Quantized Spectral Compressed Sensing: Cramer–Rao Bounds and Recovery Algorithms

Haoyu Fu [iD], *Student Member, IEEE*, and Yuejie Chi [iD], *Senior Member, IEEE*

Abstract—**Efficient estimation of wideband spectrum is of great importance for applications such as cognitive radio. Recently, sub-Nyquist sampling schemes based on compressed sensing have been proposed to greatly reduce the sampling rate. However, the important issue of quantization has not been fully addressed, particularly, for high resolution spectrum and parameter estimation. In this paper, we aim to recover spectrally sparse signals and the corresponding parameters, such as frequency and amplitudes, from heavy quantizations of their noisy complex-valued random linear measurements, e.g., only the quadrant information. We first characterize the Cramér–Rao bound under Gaussian noise, which highlights the trade-off between sample complexity and bit depth under different signal-to-noise ratios for a fixed budget of bits. Next, we propose a new algorithm based on atomic norm soft thresholding for signal recovery, which is equivalent to proximal mapping of properly designed surrogate signals with respect to the atomic norm that motivates spectral sparsity. The proposed algorithm can be applied to both the single measurement vector case, as well as the multiple measurement vector case. It is shown that under the Gaussian measurement model, the spectral signals can be reconstructed accurately with high probability, as soon as the number of quantized measurements exceeds the order of $K \log n$, where K is the level of spectral sparsity and n is the signal dimension. Finally, numerical simulations are provided to validate the proposed approaches.**

Index Terms—**Line spectrum estimation, quantization, Cramér–Rao bound, atomic norms, compressed sensing, multiple measurement vectors.**

I. INTRODUCTION

EMERGING applications in wireless communications, cognitive radio, and radar systems deal with signals of wideband or ultrawideband [1]. Spectrum sensing or signal acquisition in this regime is a fundamental challenge in signal processing, since the well-known Shannon-Nyquist sampling rate may become prohibitively high in practice. Therefore, it

H. Fu is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: fu.436@osu.edu).

Y. Chi is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15217 USA (e-mail: yuejie.chi@cmu.edu).

is highly desirable to come up with alternative approaches that have less demanding sampling requirements.

Recently, compressed sensing (CS) [2], [3] has emerged as an effective approach to allow sub-Nyquist sampling [4]–[6] when the wideband signal is approximately sparse in the spectral domain. The resulting paradigm is referred to as *Compressive Spectrum Sensing* [7], [8]. Significant focus has been put on reducing the sampling rates of the analog-to-digital converters (ADC), which only covers one aspect of the operations of ADCs. Quantization, which maps the analog samples into a finite number of bits for digital processing, is another necessary step that requires careful treatments. Most existing works, with a few exceptions, assume that the samples are quantized at a high bit level so that the quantization error is relatively small and well-behaved.

This paper aims at understanding the fundamental limits of quantization, as well as developing computationally efficient algorithms, for compressive spectrum sensing and parameter estimation, in particular in the regime of *heavy* quantization where it is no longer appropriate to model quantization errors as bounded additive noise. Examining the figure-of-merit of ADCs, two key specifications are the sampling rate and the effective number of bits (ENOB), which is the number of bits per measurement, also known as the *bit depth*. Typically, a small bit depth allows a high sampling rate, and vice versa [9]. Therefore, it is critical to understand the fundamental trade-off between sampling rate and bit depth for high-resolution spectrum estimation. Though the importance of understanding such trade-off has been realized in the context of CS [10], [11], they haven't been studied for the task of *parameter estimation* using estimation-theoretic tools.

Another motivating application is wideband spectrum sensing in bandwidth-constrained wireless networks [12], [13]. In order to reduce the communication overhead, each sensor transmits quantized messages, e.g., 1-bit messages; and it is necessary to estimate wideband spectrum from quantized measurements at the fusion center. Moreover, the quantization scheme might be unknown, due to lack of the knowledge of noise statistics or privacy constraints. Therefore, it is necessary to develop estimators that do not require exact knowledge of the quantizers.

A. Our Contributions

We study high-resolution spectrum estimation of sparse bandlimited signals from quantizations of their noisy random linear measurements. The signals of interest are spectrally sparse,

which contain a linear superposition of complex sinusoids with continuous-valued frequencies in the unit interval. In the extreme 1-bit case[1], the quantization is based on the quadrants of the complex-valued measurements. More generally, sophisticated quantization schemes such as Lloyd's quantizer [14] can be used to allow a higher bit depth. The specific form of the quantizer can be either known or unknown. In addition, the quantized measurements may be additionally contaminated by a noise model to be described later, in order to model imperfections in the quantization.

In this paper, we first derive the Cramér-Rao bound (CRB) for estimating multiple frequencies and their complex amplitudes assuming additive white Gaussian noise (AWGN) and the Lloyd's quantizer, using a fixed and deterministic CS measurement matrix. Our bounds suggest that the CRB experiences a phase transition depending on the signal-to-noise ratio (SNR) *before* quantization. In the low SNR regime it is *noise-limited*, and behaves similarly as if there was no quantization; in the high SNR regime, it is *quantization-limited*, and experiences severe performance degeneration due to quantization. Furthermore, we use the derived CRB to answer the following question: given the same budget of bits, should we use more measurements (high sample complexity) with low bit-depth, or fewer measurements (low sample complexity) with high bit-depth? We answer this question by comparing 1-bit versus 2-bit quantization schemes using the CRB, and demonstrate the answer depends on the SNR. At low SNR, 1-bit measurements are preferred, while at high SNR, 2-bit measurements are preferred.

It is well-known that maximum likelihood estimators approach the performance of CRB asymptotically at high SNR [15], however, their implementation requires exact knowledge of the likelihood function, which in our problem, includes the exact form of the quantizer and noise statistics. However, such knowledge may not be available in certain applications. Therefore, our goal is to develop estimators that do not require the knowledge of the quantization scheme. To mitigate basis mismatch [16], atomic norm [17]–[24] has been proposed recently to promote spectral sparsity via convex optimization without discretizing the frequencies onto a finite grid, which has found applications in signal denoising, interpolation of missing data, and frequency localization of spectrally-sparse signals. Existing atomic norm minimization algorithms assume unquantized measurements that are possibly contaminated by additive noise, and a direct application will lead to highly sub-optimal performance when a significant amount of the quantized measurements *saturate* [25].

In this paper, we propose a novel atomic norm soft thresholding (AST) algorithm [24] to recover spectrally-sparse signals and estimate the frequencies from their 1-bit quantized measurements. Our algorithm is based on finding the proximal mapping of properly designed surrogate signals, that are formed by linear combinations of the sample-modulated measurement vectors, with respect to the atomic norm to promote spectral sparsity. In other words, we aim to find signals that balance between the

proximity to the surrogate signals and the small atomic norm. Moreover, the frequencies can be localized without knowing the model order a priori, by examining the peak of a dual polynomial constructed from the dual solution. Alternatively, conventional subspace methods can be used to estimate the frequencies using the recovered spectral signal. The proposed algorithm can be generalized to handle quantizations of noisy random linear measurements of multiple spectrally-sparse signals [20], where each signal contains the same set of frequencies with different coefficients. The proposed algorithms do not require knowledge of the specific form of the quantizer, and therefore can be applied even when the quantizer is unknown.

When the measurement vectors are composed of i.i.d. complex Gaussian entries, under a mild separation condition that the frequencies are separated by $4/n$, it is shown that the reconstruction error scales as $\sqrt{K \log n/m}/\lambda$, where K is the level of spectral sparsity, n is the signal dimension, m is the number of measurements, and λ is a parameter that depends on the SNR before quantization, which increases as we increase SNR but saturates at high SNR. Therefore, the reconstruction error rate allows a trade-off between the sample complexity and SNR.

B. Related Work

Our work is closely related to 1-bit compressed sensing [26]–[33], which aims to recover a sparse signal from signs of random linear measurements. In particular, Plan and Vershynin [29]–[32] generalize this idea to reconstructing signals that belong to some low-dimensional set. Very recently, [34] studied a similar setup and proposed a new algorithm using projected gradient descent. The surrogate signals used in our algorithm can be traced back to [11], [32]. The difference lies in that instead of projecting the surrogate signals directly onto some low-dimensional set, we adopt the proximal mapping of the surrogate signals with respect to the atomic norm. Several algorithms have been proposed in the CS literature to deal with general quantization schemes [10], [35] and nonlinear measurement schemes [32], [34], however the focus has been on reconstruction of sparse signals in a finite dictionary, whereas our focus is on parameter estimation and reconstructing sparse signals in a parametric dictionary containing an infinite number of atoms.

There are also several conflicting evidence regarding the trade-offs between bit-depth and sample complexity [11], [36] for signal reconstruction, as they may vary for different problems when using specific algorithms. In contrast, we derive the Cramér-Rao bound for *parameter estimation* using quantized compressive random measurements, which provides an estimation-theoretic benchmark for gauging the trade-off as well as benchmarking performances. Our CRB adds to existing literature of CRB calculations for 1-bit quantized single-tone frequency estimation [37] as well as for parameter estimation using compressive measurements [38].

C. Paper Organization and Notations

The rest of this paper is organized as follows. Section II provides the problem formulation. Section III presents the

---

[1]Throughout this paper, we measure the bit depth as the number of bits used to quantizer a real number.

Cramér-Rao bound for parameter estimation and discusses the trade-off between bit depths and sample complexity. Section IV presents backgrounds on the atomic norm and the proposed algorithms with performance guarantees. Section V presents the extension when quantizations of multiple measurement vectors are available. Numerical experiments on the proposed algorithms are provided in Section VI. We conclude in Section VII.

Throughout this paper, we use boldface letters to denote vectors and matrices, e.g., $\boldsymbol{a}$ and $\boldsymbol{A}$. The Hermitian transpose of $\boldsymbol{a}$ is denoted by $\boldsymbol{a}^{\mathsf{H}}$, the transpose of $\boldsymbol{a}$ is denoted by $\boldsymbol{a}^{\mathsf{T}}$, and $\|\boldsymbol{A}\|$, $\|\boldsymbol{A}\|_{\mathrm{F}}$, $\mathrm{Tr}(\boldsymbol{A})$ denote the spectral norm, the Frobenius norm, and the trace of the matrix $\boldsymbol{A}$, respectively. An indicator function for an event A is denoted as $\mathbb{I}_A$. Denote $\mathcal{T}(\boldsymbol{u}) \in \mathbb{C}^{n \times n}$ as the Hermitian Toeplitz matrix with $\boldsymbol{u}$ as the first column. Define the inner product between two vectors $\boldsymbol{a}, \boldsymbol{b}$ as $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^{\mathsf{H}} \boldsymbol{b}$. The cardinality of a set $\mathcal{D}$ is defined as $|\mathcal{D}|$. If $\boldsymbol{A}$ is positive semidefinite (PSD), then $\boldsymbol{A} \succeq 0$. $\Re(y)$ and $\Im(y)$ denote the real and imaginary part of a complex number y, respectively. The expectation of a random variable a is written as $\mathbb{E}[a]$. Define $\odot$ as entry-wise product. Throughout this paper, we use $c, c_1, c_2, \ldots$ to denote universal constants whose values may change from line to line.

II. PROBLEM FORMULATION

Let $\boldsymbol{x}^{\star} \in \mathbb{C}^n$ be a line spectrum signal, which is composed of a small number of spectral lines, defined as

$$\boldsymbol{x}^{\star} = \sum_{k=1}^{K} c_k \boldsymbol{v}(f_k), \tag{1}$$

where K is the number of frequencies or level of sparsity, $c_k = A_k e^{j2\pi\phi_k} \in \mathbb{C}$ is the kth coefficient, $A_k > 0$ is the kth amplitude, $\phi_k \in [0, 1)$ is the kth normalized phase, $f_k \in [0, 1)$ is the kth frequency, and

$$\boldsymbol{v}(f) = \left[1 \; e^{j2\pi f} \; \cdots \; e^{j2\pi(n-1)f} \right]^{\mathsf{T}}.$$

In CS, we acquire a set of random linear measurements of $\boldsymbol{x}^{\star}$, contaminated by additive complex Gaussian noise, where each measurement is given as

$$z_i = \langle \boldsymbol{a}_i, \boldsymbol{x}^{\star} \rangle + \sigma \epsilon_i, \quad i = 1, \ldots, m, \tag{2}$$

where m is the number of measurements, $\boldsymbol{a}_i \in \mathbb{C}^n$'s are the measurement vectors composed of i.i.d. standard complex Gaussian entries $\mathcal{CN}(0, 1)$, σ is the noise level, and we further have i.i.d. $\epsilon_i \sim \mathcal{CN}(0, 1)$. In a vector notation, we write

$$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}^{\star} + \sigma\boldsymbol{\epsilon}, \tag{3}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m]^{\mathsf{H}} \in \mathbb{C}^{m \times n}$ is the measurement matrix, $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_m]^{\mathsf{T}}$, and $\boldsymbol{z} = [z_1, z_2, \ldots, z_m]^{\mathsf{T}}$. These measurements are then quantized into a finite number of bits for the ease of digital storage and processing. Denote $\mathcal{Q}(\cdot) : \mathbb{R} \mapsto \mathcal{D}$ as the quantizer that quantizes a real number into a finite alphabet $\mathcal{D}$, where the *bit depth* is the smallest number of bits necessary to represent $\mathcal{D}$, i.e., $b^{\star} = \min\{b \in \mathbb{Z}^+ : |\mathcal{D}| \le 2^b\}$. The quantized measurements $\boldsymbol{y} = [y_1, y_2, \ldots, y_m]^{\mathsf{T}}$ of $\boldsymbol{z}$ are then denoted as

$$y_i = \mathcal{Q}(\Re(z_i)) + j\mathcal{Q}(\Im(z_i)), \quad i = 1, \ldots, m, \tag{4}$$

where we apply the same quantizer $\mathcal{Q}$ to both the real part and the imaginary part of the complex-valued measurement z_i. With slight abuse of notation, we denote the quantized measurements as

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{z}). \tag{5}$$

Our goal is then to recover $\boldsymbol{x}^{\star}$, and the set of frequencies $\boldsymbol{f} = \{f_k\}_{k=1}^{K}$, from the quantized measurements $\boldsymbol{y}$, possibly without a priori knowing the sparsity level K, and the form of the quantizer $\mathcal{Q}$.

Several choices of the quantizer are of special interest. At the extreme, we consider only knowing the quadrature information of z_i, where

$$Q(a) = \mathrm{sign}(a), \quad a \in \mathbb{R}, \tag{6}$$

We refer to this quantizer as the *one-bit quantizer*, as only a single bit is used to quantize each real number.

More generally, we consider a quantizer $\mathcal{Q}(\cdot)$ that is fully characterized by the quantization intervals $\{[t_\ell, t_{\ell+1})\}_{\ell=1}^{|\mathcal{D}|-1}$, where $t_0 = -\infty$, $t_{|\mathcal{D}|} = \infty$, $\cup_{\ell=1}^{|\mathcal{D}|}[t_\ell, t_{\ell+1}) = \mathbb{R}$, as well as the representatives of each interval $\omega_\ell \in [t_\ell, t_{\ell+1})$, where

$$Q(a) = \omega_\ell, \quad \text{if} \quad a \in [t_\ell, t_{\ell+1}). \tag{7}$$

For example, the Lloyd's quantizer [14] belongs to this form. The choice of the quantization scheme plays an important role in determining the performance of parameter estimation.

III. CRAMER-RAO BOUNDS AND TRADE-OFFS

In this section, we study the effects of quantization on parameter estimation by deriving the Cramér-Rao bound assuming the quantizer, the sparsity level, and the noise level are known. In particular, the bounds are calculated for 1-bit and general quantizations, respectively, which are then used to study the trade-off between sample complexity and bit depths for a fixed bit budget.

To begin with, we assume the set of parameters, including the frequencies, amplitudes, and phases, given as $\boldsymbol{\kappa} = \{f_k, A_k, \phi_k\}_{k=1}^{K} \in \mathbb{R}^{3K}$, is deterministic but unknown, the measurement matrix $\boldsymbol{A}$ is deterministic and known. Denote the probability mass function as $p(\boldsymbol{y}|\boldsymbol{\kappa})$, which is given as

$$p(\boldsymbol{y}|\boldsymbol{\kappa}) = \prod_{i=1}^{m} p(y_i|\boldsymbol{\kappa}) = \prod_{i=1}^{m} \left[p(\Re(y_i)|\boldsymbol{\kappa}) \cdot p(\Im(y_i)|\boldsymbol{\kappa}) \right], \tag{8}$$

where the second equality follows from the fact that ϵ_i is proper. Moreover, let

$$p(\Re(y_i)|\boldsymbol{\kappa}) = \prod_{\omega \in \mathcal{D}} p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})^{\mathbb{I}_{\{\Re(y_i) = \omega\}}} \tag{9}$$

$$p(\Im(y_i)|\boldsymbol{\kappa}) = \prod_{\omega \in \mathcal{D}} p_{\Im(y_i)}(\omega|\boldsymbol{\kappa})^{\mathbb{I}_{\{\Im(y_i) = \omega\}}} \tag{10}$$

be the probability mass function of $\Re(y)$ and $\Im(y)$, respectively. The Fisher Information Matrix (FIM), denoted by $\boldsymbol{I}(\boldsymbol{\kappa}) \in \mathbb{R}^{3K \times 3K}$, is given as

$$\boldsymbol{I}(\boldsymbol{\kappa}) = \mathbb{E}\left[\left(\frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} \right) \left(\frac{\partial \log p(\boldsymbol{y}|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} \right)^T \right]. \tag{11}$$

Note that for any $1 \leq i, j \leq m$,

$$
\mathbb{E}\left[\left(\frac{\partial \log p(\Re(y_i)|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)\left(\frac{\partial \log p(\Im(y_j)|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)^T\right]
$$

$$
= \mathbb{E}\left[\frac{\partial \log p(\Re(y_i)|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right] \cdot \mathbb{E}\left[\frac{\partial \log p(\Im(y_j)|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right]^T = \mathbf{0},
$$

where the first equality follows from independence of $\Re(y_i)$ and $\Im(y_j)$, and the second equality follows from the fact

$$
\mathbb{E}\left[\left(\frac{\partial \log p(\Re(y_i)|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)\right]
$$

$$
= \mathbb{E}\left[\sum_{\omega \in \mathcal{D}} \frac{\mathbb{I}_{\{\Re(y_i)=\omega\}}}{p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})} \frac{\partial p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right]
$$

$$
= \sum_{\omega \in \mathcal{D}} \frac{\partial p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} = \frac{\partial \left(\sum_{\omega \in \mathcal{D}} p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})\right)}{\partial \boldsymbol{\kappa}} = \mathbf{0}.
$$

Thus, plugging (8) into (11) all cross-terms will be zero and we have

$$
\boldsymbol{I}(\boldsymbol{\kappa}) = \sum_{i=1}^{m} \left[\boldsymbol{I}_i^R(\boldsymbol{\kappa}) + \boldsymbol{I}_i^I(\boldsymbol{\kappa})\right], \tag{12}
$$

where

$$
\boldsymbol{I}_i^R(\boldsymbol{\kappa}) = \mathbb{E}\left[\left(\frac{\partial \log p(\Re(y_i)|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)\left(\frac{\partial \log p(\Re(y_i)|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)^T\right]
$$

$$
= \sum_{\omega \in \mathcal{D}} \frac{1}{p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})}\left(\frac{\partial p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)\left(\frac{\partial p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)^T,
$$

and $\boldsymbol{I}_i^I(\boldsymbol{\kappa})$ can be given similarly by replacing $\Re(y_i)$ with $\Im(y_i)$.

The CRB for estimating $\boldsymbol{\kappa}$, is then given as $\mathrm{CRB}(\boldsymbol{\kappa}) = \boldsymbol{I}(\boldsymbol{\kappa})^{-1}$, and the CRB for estimating the ith parameter in $\boldsymbol{\kappa}$, is given as $[\boldsymbol{I}(\boldsymbol{\kappa})^{-1}]_{i,i}$.

A. CRB for 1-Bit Quantization

Our goal is then to calculate the FIM in (12). We will explain in details the calculations for the 1-bit case. First, since $\Re(z_i) \sim \mathcal{N}(\Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle), \frac{1}{2}\sigma^2)$, then

$$
p_{\Re(y_i)}(\omega|\boldsymbol{\kappa}) = \mathbb{P}\left(\omega \cdot \Re(z_i) > 0|\boldsymbol{\kappa}\right)
$$

$$
= \frac{1}{2} + \omega \cdot \Phi\left(\frac{\Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)}{\sigma}\right), \quad \omega = \pm 1, \tag{13}
$$

where $\Phi(u) = \frac{1}{\sqrt{\pi}}\int_0^u e^{-t^2}\,dt$. Therefore, by the chain rule,

$$
\frac{\partial p_{\Re(y_i)}(\omega|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} = \frac{\omega}{\sigma}\Phi'\left(\frac{\Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)}{\sigma}\right)\frac{\partial \Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)}{\partial \boldsymbol{\kappa}}
$$

$$
= \frac{\omega}{\sqrt{\pi}\sigma}\exp\left(-\frac{\Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)^2}{\sigma^2}\right)\frac{\partial \Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)}{\partial \boldsymbol{\kappa}}. \tag{14}
$$

As a short-hand notation, denote $s_i(\boldsymbol{\kappa}) = \Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)$ and $r_i(\boldsymbol{\kappa}) = \Im(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)$. Plug (14) into $\boldsymbol{I}_i^R(\boldsymbol{\kappa})$, we have

$$
\boldsymbol{I}_i^R(\boldsymbol{\kappa}) = \frac{4\exp\left(-2s_i(\boldsymbol{\kappa})^2/\sigma^2\right)}{\pi\sigma^2\left[1 - 4\Phi^2\left(\frac{s_i(\boldsymbol{\kappa})}{\sigma}\right)\right]}\left(\frac{\partial s_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)\left(\frac{\partial s_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)^T,
$$

and similarly,

$$
\boldsymbol{I}_i^I(\boldsymbol{\kappa}) = \frac{4\exp\left(-2r_i(\boldsymbol{\kappa})^2/\sigma^2\right)}{\pi\sigma^2\left[1 - 4\Phi^2\left(\frac{r_i(\boldsymbol{\kappa})}{\sigma}\right)\right]}\left(\frac{\partial r_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)\left(\frac{\partial r_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)^T.
$$

As a remark, when $\sigma = 0$, the amplitude of the signal cannot be recovered from the 1-bit measurements due to scaling ambiguity, and the FIM becomes singular in this case. Therefore, our expressions for CRB is valid when $\sigma \neq 0$.

B. CRB for General Quantization

We now explain the calculation for a general quantization scheme. For $\omega_\ell \in \mathcal{D}$, and a corresponding interval $[t_\ell, t_{\ell+1})$, we have

$$
p_{\Re(y_i)}(\omega_\ell|\boldsymbol{\kappa}) = \mathbb{P}\left(\Re(z_i) \in [t_\ell, t_{\ell+1})|\boldsymbol{\kappa}\right)
$$

$$
= \int_{\frac{t_\ell - s_i(\boldsymbol{\kappa})}{\sigma}}^{\frac{t_{\ell+1} - s_i(\boldsymbol{\kappa})}{\sigma}} \frac{1}{\sqrt{\pi}}e^{-t^2}\,dt
$$

$$
= \Phi\left(\frac{t_{\ell+1} - s_i(\boldsymbol{\kappa})}{\sigma}\right) - \Phi\left(\frac{t_\ell - s_i(\boldsymbol{\kappa})}{\sigma}\right), \tag{15}
$$

then, following similar arguments, we have

$$
\frac{\partial p_{\Re(y_i)}(\omega_\ell|\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}
$$

$$
= \frac{1}{\sqrt{\pi\sigma^2}}\left[e^{-\frac{(t_{\ell+1}-s_i(\boldsymbol{\kappa}))^2}{\sigma^2}} - e^{-\frac{(t_\ell-s_i(\boldsymbol{\kappa}))^2}{\sigma^2}}\right]\frac{\partial \Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)}{\partial \boldsymbol{\kappa}}.
$$

Therefore, define

$$
\Gamma_i^R(\boldsymbol{\kappa}) = \sum_{\ell=1}^{|\mathcal{D}|-1} \frac{\left[e^{-\frac{(t_{\ell+1}-s_i(\boldsymbol{\kappa}))^2}{\sigma^2}} - e^{-\frac{(t_\ell-s_i(\boldsymbol{\kappa}))^2}{\sigma^2}}\right]^2}{\Phi\left(\frac{t_{\ell+1} - s_i(\boldsymbol{\kappa})}{\sigma}\right) - \Phi\left(\frac{t_\ell - s_i(\boldsymbol{\kappa})}{\sigma}\right)}, \tag{16}
$$

and

$$
\Gamma_i^I(\boldsymbol{\kappa}) = \sum_{\ell=1}^{|\mathcal{D}|-1} \frac{\left[e^{-\frac{(t_{\ell+1}-r_i(\boldsymbol{\kappa}))^2}{\sigma^2}} - e^{-\frac{(t_\ell-r_i(\boldsymbol{\kappa}))^2}{\sigma^2}}\right]^2}{\Phi\left(\frac{t_{\ell+1} - r_i(\boldsymbol{\kappa})}{\sigma}\right) - \Phi\left(\frac{t_\ell - r_i(\boldsymbol{\kappa})}{\sigma}\right)}, \tag{17}
$$

we have the following theorem for the expression of FIM in light of our derivations in the previous subsection.

Theorem 1: The Fisher information matrix $\boldsymbol{I}(\boldsymbol{\kappa})$ for estimating the unknown parameter $\boldsymbol{\kappa}$ is given as

$$
\boldsymbol{I}(\boldsymbol{\kappa}) = \frac{1}{\pi\sigma^2}\sum_{i=1}^{m}\left(\Gamma_i^R(\boldsymbol{\kappa})\frac{\partial s_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\left(\frac{\partial s_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)^T\right.
$$

$$
\left. + \Gamma_i^I(\boldsymbol{\kappa})\frac{\partial r_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\left(\frac{\partial r_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}}\right)^T\right). \tag{18}
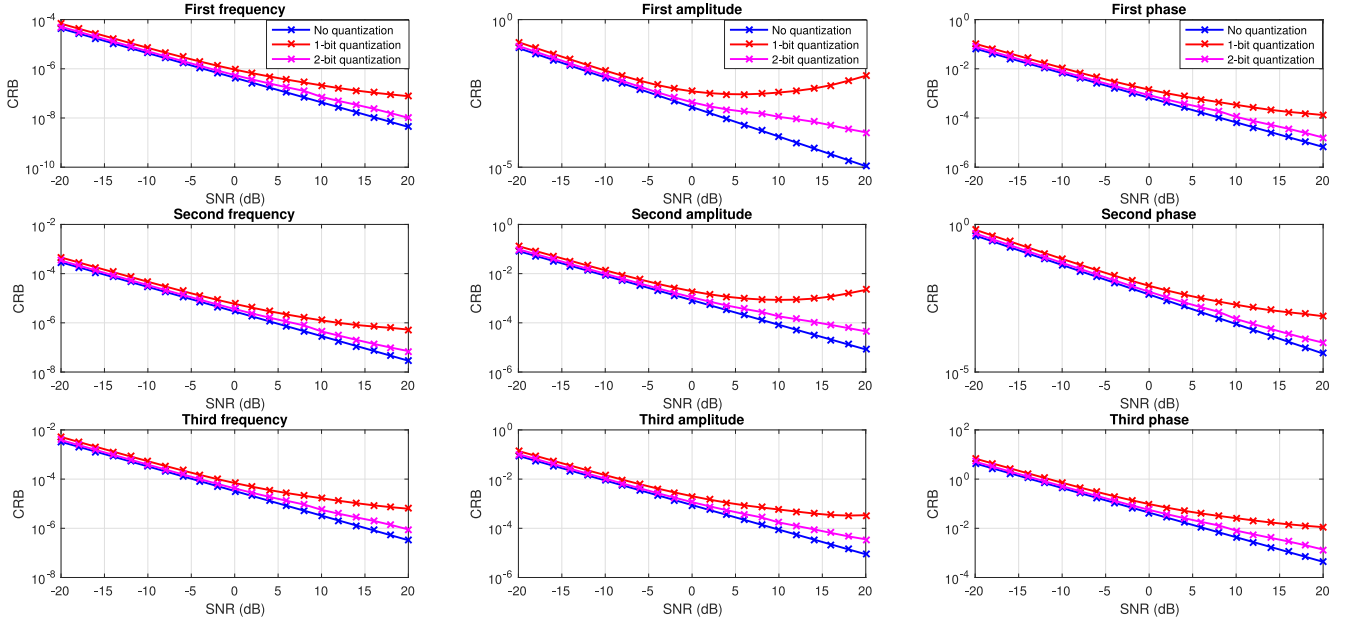$$

Fig. 1. CRB under different bit-depths with respect to SNR for a fixed number of measurements $m = 100$. Here, $n = 64$ and $K = 3$. Each row represents the CRB for estimating the frequency, amplitude, and phase of one spectral atom.

It is worth mentioning the FIM depends only on the quantization intervals, *not* the value of representatives. In contrast, the FIM using the unquantized measurements z is given as

$$I_{\text{unquantized}}(\boldsymbol{\kappa}) = \frac{2}{\sigma^2} \sum_{i=1}^{m} \frac{\partial s_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} \left(\frac{\partial s_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} \right)^T$$
$$+ \frac{\partial r_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} \left(\frac{\partial r_i(\boldsymbol{\kappa})}{\partial \boldsymbol{\kappa}} \right)^T \right). \qquad (19)$$

It remains to evaluate $\frac{\partial s_i(\boldsymbol{\kappa})}{\partial \kappa}$ and $\frac{\partial r_i(\boldsymbol{\kappa})}{\partial \kappa}$. Following the Wirtinger calculus [39], we have $\frac{\partial s_i(\boldsymbol{\kappa})}{\partial \kappa} = \frac{\partial \Re(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)}{\partial \kappa} = \frac{1}{2} \Re \left(\boldsymbol{a}_i^{\mathsf{H}} \frac{\partial \boldsymbol{x}^\star}{\partial \kappa} \right)$ and $\frac{\partial r_i(\boldsymbol{\kappa})}{\partial \kappa} = \frac{\partial \Im(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)}{\partial \kappa} = \frac{1}{2} \Im \left(\boldsymbol{a}_i^{\mathsf{H}} \frac{\partial \boldsymbol{x}^\star}{\partial \kappa} \right)$. Define $\boldsymbol{g}(f) = \frac{\partial \boldsymbol{v}(f)}{\partial f} = \left[0, j2\pi e^{j2\pi f}, \cdots, j2\pi(n-1)e^{j2\pi(n-1)f} \right]^{\mathsf{T}}$.

Then, for each of the parameters in $\boldsymbol{\kappa}$, we have, for $k = 1, \ldots, K$,

$$\frac{\partial \boldsymbol{x}^\star}{\partial f_k} = c_k \boldsymbol{g}(f_k), \qquad (20a)$$

$$\frac{\partial \boldsymbol{x}^\star}{\partial A_k} = e^{j2\pi \phi_k} \boldsymbol{v}(f_k), \qquad (20b)$$

$$\frac{\partial \boldsymbol{x}^\star}{\phi_k} = j2\pi c_k \boldsymbol{v}(f_k). \qquad (20c)$$

C. Numerical Evaluations

We now evaluate the CRB for 1-bit and 2-bit quantization schemes using the Lloyd's quantizer, and compare it against the CRB without quantization. We generate a spectrally-sparse signal $\boldsymbol{x}^\star$ of length $n = 64$ with frequencies $f_1 = 0.3$, $f_2 =$

0.325, and $f_3 = 0.8$, and complex coefficients $c_1 = 0.4e^{j2\pi \cdot 0.1}$, $c_2 = 0.15e^{j2\pi \cdot 0.55}$, and $c_3 = 0.05e^{j2\pi \cdot 0.75}$, which are selected arbitrarily.

We first fix the number of measurements as $m = 100$, and generate a measurement matrix with complex standard i.i.d. Gaussian entries. Fig. 1 shows the CRB for estimating all parameters with respect to the SNR, where it is defined as $\mathsf{SNR} = \|\boldsymbol{x}^\star\|_2^2 / \sigma^2$. It is evident that increasing the bit depth improves the performance. In the low SNR regime performance is *noise-limited*, and behaves similarly as if there was no quantization; in the high SNR regime, performance is *quantization-limited*, and experiences severe performance degeneration due to quantization.

In many situations, we cannot simultaneously have high sample complexity and high bit depth, but rather, our budget is constrained by the number of total bits, which is $B = m \cdot b^\star = m \cdot \lceil \log_2 |\mathcal{D}| \rceil$. Therefore, it is useful to understand the trade-off between sample complexity and bit depth. Here, we use the CRB as a tool to compare the 1-bit and 2-bit quantization schemes. Fix the total number of bits as $B = 100$. In the 1-bit quantization scheme, we use a measurement matrix with $m = 100$ as generated earlier. In the 2-bit quantization scheme, we only use the first $m/2$ rows of the same measurement matrix. For comparison, we also plot the CRB assuming unquantized measurements using the same measurement matrix as the 1-bit case. Fig. 2 shows the CRB for estimating all parameters with respect to the SNR. It can be seen that in the low SNR regime, 1-bit quantization is preferred, as performance is noise-limited, so higher sample complexity improves performance; in the high SNR regime, 2-bit quantization is preferred, as performance is quantization-limited, so higher bit depth improves performance. Our analysis is estimation-theoretic, and doesn't depend on the algorithm being adopted.
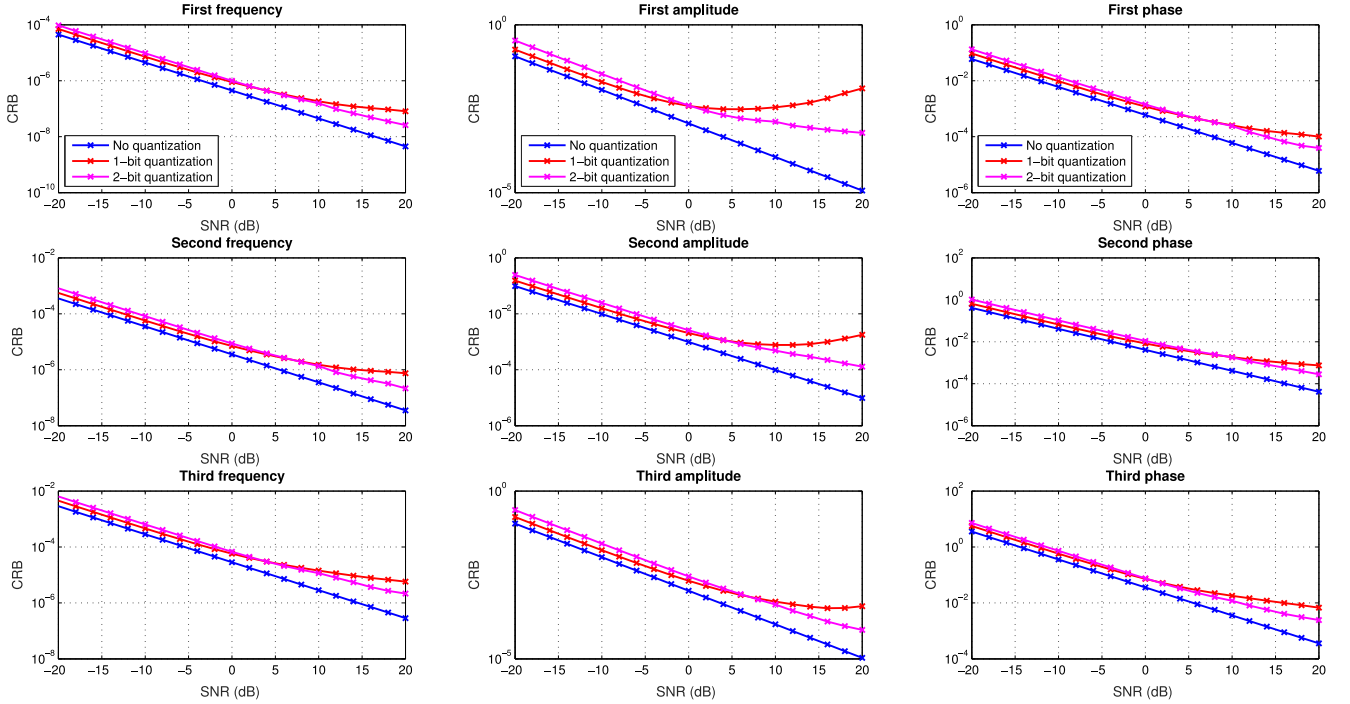
Fig. 2. CRB under different bit-depths with respect to SNR for a fixed number of bits $B = 100$. In this case, 2-bit quantization only has half the number of measurements of the 1-bit case. Here, $n = 64$ and $K = 3$. Each row represents the CRB for estimating the frequency, amplitude, and phase of one spectral atom.

IV. ATOMIC NORM SOFT THRESHOLDING FOR QUANTIZED SPECTRAL COMPRESSED SENSING

It is well-known that maximum likelihood estimators approach the performance of CRB asymptotically at high SNR [15], however, their implementation requires exact knowledge of the likelihood function, which may not be available in certain applications. Therefore, in this section, we will develop estimators that do not require the knowledge of the quantization scheme using 1-bit measurements via atomic norm minimization [17]. We first provide the backgrounds on atomic norm for line spectrum estimation, and then describe the proposed algorithms for both the single vector case and the multiple vector case with performance guarantees.

A. Backgrounds on Atomic Norms

The atomic norm is originally proposed in [17] as a unified framework of convex regularizations for solving underdetermined linear inverse problems. Subsequently, [18]–[24] has tailored it to the estimation of spectrally-sparse signals.

For the single vector case, define the *atomic set* as $\mathcal{A}_s = \left\{ e^{j\phi} \boldsymbol{v}(f) : f \in [0,1), \phi \in [0, 2\pi) \right\}$, then the atomic norm of a vector $\boldsymbol{x}$ is given as

$$\|\boldsymbol{x}\|_{\mathcal{A}} := \inf\{t > 0 : \boldsymbol{x} \in t \cdot \operatorname{conv}(\mathcal{A}_s)\}$$

$$= \inf \left\{ \sum_i |\alpha_i| \ \Big| \ \boldsymbol{x} = \sum_i \alpha_i \boldsymbol{v}(f_i) \right\}, \qquad (21)$$

where $\operatorname{conv}(\mathcal{A})$ denotes the convex hull of set $\mathcal{A}$. The atomic norm can be viewed as a continuous analog of the ℓ_1 norm over

the continuous dictionary defined by the atomic set. Therefore, by promoting signals with small atomic norms, we encourage signals that can be expressed by a small number of spectral atoms. Appealingly, as shown in [18], it is possible to calculate $\|\boldsymbol{x}\|_{\mathcal{A}}$ using an equivalent semidefinite program, which can be computed efficiently using off-the-shelf solvers:

$$\|\boldsymbol{x}\|_{\mathcal{A}} = \min_{\boldsymbol{u} \in \mathbb{C}^n, w} \left\{ \frac{1}{2n} \operatorname{Tr}(\mathcal{T}(\boldsymbol{u})) + \frac{w}{2} \left| \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \boldsymbol{x} \\ \boldsymbol{x}^{\mathsf{H}} & w \end{bmatrix} \succeq 0 \right\},$$

where $\mathcal{T}(\boldsymbol{u})$ denotes the Hermitian Toeplitz matrix with $\boldsymbol{u}$ as the first column. The dual atomic norm $\| \cdot \|_{\mathcal{A}}^*$ for a vector $\boldsymbol{q} \in \mathbb{C}^n$, as will become useful later, is given as

$$\|\boldsymbol{q}\|_{\mathcal{A}}^* = \sup_{\|\boldsymbol{x}\|_{\mathcal{A}} \leq 1} \langle \boldsymbol{q}, \boldsymbol{x} \rangle_{\mathbb{R}} = \sup_{f \in [0,1]} |\boldsymbol{q}^{\mathsf{H}} \boldsymbol{v}(f)|,$$

where the second equality follows from the fact the the extreme values are taken when $\boldsymbol{x}$ is aligned with $\boldsymbol{v}(f)$ due to convexity. From the above equation it is clear that $\|\boldsymbol{q}\|_{\mathcal{A}}^*$ can be interpreted as the largest absolute value of a polynomial of $e^{j2\pi f}$, denoted as $Q(f) = |\boldsymbol{q}^{\mathsf{H}} \boldsymbol{v}(f)|$.

B. Atomic Soft-Thresholding With Quantized Measurements

We first construct a surrogate signal from the quantized measurements as [32]

$$\boldsymbol{s} = \frac{1}{m} \sum_{i=1}^m y_i \boldsymbol{a}_i = \frac{1}{m} \boldsymbol{A}^{\mathsf{H}} \boldsymbol{y} \in \mathbb{C}^n, \qquad (22)$$
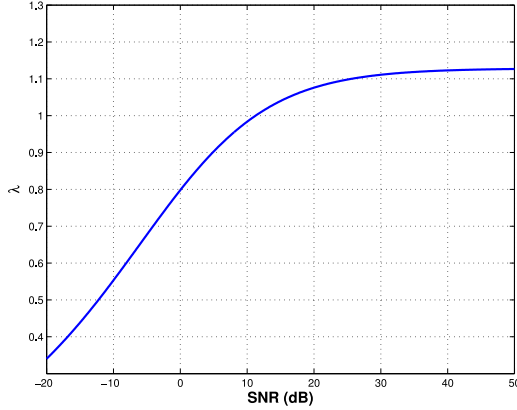
Fig. 3. Value of λ with respect to SNR before quantization.

and use the following atomic norm soft-thresholding (AST) algorithm to estimate the signal $\boldsymbol{x}$,

$$\hat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x} \in \mathbb{C}^n} \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{s}\|_2^2 + \tau \|\boldsymbol{x}\|_{\mathcal{A}}, \qquad (23)$$

which is the proximal mapping of the surrogate signal $\boldsymbol{s}$ with respect to the atomic norm, where $\tau > 0$ is a regularization parameter. One appealing feature of atomic norm minimization is that the set of frequencies can be recovered via the dual polynomial approach [24]. Namely, denote the dual variable as $\hat{\boldsymbol{q}} = (\boldsymbol{s} - \hat{\boldsymbol{x}})/\tau$, and $Q(f) = |\hat{\boldsymbol{q}}^{\mathsf{H}} \boldsymbol{v}(f)|$. Then the set of frequencies can be localized as $\hat{\mathcal{F}} = \{f : Q(f) = 1\}$. We refer interested readers to the details in [18]. Alternatively, the frequencies can be localized via performing conventional subspace methods using the estimated signal.

C. Performance Guarantees

In this section, we develop performance guarantees of the proposed AST algorithm under 1-bit quantization in the single vector case using the sign quantizer in (6). Note that in this case, it can be seen that $\boldsymbol{s}$ in (22) is an unbiased estimator of $\boldsymbol{x}^\star$ up to a scaling difference, i.e.,

$$\mathbb{E}[\boldsymbol{s}] = \lambda \frac{\boldsymbol{x}^\star}{\|\boldsymbol{x}^\star\|_2},$$

where

$$\lambda = \frac{2\|\boldsymbol{x}^\star\|_2}{\sqrt{\pi(\sigma^2 + \|\boldsymbol{x}^\star\|_2^2)}} = \frac{2}{\sqrt{\pi(1/\mathsf{SNR} + 1)}} \qquad (24)$$

depends on the SNR before quantization $\mathsf{SNR} = \|\boldsymbol{x}^\star\|_2^2/\sigma^2$. To illustrate, Fig. 3 depicts λ as a function of SNR, which is a monotonically increasing function with respect to SNR and approaches to the limit $2/\sqrt{\pi}$ as SNR goes to infinity.

Without loss of generality, we assume $\|\boldsymbol{x}^\star\|_2 = 1$. The performance of AST relies critically on the separation condition, which is defined as the minimum distance between distinct frequencies,

$$\Delta = \min_{k \neq j} |f_k - f_j| \geq \frac{4}{n}, \qquad (25)$$

where $|f_k - f_j|$ is evaluated as the wrap-around difference on the unit modulus. Under the separation condition, we have the performance guarantee of the proposed algorithm in (23), stated below.

Theorem 2: Set $\tau := \eta\sqrt{n \log n / m}$ for some constant $\eta \geq 1$. Under the separation condition, the solution $\hat{\boldsymbol{x}}$ satisfies

$$\left\| \frac{\hat{\boldsymbol{x}}}{\lambda} - \boldsymbol{x}^\star \right\|_2 \lesssim \frac{1}{\lambda} \sqrt{\frac{K \log n}{m}}$$

with high probability.

The proof of Theorem 2 can be found in Appendix A. Theorem 2 suggests that the proposed algorithm accurately recovers the signal as soon as m is on the order of $K \log n$, which is order-wise near-optimal, since at least an order of $K \log(n/K)$ measurements are needed in order to recover a sparse signal in the DFT basis [30]. Moreover, the theorem also suggests that the normalized reconstruction error is inverse proportional to λ, which plays the role of SNR *after quantization* and is a nonlinear function of the SNR before quantization. In the low SNR regime, λ scales as $1/\sqrt{\mathsf{SNR}}$, and the performance is comparable to that using unquantized measurements. However, in the high SNR regime, there is a saturation phenomenon, as evidenced by Fig. 3, and the performance does not improve as much with we increase SNR, which is also corroborated by numerical simulation in Section VI. These results are qualitatively in line with existing work on one-bit CS [30].

Remark 1: More generally, Theorem 2 can be extended to the generalized linear model following similar strategies in [29], as long as the 1-bit measurements y_i's are i.i.d. and satisfy $\mathbb{E}[y_i|\boldsymbol{a}_i] = g(\langle \boldsymbol{a}_i, \boldsymbol{x}^\star \rangle)$ for some link function $g(\cdot)$, and accordingly $\lambda = \mathbb{E}[g(\theta)\theta^{\mathsf{H}}]$ where the expectation is taken with respect to $\theta \sim \mathcal{CN}(0, 1)$. This allows us to model other complex quantization schemes with non-Gaussian noise.

V. EXTENSION TO THE MULTIPLE VECTOR CASE

In many applications, we encounter an ensemble of line spectrum signals, where each signal $\boldsymbol{x}_t \in \mathbb{C}^n$ contains a linear combination of spectral lines with the same set of frequencies $\mathcal{F}$, but with varying amplitudes, given as

$$\boldsymbol{x}_t^\star = \sum_{k=1}^{K} c_{k,t} \boldsymbol{v}(f_k), \quad 1 \leq t \leq T,$$

where $c_{k,t} \in \mathbb{C}$, and T is the number of snapshots. Denote $\boldsymbol{X}^\star = [\boldsymbol{x}_1^\star, \boldsymbol{x}_2^\star, \ldots, \boldsymbol{x}_T^\star] \in \mathbb{C}^{n \times T}$ as the signal ensemble. Similar to (3), the CS measurement of each snapshot is given as

$$\boldsymbol{z}_t = \boldsymbol{A}\boldsymbol{x}_t^\star + \sigma\boldsymbol{\epsilon}_t, \qquad (26)$$

where $\boldsymbol{\epsilon}_t = [\epsilon_{1,t}, \epsilon_{2,t}, \ldots, \epsilon_{m,t}]^{\mathsf{T}}$ contains i.i.d. standard complex Gaussian $\mathcal{CN}(0, 1)$ entries. Similar to (5), the quantized measurements of each $\boldsymbol{z}_t$ is then given as

$$\boldsymbol{y}_t = \mathcal{Q}(\boldsymbol{z}_t). \qquad (27)$$

Denote $\boldsymbol{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_T]$ and $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T]$ as the unquantized measurement ensemble and the quantized measurement ensemble, respectively. Our goal is then to recover $\boldsymbol{X}^\star$ and the set of frequencies from $\boldsymbol{Y}$, without assuming the knowledge
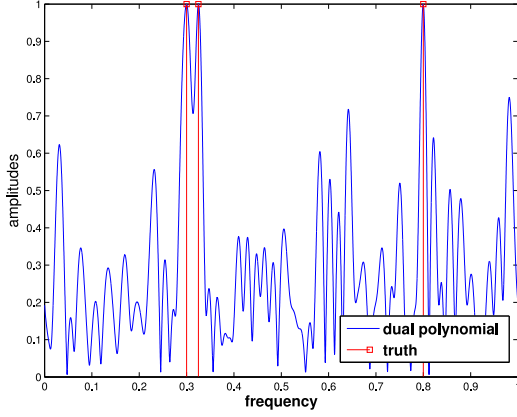
Fig. 4. Frequency localization via peaks of the dual polynomial, superimposed on the ground truth.
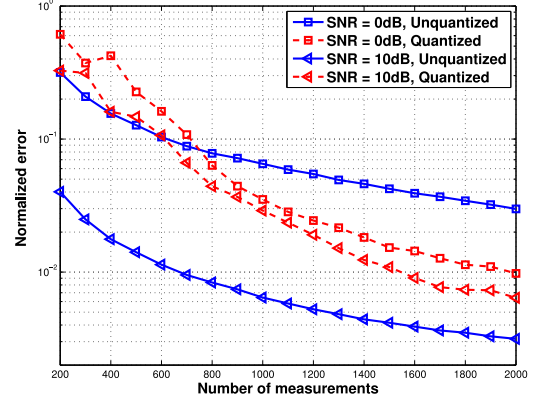

Fig. 5. Normalized reconstruction error with respect to the number of measurements at different SNRs with or without quantization.
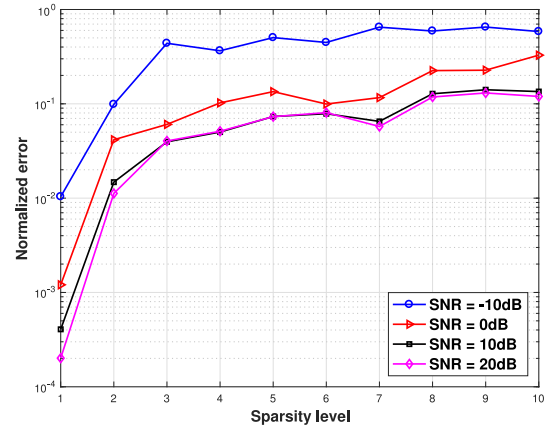

Fig. 6. Normalized reconstruction error with respect to the spectral sparsity level at different SNRs before quantization.
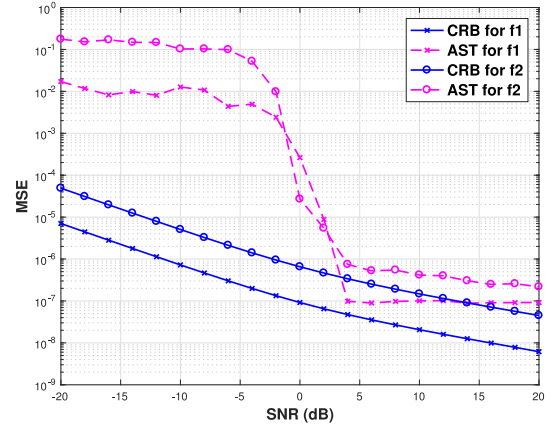

Fig. 7. Mean square error of frequency localization with respect to SNR using 1-bit measurements, CRB is provided as a benchmark: (a) first frequency; (b) second frequency.

of the sparsity level and the quantizer. The presence of multiple vectors can significantly improve the accuracy of frequency estimation.

It is possible to extend the atomic norm formulation to the multiple vector case [20]. Define the atomic set as

$$\mathcal{A}_m = \left\{ \boldsymbol{A}\left(f, \boldsymbol{b}\right) = \boldsymbol{v}\left(f\right)\boldsymbol{b} \, | \, f \in \left(0, 1\right], \boldsymbol{b} \in \mathbb{C}^{1 \times T}, \|\boldsymbol{b}\| = 1 \right\},$$

then the atomic norm is defined as

$$\|\boldsymbol{X}\|_{\mathcal{A}} = \inf \left\{ t > 0 : \boldsymbol{X} \in t \cdot \mathrm{conv}\left(\mathcal{A}_m\right) \right\}$$

$$= \inf \left\{ \sum_k |c_k| \, \Big| \, \boldsymbol{X} = \sum_k c_k \boldsymbol{A}\left(f_k, \boldsymbol{b}_k\right) \right\},$$

which can be computed similarly via solving the following semidefinite program [20]:

$$\|\boldsymbol{X}\|_{\mathcal{A}} = \min_{\boldsymbol{u} \in \mathbb{C}^n, \boldsymbol{W} \in \mathbb{C}^{T \times T}} \left\{ \frac{1}{2n}\mathrm{Tr}(\mathcal{T}(\boldsymbol{u})) + \frac{1}{2}\mathrm{Tr}(\boldsymbol{W}) \Big| \right.$$

$$\left. \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \boldsymbol{X} \\ \boldsymbol{X}^{\mathsf{H}} & \boldsymbol{W} \end{bmatrix} \succeq 0 \right\}.$$

The dual norm for some $\boldsymbol{Q} \in \mathbb{C}^{n \times T}$ is given as

$$\|\boldsymbol{Q}\|_{\mathcal{A}}^* = \sup_{\|\boldsymbol{X}\|_{\mathcal{A}} \leq 1} \langle \boldsymbol{Q}, \boldsymbol{X} \rangle_{\mathbb{R}} = \sup_{f \in [0,1]} \|\boldsymbol{Q}^{\mathsf{H}}\boldsymbol{v}(f)\|,$$

which is the largest absolute value of the polynomial $Q(f) = \|\boldsymbol{Q}^{\mathsf{H}}\boldsymbol{v}(f)\|$.

For reconstruction, we construct the surrogate signal ensemble from the quantized measurement ensemble $\boldsymbol{Y}$ as

$$\boldsymbol{S} = \frac{1}{m}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{Y} \in \mathbb{C}^{n \times T}, \qquad (28)$$

and use the following atomic norm soft-thresholding (AST) algorithm to estimate the signal ensemble $\boldsymbol{X}$,

$$\hat{\boldsymbol{X}} = \mathrm{argmin}_{\boldsymbol{X} \in \mathbb{C}^{n \times T}} \|\boldsymbol{X} - \boldsymbol{S}\|_F^2 + \tau_T \|\boldsymbol{X}\|_{\mathcal{A}}, \qquad (29)$$

where $\tau_T > 0$ is a regularization parameter. Moreover, define $\hat{\boldsymbol{Q}} = (\boldsymbol{S} - \hat{\boldsymbol{X}})/\tau_T$, and $Q(f) = \|\hat{\boldsymbol{Q}}^{\mathsf{H}}\boldsymbol{v}(f)\|$. Then the set of frequencies can be localized as $\hat{\mathcal{F}} = \{f : Q(f) = 1\}$. Alternatively, the frequencies can be localized via performing conventional subspace methods using the estimated snapshots.

VI. NUMERICAL EXPERIMENTS

In this section, we conduct numerical experiments to evaluate the performance of the proposed AST algorithms for parameter estimation using quantized compressive measurements in both the single vector case and the multiple vector case. For implementation of the AST algorithms, we used the CVX toolbox
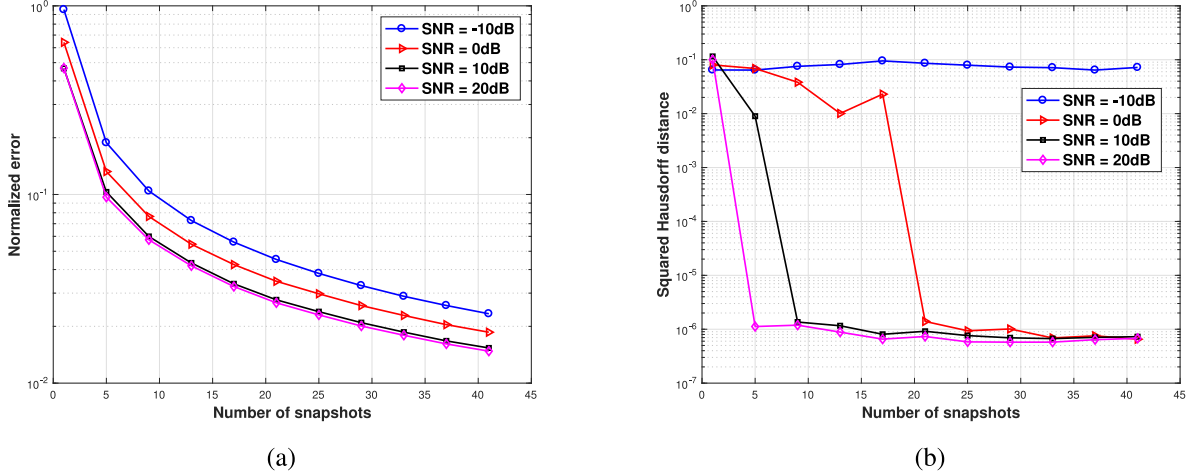
Fig. 8. Performance with respect to the number of snapshots at different SNRs using 1-bit measurements: (a) signal reconstruction error; (b) frequency estimation error measured in Hausdorff distance.

[40]. There're several other fast solvers developed for atomic norm minimization that are more scalable to large problems, including ADMM [20], [24], ADCG [41], and CoGent [42], to name a few.

A. Single Vector Case

Let $n = 64$ and $K = 3$. The set of frequencies is located at $\boldsymbol{f} = \{0.3, 0.325, 0.8\}$, where the first two frequencies are separated barely more than $1/n$, the Rayleigh limit. The number of bits is set as $m = 1000$, where the measurement vectors are generated with i.i.d. $\mathcal{CN}(0, 1)$ entries. The measurements are quantized according to (6). Fig. 4 shows the amplitude of the constructed dual polynomial by solving (23), where its peaks can be used to localize the frequencies. It can be seen that it matches accurately with the ground truth.

Next, we compare the performance of signal reconstruction using atomic norm with unquantized measurements $\boldsymbol{z}$, by running the algorithm:

$$\hat{\boldsymbol{x}}_{\mathsf{UQ}} = \operatorname{argmin}_{\boldsymbol{x} \in \mathbb{C}^n} \frac{1}{2} \|\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \tilde{\tau} \|\boldsymbol{x}\|_{\mathcal{A}},$$

where $\tilde{\tau}$ is a properly tuned regularization parameter. The normalized reconstruction error is defined as $\sin^2(\angle \hat{\boldsymbol{x}}, \boldsymbol{x}) = 1 - |\langle \hat{\boldsymbol{x}}, \boldsymbol{x}^\star \rangle|^2 / (\|\hat{\boldsymbol{x}}\|_2^2 \|\boldsymbol{x}^\star\|_2^2)$, where $\hat{\boldsymbol{x}}$ is the reconstructed signal using either algorithm. Fig. 5 shows the normalized reconstruction error at different SNRs with comparisons to that using the quantized measurements and the AST algorithm (23), where SNR is defined again as $\mathsf{SNR} = \|\boldsymbol{x}^\star\|_2^2 / \sigma^2$. It can be seen that the reconstruction accuracy improves as we increase the SNR as well as the number of measurements, validating the theoretical analysis. In particular, at low SNR, using quantized measurements can potentially achieve better reconstruction quality with much fewer measurement budgets in bits. It can also be seen that improving the SNR before quantization does not have as strong impact as for the unquantized case.

Next, we examine the performance of the proposed algorithm as a function of the spectral sparsity level. Fix $n = 64$ and

$m = 1000$. At each run, we randomly generate K different frequencies that satisfy the separate condition. Fig. 6 shows the normalized reconstruction error as a function of the sparsity level at various SNR, averaged over 200 Monte Carlo simulations. It can be seen that the reconstruction error is higher when the spectral sparsity level is higher, and the SNR is lower. Moreover, it can be seen that the reconstruction error stops to decrease when the SNR is relatively high, indicating a saturation effect due to quantization, as predicted by our theory.

We further compare the performance of frequency localization using the proposed algorithm with the CRB. Fix $n = 64$ and $m = 1000$. We generate the ground signal with frequencies $f_1 = 0.3$, $f_2 = 0.325$ and amplitudes $c_1 = 0.4e^{j2\pi \cdot 0.1}$, $c_2 = 0.15e^{j2\pi \cdot 0.55}$. Fig. 7 shows the average mean squared error for each frequency over 200 Monte Carlo simulations, against the corresponding CRB calculated using the formulas in Section III. The frequencies are estimated by using the MATLAB function `rootmusic` by assuming the correct model order, that is $K = 2$. The performance of the proposed algorithm exhibits a threshold effect where it approaches that of CRB as soon as SNR is large enough. However, further increasing the SNR doesn't seem to improve the performance, which coincides with the saturation effect discussed earlier.

B. Multiple Vector Case

We evaluate the performance of the AST algorithm (29) in the multiple vector case. We follow the same setup as Fig. 4, where $n = 64$, the set of frequencies $\boldsymbol{f} = \{0.3, 0.325, 0.8\}$, and the number of measurements for each snapshot is $m = 50$. The coefficients of each snapshot in $\boldsymbol{X}$ is generated independently using the standard complex Gaussian distribution. The SNR per snapshot is defined as $\mathsf{SNR} = \|\boldsymbol{X}\|_F^2 / (T\sigma^2)$, where T is the number of snapshots. We set the regularization parameter $\tau_T = \sqrt{n \log n / (10 \cdot mT)}$ in the experiment. The normalized reconstruction error is defined as $\sin^2(\angle \boldsymbol{X}; \hat{\boldsymbol{X}})$, where $\hat{\boldsymbol{X}}$ is the recovered signal containing multiple snapshots, and $\angle$ denotes the angle between the subspace spanned by $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$. Once $\hat{\boldsymbol{X}}$

is obtained, we estimate the frequencies by using the MATLAB function `rootmusic` by assuming the correct model order, that is $K = 3$. The accuracy of frequency estimation is evaluated by examining the Hausdorff distance between the recovered frequencies $\hat{\boldsymbol{f}}$ and the ground truth $\boldsymbol{f}$ as

$$d_H(\boldsymbol{f}, \hat{\boldsymbol{f}}) = \max \left\{ \sup_{f \in \boldsymbol{f}} \inf_{\hat{f} \in \hat{\boldsymbol{f}}} \|f - \hat{f}\|_2, \sup_{\hat{f} \in \hat{\boldsymbol{f}}} \inf_{f \in \boldsymbol{f}} \|f - \hat{f}\|_2 \right\}.$$

Fig. 8 shows the recovery performance with respect to the number of snapshots at different SNRs, averaged over 50 Monte Carlo simulations, where (a) depicts the normalized reconstruction error, and (b) depicts the squared Hausdorff distance. At a fixed SNR, it can be seen that both the normalized reconstruction error and frequency estimation error reduce, highlighting the benefit of having multiple snapshots. In particular, having multiple snapshots allows better frequency recovery once the number of snapshots is large enough. Moreover, performance improves as we increase the SNR.

VII. CONCLUDING REMARKS

In this paper, we examined the effect of (heavy) quantization in spectral compressed sensing that is useful for understanding wideband spectral signal acquisition and processing. Our contributions are two-fold. We first derived the Cramér-Rao bound for parameter estimation with multiple complex sinusoids using quantized compressed linear measurements. This bound is instrumental in describing the trade-offs between bit depth and sample complexity at different SNR regimes. Such an estimation-theoretical perspective is independent of the algorithm and hasn't been exploited in the previous literature. Secondly, we developed algorithms for spectral-sparse signal recovery using quantized measurements via atomic norm minimization, which do not require knowledge of the quantizer in recovery. Under a mild separation condition, we establish that we can accurately recover a spectrally-sparse signal from the signs of $O(K \log n)$ random linear measurements. The proposed algorithm also can be extended to handle multiple signal snapshots. This generalizes the literature on one-bit compressed sensing to the important class of spectrally sparse signals using atomic norms, and we carefully examined the performance of the proposed algorithms via numerical experiments.

An alternative convex relaxation for spectrally-sparse signal recovery is based on Hankel matrix enhancement and nuclear norm minimization [43], [44]. In the single vector case, instead of imposing the atomic norm regularizer as in (23), one may consider

$$\hat{\boldsymbol{x}} = \operatorname{argmin}_{\boldsymbol{x} \in \mathbb{C}^n} \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{s}\|_2^2 + \tau_H \|\mathcal{H}(\boldsymbol{x})\|_*. \quad (30)$$

Here, $\mathcal{H}(\boldsymbol{x})$ denotes a Hankel matrix given as

$$\mathcal{H}(\boldsymbol{x}) = \begin{bmatrix} x_1 & x_2 & & \\ x_2 & & & \ddots \\ \vdots & & \ddots & \\ x_{n_1} & x_{n_1+1} & \cdots & x_n \end{bmatrix}, \quad (31)$$

where n_1 is set as $\lfloor n/2 \rfloor$ to make the matrix $\mathcal{H}(\boldsymbol{x})$ as square as possible, $\| \cdot \|_*$ is the nuclear norm, and τ_H is a regularization parameter. Our preliminary numerical simulations suggest this method is also effective for promoting spectral sparsity, but a detailed study is beyond the scope of the current paper. We leave the thorough analysis of (30) to future work.

Since the Cramér-Rao bounds assume perfect knowledge of the quantizers, they may not be indicative to benchmark the performance of the atomic norm minimization algorithms as proposed in this paper, since these algorithms do not make use of such knowledge. In the future, it might be interesting to develop estimation-theoretical bounds that only assume partial or little knowledge about the quantizer.

APPENDIX A
PROOF OF THEOREM 2

An alternative way to represent the atomic decomposition is to write it as an integration of certain point measure [23]. Define the representing measure of $\boldsymbol{x}^\star$ as

$$\mu(f) = \sum_{k=1}^{K} c_k \delta(f - f_k),$$

where $\delta(\cdot)$ is the delta function. Then we can rewrite $\boldsymbol{x}^\star$ as

$$\boldsymbol{x}^\star = \int_0^1 \boldsymbol{v}(f) d\mu(f) = \sum_{k=1}^{K} c_k \boldsymbol{v}(f_k). \quad (32)$$

Correspondingly, denote $\hat{\mu}(f)$ as the representing measure for the solution $\hat{\boldsymbol{x}}$ of (23), which means $\hat{\boldsymbol{x}} = \int_0^1 \boldsymbol{v}(f) d\hat{\mu}(f)$.

Denote the reconstruction error as $\boldsymbol{e} = \lambda \boldsymbol{x}^\star - \hat{\boldsymbol{x}}$, and its representing measure is $\gamma = \lambda \mu - \hat{\mu}$. With these definitions, applying [23, Lemma 1], we can bound the error as [23]

$$\|\boldsymbol{e}\|_2^2 \le \|\boldsymbol{e}\|_{\mathcal{A}}^* \left(\int_F |\gamma| (df) + I_0 + I_1 + I_2 \right), \quad (33)$$

where $I_\ell = \sum_{k=1}^{K} I_\ell^k$, for $\ell = 0, 1, 2$, with $I_0^k = \left| \int_{N_k} \gamma (df) \right|$, $I_1^k = n \left| \int_{N_k} (f - f_k) \gamma (df) \right|$, $I_2^k = \frac{n^2}{2} \int_{N_k} (f - f_k)^2 |\gamma| (df)$, where $N_k = \{f \in \mathbb{T} : d(f, f_k) \le 0.16/n\}$ as the neighborhoods around each frequency, and $F = \mathbb{T} \setminus \cap_{k=1}^{K} N_k$.

To bound the first term in (33), let us denote the deviation

$$\boldsymbol{w} = \boldsymbol{s} - \mathbb{E}[\boldsymbol{s}] = \boldsymbol{s} - \lambda \boldsymbol{x}^\star, \quad (34)$$

where $\mathbb{E}[\boldsymbol{w}] = 0$. We have

$$\|\boldsymbol{e}\|_{\mathcal{A}}^* \le \|\boldsymbol{w}\|_{\mathcal{A}}^* + \|\boldsymbol{s} - \hat{\boldsymbol{x}}\|_{\mathcal{A}}^*$$
$$\le \|\boldsymbol{w}\|_{\mathcal{A}}^* + \tau, \quad (35)$$

where the first line follows from the triangle inequality, and the second line follows from the optimality condition of the AST algorithm in (23) in the following lemma.

Lemma 1 (Optimality conditions [24]): $\hat{\boldsymbol{x}}$ is the solution of (23) if and only if $\|\boldsymbol{s} - \hat{\boldsymbol{x}}\|_{\mathcal{A}}^* \le \tau$, and $\langle \boldsymbol{s} - \hat{\boldsymbol{x}}, \hat{\boldsymbol{x}} \rangle = \tau \|\hat{\boldsymbol{x}}\|_{\mathcal{A}}$.

Therefore, if we set $\tau \ge \eta \|\boldsymbol{w}\|_{\mathcal{A}}^*$, where $\eta \ge 1$ is some constant, then plugging this into (35) we can show that

$$\|\boldsymbol{e}\|_{\mathcal{A}}^* \le (\eta^{-1} + 1)\tau \le 2\tau. \quad (36)$$

The second term in (33) can be bounded in exactly the same manner as in [23], as long as (36) holds. In effect, [23] proved the following bound, under the separation condition, with high probability we have

$$\left(\int_F |\gamma| (df) + I_0 + I_1 + I_2 \right) \leq C \frac{K\tau}{n}. \tag{37}$$

The following lemma bounds $\|\boldsymbol{w}\|_{\mathcal{A}}^*$, whose proof is provided in Appendix B.

Lemma 2: With probability at least $1 - 1/(\pi n \log n)$, we have

$$\|\boldsymbol{w}\|_{\mathcal{A}}^* \leq C \cdot \sqrt{\frac{n \log n}{m}},$$

where C is some universal constant.

Therefore, set $\tau = C\eta\sqrt{n \log n/m}$, and plug (36) and (37) into (33), we have

$$\|\boldsymbol{e}\|_2^2 \leq C' \cdot \frac{K\tau^2}{n} \leq C' \frac{K \log n}{m}. \tag{38}$$

which is equivalent to

$$\left\| \frac{\hat{\boldsymbol{x}}}{\lambda} - \boldsymbol{x}^\star \right\|_2 \lesssim \frac{1}{\lambda} \sqrt{\frac{K \log n}{m}}.$$

The proof is complete.

APPENDIX B
PROOF OF LEMMA 2

By definition, we can write $\|\boldsymbol{w}\|_{\mathcal{A}}^*$ as

$$\begin{aligned}
\|\boldsymbol{w}\|_{\mathcal{A}}^* &= \sup_{f \in [0,1)} |\langle \boldsymbol{s} - \lambda \boldsymbol{x}^\star, \boldsymbol{v}(f) \rangle| \\
&= \sup_{f \in [0,1)} |\langle \boldsymbol{s}, \boldsymbol{v}(f) \rangle - \mathbb{E}[\langle \boldsymbol{s}, \boldsymbol{v}(f) \rangle]| \\
&= \sup_{f \in [0,1)} |g_{\boldsymbol{x}^\star}(f) - \mathbb{E}[g_{\boldsymbol{x}^\star}(f)]| \tag{39}
\end{aligned}$$

where

$$g_{\boldsymbol{x}^\star}(f) := \langle \boldsymbol{s}, \boldsymbol{v}(f) \rangle = \frac{1}{m} \sum_{i=1}^m y_i \langle \boldsymbol{a}_i, \boldsymbol{v}(f) \rangle.$$

To proceed, we use the following symmetrization bound, which is the complex-valued version of [30, Lemma 5.1].

Lemma 3: Let $\{\epsilon_i\}_{i=1}^m$ be a sequence of independent complex-valued random variables, where $\epsilon_i \sim \epsilon = e^{j 2\pi \theta}$, where θ uniformly distributed between $[0, 1)$. Then

$$\begin{aligned}
\mu :&= \mathbb{E}\left[\sup_{f \in [0,1)} |g_{\boldsymbol{x}^\star}(f) - \mathbb{E}[g_{\boldsymbol{x}^\star}(f)]| \right] \\
&\leq 2\mathbb{E}\left[\sup_{f \in [0,1)} \frac{1}{m} \left| \sum_{i=1}^m \epsilon_i y_i \langle \boldsymbol{a}_i, \boldsymbol{v}(f) \rangle \right| \right]. \tag{40}
\end{aligned}$$

Furthermore, we have the deviation inequality

$$\mathbb{P}\left\{ \sup_{f \in [0,1)} |g_{\boldsymbol{x}^\star}(f) - \mathbb{E}[g_{\boldsymbol{x}^\star}(f)]| \geq 2\mu + t \right\}$$
$$\leq 4\mathbb{P}\left\{ \sup_{f \in [0,1)} \frac{1}{m} \left| \sum_{i=1}^m \epsilon_i y_i \langle \boldsymbol{a}_i, \boldsymbol{v}(f) \rangle \right| > \frac{t}{2} \right\}. \tag{41}$$

Before applying Lemma 3, note that by symmetrization and rotational invariance, $\epsilon_i y_i \boldsymbol{a}_i$ have the same i.i.d. distribution of $\sqrt{2}\boldsymbol{a}_i$. Therefore, the following quantities are equivalent in distribution:

$$\sup_{f \in [0,1)} \frac{1}{m} \left| \sum_{i=1}^m \epsilon_i y_i \langle \boldsymbol{a}_i, \boldsymbol{v}(f) \rangle \right| \sim \frac{\sqrt{2}}{m} \sup_{f \in [0,1)} \left| \sum_{i=1}^m \langle \boldsymbol{a}_i, \boldsymbol{v}(f) \rangle \right|$$
$$\sim \sqrt{\frac{2}{m}} \sup_{f \in [0,1)} |\langle \boldsymbol{g}, \boldsymbol{v}(f) \rangle|,$$

where $\boldsymbol{g}$ is a vector composed of i.i.d. $\mathcal{CN}(0, 1)$.

Applying (41) in Lemma 3 to (39), we have

$$\mathbb{P}\left(\|\boldsymbol{w}\|_{\mathcal{A}}^* \geq 2\mu + t \right) \leq 4\mathbb{P}\left(\sqrt{\frac{2}{m}} \sup_{f \in [0,1)} |\langle \boldsymbol{g}, \boldsymbol{v}(f) \rangle| \geq \frac{t}{2} \right). \tag{42}$$

From (40) in Lemma 3, we have

$$\mu = \mathbb{E}[\|\boldsymbol{w}\|_{\mathcal{A}}^*] \leq 2\sqrt{\frac{2}{m}} \mathbb{E}\left[\sup_{f \in [0,1)} |\langle \boldsymbol{g}, \boldsymbol{v}(f) \rangle| \right]$$
$$\leq C\sqrt{\frac{n \log n}{m}}, \tag{43}$$

where the second line follows from [24, Appendix C,D] as

$$\mathbb{E}\left[\sup_{f \in [0,1]} |\langle \boldsymbol{g}, \boldsymbol{v}(f) \rangle| \right] \leq C_1 \sqrt{n \log(n)}.$$

Moreover, from [24, Appendix C], we have

$$\sup_{f \in [0,1)} |\langle \boldsymbol{g}, \boldsymbol{v}(f) \rangle| \leq C_2 \cdot \sqrt{n \log n}$$

hold with probability at least $1 - 1/(\pi n \log n)$. Set $t = 2C_2\sqrt{n \log n}$ and plug in the above two inequalities in (4), we have that

$$\|\boldsymbol{w}\|_{\mathcal{A}}^* \leq C \cdot \sqrt{\frac{n \log n}{m}}$$

holds with probability at least $1 - 1/(\pi n \log n)$.

REFERENCES

[1] Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 4, 2007, pp. IV–1357–IV–1360.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[4] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[5] M. Mishali, Y. C. Eldar, and A. J. Elron, "Xampling: Signal acquisition and processing in union of subspaces," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4719–4734, Oct. 2011.

[6] M. Wakin *et al.* "A nonuniform sampler for wideband spectrally-sparse environments," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 2, no. 3, pp. 516–529, Sep. 2012.

[7] 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.

[8] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wideband spectrum sensing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. ,* 2009, pp. 2337–2340.

[9] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[10] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Appl. Comput. Harmonic Anal.*, vol. 31, no. 3, pp. 429–443, 2011.

[11] M. Slawski and P. Li, "Linear signal recovery from b-bit-quantized linear measurements: Precise analysis of the trade-off between bit depth and number of measurements," arXiv:1607.02649.

[12] O. Mehanna and N. Sidiropoulos, "Frugal sensing: Wideband power spectrum sensing from few bits," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2693–2703, May 2013.

[13] Y. Chi and H. Fu, "Subspace learning from bits," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4429–4442, Sep. 2017.

[14] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[15] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.

[16] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[17] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805–849, 2012.

[18] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.

[19] Y. Chi and Y. Chen, "Compressive two-dimensional harmonic retrieval via atomic norm minimization," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1030–1042, Feb. 2015.

[20] Y. Li and Y. Chi, "Off-the-grid line spectrum denoising and estimation with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1257–1269, Mar. 2016.

[21] R. Heckel and M. Soltanolkotabi, "Generalized line spectral estimation via convex optimization," arXiv:1609.08198.

[22] Z. Yang and L. Xie, "Exact joint sparse frequency recovery via optimization methods," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5145–5157, Oct. 2016.

[23] G. Tang, B. N. Bhaskar, and B. Recht, "Near minimax line spectral estimation," *IEEE Trans. Inform. Theory*, vol. 61, no. 1, pp. 499–512, Jan. 2015.

[24] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987–5999, Dec. 2013.

[25] G. Gray and G. Zeoli, "Quantization and saturation noise due to analog-to-digital conversion," *IEEE Trans. Aerosp. Electron. Syst.*, no. 1, pp. 222–223, Jan. 1971.

[26] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *Proc. 42nd Annu. Conf. Inform. Sci. Syst.,* 2008, pp. 16–21.

[27] A. Gupta, R. Nowak, and B. Recht, "Sample complexity for 1-bit compressed sensing and sparse classification," in *Proc. IEEE Int. Symp. Inform. Theory*. 2010, pp. 1553–1557.

[28] M. Yan, Y. Yang, and S. Osher, "Robust 1-bit compressive sensing using adaptive outlier pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3868–3875, Jul. 2012.

[29] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Commun. Pure Appl. Math.*, vol. 66, no. 8, pp. 1275–1297, 2013.

[30] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 482–494, Jan. 2013.

[31] Y. Plan and R. Vershynin, "The generalized Lasso with non-linear observations," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1528–1537, Mar. 2016.

[32] Y. Plan, R. Vershynin, and E. Yudovina, "High-dimensional estimation with geometric constraints," *Inf. Infer.: A J. IMA*, vol. 6, no. 1, pp. 1–40, 2017.

[33] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. 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.

[34] S. Oymak and M. Soltanolkotabi, "Fast and reliable parameter estimation from nonlinear observations," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2276–2300, 2017.

[35] P. T. Boufounos, L. Jacques, F. Krahmer, and R. Saab, "Quantization and compressive sensing," in *Proc. Compressed Sens. Appl.*, Springer, 2015, pp. 193–237.

[36] J. N. Laska and R. G. Baraniuk, "Regime change: Bit-depth versus measurement-rate in compressive sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3496–3505, Jul. 2012.

[37] A. Host-Madsen and P. Handel, "Effects of sampling and quantization on single-tone frequency estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 650–662, Mar. 2000.

[38] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of fisher information and the Cramér–Rao bound for nonlinear parameter estimation after compressed sensing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.,* 2013, pp. 6630–6634.

[39] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[40] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming." [Online]. Available: http://cvxr.com/cvx/

[41] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM J. Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[42] N. Rao, P. Shah, and S. Wright, "Forward–backward greedy algorithms for atomic norm regularization," *IEEE Trans. Signal Process.*, vol. 63, no. 21, pp. 5798–5811, Nov. 2015.

[43] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6576–6601, Oct. 2014.

[44] J.-F. Cai, X. Qu, W. Xu, and G.-B. Ye, "Robust recovery of complex exponential signals from random gaussian projections via low rank Hankel matrix reconstruction," *Appl. Comput. Harmonic Anal.*, vol. 41, no. 2, pp. 470–490, 2016.
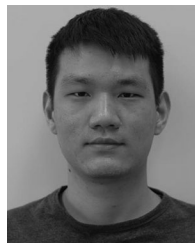
Haoyu Fu (S'16) received the B.Eng. degree from the Wuhan University of Technology, Wuhan, China, in 2014. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, USA. His research interests include compressed sensing, statistical signal processing, optimization, and machine learning.
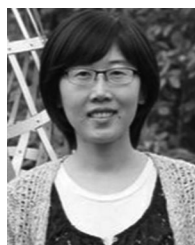
Yuejie Chi (S'09–M'12–SM'17) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, USA, in 2012, and the B.E. (Hon.) degree in electrical engineering from Tsinghua University, Beijing, China, in 2007. Since January 2018, she has been an Associate Professor with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, after spending five years at The Ohio State University, Columbus, OH, USA. Her research interests include signal processing, machine learning, large-scale optimization, and their applications in data science, inverse problems, imaging, and sensing systems. She is the recipient of the IEEE Signal Processing Society Young Author Best Paper Award in 2013 and the Best Paper Award at the IEEE International Conference on Acoustics, Speech, and Signal Processing in 2012. She received the NSF CAREER Award in 2017, Young Investigator Program Awards from AFOSR and ONR in 2015, the Ralph E. Powe Junior Faculty Enhancement Award from Oak Ridge Associated Universities in 2014, the Google Faculty Research Award in 2013, the and Roberto Padovani scholarship from Qualcomm Inc. in 2010. Since January 2016, she has been an Elected Member of the MLSP and SPTM Technical Committees of the IEEE Signal Processing Society.