcal{P}_M$ changes by removing an inequality $\mathbf{a}_m^T\mathbf{x} \leq c_m$, then we denote this inequality as *active*. Define the set $\mathcal{S}_M$ as the set of indices that correspond to *active* inequalities defining $\mathcal{P}_M$, such that $m \in \mathcal{S}_M$ implies that $\mathbf{a}_m^T\mathbf{x} \leq c_m$ is *active*, and $\mathcal{P}_M = \left\{\mathbf{x} \,|\, \mathbf{a}_m^T\mathbf{x} \leq c_m,\ m \in \mathcal{S}_M\right\}$. Note that $n + 1 \leq |\mathcal{S}_M| \leq M$, 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 $\{\mathbf{a}_m\}_{m\in\mathcal{S}_M}$ 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 $\mathbf{a}_m^T\mathbf{x} = c_m$, for $m \in \mathcal{S}_M$. Since the set of constraints are consistent, any n vectors of $\{\mathbf{a}_m\}_{m\in\mathcal{S}_M}$ are linearly independent, $|\mathcal{S}_M| \geq n + 1$, and $\mathbf{r}_x \in \mathcal{P}_M$, then the unique solution to $\mathbf{a}_m^T\mathbf{x} = c_m$, for $m \in \mathcal{S}_M$, is $\mathbf{x} = \mathbf{r}_x$. The linear independence condition for $\{\mathbf{a}_m\}_{m\in\mathcal{S}_M}$ is guaranteed with high probability if the random vectors $\{\mathbf{q}_m\}_{m=1}^{M}$ are drawn from a discrete distribution with large K (probability increases with K), and with probability one if $\{\mathbf{q}_m\}_{m=1}^{M}$ are drawn from a continuous distribution.

REFERENCES

[1] O. Mehanna, N. D. Sidiropoulos, and E. Tsakonas, "Model-based power spectrum sensing from a few bits," presented at the 21st Eur. Signal Process. Conf., Marrakech, Morroco, Sep. 9–13, 2013.

[2] O. Mehanna and N. D. Sidiropoulos, "Adaptive thresholding for distributed power spec- trum sensing," presented at the 38th Int. Conf. Acoust., Speech, Signal Process., Vancouver, Canada, May 26–31, 2013.

[3] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

[4] Q. Zhao and A. Swami, "A survey of dynamic spectrum access: signal processing and networking perspectives," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, vol. 4, pp. 1349–1352.
[5] E. J. Msechu and G. B. Giannakis, "Sensor-centric data reduction for estimation with WSNs via censoring and quantization," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 400–414, Jan. 2012.
[6] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[7] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, vol. 4, pp. 1357–1360.
[8] J. A. Bazerque and G. B. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1847–1862, Mar. 2010.
[9] F. Zeng, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, Feb. 2011.
[10] D. D. Arianandа and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.
[11] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wideband spectrum sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process, (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2337–2340.
[12] D. D. Arianandа and G. Leus, "Power spectrum blind sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 443–446, Aug. 2011.
[13] M. Lexa, M. Davies, J. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3884–3887.
[14] V. Havary-Nassab, S. Hassan, and S. Valaee, "Compressive detection for wideband spectrum sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Dallas, TX, USA, Mar. 2010, pp. 3094–3097.
[15] D. Romero, G. Leus, and R. Lopez-Valcarce, "Compressive wideband spectrum sensing with spectral prior information," presented at the 38th Int. Conf. Acoust., Speech, Signal Process. (ICASSP), Vancouver, Canada, May 26–31, 2013.
[16] O. Mehanna and N. D. Sidiropoulos, "Frugal sensing: wideband power spectrum sensing from few bits," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2693–2703, May 2013.
[17] P. Boufounos and R. Baraniuk, "1-bit compressive sensing," presented at the Conf. Inf. Sci. Syst. (CISS), Princeton, NJ, USA, Mar. 2008.
[18] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, Apr. 2013.
[19] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Commun. Pure Appl. Math., doi: 10.1002/cpa.21442*.
[20] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, "Exponential decay of reconstruction error from binary measurements of sparse signals," *ArXiv Preprint*, Jul. 2014 [Online]. Available: arxiv.org/abs/1407.8246
[21] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 149–152, Feb. 2010.
[22] E. Tsakonas, J. Jalden, N. D. Sidiropoulos, and B. Ottersten, "Sparse conjoint analysis through maximum likelihood estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5704–5715, Nov. 2013.
[23] J. A. Bazerque, G. Mateos, and G. B. Giannakis, "Group-lasso on splines for spectrum cartography," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4648–4663, Oct. 2011.
[24] U. Kamilov, A. Bourquard, A. Amini, and M. Unser, "One-bit measurements with adaptive thresholds," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 607–610, Oct. 2012.
[25] P. Billingsley, *Probability and Measure*, 2nd ed. New York, NY, USA: Wiley, 1986.
[26] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.
[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[28] W. Newey and D. McFadden, "Chapter 35: Large sample estimation and hypothesis testing," in *Handbook of Econometrics*. New York, NY, USA: Elsevier Science, vol. 4, pp. 2111–2245.
[29] J. D. Gorman and A. O. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 1285–1301, Nov. 1990.
[30] Z. Ben-Haim and Y. C. Eldar, "On the constrained Cramér-Rao bound with a singular Fisher information matrix," *IEEE Signal Process. Lett.*, vol. 16, no. 6, pp. 453–456, Jun. 2009.
[31] Z. Ben-Haim and Y. C. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.
[32] M. Dyer and A. Frieze, "On the complexity of computing the volume of a polyhedron," *SIAM J. Comput.*, vol. 17, no. 5, pp. 967–974, 1988.
[33] J. Elzinga and T. G. Moore, "A central cutting plane algorithm for the convex programming problem," *Math. Programm. Stud.*, vol. 8, pp. 134–145, 1975.
[34] S. Boyd and L. Vandenberghe, "Localization and cutting-plane methods," EE364b Lecture Notes Stanford Univ., Stanford, CA, USA, 2007.
[35] J. L. Goffin, Z. Q. Luo, and Y. Ye, "Complexity analysis of an interior cutting plane method for convex feasibility problems," *SIAM J. Optim.*, vol. 6, pp. 638–652, 1996.

Omar Mehanna (S'05) received the B.Sc. degree in Electrical Engineering from Alexandria University, Egypt, in 2006, the M.Sc. degree in Electrical Engineering from Nile University, Egypt, in 2009, and the Ph.D. degree in Electrical Engineering from University of Minnesota in 2014. He is currently a Senior Systems Engineer at Qualcomm Technologies, Inc. His current research interests are in cognitive radio and coordinated multi-point communications.

Nicholas D. Sidiropoulos (F'09) received the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Greece, and M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland—College Park, in 1988, 1990 and 1992, respectively. He served as Assistant Professor at the University of Virginia (1997–1999); Associate Professor at the University of Minnesota—Minneapolis (2000–2002); Professor at the Technical University of Crete, Greece (2002–2011); and Professor at the University of Minnesota—Minneapolis (2011-). His current research focuses primarily on signal and tensor analytics, with applications in cognitive radio, big data, and preference measurement. He received the NSF/CAREER award (1998), the IEEE Signal Processing Society (SPS) Best Paper Award (2001, 2007, 2011), and the IEEE SPS Meritorious Service Award (2010). He has served as IEEE SPS Distinguished Lecturer (2008–2009), and Chair of the IEEE Signal Processing for Communications and Networking Technical Committee (2007–2008). He received the Distinguished Alumni Award of the Department of Electrical and Computer Engineering, University of Maryland, College Park (2013).