Compressive estimation in AWGN: general observations and a case study

Dinesh Ramasamy, Sriram Venkateswaran, Upamanyu Madhow
Department of ECE, University of California Santa Barbara, CA 93106, USA
Email: {dineshr, sriram, madhow}@ece.ucsb.edu

***Abstract*—Compressive random projections followed by ℓ_1 reconstruction is by now a well-known approach to capturing sparsely distributed information, but applying this approach via discretization to estimation of continuous-valued parameters can perform poorly due to basis mismatch. However, we show in this paper it is still possible to capture the information required for effective estimation using a small number of random projections. We characterize the isometries required for preserving the geometric structure of estimation in additive white Gaussian noise (AWGN) under such compressive measurements. Under these conditions, estimation-theoretic quantities such as the Cramer-Rao Lower Bound (CRLB) are preserved, except for attenuation of the Signal-to-Noise Ratio (SNR) by the dimensionality reduction factor. For the canonical problem of frequency estimation of a single sinusoid based on N uniformly spaced samples, we show that the required isometries hold for $M = O(\log N)$ random projections, and that the CRLB scales as predicted. While we prove isometry results for a single sinusoid, we present an algorithm to estimate *multiple* sinusoids from compressive measurements. Our algorithm combines coarse estimation on a grid with iterative Newton updates and avoids the error floors incurred by prior algorithms which apply standard compressed sensing with an oversampled grid. Numerical results are provided for spatial frequency (equivalently, angle of arrival) estimation for large (32×32) two-dimensional arrays.**

I. Introduction

Compressed sensing is by now firmly established as an effective means of extracting sparsely distributed information from high-dimensional observations: a canonical approach is to take a small number (logarithmic in the dimension of the observation) of random projections, and to then reconstruct the signal using these "compressive measurements" using ℓ_1 optimization (a number of alternative reconstruction techniques have also been developed). In this paper, we explore the use of compressive measurements for estimation of continuous-valued parameters from high-dimensional observations perturbed by AWGN.

Consider, for example, the problem of estimating the frequencies of a noisy mixture of sinusoids: this is a canonical problem with numerous applications such as Angle-of-Arrival estimation using arrays and pitch detection. The number of sinusoids in the mixture is often much smaller than the number of available samples, leading to a sparse signal structure in the frequency domain. While such sparsity immediately suggests the use of compressive measurements, standard compressed sensing requires sparsity in a finite basis, while the frequencies come from a continuum, so that sparsity is over an infinite basis. Reconstruction using standard compressed sensing techniques by restricting frequencies to a finite grid leads to error floors due to "basis mismatch" [1]. The results in this paper indicate, however, that compressive measurements, along with appropriately designed estimation algorithms, provide an effective framework for dimensionality reduction while avoiding discretization artifacts.

Our results are summarized as follows: Suppose a discrete time signal $\mathbf{x}(\boldsymbol{\theta})$ parametrized by a K dimensional $\boldsymbol{\theta} = (\theta_1, ..., \theta_K)^T$ is observed in discrete time AWGN. The performance of coarse-grained estimation of the parameter (which can be viewed as a detection problem) depends on Euclidean distances of the type $\|\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')\|$, while the Cramer-Rao lower bound depends on linear combinations of the partial derivatives $\{\partial \mathbf{x}(\boldsymbol{\theta})/\partial \theta_k\}$ (i.e., the tangent planes). If we project down to an M-dimensional space along randomly chosen unit vectors, we expect to capture a fraction M/N of the signal energy, but the noise remains roughly white with the same variance per dimension, so that the SNR is scaled by M/N. Thus, if the projection preserves the Euclidean geometry governing detection and estimation performance, then the Fisher information and CRLB are as in the original problem, except that the SNR is scaled by M/N. We characterize the isometries required for preserving this geometric structure, and note that prior results on random projections on manifolds [2] appear to indicate that such isometries can be achieved under rather general conditions. However, a computational characterization of the manifold in order to obtain explicit estimates of the number of projections required is difficult. We therefore focus on developing a thorough, self-contained understanding of the canonical problem of frequency estimation.

We show that $M = O(\log N)$ compressive measurements suffice to preserve the relevant geometries of the frequency estimation problem. Our approach is similar to that in [2], [3]. For standard compressed sensing, the restricted isometry property (RIP) on norm preservation for sparse signals was shown in [3] to be a consequence of the Johnson-Lindenstrauss (JL) lemma, which specifies how the geometry of points in high-dimensional spaces is preserved under randomized mappings to lower dimensions. We infer the isometries required for estimation of a frequency lying in a continuum by extending

This work was supported by the National Science Foundation through the grant CNS-0832154 and by the Institute for Collaborative Biotechnologies through the grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

the JL lemma (that applies to a finite set of points) to the continuum of frequencies using covering arguments similar to those in [2] (which considers a more general manifold setting). Once we establish these isometries, we can immediately infer that the CRLB for frequency estimation with compressive measurements $O(\sigma^2/N^2M)$, applying the SNR scaling M/N to the CRLB $O(\sigma^2/N^3)$ for the original problem. While these results are proved for a single sinusoid, we propose an algorithm to estimate *multiple* sinusoids from compressive measurements. Our algorithm, which combines estimation on a coarse-grained grid, employs iterative Newton updates to avoid error floors, and is shown to approach the CRLB. Our numerical results illustrate the efficacy of our algorithm for estimation of two-dimensional spatial frequencies, or angles of arrival, motivated by millimeter wave applications in which very large (e.g., 32×32) arrays can be realized in compact form factor.

Related Work: There is a rich body of work devoted to the theory of compressed sensing [4], [5]. It was shown in [2] that a class of manifolds can be stably embedded using compressive measurements via covering arguments. As already mentioned, we use similar arguments in our proofs. Parameter estimation from compressive measurements was discussed in [6], but without any assumptions on the measurement noise. We observe that the results in [2], [6] can be applied to provide a general framework for continuous parameter estimation in AWGN. However, the connection to the CRLB was not made in these references, and using these results to compute the number of measurements needed for a particular application, including for the frequency estimation problem considered here, is not straightforward. The poor performance of standard ℓ_1 reconstruction after a naïve discretization of a continuous parameter was discussed in [1]. Recovery algorithms for sparse frequency estimation are considered in [7] [8], but these consider oversampled grids and specify the number of measurements required to recover the signal in terms of the size of the grid used. In contrast, we focus on the problem of estimating the original continuous-valued frequency, and provide algorithms that can bootstrap with a much coarser grid, while avoiding error floors due to gridding by using Newton methods. The estimation algorithm improves on our prior work [9] and also extends it to multi-dimensional frequencies.

II. Cramer-Rao bounds for compressive parameter estimation

We wish to estimate a parameter $\boldsymbol{\theta} \in \mathbb{R}^K$ from M random projections of the signal manifold $\mathbf{x}(\boldsymbol{\theta}) \in \mathbb{R}^N$ of the form

$$y_i = \mathbf{w}_i^T \left(\mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}_i\right), \quad i = 1, 2, \dots M, \tag{1}$$

where $\mathbf{w}_i$ contains the projection weights whose entries are chosen i.i.d. from $\text{Uniform}\{\pm 1/\sqrt{M}\}$ and $\mathbf{z}_i \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_N)$. Furthermore, we assume that the measurement noise $\mathbf{z}_i$ is independent across i. We have set the variance of the elements of $\mathbf{w}_i$ to $1/M$ for convenience. Stacking these observations, we get

$$\mathbf{y} = \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}, \tag{2}$$

where the ith row of $\boldsymbol{A}$ is $\mathbf{w}_i^T$ and $\mathbf{z} \sim \mathcal{N}(0, (N\sigma^2/M)\mathbb{I}_M)$. The preceding normalization is chosen to preserve signal norms on average ($\mathbb{E}\left[\|\boldsymbol{A}\mathbf{x}(\boldsymbol{\theta})\|^2\right] = \|\mathbf{x}(\boldsymbol{\theta})\|^2$), but amplifies the noise variance per dimension by N/M, which is the SNR penalty for dimension reduction.

In order to estimate $\boldsymbol{\theta}$ accurately, we require that $\boldsymbol{A}$ preserve norms in a stronger sense, called ϵ-isometry, defined as follows.

Definition 1. ***ϵ-isometry property [2]:*** *Consider a set $\mathcal{H}$ with elements in $\mathbb{R}^N$. We say that a matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ has an ϵ-isometry property ($\epsilon > 0$) for the set $\mathcal{H}$, if for all $\mathbf{v} \in \mathcal{H}$ and some $C > 0$, we have,*

$$1 - \epsilon \ \leq \ C\frac{\|\boldsymbol{A}\mathbf{v}\|}{\|\mathbf{v}\|} \ \leq \ 1 + \epsilon \ .$$

Remark: It is worth emphasizing that, for the RIP in standard compressed sensing [3] , $\mathcal{H}$ is the set of all $2K$-sparse vectors, whereas we require that ϵ-isometry hold for pairwise differences of points on the manifold and the tangent planes.
Intuitive interpretation: For estimation of a parameter that lies in a discrete grid, performance in AWGN depends on distances of the form $\|\mathbf{x}(\boldsymbol{\theta}_1) - \mathbf{x}(\boldsymbol{\theta}_2)\|$, normalized by the noise standard deviation. If the measurement matrix $\boldsymbol{A}$ satisfies ϵ-isometry for vectors of the form $\mathbf{x}(\boldsymbol{\theta}_1) - \mathbf{x}(\boldsymbol{\theta}_2), \forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, then the performance is as in the original system, except for the amplification of the noise variance. However, for the continuous-valued parameter estimation problem, we must go further, and ask that (appropriately normalized) distances be preserved as $\boldsymbol{\theta}_1 \rightarrow \boldsymbol{\theta}_2$. We rigorize this intuition by showing that, if $\boldsymbol{A}$ provides an ϵ-isometry for all tangent planes – vectors of the form $\sum_k a_k \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_k}$ for some $a_k \in \mathbb{R}$ – then the Fisher Information Matrix (FIM) with compressive measurements is a scaled version of the FIM with all the samples, with the scale factor being the SNR penalty M/N.

Denoting the FIM for the model in (2) by I^A, we can show that its $(m,n)^{th}$ sample is given by

$$I^A_{m,n} = \frac{M}{N\sigma^2}\left\langle \boldsymbol{A}\frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m}, \boldsymbol{A}\frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_n} \right\rangle.$$

We denote the FIM obtained when we have access to all the samples (setting $\mathbf{w}_i$ to the unit vector with a 1 in the ith position for $1 \leq i \leq N$ in (1)) by I^{all}. The $(m,n)th$ entry in I^{all} is given by

$$I^{all}_{m,n} = \frac{1}{\sigma^2}\left\langle \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m}, \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_n} \right\rangle.$$

In relating the Cramer-Rao lower bounds (CRLB) obtained in these cases, we are interested in the behavior of quadratic forms $\mathbf{a}^T I \mathbf{a}$, where I is a generic FIM and $\mathbf{a}$ is an arbitrary vector. With some manipulation, we can simplify these as follows:

$$\begin{aligned}
\mathbf{a}^T I^A \mathbf{a} &= \frac{M}{N\sigma^2}\left\| \boldsymbol{A}\sum_m a_m \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m} \right\|^2, \\
\mathbf{a}^T I^{all} \mathbf{a} &= \frac{1}{\sigma^2}\left\| \sum_m a_m \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m} \right\|^2.
\end{aligned}$$

If $\boldsymbol{A}$ preserves norms of vectors in the tangent plane up to ϵ, we get

$$\frac{M}{N}\mathbf{a}^T I^{all}\mathbf{a}(1-\epsilon)^2 \leq \mathbf{a}^T I^A \mathbf{a} \leq \frac{M}{N}\mathbf{a}^T I^{all}\mathbf{a}(1+\epsilon)^2 \quad \forall \mathbf{a}. \quad (3)$$

This gives the following theorem.

Theorem 1. *Let $\mathbf{y} = \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta})+\mathbf{z}$ be compressive measurements with the entries in $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ chosen i.i.d. with variance $1/M$ and $\mathbf{z} \sim \mathcal{N}(0, N\sigma^2/M)$. We denote the Fisher Infomation Matrix with this model by I^A and the corresponding FIM with all samples by I^{all} (obtained by replacing $\boldsymbol{A}$ with $\mathbb{I}_N$ in the model and, therefore, setting $M = N$). If $\boldsymbol{A}$ provides an ϵ-isometry for all vectors of the form $\sum_k a_k \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_k}$, we have*

$$\frac{M}{N}(1-\epsilon)^2 I^{all} \preceq I^A \preceq \frac{M}{N}(1+\epsilon)^2 I^{all}. \quad (4)$$

Thus, compressive measurement matrices that provide ϵ-isometry for all tangent planes preserve estimation bounds, except for an SNR penalty M/N.

It is interesting to note that, in the process of proving Theorem 3.1 in [2], the authors show that compressive measurements provide ϵ-isometry for all tangent planes with probability $1-\rho$ provided

$$M = O\left(\epsilon^{-2}\log(1/\rho)K\log\left(NVR\tau^{-1}\epsilon^{-1}\right)\right), \quad (5)$$

where $0 < \epsilon, \rho < 1$ and V, R, τ are properties of the manifold ($1/\tau$ is the condition number, R is the geodesic covering regularity and V is the volume). Taken in conjunction with our result, (5) shows that compressive measurements preserve estimation bounds in rather general settings. However, to the best of our knowledge, it is difficult to specify how V, R, τ scale with N and K in general, hence we focus here on a self-contained derivation of the required isometries for the specific setting of compressive frequency estimation.

III. Compressive frequency estimation of a single sinusoid

We denote an N dimensional sinusoid of frequency ω by $\mathbf{x}(\omega)$ and define it as

$$\mathbf{x}(\omega) = [h_0 \;\; h_1 e^{j\omega} \;\; \cdots \;\; h_{N-1}e^{j\omega(N-1)}]^T,$$

where the h_n are *known* windowing weights normalized so that $\sum |h_n|^2 = 1$ and $\max_n |h_n|^2 < 1$. Such windows are used to reduce spectral leakage and two examples are the Hamming and Chebyshev windows. We make compressive measurements of the form

$$\mathbf{y} = \boldsymbol{A}\left(g\mathbf{x}(\omega)\right) + \mathbf{z}, \quad (6)$$

where g is the complex gain of the sinusoid, $\boldsymbol{A}$ is an $M \times N$ measurement matrix ($M \ll N$) whose entries are chosen uniformly at random from $\{\pm 1/\sqrt{M}, \pm j/\sqrt{M}\}$ independent of each other and $\mathbf{z} \sim \mathcal{CN}(0, \sigma^2\mathbb{I}_M)$ is the measurement noise.

We note a slight change of notation from the previous section: (1) the set of parameters defining the manifold are the gain g and the frequency ω. However, we denote the sinusoidal manifold by $g\mathbf{x}(\omega)$, with $\mathbf{x}(\omega)$ denoting the sinusoid, rather than $\mathbf{x}(g, \omega)$ for clarity. (2) The gain g and the manifold $\mathbf{x}(\omega)$ are complex-valued, unlike the real-valued manifolds we considered earlier.

Using the intuition of good measurement matrices for parameter estimation that we developed earlier, we see that $\boldsymbol{A}$ must satisfy

$$\|g_1\boldsymbol{A}\mathbf{x}(\omega_1) - g_2\boldsymbol{A}\mathbf{x}(\omega_2)\| \approx \|g_1\mathbf{x}(\omega_1) - g_2\mathbf{x}(\omega_2)\|.$$

Equivalently, $\boldsymbol{A}$ must provide an ϵ-isometry for the subspace spanned by $\mathbf{x}(\omega_1)$ and $\mathbf{x}(\omega_2)$ for different choices of ω_1 and ω_2. We now show that $M = O(\log N)$ measurements suffice for this condition to hold.

A. Isometry for subspace of two sinusoids

Let $H(\omega)$ be the DTFT of $\{|h_n|^2\}$. We define ζ_h as

$$\zeta_h = -\frac{N^{-2}}{2!}\left.\frac{\partial^2|H(\omega)|^2}{\partial\omega^2}\right|_{\omega=0} \quad (7)$$

and note that for $h_n = 1/\sqrt{N}\ \forall n$, $\zeta_h \uparrow 1/12$ rapidly as N increases.

Theorem 2. *Let $\boldsymbol{A}$ be an $M \times N$ measurement matrix whose entries are drawn i.i.d. from Uniform$\{\pm 1/\sqrt{M}, \pm j/\sqrt{M}\}$. For any two frequencies ω_1, ω_2 such that $|\omega_1 - \omega_2| > \delta/N^{1.5}$ and $\epsilon > 0$,*

$$\left|\frac{\|g_1\boldsymbol{A}\mathbf{x}(\omega_1) - g_2\boldsymbol{A}\mathbf{x}(\omega_2)\|}{\|g_1\mathbf{x}(\omega_1) - g_2\mathbf{x}(\omega_2)\|} - 1\right| < \epsilon \;\; \forall g_1, g_2 \in \mathbb{C} \quad (8)$$

with high probability when $M = O\left(\epsilon^{-2}\log\left(N\epsilon^{-1}\delta^{-1}\zeta_h^{-1}\right)\right)$.

We note that it is possible to prove a result similar to Theorem 2 for frequencies that are arbitrarily close (proof omitted owing to lack of space). However, the CRLB on the variance of the frequency estimate is $O(\sigma^2/N^3)$ even with all N measurements [10], so that we can only hope to achieve frequency estimation errors on the order of $1/N^{1.5}$ at any finite SNR.

Proof sketch: The proof employs techniques used in [2]. For a finite collection of points $\mathcal{Q}$, by employing the Chernoff bound on the deviations of $\|\boldsymbol{A}\mathbf{x}\|^2$ and following it up with the union bound, we can show that $M = O(\epsilon_0^{-2}\log|\mathcal{Q}|)$ measurements suffice to give an ϵ_0-isometry for $\mathcal{Q}$. The goal is to chose a "fine enough" sampling $\mathcal{Q}$ of the manifold so that the ϵ_0-isometry of the point cloud can be extended to an $8\epsilon_0$-isometry for all points of interest. $|\mathcal{Q}|$ will determine the number of measurements M required.

From (8), we see that we need an ϵ-isometry for the span of $\mathbf{X}(\omega_1, \omega_2) = [\mathbf{x}(\omega_1) \;\; \mathbf{x}(\omega_2)]$ for all $|\omega_1 - \omega_2| > \delta/N^{1.5}$. We first discretize the frequencies in $[0, 2\pi]$ with R uniform samples to form the set F. We then establish a $2\epsilon_0$ isometry for the span of $\mathbf{X}(q_1, q_2)$ $q_1, q_2 \in F$. We extend this to an $\epsilon \equiv 8\epsilon_0$ isometry for the span of $\mathbf{X}(\omega_1, \omega_2)$ for $|\omega_1 - \omega_2| > \delta/N^{1.5}$ by choosing R large enough (fine frequency discretization). By characterizing the smallest singular value of $\mathbf{X}(\omega_1, \omega_2)$ for

$|\omega_1 - \omega_2| > \delta/N^{1.5}$, we find that $R = O(N^2\delta^{-1}\zeta_h^{-0.5}\epsilon_0^{-1})$ suffices. From the covering argument used in [3], an ϵ_0 isometry of $(6\epsilon_0^{-1})^4$ well chosen points in the span of $\mathbf{X}(q_1, q_2)$ can be extended to a $2\epsilon_0$ isometry for all points in the span of $\mathbf{X}(q_1, q_2)$. There are $R^2/2$ pairs of frequencies in F ($q_1 > q_2$) and to give the $2\epsilon_0$ isometry that we need for the span of $\mathbf{X}(q_1, q_2)$ for all $q_1, q_2 \in F$ we see that $|\mathcal{Q}| = R^2(6\epsilon_0^{-1})^4/2$ points need to be given an ϵ_0 isometry. Substituting for R, we find that $M = O\left(\epsilon^{-2}\log\left(N\epsilon^{-1}\delta^{-1}\zeta_h^{-1}\right)\right)$ measurements suffice.

B. Isometry for tangent planes

We quantify the number of measurements needed for $\boldsymbol{A}$ to provide an isometry for the tangent planes of the sinusoidal manifold (given by the span of $[\mathbf{x}(\omega)\ \partial\mathbf{x}(\omega)/\partial\omega]\ \forall\omega$). In the process, we find that $M = O\left(\log N\delta^{-1}\right)$ measurements suffice for providing isometries for both the tangent planes and for Theorem 2. Using this isometry, we infer that the FIM with compressive measurements is M/N times the FIM with access to all the samples.

Theorem 3. *Let $\boldsymbol{A}$ be an $M \times N$ measurement matrix whose entries are drawn i.i.d. from Uniform$\{\pm 1/\sqrt{M}, \pm j/\sqrt{M}\}$. For all $\omega \in [0, 2\pi]$, $g_1, g_2 \in \mathbb{C}$ and $\epsilon > 0$,*

$$\left| \frac{\|g_1\boldsymbol{A}\mathbf{x}(\omega) + g_2\boldsymbol{A}\partial\mathbf{x}(\omega)/\partial\omega\|}{\|g_1\mathbf{x}(\omega) + g_2\partial\mathbf{x}(\omega)/\partial\omega\|} - 1 \right| < \epsilon \qquad (9)$$

with high probability when $M = O\left(\epsilon^{-2}\log N\epsilon^{-1}\right)$.

Proof sketch: The proof proceeds along the same lines of Theorem 2. We find that $\boldsymbol{A}$ needs to provide an ϵ_0-isometry for $O(N^{1.5})$ points in order to give an $8\epsilon_0$-isometry for all tangent planes. We note that this is smaller than the number of samples for which $\boldsymbol{A}$ had to preserve norms for Theorem 2 to hold, given by $O\left(N^4\right)$. Thus, by applying the union bound to all samples required for Theorems 2 and 3, we find, as before, that $M = O\left(\epsilon^{-2}\log\left(N\epsilon^{-1}\delta^{-1}\zeta_h^{-1}\right)\right)$ measurements suffice to guarantee an ϵ isometry for the spans of $\mathbf{X}(\omega_1, \omega_2)$, $|\omega_1 - \omega_2| > \delta/N^{1.5}$ and $[\mathbf{x}(\omega)\ \partial\mathbf{x}(\omega)/\partial\omega], \forall\omega$.

CRLB for compressive frequency estimation: Combining Theorem 3 with Theorem 1, we infer that the CRLB for compressive frequency estimation is $O(\sigma^2/N^2M)$; this is obtained by scaling the well-known [10] CRLB for the original system, $O(\sigma^2/N^3)$, by the SNR attenuation factor M/N.

IV. Frequency estimation algorithm

In this section, we propose an algorithm to estimate frequencies from compressive measurements. We generalize the setup we have been considering in two ways: (1) we allow the signal to contain multiple sinusoids and (2) we investigate two-dimensional frequencies, motivated by the problem of AoA estimation using large arrays explained in [9]. The (m, n)-th sample we obtain (without compressive measurements) takes the form $ge^{j(\omega_x m + \omega_y n)}$, $0 \le m, n \le N_{1\mathrm{D}} - 1$ where $\boldsymbol{\omega} = (\omega_x, \omega_y)$ is the two-dimensional frequency we wish to estimate. We denote a suitably vectorized version of $e^{j(\omega_x m + \omega_y n)}/\sqrt{N}$, where $N = N_{1\mathrm{D}}^2$ by $\mathbf{x}(\boldsymbol{\omega})$ and observe that $\|\mathbf{x}(\boldsymbol{\omega})\| = 1$. We denote the kth among K frequencies by $\boldsymbol{\omega}_k$ and let $\mathbf{S}(\boldsymbol{\omega}) = \boldsymbol{A}\mathbf{x}(\boldsymbol{\omega})$. Making compressive measurements using a matrix $\boldsymbol{A}$ whose entries are i.i.d. Uniform$\{\pm 1/\sqrt{M}, \pm j/\sqrt{M}\}$, we get

$$\mathbf{y} = \sum_k g_k\mathbf{S}(\boldsymbol{\omega}_k) + \mathbf{z}\ ;\ \mathbf{z} \sim \mathcal{CN}(\mathbf{0}, (N\sigma^2/M)\mathbb{I}_M).$$

Single Sinusoid: We first present our algorithm for the case when there is only one frequency ($K = 1$). The GLRT estimate $\hat{\boldsymbol{\omega}}_1$ is given by the spatial frequency that maximizes the normalized correlation $J(\mathbf{y}, \boldsymbol{\omega})$:

$$J(\mathbf{y}, \boldsymbol{\omega}) = \left|\mathbf{S}^H(\boldsymbol{\omega})\mathbf{y}\right|^2 / \|\mathbf{S}(\boldsymbol{\omega})\|^2 .$$

The first step is to estimate $\boldsymbol{\omega}_1$ coarsely by picking the maximum of $J(\mathbf{y}, \boldsymbol{\omega})$ from a discrete set of frequencies $(2\pi m/R, 2\pi n/R)$ with $0 \le m, n \le R-1$, for some $R \ge N_{1\mathrm{D}}$. We denote this estimate by $\hat{\boldsymbol{\omega}}_1$. We refine $\hat{\boldsymbol{\omega}}_1$ using Newton's method to find a local optimum of $J(\mathbf{y}, \boldsymbol{\omega})$. Denoting the estimate of $\boldsymbol{\omega}_1$ at the start of the i-th Newton step by $\boldsymbol{\omega}_1^{(i)}$, the estimate after this step $\boldsymbol{\omega}_1^{(i+1)}$ is given by

$$\boldsymbol{\omega}_1^{(i+1)} = \boldsymbol{\omega}_1^{(i)} - \left(H_\nabla J(\boldsymbol{\omega}_1^{(i)})\right)^{-1}\nabla J(\boldsymbol{\omega}_1^{(i)}), \qquad (10)$$

where ∇J denotes the gradient of $J(\mathbf{y}, \boldsymbol{\omega})$ and $H_\nabla J$ its Hessian with respect to $\boldsymbol{\omega}$. When $\boldsymbol{\omega}_1^{(i)}$ is in a strictly concave region ($H_\nabla J$ is negative definite), a Newton step will take us closer to the maximum of $J(\mathbf{y}, \boldsymbol{\omega})$. When this is not the case, we exploit the observation that $\max J(\mathbf{y}, \boldsymbol{\omega}) \approx \|\mathbf{y}\|^2$ and modify the rule to

$$\boldsymbol{\omega}_1^{(i+1)} = \boldsymbol{\omega}_1^{(i)} + \frac{\|\mathbf{y}\|^2 - J(\mathbf{y}, \boldsymbol{\omega}_1^{(i)})}{\|\nabla J(\mathbf{y}, \boldsymbol{\omega}_1^{(i)})\|^2}\nabla J(\mathbf{y}, \boldsymbol{\omega}_1^{(i)}). \qquad (11)$$

Multiple sinusoids: When $K > 1$ beams are present, we obtain estimates of $\{\boldsymbol{\omega}_k\}_{k=1}^{k=K}$ via a matching pursuit that seeks the best greedy explanation of the observation $\mathbf{y}$ over the continuum of frequencies in $[0, 2\pi]^2$. The kth pursuit step that adds the kth frequency from the continuum consists of two stages: coarse detection of a new frequency from the $(2\pi m/R, 2\pi n/R)$ grid followed by a local refinement of all the frequencies that have been detected so far.

Suppose that $k-1$ frequencies $\{\hat{\boldsymbol{\omega}}_1, \cdots, \hat{\boldsymbol{\omega}}_{k-1}\}$ have been estimated so far. We want to detect the kth frequency and add it to our list of estimated frequencies. The frequencies estimated so far can explain measurements that lie in the span of $\mathcal{B}_{k-1} = [\mathbf{S}(\hat{\boldsymbol{\omega}}_1)\ \cdots\ \mathbf{S}(\hat{\boldsymbol{\omega}}_{k-1})]$. So we compute $\mathbf{r}_{k-1} = \mathcal{B}_{k-1}^{\perp}\mathbf{y}$, where $\mathcal{B}_{k-1}^{\perp}$ is the matrix representing the projection onto the subspace orthogonal to the span of $\mathcal{B}_{k-1}$. The first stage involves greedily picking $\hat{\boldsymbol{\omega}}_k$ to be the frequency on the $(2\pi m/R, 2\pi n/R)$ grid that best fits the current residual measurement $\mathbf{r}_{k-1}$:

$$\hat{\boldsymbol{\omega}}_k = \underset{\boldsymbol{\omega} \in \left(\frac{2\pi m}{R}, \frac{2\pi n}{R}\right)}{\arg\max}\ J(\mathbf{r}_{k-1}, \boldsymbol{\omega}).$$

Once a coarse estimate of the kth frequency is made, we refine all the $1, \ldots, k$ frequencies that have been detected so far as follows. While refining the lth frequency, we first project out

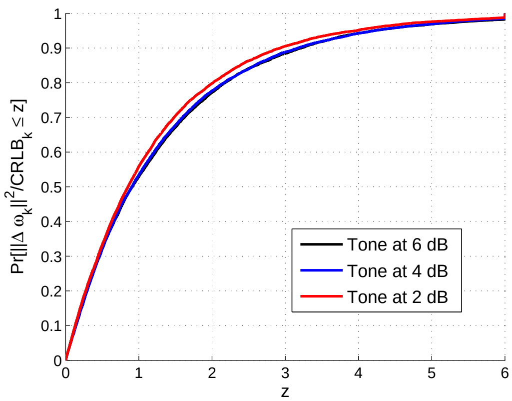

Fig. 1. Empirical CDF of ratios of squared frequency error and its CRLB (i.e., $\|\boldsymbol{\omega}_k - \hat{\boldsymbol{\omega}}_k\|^2/\mathrm{CRLB}_k$) for a mixture of $K = 3$ sinusoids of SNR 6, 4, 2 dB per sample from $M = 48$ compressive measurements taken from a 32×32 array($N = 1024$ element array). The coarse estimates are made using an $R = 2N_{1\mathrm{D}}$ grid. CRLB_k is the Cramer-Rao lower bound on the mean-square error when observing all N samples with an SNR penalty M/N.

the responses of the $k-1$ other frequencies from $\mathbf{y}$ to obtain $\tilde{\mathbf{y}}_l$. Assuming that $\boldsymbol{\omega}_l$ is the only frequency present, we refine $\hat{\boldsymbol{\omega}}_l$ by maximizing $J(\tilde{\mathbf{y}}_l, \boldsymbol{\omega})$ in the neighborhood of $\hat{\boldsymbol{\omega}}_l$ using the same procedure as in the single frequency case. We refine all the k frequencies sequentially and this constitutes a single round of Newton refinements. We go through multiple rounds (typically 3) of refinement and note that one update of the form (10) or (11) per frequency per round suffices. Once we have detected and refined K frequencies we terminate the pursuit.

Simulation Results: We demonstrate the effectiveness of the above algorithm, comparing it with the CRLB for compressive estimation of angles of arrival using a 1024 element 2D array (i.e., $N_{1\mathrm{D}} = 32$). The spatial frequency for an angle of arrival (θ, ϕ) is given by $\boldsymbol{\omega} = 2\pi d \sin\theta(\cos\phi, \sin\phi)$, where d is the inter-element spacing normalized by carrier wavelength. We consider a mixture of $K = 3$ such spatial sinusoids whose amplitudes $|g_k|^2/\sigma^2$ are set to 6, 4 and 2 dB respectively. We choose the spatial frequencies $\boldsymbol{\omega}_k$ and phases of g_k uniformly at random from $[0, 2\pi]^2$ and $[0, 2\pi]$ respectively and make $M = 48$ compressive measurements. The coarse estimates are obtained using an $R \times R$ grid of the frequencies with $R = 2N_{1\mathrm{D}}$. We run 2000 such trials and constrain each Newton update step to a maximum of π/R. Each refinement stage consists of three round robin refinements with one Newton update per frequency. In Figure 1, we plot the empirical CDFs of the ratio of the squared frequency estimation errors $\|\boldsymbol{\omega}_k - \hat{\boldsymbol{\omega}}_k\|^2$ and the corresponding bounds, CRLB_k. The CRLB for each parameter setting is estimated simply by scaling the CRLB with all measurements using the SNR attenuation factor M/N, rather than accounting for the specific measurement matrix used. We see that the median of the ratio $\|\boldsymbol{\omega}_k - \hat{\boldsymbol{\omega}}_k\|^2/\mathrm{CRLB}_k$ is approximately 1 for all the tones independent of their SNRs. This shows that our algorithm is efficient (and does not incur discretization artifacts) and that it effectively overcomes the interference among tones. It also indicates that that the CRLB estimate provided by Theorem 1 provides a practical benchmark for the performance of compressive estimation.

V. Conclusions

We have identified isometry conditions required to preserve the geometry of estimation in AWGN under compressive measurements, and shown that fundamental estimation-theoretic quantities such as the Fisher information and the CRLB are preserved under these conditions, up to scaling by a natural SNR attenuation factor. Thus, the poor performance of standard compressed sensing for such problems, as observed in prior work, stem from naïve discretization; dimensionality reduction via randomized projections is not to blame. We show that the desired ϵ-isometry conditions do hold for compressive frequency estimation for a single noisy sinusoid of length N, with $M = O(\log N)$ random projections, and thereby infer the CRLB for this problem. Our algorithm for frequency estimation, when applied to a mixture of sinusoids, approaches the CRLB for each individual sinusoid; thus, it avoids gridding errors, despite being initiated by search over a coarse grid, and is not interference-limited. The combination of coarse estimation and Newton refinements in our algorithm parallels the nature of the isometry requirements that we have identified: isometries preserving vector differences, corresponding to large-scale "discrete" geometry determining coarse estimation (or detection) performance, and isometries preserving tangent planes, which determines infinitesimal variations in geometry corresponding to the small-scale "continuous" variations captured by estimation-theoretic bounds. An interesting topic for future research is to provide tight estimates of the number of measurements required to maintain isometry for a mixture of sinusoids, and to explore whether the general characterization of random projections on manifolds in [2] yields tight and computable estimates for applications of interest.

References

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

[2] R. Baraniuk and M. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, no. 1, p. 5177, 2009.

[3] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, p. 253263, 2008.

[4] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289 –1306, april 2006.

[5] E. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406 –5425, dec. 2006.

[6] M. Wakin, "Manifold-based signal recovery and parameter estimation from compressive measurements," *Arxiv preprint arxiv:1002.1247*, 2010.

[7] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, Submitted.

[8] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59–73, Jul. 2011.

[9] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive adaptation of large steerable arrays," in *Information Theory and Applications Workshop (ITA), 2012*, feb. 2012, pp. 234 –239.

[10] D. Rife and R. Boorstyn, "Single tone parameter estimation from discrete-time observations," *Information Theory, IEEE Transactions on*, vol. 20, no. 5, pp. 591 – 598, sep 1974.