Compressive parameter estimation in AWGN

Dinesh Ramasamy, Sriram Venkateswaran and Upamanyu Madhow

Abstract—**Compressed sensing is by now well-established as an effective tool for extracting sparsely distributed information, where sparsity is a *discrete* concept, referring to the number of dominant nonzero signal components in some basis for the signal space. In this paper, we establish a framework for estimation of *continuous-valued* parameters based on compressive measurements on a signal corrupted by additive white Gaussian noise (AWGN). While standard compressed sensing based on naive discretization has been shown to suffer from performance loss due to basis mismatch, we demonstrate that this is not an inherent property of compressive measurements. Our contributions are summarized as follows: (a) We identify the isometries required to preserve fundamental estimation-theoretic quantities such as the Ziv-Zakai bound (ZZB) and the Cramér-Rao bound (CRB). Under such isometries, compressive projections can be interpreted simply as a reduction in "effective SNR." (b) We show that the threshold behavior of the ZZB provides a criterion for determining the minimum number of measurements for "accurate" parameter estimation. (c) We provide detailed computations of the number of measurements needed for the isometries in (a) to hold for the problem of frequency estimation in a mixture of sinusoids. We show via simulations that the design criterion in (b) is accurate for estimating the frequency of a single sinusoid.**

I. INTRODUCTION

Compressed sensing has proven remarkably successful in exploiting sparsity to extract information from signals with only a small number of measurements. The standard approach has two stages. First, take multiple random projections of the signal, with the number of projections growing linearly with the sparsity and only logarithmically with the dimensionality of the signal. Then, use one among a variety of recovery algorithms, such as ℓ_1 reconstruction/Orthogonal Matching Pursuit (OMP), to estimate the signal from the random projections. In this standard framework, sparsity is an inherently *discrete* concept: the number of nonzero signal components in some basis has to be small compared to the dimension of the signal. In this paper, we investigate the effectiveness of compressive measurements in estimating *continuous valued* parameters from signals that are corrupted by AWGN, when the dimensionality of the parameter set is much smaller than the signal dimension.

It is possible to apply standard compressed sensing to continuous-valued parameter estimation, but it does not per-

form well. Consider the fundamental problem of estimating the frequencies in a mixture of sinusoids. Typically, the number of sinusoids is much smaller than the number of samples and, therefore, the signal is sparse in the frequency domain. However, the conventional compressed sensing framework does not apply directly, since it requires the signal to be sparse over a *finite* basis, whereas the frequencies could lie anywhere on a continuum. Straightforward application of compressed sensing recovery algorithms after discretizing the set of frequencies has been shown to result in error floors due to "basis mismatch" and the consequent spectral leakage [1]. This observation raises some fundamental questions. Do compressive measurements preserve all the information needed for continuous valued parameter estimation? If so, under what conditions? How many measurements do we require to satisfy these conditions? In this paper, we establish a systematic framework that addresses these questions for parameter estimation based on signals corrupted by AWGN.

Contributions: We first identify fundamental structural properties for compressive estimation in AWGN, and then illustrate them by explicit computation for frequency estimation for a mixture of sinusoids.

Isometries for estimation: Suppose we want to estimate a K dimensional parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)$ from projections of an N dimensional signal $\mathbf{x}(\boldsymbol{\theta})$ in AWGN. When we make all N measurements, fundamental bounds on the estimation error variance, such as the Ziv-Zakai bound (ZZB) and the Cramér Rao bound (CRB), relate the geometry of the signal manifold to the best achievable performance. From the ZZB, we can infer that "coarse" estimation depends on the pairwise distances $\|\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')\| \; \forall \boldsymbol{\theta}, \boldsymbol{\theta}'$, while the CRB tells us that "fine" estimation depends on norms of linear combinations of the partial derivatives $\{\partial \mathbf{x}/\partial \theta_k\}$ (vectors in the tangent plane, which are the limit of differences as $\boldsymbol{\theta} \to \boldsymbol{\theta}'$). We extend these observations to compressive estimation by replacing the signal manifold $\mathbf{x}(\boldsymbol{\theta})$ by $\boldsymbol{A}\mathbf{x}(\boldsymbol{\theta})$, where $\boldsymbol{A}$ is the compressive measurement matrix containing the random projection weights. We identify the isometries required to ensure that the geometry (and hence the structure of the ZZB and CRB) is roughly unaltered after the compressive projection. We also note that, if these isometries hold, then the only consequence of compressive projection onto a subspace of dimension M is an SNR penalty of M/N. This is because each random projection captures $1/N$ of the signal energy on average (normalizing such that the noise variance is unchanged). Specifically, we show that if the measurement matrix $\boldsymbol{A}$ satisfies the pairwise isometry property (PIP) ($\|\boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}) - \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}')\| \approx \sqrt{M/N} \|\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')\|$), the ZZB with compressive measurements is approximately equal to the ZZB with all N measurements, except for the SNR penalty of M/N. We prove an analogous result for the CRB when $\boldsymbol{A}$

guarantees tangent plane isometry ($\|\boldsymbol{A}\sum_k a_k \partial \mathbf{x}(\boldsymbol{\theta})/\partial\theta_k\| \approx \sqrt{M/N}\|\sum_k a_k \partial\mathbf{x}(\boldsymbol{\theta})/\partial\theta_k\|$, $\forall a_k$), which is a weaker requirement than pairwise isometry.

Number of measurements: When the preceding isometries hold, we can use their relationship to the ZZB/CRB to obtain a tight prediction on the number of measurements necessary for successful compressive estimation. It is known that nonlinear estimation problems exhibit a threshold behavior with the SNR which is closely mirrored by the threshold behavior of the ZZB. We employ this observation to predict the number of measurements required to avoid performance floors, since the the effective SNR with compressive measurements increases linearly with the number of measurements.

Computations for sinusoidal mixtures: While the preceding results reveal the structure of compressive estimation, computation of the number of measurements required to achieve the desired isometries and to avoid performance floors requires a problem-specific analysis. To this end, we consider the fundamental problem of frequency estimation for a mixture of sinusoids. For estimating K frequencies from N samples, we show that: (a) $O\left(K\log(NK\delta^{-1})\right)$ measurements suffice to provide tangent plane isometries, where δ depends on the frequency separation between the sinusoids in the mixture (δ vanishes when any two of the K frequencies approach one another). (b) $O\left(K\log(NK\delta^{-1})\right)$ measurements suffice to provide pairwise isometries between two sets of frequencies $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_K)$ and $\boldsymbol{\omega}' = (\omega'_1, \omega'_2, \ldots, \omega'_K)$ that are "well-separated." Here δ depends only on the frequency separation between the sinusoids in the mixture of $2K$ sinusoids $(\boldsymbol{\omega}, \boldsymbol{\omega}')$, and vanishes when any two frequencies in $(\boldsymbol{\omega}, \boldsymbol{\omega}')$ approach one another. Therefore, with $O\left(K\log(NK\delta^{-1})\right)$ compressive measurements, we can preserve the "well-separated" geometry of the frequency estimation problem. The tangent plane isometry results (a) indicate that when the K frequencies in $\boldsymbol{\omega}$ themselves are "well-separated", compressive measurements preserve the "fine" geometry of the frequency estimation problem (and therefore the CRB). We strengthen these results for a single sinusoid ($K = 1$), exploiting the continuity of the sinusoidal manifold to show that $O(\log N)$ measurements suffice to guarantee pairwise isometry between sinusoids at *any* two frequencies ω, ω' (by merging the "well-separated" and "fine" regimes). We also show that the criterion for prediction of the number of measurements, based on the threshold behavior of the ZZB, is tight, by evaluating the performance of an algorithm which closely approximates the MAP estimator. The algorithm works in two stages: first, from a discrete set of frequencies, we pick the one that fits the observations best and, then, we perform local refinements using Newton's method.

II. RELATED WORK

The goal of standard compressed sensing [2], [3] is to recover signals which are sparse over a finite basis with significantly fewer measurements than the dimension of the observation space. Signal recovery requires that the measurement matrix must satisfy the Restricted Isometry Property (RIP): the distance between any two sparse signals must be roughly

invariant under the action of the matrix. If the RIP is satisfied, sparse signals can be recovered efficiently using techniques such as Orthogonal Matching Pursuit (OMP) and ℓ_1-norm minimization. Reference [4] used the Johnson-Lindenstrauss (JL) lemma to provide a simple proof that $O(K\log N)$ random projections suffice to establish RIP for recovering K-sparse vectors in $\mathbb{R}^N$. We briefly summarize the key ideas, since we use an analogous approach in establishing pairwise isometry for the mixture of sinusoids example discussed in this paper. The JL lemma states that, to approximately preserve the pairwise distances between P points after random projections (with the weights chosen from appropriate distributions, such as i.i.d. ± 1 [5]), we need $O(\log P)$ such projections. However, to provide an RIP for compressive measurement matrices, the distances between any two K-sparse vectors must be preserved. Since the number of such vectors is infinite, the JL lemma cannot be applied directly. However, the desired RIP result is established in [4] by discretizing the set of K-sparse vectors sufficiently finely, applying the JL lemma to the resulting discrete set of points, and then exploiting continuity to provide isometries for the remaining points.

For compressive estimation of continuous-valued parameters, sparsity corresponds to the dimension of the parameter space K being significantly smaller than that of the observation space N. This problem was perhaps first investigated in [6], which identifies that the analogue of the RIP here is the pairwise isometry property considered in the present paper. However, it does not relate this property to estimation-theoretic bounds as done here. Reference [6] also shows that compressive measurements guarantee pairwise ϵ-isometry for a signal manifold with probability $1-\rho$, as long as the number of measurements M satisfies

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

where V, R, τ are properties of the signal manifold ($1/\tau$ is the condition number which is a generalization of the radius of curvature, R is the geodesic covering regularity and V is the volume). However, to the best of our knowledge, it is difficult to specify how $\{\tau, V, R\}$ scale with the parameters N and K in general. In this paper, therefore, we provide a self-contained derivation of the number of measurements required to preserve these isometries when the signal manifold consists of a mixture of sinusoids in Section VII. Compressive parameter estimation has also been studied in [7]; however, since the noise model there is adversarial, the results are pessimistic for many practical applications in which a Gaussian model for the noise is a good fit.

Algorithms to estimate the frequencies in a mixture of sinusoids from compressive measurements are proposed and evaluated in [8], [9]. Both of these papers assume that the sinusoids have a minimum frequency separation and [9] further assumes that the frequencies come from an oversampled DFT grid. They propose variants of standard compressed sensing algorithms, such as Orthogonal Matching Pursuit (OMP) and Iterative Hard Thresholding (IHT), which rely on the sinusoids' frequencies not being too close. As mentioned earlier, restricting the frequency estimation to a discrete grid in this fashion results in performance floors, as studied in great

detail in [1]. However, as shown in this paper and in our earlier conference papers [10], [11], it is possible to avoid such performance floors, and to attain the CRB, by local refinements based on Newton-like algorithms after grid-based coarse estimation. A one-shot quadratic refinement is also proposed in [8] to improve estimates of off-grid frequencies.

We characterize the structure of compressive estimation here in terms of that of the original problem. However, in many cases, an estimation-theoretic understanding of the original problem is incomplete: in particular, for the mixture of sinusoids model, a characterization of the difficulty of the problem in terms of the minimum separation of frequencies in $\boldsymbol{\omega}$ remains an ongoing effort [12], [13], [14], as discussed in more detail below.

The problem of estimating frequencies in a mixture of sinusoids from noise-free compressive measurements is studied in [14]. While the frequencies can come from the $[0, 2\pi)$ continuum, [14] requires that they are "well-separated" (four times the DFT spacing of $2\pi/N$). When this condition is met, it is shown that atomic-norm denoising (cast as a semi-definite program) correctly estimates the frequencies in the mixture. The same $4 \times (2\pi/N)$ frequency separation is shown to be necessary for recovering frequencies over a continuum with noisy measurements of all N samples (not compressive) in [12], [13]. It is interesting to note that even when all N samples are observed, the same minimum frequency separation is necessary for stable recovery. This falls in line with the observations that we make on the equivalence (except for an SNR penalty) of the "difficulty" in estimation using compressive measurements and uncompressed measurements (all N samples) by relating corresponding estimation error bounds.

To the best of our knowledge, other than our conference paper [10], this is the first paper to relate isometry conditions to estimation-theoretic bounds for compressive parameter estimation, and to show that, in the AWGN setting, the only effect of compressive measurements when appropriate isometry conditions are satisfied is an SNR penalty of M/N. The M/N SNR penalty due to compressive measurements has also been noted in [15], but we go further and make the connection between isometries and estimation bounds. Isometries and SNR loss for signal detection were considered in [16], but we believe that the present paper is the first to address these for the general problem of parameter estimation in AWGN.

This paper goes beyond the results in our conference paper [10] in multiple ways. First, we establish a connection between the pairwise isometry property and the Ziv-Zakai bound. We then show how the connections between the ZZB and CRB, together with the isometry conditions, can be used to predict the number of measurements required for accurate compressive estimation. We also characterize the number of measurements needed to provide isometry guarantees for a mixture of sinusoids unlike [10], which only deals with a single sinusoid. Finally, the isometry guarantees provided in [10] for a pair of sinusoids require their frequencies to be "well-separated". Here, we close the gap and provide such an isometry for any pair of frequencies.

In the algorithm description and numerical illustrations in this paper, we restrict attention to a single sinusoid in order to illustrate the fundamental features of compressive estimation. However, as described in detail in our conference papers [10], [11], our algorithmic approach (discrete grid followed by Newton refinement) extends easily to estimate the frequencies of multiple sinusoids. While the latter is a canonical problem of fundamental interest, it is worth noting that an important application that motivates us is the problem of adapting very large antenna arrays [17], [11]. Compressive parameter estimation in this context exploits the relatively small number of dominant multipath rays in order to estimate the spatial frequencies (and hence the angles of arrival) for these rays.

While we focus on compressive estimation based on a finite-dimensional signal, there has been significant research on the processing of continuous time signals exhibiting some measure of sparsity, sometimes termed "finite rate of innovation" (FRI) signals [18]. Sampling strategies for parameter estimation for such signals are studied in [19], using the CRB as the performance metric. The benefits of such good sampling strategies coupled with compressive processing at the analog front end are investigated in [20]. The isometry conditions derived in the present paper could potentially provide a systematic framework for design of compressive analog front ends for FRI signals.

Outline: We begin in Section III by stating the compressive parameter estimation problem in AWGN and the isometry properties needed for successful estimation. In Section IV we review bounds on parameter estimation in AWGN. The relationship between these estimation error bounds (CRB/ZZB) and the isometry properties are brought out in Section V. In Section VI, we consider the problem of estimating the frequency of a sinusoid. We show how the threshold behavior of the ZZB can predict the number of compressive measurements needed to avoid error floors. Section VII derives the number of measurements needed to guarantee these isometry conditions for the problem of frequency estimation from a mixture of K sinusoids and concludes by focussing on the single sinusoid case ($K = 1$).

III. COMPRESSIVE MEASUREMENTS

We begin by presenting the model for compressive measurements and providing the intuition behind two isometry conditions that are necessary for successful parameter estimation.

Consider the problem of estimating $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^K$ from noisy measurements of a differentiable manifold $\mathbf{x}(\boldsymbol{\theta}) \in \mathbb{C}^N$. The conventional estimation problem involves measuring all N elements of $\mathbf{x}(\boldsymbol{\theta})$ individually. In vector notation, the measurements are given by:

$$\mathbf{y} = \mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbb{I}_N\right). \qquad (2)$$

In contrast, with compressive measurements, we only observe $M \ll N$ noisy projections of the manifold $\mathbf{x}(\boldsymbol{\theta})$. Therefore, we have

$$\mathbf{y} = \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2 \mathbb{I}_M\right), \qquad (3)$$

where $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, which specifies the projection weights, is called the compressive measurement matrix. The elements

of $\boldsymbol{A}$ are chosen independently from zero-mean distributions of variance $1/N$ for which certain concentration results (we comment on this later) are available. Examples of such distributions include Uniform$\{\pm 1/\sqrt{N}\}$, Gaussian and Uniform$\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$. When the matrix $\boldsymbol{A}$ satisfies certain isometry conditions, we can successfully estimate $\boldsymbol{\theta}$ from $M \ll N$ measurements. We first explain why these conditions are helpful intuitively and then define them formally.

The Maximum Likelihood (ML) estimator [21] of $\boldsymbol{\theta}$ for the model in (3) is given by

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}'} \ \|\mathbf{y} - \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}')\| \tag{4}$$

$$= \arg\min_{\boldsymbol{\theta}'} \ \|\boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}) - \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}') + \mathbf{z}\| . \tag{5}$$

If the number of measurements is too small and $\boldsymbol{A}$ has a large nullspace, it is possible that $\|\boldsymbol{A}(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))\| \approx 0$ even when $\|\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')\|$ is large. Thus, with small amounts of noise $\mathbf{z}$, the optimizing parameter $\hat{\boldsymbol{\theta}}$ could be drastically different from the true parameter $\boldsymbol{\theta}$, resulting in large errors. This problem can be avoided if the matrix $\boldsymbol{A}$ preserves the geometry of the estimation problem by ensuring that the distance between $\mathbf{x}(\boldsymbol{\theta})$ and $\mathbf{x}(\boldsymbol{\theta}')$ remains approximately unaltered under its action. Specifically, if we have,

$$\|\boldsymbol{A}(\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}'))\| \propto \|\mathbf{x}(\boldsymbol{\theta}) - \mathbf{x}(\boldsymbol{\theta}')\|, \quad \forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta, \tag{6}$$

we see from (5) that the ML estimate at high SNR from M compressive measurements roughly coincides with the estimate we would have obtained with (2), where we have access to *all* N measurements of $\mathbf{x}(\boldsymbol{\theta})$. The *pairwise ϵ-isometry property* captures this idea of distance preservation precisely.

Pairwise ϵ-isometry property: The matrix $\boldsymbol{A}$ satisfies the pairwise ϵ-isometry property ($\epsilon < 1$) for the signal model $\mathbf{x}(\boldsymbol{\theta})$ if

$$\sqrt{\frac{M}{N}}(1 - \epsilon) \leq \frac{\|\boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}_1) - \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}_2)\|}{\|\mathbf{x}(\boldsymbol{\theta}_1) - \mathbf{x}(\boldsymbol{\theta}_2)\|} \leq \sqrt{\frac{M}{N}}(1 + \epsilon)$$
$$\forall \ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta. \tag{7}$$

We now motivate the isometry constants $\sqrt{M/N}(1-\epsilon)$ and $\sqrt{M/N}(1+\epsilon)$. Let $\mathbf{w}_i^H$ denote the i-th row of $\boldsymbol{A}$. Consider a single random projection of a signal $\mathbf{v}$ onto the weights $\mathbf{w}_i$ that have been chosen independently from zero-mean distributions of variance $1/N$. The average energy in the projection is $1/N$ of the energy in the signal $\mathbf{v}$: $\mathbb{E}\left|\mathbf{w}_i^H\mathbf{v}\right|^2 = (1/N)\|\mathbf{v}\|^2$. Thus, M compressive measurements capture M/N of the signal energy on average: $\mathbb{E}\|\boldsymbol{A}\mathbf{v}\|^2 = (M/N)\|\mathbf{v}\|^2$. Thus, for compressive measurements, it is natural to define the pairwise isometry property with the constants $\sqrt{M/N}(1 - \epsilon)$ and $\sqrt{M/N}(1 + \epsilon)$.

When the elements of $\boldsymbol{A}$ are drawn from appropriate distributions, for any particular realization of the measurement matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\mathbf{v}\|^2$ concentrates around its expected value $(M/N)\|\mathbf{v}\|^2$ with high probability. Specifically, for any $\mathbf{v} \in \mathbb{C}^N$:

$$\Pr\left[\left|\frac{N}{M}\|\boldsymbol{A}\mathbf{v}\|^2 - \|\mathbf{v}\|^2\right| > \delta\right] < C\exp\left(-M\,c(\delta)\right), \tag{8}$$

with constants C and $c(\delta)$ that depend only on the distribution from which the elements of $\boldsymbol{A}$ are picked from. For example, when the elements of $\boldsymbol{A}$ are picked i.i.d from Uniform$\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$ or Uniform$\{\pm 1/\sqrt{N}\}$ or $\mathcal{N}(0, 1/N)$, we can show that $C = 4$ and

$$c(\delta) = \delta^2/4 - \delta^3/6. \tag{9}$$

These concentration results are typically used to prove the pairwise isometry property (7). Refer [22] for a class of distributions (this includes all sub-gaussian distributions) for which such results are available.

We note that a particular instance of a randomly generated measurement matrix need not satisfy the pairwise isometry property for the signal manifold $\mathbf{x}(\boldsymbol{\theta})$. However, when the number of measurements M is sufficiently large, [6] shows that the pairwise ϵ-isometry property can be satisfied with arbitrarily high probability (the proof involves the use of concentration results (8) on carefully chosen samples on the manifold).

A weaker notion of distance preservation is the *tangent plane isometry property* that is particularly useful when we wish to refine an estimate $\hat{\boldsymbol{\theta}}$ that is "close" to the true parameter value. In this case, since we are interested only in the ML cost surface around the true parameter $\boldsymbol{\theta}$, it suffices to preserve the geometry of the estimation problem in the vicinity of $\boldsymbol{\theta}$ by ensuring that the distances between $\mathbf{x}(\boldsymbol{\theta}')$ and $\mathbf{x}(\boldsymbol{\theta})$ for $\boldsymbol{\theta}' \to \boldsymbol{\theta}$ are preserved under the action of $\boldsymbol{A}$. This is captured by the tangent plane isometry property defined as follows.

Tangent plane ϵ-isometry property: The matrix $\boldsymbol{A}$ satisfies the tangent plane ϵ-isometry property ($\epsilon < 1$) for the signal model $\mathbf{x}(\boldsymbol{\theta})$ if

$$\sqrt{\frac{M}{N}}(1 - \epsilon) \leq \frac{\|\boldsymbol{A}\sum a_m(\partial\mathbf{x}(\boldsymbol{\theta})/\partial\theta_m)\|}{\|\sum a_m(\partial\mathbf{x}(\boldsymbol{\theta})/\partial\theta_m)\|} \leq \sqrt{\frac{M}{N}}(1 + \epsilon)$$
$$\forall \ [a_1, a_2, \ldots, a_K]^T \in \mathbb{R}^K\backslash\{\mathbf{0}\}, \ \forall \boldsymbol{\theta} \in \Theta \tag{10}$$

By letting $\boldsymbol{\theta}_2 \to \boldsymbol{\theta}_1$ in the definition of the pairwise ϵ-isometry property, we see that a matrix $\boldsymbol{A}$ which satisfies the pairwise isometry property for the signal model $\mathbf{x}(\boldsymbol{\theta})$ also satisfies the tangent plane isometry, thereby confirming that tangent plane isometry is a weaker notion of distance preservation.

IV. PARAMETER ESTIMATION IN AWGN

We now review classical bounds on parameter estimation in AWGN that we relate to the isometry properties in the next section.

Consider the problem of estimating a parameter $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^K$ from noisy observations of the differentiable manifold $\mathbf{s}(\boldsymbol{\theta}) \in \mathbb{C}^M$. The observations are given by:

$$\mathbf{y} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{z}, \ \mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbb{I}_M). \tag{11}$$

For this measurement model,

$$p(\mathbf{y}|\boldsymbol{\theta}) = (\pi\sigma^2)^{-M}\exp\left(\|\mathbf{y} - \mathbf{s}(\boldsymbol{\theta})\|^2/\sigma^2\right). \tag{12}$$

For the observations $\mathbf{y}$, let $\hat{\boldsymbol{\theta}}(\mathbf{y})$ be an estimate of $\boldsymbol{\theta}$. Given a weight vector $\mathbf{a} \in \mathbb{R}^K$, classical bounds establish lower limits

on the error in estimating $\mathbf{a}^T\boldsymbol{\theta}$, given by $\mathbb{E}\left(\mathbf{a}^T\hat{\boldsymbol{\theta}}(\mathbf{y}) - \mathbf{a}^T\boldsymbol{\theta}\right)^2$, for a class of estimators $\hat{\boldsymbol{\theta}}(\mathbf{y})$. What we have left unspecified is the set of quantities we take the expectation over, and depending on this, the bounds fall into one of two categories:

Deterministic, but unknown, parameters: One class of bounds do not use the prior distribution of $\boldsymbol{\theta}$, so that the parameter to be estimated $\boldsymbol{\theta}$ is best thought of as a deterministic but unknown quantity. The most popular such bound is the Cramér Rao Bound (CRB). For the CRB, the expectation is taken over the conditional distribution $p(\mathbf{y}|\boldsymbol{\theta})$, so that the bound is on $\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\left(\mathbf{a}^T\hat{\boldsymbol{\theta}}(\mathbf{y}) - \mathbf{a}^T\boldsymbol{\theta}\right)^2$. The CRB typically depends on the parameter $\boldsymbol{\theta}$ and the most common version, which is what we use here, applies to estimators $\hat{\boldsymbol{\theta}}(\mathbf{y})$ that are unbiased [1].

Bayesian bounds: When we know the prior distribution $p(\boldsymbol{\theta})$ from which $\boldsymbol{\theta}$ is chosen, we can incorporate this information into the bounds. Such bounds are called Bayesian bounds and, in these cases, the expectation is taken over the joint distribution $p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$. They establish lower limits on the Mean-Squared-Error (MSE) in estimating $\mathbf{a}^T\boldsymbol{\theta}$, given by $\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\left(\mathbf{a}^T\hat{\boldsymbol{\theta}}(\mathbf{y}) - \mathbf{a}^T\boldsymbol{\theta}\right)^2$. Among the Bayesian bounds, we are primarily concerned with the Ziv-Zakai bound (ZZB) (we also briefly describe a version of the CRB, called the Bayesian CRB). Neither of these bounds (ZZB/BCRB) require the estimator to be unbiased.

The Ziv-Zakai Bound is known to be an accurate predictor of best possible estimation performance over a wide range of SNRs. Roughly speaking, it takes into account two sources of error: coarse error, when the estimate is not close to the true value of the parameter (essentially, making an error in hypothesis testing after binning the parameter space); and fine-grained error (the mean squared error from the true value when the estimate is in the right bin). At high SNR, the probability of the estimate falling into the wrong bin becomes negligible, and the Cramér Rao bound (CRB), which characterizes only fine-grained error, provides an excellent prediction of performance, while being easier to compute than the ZZB. We now state these bounds.

A. Cramér Rao Bound[21], [23]

Let $\mathbf{a} \in \mathbb{R}^K$. The variance of any unbiased estimator of $\mathbf{a}^T\boldsymbol{\theta}$, given by $\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\left(\mathbf{a}^T\hat{\boldsymbol{\theta}}(\mathbf{y}) - \mathbf{a}^T\boldsymbol{\theta}\right)^2$, is lower bounded by $\mathbf{a}^T F^{-1}(\boldsymbol{\theta})\mathbf{a}$, where $F(\boldsymbol{\theta})$ is the Fisher Information Matrix (FIM). The (m, n)th element of the FIM is given by:

$$F_{m,n}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\left\{\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_m}\frac{\partial \ln p(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_n}\right\}. \quad (13)$$

For parameter estimation in AWGN (12), this simplifies to [21]

$$F_{m,n}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\Re\left\{\left(\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \theta_m}\right)^H \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \theta_n}\right\}, \quad (14)$$

where $\Re\{b\}$ denotes the real part of the complex number b.

[1] An unbiased estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ is one which satisfies $\mathbb{E}_{\mathbf{y}|\boldsymbol{\theta}}\{\hat{\boldsymbol{\theta}}(\mathbf{y})\} = \boldsymbol{\theta}$ for all $\boldsymbol{\theta}$

B. Bayesian Bounds on Mean Square Error

To describe the Bayesian bounds, it is convenient to define the MSE matrix, $R(\hat{\boldsymbol{\theta}})$ of the estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$. The m, n-th element of the MSE matrix $R(\hat{\boldsymbol{\theta}})$ is given by $R_{m,n}(\hat{\boldsymbol{\theta}}) = \mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}\{(\hat{\theta}_m - \theta_m)(\hat{\theta}_n - \theta_n)\}$. For a vector $\mathbf{a} \in \mathbb{R}^K$, the ZZB and BCRB provide bounds on $\mathbb{E}_{\mathbf{y},\boldsymbol{\theta}}(\mathbf{a}^T\hat{\boldsymbol{\theta}}(\mathbf{y}) - \mathbf{a}^T\boldsymbol{\theta})^2$ which is simply $\mathbf{a}^T R(\hat{\boldsymbol{\theta}})\mathbf{a}$.

1) Bayesian Cramér Rao Bound[21], [23]: For any weight vector $\mathbf{a} \in \mathbb{R}^K$ and estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$ (not necessarily unbiased), the Bayesian Cramér Rao Bound (BCRB) lower bounds the MSE $\mathbf{a}^T R(\hat{\boldsymbol{\theta}})\mathbf{a}$ by $\mathbf{a}^T B^{-1}\mathbf{a}$, where B is the Bayesian Information Matrix (BIM). The (m, n)th element of B is given by:

$$B_{m,n} = \mathbb{E}_{\boldsymbol{\theta}}\{F_{m,n}(\boldsymbol{\theta})\} + \mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \theta_m}\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \theta_n}\right\}. \quad (15)$$

2) (Extended) Ziv-Zakai Bound[24]: Since the ZZB is not as widely used as the CRB, we provide a brief review in Appendix A. Here, we simply state the bound. The ZZB bounds the MSE $\mathbf{a}^T R(\hat{\boldsymbol{\theta}})\mathbf{a}$ and, for the AWGN measurement model (11), it is given by:

$$\mathbf{a}^T R(\hat{\boldsymbol{\theta}})\mathbf{a} \geq \frac{1}{2}\int_0^\infty \mathcal{V}\left\{\max_{\boldsymbol{\delta}:\mathbf{a}^T\boldsymbol{\delta}=h}\int_{\boldsymbol{\phi}\in\mathbb{R}^K}(p(\boldsymbol{\phi})+ \right.$$
$$\left. p(\boldsymbol{\phi}+\boldsymbol{\delta}))\, f(\boldsymbol{\phi},\boldsymbol{\phi}+\boldsymbol{\delta})\, d\boldsymbol{\phi}\right\}h\, dh, \quad \forall\hat{\boldsymbol{\theta}}(\mathbf{y}) \quad (16)$$

where $\mathcal{V}\{\ \}$ is the valley filling operation, defined as $\mathcal{V}\{g(h)\} = \max_{r\geq 0}g(h+r)$, and $f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is the probability of error for the optimal detection rule in the following *hypothesis testing problem*:

$$H_1 \ : \ \mathbf{y} = \mathbf{s}(\boldsymbol{\theta}_1) + \mathbf{z}, \ \Pr(H_1) = \frac{p(\boldsymbol{\theta}_1)}{p(\boldsymbol{\theta}_1) + p(\boldsymbol{\theta}_2)}$$
$$H_2 \ : \ \mathbf{y} = \mathbf{s}(\boldsymbol{\theta}_2) + \mathbf{z}, \ \Pr(H_2) = \frac{p(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_1) + p(\boldsymbol{\theta}_2)}. \quad (17)$$

Since $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbb{I}_M)$, this detection error probability is given by[21]:

$$f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{p(\boldsymbol{\theta}_1)}{p(\boldsymbol{\theta}_1) + p(\boldsymbol{\theta}_2)}Q\left(\frac{d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}\ln\frac{p(\boldsymbol{\theta}_1)}{p(\boldsymbol{\theta}_2)}\right)$$
$$+ \frac{p(\boldsymbol{\theta}_2)}{p(\boldsymbol{\theta}_1) + p(\boldsymbol{\theta}_2)}Q\left(\frac{d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\sqrt{2}\sigma} - \frac{\sigma}{\sqrt{2}d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}\ln\frac{p(\boldsymbol{\theta}_1)}{p(\boldsymbol{\theta}_2)}\right). \quad (18)$$

In the above expression, $Q(\ )$ stands for the CCDF of the standard normal distribution $\mathcal{N}(0, 1)$ and

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \|\mathbf{s}(\boldsymbol{\theta}_1) - \mathbf{s}(\boldsymbol{\theta}_2)\|. \quad (19)$$

Remark: While the expression for the ZZB is complicated, we only need two simple observations to prove the result we are interested in:

- With compressive measurements, the signal manifold $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta})$ and the measurement matrix $\boldsymbol{A}$ enters the ZZB *only* through the pairwise SNRs $d^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\sigma^2$.
- The minimum probability of detection error $f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ for the binary hypothesis testing problem (17) is a non-increasing function of the pairwise SNR $d^2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\sigma^2$.

We revisit these observations in Section V.

C. Threshold behavior of ZZB

The ZZB typically exhibits a *threshold behavior* with SNR [23]. When the SNR is very low, the measurements carry little information about the parameters we wish to estimate. Since the ZZB accounts for errors of "all magnitudes", it is usually large (depending primarily on the prior $p(\boldsymbol{\theta})$) and insensitive to small changes in SNR in this regime. However, at high SNRs, the variation of the ZZB with SNR is predictable. When the SNR and the ZZB are both expressed on a logarithmic scale, the ZZB falls off linearly with SNR, provided that the SNR is above a certain value, which is called the (asymptotic) *ZZB threshold* [24]. When the SNR exceeds the ZZB threshold, "large" estimation errors are unlikely, which is exactly when we would declare estimation of a continuous-valued parameter to be successful.

V. Relating the isometries to estimation bounds

We are now ready to relate the estimation error bounds for the compressive estimation problem to the corresponding bounds when we make all N measurements, provided that the compressive measurement matrix $\boldsymbol{A}$ satisfies appropriate isometry conditions.

Consider the general problem of estimating $\boldsymbol{\theta}$ from L measurements

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}, \quad \boldsymbol{\theta} \in \Theta \qquad (20)$$

where $\boldsymbol{\Phi}$ is any $L \times N$ complex-valued matrix and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbb{I}_L)$. The compressive estimation problem is subsumed in this model (obtained by setting $\boldsymbol{\Phi} = \boldsymbol{A}$, whose elements are chosen i.i.d. from a zero-mean distribution of variance $1/N$ for which concentration results of the form (8) are available), as is the conventional problem of estimating $\boldsymbol{\theta}$ from all N measurements (obtained by setting $\boldsymbol{\Phi} = \mathbb{I}_N$, the $N \times N$ identity matrix). Note that, in both these cases, the per-measurement SNR $(1/L) \sum_{k=1}^{k=L} \mathbb{E}|y_k|^2/\sigma^2$ is the same, since the rows of $\boldsymbol{A}$ have unit norm in expectation.

We prove two theorems that connect the fundamental estimation-theoretic bounds to the isometries defined in the previous section. First, we make a connection between the ZZB and the pairwise isometry property. As we observed in the remark under the statement of the ZZB, for the manifold $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta})$, the ZZB depends on the matrix $\boldsymbol{\Phi}$ only through the set of pairwise SNRs $\|\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}_1) - \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}_2)\|^2/\sigma^2 \ \forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$. When the compressive measurement matrix $\boldsymbol{A}$ satisfies the pairwise isometry property (7), the pairwise SNRs with $\boldsymbol{\Phi} = \boldsymbol{A}$ are approximately M/N times the corresponding values with $\boldsymbol{\Phi} = \mathbb{I}_N$. Thus, the ZZB with compressive measurements is approximately the same as the ZZB with all N measurements, but at an SNR penalty of M/N. Theorem 2 proves this intuition rigorously.

Likewise, we can connect the CRB to the tangent-plane isometry property. We can show that the CRB depends on the measurement matrix only through norms of the vectors $\boldsymbol{\Phi} \sum_m a_m (\partial \mathbf{x}(\boldsymbol{\theta})/\partial \theta_m)$. Thus, if $\boldsymbol{A}$ satisfies the tangent-plane isometry (10), the CRB with M compressive measurements is approximately equal to the CRB with all N measurements, but

at an SNR that is lower by M/N. We prove this in Theorem 1.

While the connections established here between estimation-theoretic bounds and the corresponding isometries apply generally to compressive estimation in AWGN, showing that these isometries indeed hold requires a problem-specific analysis, as we illustate for sinusoidal mixtures in later sections. As with standard compressed sensing, the goal of such analyses is to characterize the number of measurements required for such isometries to hold with high probability for random measurement matrices.

A. Cramér Rao Bound

Let $F(\boldsymbol{\Phi}, \boldsymbol{\theta})$ denote the Fisher Information Matrix for the measurement model (20). For this measurement model the expression for FIM is given by (14) with $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta})$:

$$F_{m,n}(\boldsymbol{\Phi}, \boldsymbol{\theta}) = \frac{2}{\sigma^2} \Re \left\{ \left(\boldsymbol{\Phi} \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m} \right)^H \boldsymbol{\Phi} \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_n} \right\}. \quad (21)$$

Theorem 1. *Let $\boldsymbol{A}$ be an $M \times N$ measurement matrix which satisfies the tangent plane ϵ-isometry property (10) for the signal manifold $\mathbf{x}(\boldsymbol{\theta})$. Then, the Fisher Information Matrix $F(\boldsymbol{A}, \boldsymbol{\theta})$, with compressive measurements (3) is related to the FIM with all N measurements as follows:*

$$\begin{aligned} F(\boldsymbol{A}, \boldsymbol{\theta}) \preceq F\left(\sqrt{\tfrac{M}{N}}(1+\epsilon)\mathbb{I}_N, \boldsymbol{\theta} \right) \\ F(\boldsymbol{A}, \boldsymbol{\theta}) \succeq F\left(\sqrt{\tfrac{M}{N}}(1-\epsilon)\mathbb{I}_N, \boldsymbol{\theta} \right) \end{aligned} \quad \forall \boldsymbol{\theta} \in \Theta. \quad (22)$$

Proof: Consider the quadratic form $\mathbf{a}^T F(\boldsymbol{\Phi}, \boldsymbol{\theta})\mathbf{a}$ for any $\mathbf{a} = [a_1 \ \cdots \ a_K]^T \in \mathbb{R}^K$. We see that

$$\mathbf{a}^T F(\boldsymbol{\Phi}, \boldsymbol{\theta})\mathbf{a} = \frac{2}{\sigma^2} \left\| \boldsymbol{\Phi} \sum_m a_m \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m} \right\|^2. \quad (23)$$

Since the compressive measurement matrix $\boldsymbol{A}$ satisfies the tangent plane ϵ-isometry property (10) for the signal model $\mathbf{x}(\boldsymbol{\theta})$, we have that for all $\boldsymbol{\theta} \in \Theta$ and $\mathbf{a} \in \mathbb{R}^K$,

$$\left\| \boldsymbol{A} \sum_m a_m \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m} \right\|^2 \leq \frac{M}{N}(1+\epsilon)^2 \left\| \sum_m a_m \frac{\partial \mathbf{x}(\boldsymbol{\theta})}{\partial \theta_m} \right\|^2. \quad (24)$$

Multiplying both sides by $2/\sigma^2$, we see that the LHS is $\mathbf{a}^T F(\boldsymbol{A}, \boldsymbol{\theta})\mathbf{a}$, while the RHS corresponds to $\mathbf{a}^T F\left(\sqrt{M/N}(1+\epsilon)\mathbb{I}_N, \boldsymbol{\theta} \right) \mathbf{a}$. Therefore, we have that $\forall \boldsymbol{\theta} \in \Theta$,

$$\mathbf{a}^T F(\boldsymbol{A}, \boldsymbol{\theta})\mathbf{a} \leq \mathbf{a}^T F\left(\sqrt{M/N}(1+\epsilon)\mathbb{I}_N, \boldsymbol{\theta} \right) \mathbf{a}, \ \forall \mathbf{a} \in \mathbb{R}^K. \quad (25)$$

This establishes the required upper bound on $F(\boldsymbol{A}, \boldsymbol{\theta})$. The proof for the lower bound is analogous. ∎

B. Bayesian Cramér Rao Bound

Let $B(\boldsymbol{\Phi})$ denote the Bayesian Information Matrix for the measurement model (20). Let $p(\boldsymbol{\theta})$ be the prior on $\boldsymbol{\theta}$. For this

measurement model the expression for BIM is given by (15) with $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta})$:

$$B_{m,n}(\boldsymbol{\Phi}) = \mathbb{E}_{\boldsymbol{\theta}}\left\{F_{m,n}(\boldsymbol{\Phi},\boldsymbol{\theta})\right\} + \mathbb{E}_{\boldsymbol{\theta}}\left\{\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \theta_m}\frac{\partial \ln p(\boldsymbol{\theta})}{\partial \theta_n}\right\}. \tag{26}$$

Corollary (of Theorem 1). *Let $\boldsymbol{A}$ be an $M \times N$ measurement matrix which satisfies the tangent plane ϵ-isometry property (10) for the signal manifold $\mathbf{x}(\boldsymbol{\theta})$. Then, the Bayesian Information matrix $B(\boldsymbol{A})$ with compressive measurements (3) is related to the BIM with all N measurements as follows:*

$$B\left(\sqrt{\frac{M}{N}}(1-\epsilon)\mathbb{I}_N\right) \preceq B(\boldsymbol{A}) \preceq B\left(\sqrt{\frac{M}{N}}(1+\epsilon)\mathbb{I}_N\right) \tag{27}$$

Proof: Let $\mathbf{a} \in \mathbb{R}^K$. We see that $\mathbf{a}^T B(\boldsymbol{A})\mathbf{a}$ depends on the measurement matrix $\boldsymbol{A}$ only through quadratic forms of the FIM i.e., $\mathbf{a}^T F(\boldsymbol{A},\boldsymbol{\theta})\mathbf{a}$. When the tangent plane isometry condition (10) is satisfied, we have from Theorem 1 that $\mathbf{a}^T F(\boldsymbol{A},\boldsymbol{\theta})\mathbf{a}$ is bounded by $\mathbf{a}^T F(\sqrt{M/N}(1\pm\epsilon)\mathbb{I}_N,\boldsymbol{\theta})\mathbf{a}$ for all $\mathbf{a}, \boldsymbol{\theta}$. It immediately follows that the quadratic forms of $B(\boldsymbol{A})$ are bounded by the corresponding quadratic forms of $B(\sqrt{M/N}(1\pm\epsilon)\mathbb{I}_N)$. ∎

C. Ziv-Zakai Bound

Let $Z(\boldsymbol{\Phi},\mathbf{a})$ denote the ZZB corresponding to the Mean-Squared-Error in estimating $\mathbf{a}^T\boldsymbol{\theta}$ for the measurement model (20). The expression for $Z(\boldsymbol{\Phi},\mathbf{a})$ is given by the right hand side of (16), with $d(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \|\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}_1) - \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}_2)\|$ (obtained by setting $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta})$).

Note that in (16), $f(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ is the probability of detection error for the hypothesis testing problem (17) with $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta})$. We capture the dependence of this probability on the matrix $\boldsymbol{\Phi}$ by defining $g(\boldsymbol{\Phi},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = f(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ when $\mathbf{s}(\boldsymbol{\theta}) = \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta})$.

Theorem 2. *Let $\boldsymbol{A}$ be an $M \times N$ measurement matrix which satisfies the pairwise ϵ-isometry property (7) for the signal manifold $\mathbf{x}(\boldsymbol{\theta})$. Then, the ZZB $Z(\boldsymbol{A},\mathbf{a})$, with the compressive measurements in (3), is related to the ZZB with all N measurements as*

$$Z\left(\sqrt{\frac{M}{N}}(1+\epsilon)\mathbb{I}_N,\mathbf{a}\right) \leq Z(\boldsymbol{A},\mathbf{a}) \leq Z\left(\sqrt{\frac{M}{N}}(1-\epsilon)\mathbb{I}_N,\mathbf{a}\right). \tag{28}$$

Proof: As we observed in the remark at the end of the definition of the ZZB, $g(\boldsymbol{\Phi},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ is a non-increasing function of the pairwise SNR $\|\boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}_1) - \boldsymbol{\Phi}\mathbf{x}(\boldsymbol{\theta}_2)\|^2/\sigma^2$. When $\boldsymbol{A}$ satisfies the pairwise ϵ-isometry property (7), we can bound all the pairwise SNRs as follows:

$$\|\boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}_1) - \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}_2)\|^2/\sigma^2 \leq \frac{M}{N}(1+\epsilon)^2\|\mathbf{x}(\boldsymbol{\theta}_1) - \mathbf{x}(\boldsymbol{\theta}_2)\|^2/\sigma^2$$
$$\forall \boldsymbol{\theta}_1,\boldsymbol{\theta}_2 \in \Theta. \tag{29}$$

Combining these facts, we get $g(\boldsymbol{A},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \geq g(\sqrt{M/N}(1+\epsilon)\mathbb{I}_N,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, which is the probability of detection error with all N measurements, but at an SNR penalty of $(M/N)(1+\epsilon)^2$. Substituting these pointwise bounds in the expression for $Z(\boldsymbol{A},\mathbf{a})$, we have that $Z(\boldsymbol{A},\mathbf{a}) \geq Z\left(\sqrt{\frac{M}{N}}(1+\epsilon)\mathbb{I}_N,\mathbf{a}\right)$. The other inequality can be proved similarly. ∎

D. Number of measurements needed

These theorems show that, when the compressive measurement matrix $\boldsymbol{A}$ satisfies the pairwise isometry property, the CRB and the ZZB are well approximated by $\mathbf{a}^T F^{-1}(\sqrt{M/N}\mathbb{I}_N,\boldsymbol{\theta})\mathbf{a}$ and $Z(\sqrt{M/N}\mathbb{I}_N,\mathbf{a})$ respectively (for any $\mathbf{a}$). Thus, the estimation performance with the measurement matrix $\boldsymbol{\Phi} = \boldsymbol{A}$ is roughly the same as that with $\boldsymbol{\Phi} = \sqrt{M/N}\mathbb{I}_N$ (all N measurements, but with the signal component scaled by $\sqrt{M/N}$). Note that observations with $\boldsymbol{\Phi} = \sqrt{M/N}\mathbb{I}_N$ and per-sample noise variance σ^2 are *equivalent* to observations $\boldsymbol{\Phi} = \mathbb{I}_N$ (conventional measurements) but with an increased per-sample noise variance $\sigma^2(N/M)$ (easily seen by multiplying the observations with $\boldsymbol{\Phi} = \sqrt{M/N}\mathbb{I}_N$ by $\sqrt{N/M}$). Putting these observations together, we get a simple procedure for estimating the number of measurements M required for successful compressive estimation:

(1) For the case when we make all N measurements, $\mathbf{y} = \mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}$ with $\mathbf{z} \sim \mathcal{CN}(0,\sigma^2\mathbb{I}_N)$, compute the ZZB as a function of σ^2. Find the ZZB threshold as described in Section IV (the value of σ^2 *below* which $\log$ ZZB falls off linearly with $\log\sigma^2$). Denote this threshold by σ_t^2.

(2) Making M compressive measurements $\mathbf{y} = \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}$ with $\mathbf{z} \sim \mathcal{CN}(0,\sigma_0^2\mathbb{I}_M)$ is roughly equivalent to making the observations $\tilde{\mathbf{y}} = \mathbf{x}(\boldsymbol{\theta}) + \tilde{\mathbf{z}}$ with $\tilde{\mathbf{z}} \sim \mathcal{CN}(0,\sigma_0^2(N/M)\mathbb{I}_M)$ when $\boldsymbol{A}$ satisfies the pairwise isometry property. Thus, the number of measurements needed for successful compressive estimation is given by:

$$\sigma_0^2\frac{N}{M} < \sigma_t^2 \quad \text{or} \quad M > N\left(\frac{\sigma_0^2}{\sigma_t^2}\right) \tag{30}$$

We reiterate that the above SNR criterion is not the only condition for successful compressive estimation: the number of measurements M must be large enough for the matrix $\boldsymbol{A}$ to satisfy the pairwise isometry property, so that we can invoke the SNR penalty arguments.

In the next section, we illustrate these ideas by considering the example of frequency estimation of a single sinusoid. But before that, we comment on the generality of the model we have considered so far.

Remarks on model generality: While we describe our results in the context of the measurement model (3), they extend easily to variants commonly encountered in the compressed sensing literature, two of which we now discuss.

• For applications such as Direction of Arrival (DoA) estimation using large arrays [11], compressive measurements are acquired sequentially in time and every measurement is corrupted by independent *measurement* noise. Thus, the measurements satisfy

$$y_l = \mathbf{w}_l^T(\mathbf{x}(\boldsymbol{\theta}) + \tilde{\mathbf{z}}_l) = \mathbf{w}_l^T\mathbf{x}(\boldsymbol{\theta}) + z_l, \tag{31}$$

where $\tilde{\mathbf{z}}_l \sim \mathcal{CN}(\mathbf{0},\sigma^2\mathbb{I}_N)$ and $z_l = \mathbf{w}_l^T\tilde{\mathbf{z}}_l \sim \mathcal{CN}(0,\sigma^2\|\mathbf{w}_l\|^2)$. The key point here is that $\tilde{\mathbf{z}}_1,\ldots,\tilde{\mathbf{z}}_M$ are i.i.d. and as a result $z_1,\ldots,z_M$ are independent. Letting $\boldsymbol{A}$ denote the matrix with rows $\mathbf{w}_l^T$, $\mathbf{y} = [y_1 \cdots y_M]^T$ and $\mathbf{z} = [z_1 \cdots z_M]^T$, we have:

$$\mathbf{y} = \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{CN}(\mathbf{0},\sigma^2\mathbf{K}_1), \tag{32}$$

where $\mathbf{K}_1$ is a diagonal matrix whose diagonal entries are $\|\mathbf{w}_l\|^2$, $l = 1, \ldots, M$.

• For other applications, when we have access to a *single* noisy version of $\mathbf{x}(\boldsymbol{\theta})$ and compressive measurements are merely used as a *dimensionality reduction* tool, we have

$$\mathbf{y} = \boldsymbol{A}\left(\mathbf{x}(\boldsymbol{\theta}) + \tilde{\mathbf{z}}\right), \quad \tilde{\mathbf{z}} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbb{I}_N). \tag{33}$$

The same equation holds for the case when there are errors in modeling the manifold $\mathbf{x}(\boldsymbol{\theta})$ (given by $\tilde{\mathbf{z}}$) and we make M sequential *noiseless* projections. Letting $\mathbf{z} = \boldsymbol{A}\tilde{\mathbf{z}}$ we have

$$\mathbf{y} = \boldsymbol{A}\mathbf{x}(\boldsymbol{\theta}) + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{K}_2), \tag{34}$$

where $\mathbf{K}_2 = \boldsymbol{A}\boldsymbol{A}^H$.

Neither $\mathbf{K}_1$ and $\mathbf{K}_2$ are the identity matrix, hence these measurement models do not fit directly into the framework in (3). However, we can extend our results easily to these models by considering the whitened observations $\tilde{\mathbf{y}}_i = \mathbf{K}_i^{-1/2}\mathbf{y}$, $i = 1, 2$, and establishing bounds on the singular values of $\mathbf{K}_i$. When the elements of $\boldsymbol{A}$ are chosen from a zero-mean distribution of variance $1/N$ (for which concentration results of the form (8) are available), the singular values of $\mathbf{K}_i$ concentrate around 1. As a result, an ϵ-isometry (tangent plane or pairwise) for $\boldsymbol{A}$ can be shown to translate to a mildly weaker $\epsilon_{\text{eff},i}$-isometry ($\epsilon_{\text{eff},i} \geq \epsilon$) for $\boldsymbol{A}_{\text{eff},i} = \mathbf{K}_i^{-1/2}\boldsymbol{A}$, the effective measurement matrix for the whitened measurements $\tilde{\mathbf{y}}_i$. All of our results now apply by simply replacing ϵ with $\epsilon_{\text{eff},i}$. This equivalence of the measurement model (33) and the general compressive model (3) has also been investigated in detail in [15]. The proof for the conditioning of both $\mathbf{K}_1$ and $\mathbf{K}_2$ involves using the concentration result (8) for $\sqrt{N/M}\boldsymbol{A}^H$ (see [22] for $\mathbf{K}_2$).

The concentration results for the singular values of $\mathbf{K}_2 = \boldsymbol{A}\boldsymbol{A}^H$ (which are the square of the singular values of $\boldsymbol{A}^H$) needs N to be somewhat larger than M. This is not an issue, since this is the regime of interest for compressive estimation. The diagonal matrix $\mathbf{K}_1$, on the other hand, is well-conditioned for much larger values of M (potentially larger than N).

VI. DESIGNING COMPRESSIVE ESTIMATION STRATEGIES

In this section, we illustrate, using the example of frequency and phase estimation for a single sinusoid, how to apply the preceding results to design compressive estimation strategies. We describe an algorithm which attains the CRB given "enough" compressive measurements, and show how to determine how many measurements are enough, based on the threshold behavior of the ZZB. We implicitly assume that we have enough measurements for the appropriate isometries to hold; detailed analytical characterization of the number of measurements required for this purpose is deferred to later sections.

The measurements are given by

$$\mathbf{y} = e^{j\phi}\boldsymbol{\Phi}\mathbf{x}(\omega) + \mathbf{z} \tag{35}$$

where $\mathbf{x}(\omega) = \left[e^{-j\omega(N-1)/2} \ e^{-j\omega(N-3)/2} \ \cdots \ e^{j\omega(N-1)/2}\right]^T$ is an N-dimensional sinusoid with frequency ω, ϕ is its phase, $\boldsymbol{\Phi}$ is an $L \times N$ complex valued measurement matrix and $\mathbf{z} \sim \mathcal{CN}(0, \sigma^2 \mathbb{I}_L)$. The parameters to be estimated

ϕ and ω are both distributed uniformly over $[0, 2\pi]$. Note that there is a slight change in notation from the previous section. Earlier, we denoted the parameter to be estimated by $\boldsymbol{\theta} = [\omega \ \phi]^T$ and the signal manifold $\mathbf{x}(\boldsymbol{\theta}) = e^{j\phi}\left[e^{-j\omega(N-1)/2} \ e^{-j\omega(N-3)/2} \ \cdots \ e^{j\omega(N-1)/2}\right]^T$. We now separate the contributions from the phase and frequency and use $\mathbf{x}(\omega)$ to denote a sinusoid with frequency ω and zero phase ($\phi = 0$).

When we make all N measurements (setting $\boldsymbol{\Phi} = \mathbb{I}_N$ in (35)), the CRB is well known [25]. The FIM in estimating $\boldsymbol{\theta} = [\omega \ \phi]$ is

$$F(\mathbb{I}_N, \boldsymbol{\theta}) = \frac{2}{\sigma^2} \begin{bmatrix} N(N^2 - 1)/12 & 0 \\ 0 & N \end{bmatrix} \forall \boldsymbol{\theta}. \tag{36}$$

In particular, the CRB on the variance of the frequency estimate (computed as $\mathbf{a}^T F^{-1}(\mathbb{I}_N, \boldsymbol{\theta})\mathbf{a}$ with $\mathbf{a} = [1 \ 0]$) is $\text{CRB}(\mathbb{I}_N, \boldsymbol{\theta}) = 6\sigma^2/(N(N^2 - 1))$. Note that the CRB is independent of $\boldsymbol{\theta}$.

We must be careful in computing the ZZB because the noiseless signal is a periodic function (with period 2π) of both the phase and the frequency. Thus, the errors in estimating these parameters must be appropriately defined (i.e., the difference between 0 and $2\pi - \epsilon$ is ϵ for small ϵ). The ZZB on the "periodic-MSE" of the frequency estimate is given by (using (27) in [26])

$$Z(\mathbb{I}_N, \mathbf{a}) = \int_0^\pi \max_{\phi' \in [0, 2\pi]} Q\left(\frac{\|\mathbf{x}(0) - e^{j\phi'}\mathbf{x}(h)\|}{\sqrt{2}\sigma}\right) h \ dh$$
$$= \int_0^\pi Q\left(\sqrt{\frac{N}{\sigma^2}\left(1 - \left|\frac{\sin(Nh/2)}{N\sin(h/2)}\right|\right)}\right) h \ dh. \tag{37}$$

Suppose now that we make M compressive measurements (setting $\boldsymbol{\Phi} = \boldsymbol{A}$), choosing M large enough so that the measurement matrix $\boldsymbol{A}$ satisfies the pairwise ϵ-isometry property for the $\{e^{j\phi}\mathbf{x}(\omega)\}$ signal model (the number of compressive measurements needed to establish pairwise isometries for this signal model is analytically characterized in Section VII-A). Then, from Section V, we know that the Fisher information with compressive measurements $F(\boldsymbol{A}, \boldsymbol{\theta})$ is well-approximated by $F\left(\sqrt{M/N}\ \mathbb{I}_N, \boldsymbol{\theta}\right)$, the Fisher information with all N measurements at an M/N SNR penalty. Given that we know $F(\mathbb{I}_N, \boldsymbol{\theta})$, computing $F(\sqrt{M/N}\ \mathbb{I}_N, \boldsymbol{\theta})$ is easy: we simply replace σ^2 in (36) by $\sigma^2(N/M)$.

When $\boldsymbol{A}$ satisfies the pairwise isometry property, we can show that the ZZB with periodic-MSE also satisfies Theorem 2. Therefore, the above arguments regarding the increase in the noise level by a factor of N/M hold true for the ZZB with periodic distortion too. Thus, we get the CRB and the ZZB with compressive measurements to be

$$\text{CRB}(\boldsymbol{A}, \boldsymbol{\theta}) \approx \text{CRB}(\sqrt{M/N}\ \mathbb{I}_N, \boldsymbol{\theta}) = 6\sigma^2/(M(N^2 - 1)) \ \forall \boldsymbol{\theta} \tag{38}$$

$$Z(\boldsymbol{A}, \mathbf{a}) \approx Z(\sqrt{M/N}\ \mathbb{I}_N, \mathbf{a})$$
$$= \int_0^\pi Q\left(\sqrt{\frac{M}{\sigma^2}\left(1 - \left|\frac{\sin(Nh/2)}{N\sin(h/2)}\right|\right)}\right) h \ dh. \tag{39}$$
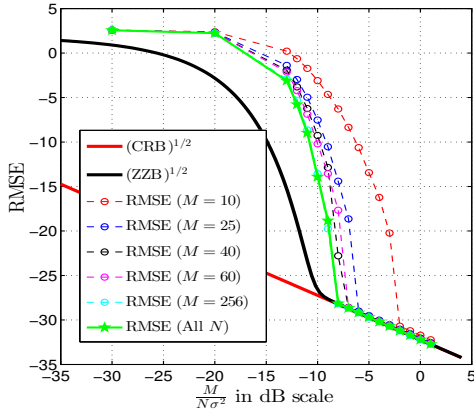
Fig. 1. RMSE in dB scale for 5 compressive measurement matrices ($\mathbf{\Phi} = \mathbf{A}$) with $M = 10, 25, 40, 60, 256$ and the all N measurements case ($\mathbf{\Phi} = \mathbb{I}_N$) plotted against effective per sample SNR $M/(N\sigma^2)$. Overlaid are plots of $\sqrt{\text{CRB}}$ and $\sqrt{\text{ZZB}}$ for all N measurements ($\mathbf{\Phi} = \mathbb{I}_N$) corresponding to this effective SNR. The length of the sinusoid $\mathbf{x}(\omega)$ is $N = 256$.

Fig. 2. *Bounds* on pairwise SNR variation due to pairwise isometry constant ϵ (7) for the compressive measurement matrices used in Fig. 1. Isometry constant ϵ corresponds to the manifold $\{ge^{j\phi}\mathbf{x}(\omega)\}$ where $g \in \mathbb{R}^+$ and $\phi, \omega \in [0, 2\pi]$.

We now illustrate how to predict the number of measurements needed for successful compressive estimation based on the threshold behavior of the ZZB. Consider frequency estimation of a $N = 256$ sinusoid from all N measurements ($\mathbf{\Phi} = \mathbb{I}_N$) at a noise level σ^2. In Fig. 1, we plot the CRB and the ZZB for this estimation problem as a function of the per-measurement SNR $\overset{\triangle}{=} 1/\sigma^2$. For SNRs that are smaller than -30dB, we see from Fig. 1 that the ZZB is insensitive to changes in SNR, unlike the CRB which exhibits a linear falloff for all SNRs. However, when the SNR exceeds -10dB, the ZZB exhibits a linear falloff with SNR.

If we now make M compressive measurements ($\mathbf{\Phi} = \mathbf{A}$), the results of the previous section tell us that the effective SNR is given by $(1/\sigma^2)(M/N)$. We expect "good" estimation performance when this effective SNR exceeds the ZZB threshold, which translates to the following rule of thumb for the number of compressive measurements required:

$$M > N\sigma^2 \times \text{ZZB threshold SNR}, \qquad (40)$$

Note that the ZZB threshold is computed for the *original* system with all N measurements ($\mathbf{\Phi} = \mathbb{I}_N$), independent of the compressive measurement matrix $\mathbf{A}$ and the noise level σ^2. For our specific example of a sinusoid of length $N = 256$, the preceding prescription translates to $M > N\sigma^2/10$, since the ZZB threshold is -10 dB.

We now describe an algorithm whose performance closely follows these predictions: the algorithm approaches the CRB (for a given effective SNR) when the effective SNR exceeds the *ZZB threshold*. This illustrates the efficiency of the algorithm, as well as the accuracy of our design guideline of "sufficient effective SNR."

Algorithm: Suppose that for the purposes of algorithm design, we ignore the fact that the unknown phase rotation $e^{j\phi}$ has unit amplitude and estimate the complex gain g and the frequency ω according to the model

$$\mathbf{y} = g\mathbf{\Phi}\mathbf{x}(\omega) + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbb{I}). \qquad (41)$$

The ML estimates of the gain and frequency $(\hat{g}, \hat{\omega})$ are
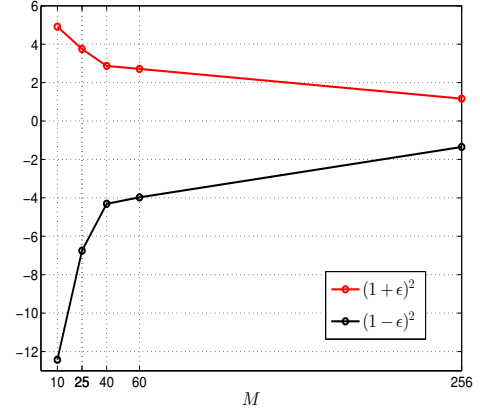
obtained by optimizing the function

$$S(g, \omega) = \Re\left\{\mathbf{y}^H g\mathbf{\Phi}\mathbf{x}(\omega)\right\} - 0.5|g|^2\|\mathbf{\Phi}\mathbf{x}(\omega)\|^2, \qquad (42)$$

over $g \in \mathbb{C}, \omega \in [0, 2\pi]$ and $\Re\{a\}$ denotes the real part of the complex number a. Performing a direct optimization over g and ω is difficult. Therefore, we resort to a two stage procedure, consisting of a detection phase and a refinement phase, which we describe now.

(i) *Detection phase*: First, we notice that for any ω, the optimizing g is given by $(\mathbf{\Phi}\mathbf{x}(\omega))^H \mathbf{y}/\|\mathbf{\Phi}\mathbf{x}(\omega)\|^2$. Substituting this in the cost function $S(g, \omega)$, we see that the ML estimate of the frequency $\hat{\omega}$ should optimize $G(\omega) = \max_{g \in \mathbb{C}} S(g, \omega) = 0.5|\mathbf{y}^H\mathbf{\Phi}\mathbf{x}(\omega)|^2/\|\mathbf{\Phi}\mathbf{x}(\omega)\|^2$. We obtain a coarse frequency estimate by discretizing the frequencies uniformly into a set $F = \{0, 2\pi/(4N), \ldots, 2\pi(4N - 1)/(4N)\}$ of size $4N$ and then choosing $q^\star \in F$ that maximizes $G(q), q \in F$. Since the frequency estimation error is substantial (on the order of $1/N$), we call this the detection phase. The gain estimate is given by $\hat{g} = (\mathbf{\Phi}\mathbf{x}(q^\star))^H \mathbf{y}/\|\mathbf{\Phi}\mathbf{x}(q^\star)\|^2$

(ii) *Refinement phase*: In the second stage, we iteratively refine the gain and frequency estimates. Suppose that after the nth round of optimization, the gain and frequency estimates are given by $\hat{g}_n$ and $\hat{\omega}_n$ respectively (starting off with the estimates from the detection phase). In the $n + 1$th round, we refine the frequency estimate by fixing the gain to $\hat{g}_n$ and locally optimizing $S(\hat{g}_n, \omega)$ around $\hat{\omega}_n$ using Newton's method:

$$\hat{\omega}_{n+1} = \hat{\omega}_n - \frac{\partial S(\hat{g}_n, \hat{\omega}_n)/\partial \omega}{\partial^2 S(\hat{g}_n, \hat{\omega}_n)/\partial \omega^2}, \qquad (43)$$

where

$$\frac{\partial S(g, \omega)}{\partial \omega} = \Re\left\{(\mathbf{y} - g\mathbf{\Phi}\mathbf{x}(\omega))^H g\mathbf{\Phi}(d\mathbf{x}(\omega)/d\omega)\right\}, \quad (44)$$

$$\frac{\partial^2 S(g, \omega)}{\partial \omega^2} = \Re\left\{(\mathbf{y} - g\mathbf{\Phi}\mathbf{x}(\omega))^H g\mathbf{\Phi}\left(d^2\mathbf{x}(\omega)/d\omega^2\right)\right\}$$
$$- |g|^2\|\mathbf{\Phi}\left(d\mathbf{x}(\omega)/d\omega\right)\|^2. \qquad (45)$$

Next, fixing the frequency estimate to $\hat{\omega}_{n+1}$, we get the updated gain after the $n + 1$th round to be $\hat{g}_{n+1} =$

$\left(\mathbf{\Phi}\mathbf{x}(\hat{\omega}_{n+1})\right)^H \mathbf{y}\big)/\|\mathbf{\Phi}\mathbf{x}(\hat{\omega}_{n+1})\|^2$. Our numerical results are based on applying three such rounds of iterative optimization. **Results:** We simulate the performance of the algorithm with $M = 10, 25, 40, 60$ and 256 compressive measurements across *effective per measurement SNRs* $M/(N\sigma^2)$ ranging from -30dB to 1dB using 5×10^4 trials (for each M, we use the same measurement matrix $\boldsymbol{A}$ for all SNR values). The elements of $\boldsymbol{A}$ are picked i.i.d from Uniform$\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$. We plot the Root-Mean-Squared-Error (RMSE) of the frequency estimate versus the effective SNR $M/(N\sigma^2)$ along with the CRB and ZZB in Fig. 1. We define the effective SNR beyond which the RMSE of the estimate exhibits a linear falloff with SNR in the log-log plot (similar to the ZZB at high SNRs) as the *RMSE threshold*. From our earlier discussions on the number of measurements needed for successful compressive estimation, we expect the RMSE threshold to exceed the ZZB threshold. From Fig. 1, we see that the RMSE thresholds for $M = 10, 25, 40, 60$ and 256 measurements are $-2, -6, -7, -7$ and -8dB respectively. All the RMSE thresholds are larger than the ZZB threshold of -10dB as expected. We also evaluate the algorithm for the all N measurements case ($\mathbf{\Phi} = \mathbb{I}_N$) and find that the RMSE threshold in this case is -8dB.

Differences in the isometry constant ϵ explain why the RMSE thresholds are different for different measurement matrices $\boldsymbol{A}$. With increasing number of measurements M, the isometry constant decreases. This trend is shown in Fig. 2 where we plot the bounds on the deviation of the pairwise SNRs from M/N, corresponding to $(1 \pm \epsilon)^2$, for the measurement matrices $\boldsymbol{A}$ used in our simulations. (*Note*: These isometry constants correspond to the manifold $\{ge^{j\phi}\mathbf{x}(\omega) : g \in \mathbb{R}^+, \phi, \omega \in [0, 2\pi)\}$ because the algorithm does not use the fact that $g = 1$.) When we take few compressive measurements, pairwise SNRs can deteriorate significantly (ϵ is large) and, as a result, the RMSE threshold increases. The bounds on pairwise SNR variation (in Fig. 2) when we make 40 and 60 measurements do not differ by much. This illustrates the diminishing improvements in isometry per measurement beyond a point. For the all-N measurements case ($\mathbf{\Phi} = \mathbb{I}_N$), the isometry constant $\epsilon = 0$ by definition and therefore the RMSE threshold is close to the ZZB threshold.

When we set $M = N = 256$, the degradation in pairwise SNRs is smaller than 2dB. However, for this extreme case, the RMSE threshold is merely 1dB smaller than that for $M = 40$. This indicates that, for our example of frequency estimation for sinusoid of length $N = 256$, the isometry constant is small enough when we make 40 or more compressive measurements.

To summarize, when the number of measurements M is large enough for the isometry constant ϵ to be small, the number of measurements M necessary obeys the rule of thumb in (40), based on ZZB threshold computations for the original system. For our example $N = 256$ sinusoid, this translates to the rule of thumb $M \geq \max\{40, 25.6\sigma^2\}$.

VII. Isometry conditions for frequency estimation from compressive measurements

In the single sinusoid example in the previous section, we assume that there are enough measurements to guarantee the required isometries. In this section, we seek to analytically characterize the number of measurements required to provide such guarantees. We show that, for a mixture of K sinusoids, the number of measurements required depends on the conditioning of appropriately defined matrices, which in turn depends on the separation between the frequencies in the mixture. We return to the special case of a single sinusoid, for which we can prove stronger results, at the end of this section

Consider a manifold of signals which are linear combinations of K complex sinusoids $\sum_{l=1}^{K} g_l\mathbf{x}(\omega_l)$, where $g_l \in \mathbb{C}$ are complex gains and

$$\mathbf{x}(\omega) = \left[h_1 e^{-j\omega(N-1)/2} \;\; \cdots \;\; h_N e^{j\omega(N-1)/2} \right]^T \quad (46)$$

is a windowed sinusoid, with window weights given by $\{h_n\}$. The example in the previous section is a special case with $K = 1$ and an all-ones window. Without loss of generality, we assume that the window weights are normalized so that $\sum_n |h_n|^2 = 1$. To avoid trivialities, we assume that more than one of the h_n's are non-zero.

Suppose that we make M compressive measurements of the form

$$\mathbf{y} = \boldsymbol{A} \sum_{l=1}^{l=K} g_l\mathbf{x}(\omega_l) + \mathbf{z}, \quad (47)$$

and we wish to estimate the gains $\mathbf{g} = [g_1 \;\; \cdots \;\; g_K]^T$ and the frequencies $\boldsymbol{\omega} = [\omega_1 \;\; \cdots \;\; \omega_K]^T$. Therefore, in the notation of the preceding sections the parameter to be estimated is $\boldsymbol{\theta} = (\mathbf{g}, \boldsymbol{\omega})$.

Tangent plane isometry for a mixture of K sinusoids: Our first goal is to quantify the number of measurements needed to preserve the CRB for a given frequency support $\boldsymbol{\omega}$ (i.e., for all $\boldsymbol{\theta}$ that share this frequency support). We show that this is equivalent to guaranteeing ϵ-isometry for a *set* of tangent planes as follows. For any specific value of the unknown parameters – gain magnitude $\{|g_l|\}$, phases $\{g_l/|g_l|\}$ and frequencies $\{\omega_l\}$ (we split the complex gain in this manner in order to restrict attention to real parameters) – Theorem 1 guarantees that the CRB can be preserved (up to the M/N SNR penalty) by ensuring ϵ-isometry for the plane tangent to the manifold at this set of parameters. Therefore, to preserve the CRB for the frequency support $\boldsymbol{\omega}$, we need to guarantee ϵ-isometry for tangent-planes for all values that the gain magnitudes $\{|g_l|\}$ and the phases $\{g_l/|g_l|\}$ can take. We can show that the union of all such tangent planes is a subset of the span of the matrix $\mathbf{T}(\boldsymbol{\omega})$ (in $\mathbb{C}^N$), defined as

$$\mathbf{T}(\boldsymbol{\omega}) = \left[\mathbf{x}(\omega_1) \;\; \cdots \;\; \mathbf{x}(\omega_K) \;\; \tau\frac{d\mathbf{x}(\omega_1)}{d\omega} \;\; \cdots \;\; \tau\frac{d\mathbf{x}(\omega_K)}{d\omega} \right] \quad (48)$$

where $\tau = 1/\|d\mathbf{x}(\omega)/d\omega\|$ (note that τ does not depend on ω). Therefore, if the compressive measurement matrix $\boldsymbol{A}$ satisfies

$$\sqrt{\frac{M}{N}}(1-\epsilon) \leq \frac{\|\boldsymbol{A}\mathbf{T}(\boldsymbol{\omega})\mathbf{q}\|}{\|\mathbf{T}(\boldsymbol{\omega})\mathbf{q}\|} \leq \sqrt{\frac{M}{N}}(1+\epsilon) \;\; \forall \mathbf{q} \in \mathbb{C}^{2K}, \quad (49)$$

we can preserve the CRB (up to the SNR penalty) for a given frequency support $\boldsymbol{\omega}$. Furthermore, if the above relationship holds, we say that $\boldsymbol{A}$ satisfies the tangent plane ϵ-isometry property at $\boldsymbol{\omega}$.

Our first result is to show that the smallest singular value of the matrix $\mathbf{T}(\boldsymbol{\omega})$, given by $\delta = \min_{\mathbf{q} \in \mathbb{C}^{2K}} \|\mathbf{T}(\boldsymbol{\omega})\mathbf{q}\|/\|\mathbf{q}\|$,

compactly characterizes the number of measurements needed to preserve tangent plane ϵ-isometry.

Theorem 3. *Let $\boldsymbol{A}$ be an $M \times N$ measurement matrix whose entries are drawn i.i.d. from* Uniform $\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$. *Let $\mathbf{T}(\boldsymbol{\omega})$ denote the tangent plane matrix (48) of sinusoids (46) with frequencies $\boldsymbol{\omega} = (\omega_1 \ldots \omega_K) \in \mathbb{R}^K$. Let $\Lambda_T(\delta) = \{\boldsymbol{\omega} : \text{smallest singular value of } \mathbf{T}(\boldsymbol{\omega}) \geq \delta\}$. Then, for any $\epsilon > 0$, we have*

$$1 - \epsilon \leq \sqrt{\frac{N}{M}} \frac{\|\boldsymbol{A}\mathbf{T}(\boldsymbol{\omega})\mathbf{q}\|}{\|\mathbf{T}(\boldsymbol{\omega})\mathbf{q}\|} \leq 1 + \epsilon, \forall \boldsymbol{\omega} \in \Lambda_T(\delta), \mathbf{q} \in \mathbb{C}^{2K} \quad (50)$$

with high probability when $M = O(\epsilon^{-2} K \log(NK \epsilon^{-1} \delta^{-1}))$.

Remarks:

• The theorem states that the minimum number of measurements scales as the inverse of the smallest singular value δ. The singular values of $\mathbf{T}(\boldsymbol{\omega})$ are the square roots of the eigenvalues of $\mathbf{T}^H(\boldsymbol{\omega})\mathbf{T}(\boldsymbol{\omega})$, whose entries can be shown to depend only on the set of frequency *differences* $\omega_i - \omega_j, 1 \leq i, j \leq K$. Therefore, δ depends only on the set of frequency differences.

• The smallest singular value δ tends to zero when any two of the K frequencies (say ω_i and ω_j) get close, since the columns $\mathbf{x}(\omega_i)$ and $\mathbf{x}(\omega_j)$ (and hence the columns $d\mathbf{x}(\omega_i)/d\omega$ and $d\mathbf{x}(\omega_j)/d\omega$) approach each other, and the matrix $\mathbf{T}(\boldsymbol{\omega})$ becomes poorly conditioned. It is a natural question, therefore, to ask whether it is possible to provide a lower bound on δ, and hence an upper bound on the number of measurements required to give tangent plane isometries, by ensuring that the spacing between the constituent frequencies is large enough (larger than say $\Delta\omega$). We leave this as a topic for further investigation, since that characterization of the smallest singular value δ in terms of the minimum frequency separation $\Delta\omega$ is a feature of the original system with a full set of measurements rather than a problem inherent to compressive estimation. It is interesting to note that prior work on non-compressive frequency estimation [12], [13], [27], while not directly working with the parameter δ, also requires a minimum frequency separation for successful estimation (e.g., a separation of around four times the DFT spacing of $2\pi/N$) using N measurements.

• When the frequency support $\boldsymbol{\omega}$ is "*roughly*" known ahead of time (say $\boldsymbol{\omega} \approx \boldsymbol{\omega}_0$), such as in tracking scenarios encountered in radar (where frequencies correspond to directions of arrival), $\boldsymbol{A}$ need *only* preserve the norms of vectors in the span of $\mathbf{T}(\boldsymbol{\omega})$ for $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ (not all $\boldsymbol{\omega} \in \Lambda_T(\delta)$). Typically, the number of sinusoids K in the mixture is small. So, one can do better than the M/N SNR penalty that would be incurred if a compressive measurement matrix is used: In such a scenario, it may even be possible to preserve the CRB with no SNR degradation whatsoever. The equivalent problem of direction-of-arrival estimation is studied in [28]. The precise conditions on $\boldsymbol{A}$ so that the CRB is preserved with no SNR penalty are stated in [28]. This, however, requires knowing the very frequencies that we wish to estimate. Of course, this is not applicable to the one-shot estimation problem considered here, where we wish to preserve the CRB (up to the SNR penalty M/N) with a few measurements M, irrespective of what the particular realization of $\boldsymbol{\omega}$ is.

Pairwise isometry for a mixture of K sinusoids: Consider now the problem of quantifying the number of measurements needed to guarantee pairwise ϵ-isometry for a mixture of K sinusoids. We denote the matrix containing the sinusoids $[\mathbf{x}(\omega_1) \ \mathbf{x}(\omega_2) \ \ldots \mathbf{x}(\omega_K)]$ by $\mathbf{X}(\boldsymbol{\omega})$. From the definition of pairwise isometry in Section III, compressive measurements must preserve the ML cost structure, thereby implying that

$$\|\boldsymbol{A}\mathbf{X}(\boldsymbol{\omega})\mathbf{g} - \boldsymbol{A}\mathbf{X}(\boldsymbol{\omega}')\mathbf{g}'\| \approx \sqrt{M/N}\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g} - \mathbf{X}(\boldsymbol{\omega}')\mathbf{g}'\|, \quad (51)$$

for pairs of $(\mathbf{g}, \boldsymbol{\omega})$ and $(\mathbf{g}', \boldsymbol{\omega}')$ of interest. We are typically interested in all values of the gains $\mathbf{g}, \mathbf{g}'$ but may restrict the set of frequencies $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ to each come from a set Θ (for example, the set of K frequencies that are separated pairwise by at least $\Delta\omega$).

To simplify the problem, we only consider $\boldsymbol{\omega}$ and $\boldsymbol{\omega}'$ that are "well-separated" (we comment on why this helps later). For example, we may restrict $\boldsymbol{\omega}'$ to $\Theta'(\boldsymbol{\omega}) = \Theta \backslash B(\boldsymbol{\omega}, \mu)$, where $B(\boldsymbol{\omega}, \mu)$ is a small ball of frequencies around $\boldsymbol{\omega}$. (A possible definition for the ball $B(\boldsymbol{\omega}, \mu)$ can be $B(\boldsymbol{\omega}, \mu) = \{\boldsymbol{\omega}' : \min_{1 \leq i,j \leq K} |\omega_i' - \omega_j| \leq \mu\}$). Suppose that we make enough measurements to guarantee pairwise ϵ-isometry for all $\boldsymbol{\omega} \in \Theta$ and $\boldsymbol{\omega}' \in \Theta'(\boldsymbol{\omega})$, no matter what value $\boldsymbol{\omega}$ takes. This implies that for any set of frequencies $\boldsymbol{\omega} \in \Theta$, we have preserved the cost-structure of the estimation problem at hypothesis frequencies $\boldsymbol{\omega}'$ that are "far-away" ($\boldsymbol{\omega}'$ outside $B(\boldsymbol{\omega}, \mu)$). Roughly, a good estimation algorithm should not incur frequency errors larger than μ at high SNRs.

We introduce some notation for the following discussion. Let $\tilde{\boldsymbol{\omega}} = [\boldsymbol{\omega} \ \boldsymbol{\omega}'], \tilde{\mathbf{g}} = [\mathbf{g} \ -\mathbf{g}']$ denote vectors of length $2K$ concatenating the gains and frequencies. Also let $\mathbf{X}(\tilde{\boldsymbol{\omega}}) = [\mathbf{X}(\boldsymbol{\omega}) \ \mathbf{X}(\boldsymbol{\omega}')]$ denote the $N \times 2K$ matrix containing all the sinusoids. Note that $\tilde{\mathbf{g}}$ can take any value in $\mathbb{C}^{2K}$ but $\tilde{\boldsymbol{\omega}}$ has a special structure: its first K entries $\boldsymbol{\omega}$ must belong to Θ and its last K entries come from a set $\Theta'(\boldsymbol{\omega})$ that depend on the first K values. As shorthand, we say that $\tilde{\boldsymbol{\omega}} \in \tilde{\Theta} = \{[\boldsymbol{\omega} \ \boldsymbol{\omega}'] : \boldsymbol{\omega} \in \Theta, \boldsymbol{\omega}' \in \Theta'(\boldsymbol{\omega})\}$. With this notation, the above pairwise isometry condition for a mixture of K sinusoids, which we desire can be written as

$$\sqrt{\frac{M}{N}}(1 - \epsilon) \leq \frac{\|\boldsymbol{A}\mathbf{X}(\tilde{\boldsymbol{\omega}})\tilde{\mathbf{g}}\|}{\|\mathbf{X}(\tilde{\boldsymbol{\omega}})\tilde{\mathbf{g}}\|} \leq \sqrt{\frac{M}{N}}(1 + \epsilon) \ \forall \tilde{\mathbf{g}} \in \mathbb{C}^{2K}, \quad (52)$$

for a particular $\tilde{\boldsymbol{\omega}} \in \tilde{\Theta}$. If the matrix $\boldsymbol{A}$ satisfies this relationship, we say that $\boldsymbol{A}$ guarantees ϵ-isometry (*just isometry, not pairwise*) for the frequency support $\tilde{\boldsymbol{\omega}}$ ($2K$ sinusoids).

Our goal is to quantify the number of measurements necessary for (52) to hold for all $\tilde{\boldsymbol{\omega}} \in \tilde{\Theta}$. While solving this problem in its entirety is difficult, we can break it down into two subproblems, the first of which we tackle. We explain the solution to this subproblem and then comment on the other. In analogy with our previous discussion of tangent plane isometry, let $\Lambda_p(\delta)$ denote the set of all frequencies $\tilde{\boldsymbol{\omega}}$ (chosen from anywhere in $\mathbb{R}^{2K}$, not just $\tilde{\Theta}$) such that the smallest singular value of $\mathbf{X}(\tilde{\boldsymbol{\omega}})$ is at least as large as δ. Suppose that we want $\boldsymbol{A}$ to guarantee ϵ-isometry for all $\tilde{\boldsymbol{\omega}} \in \Lambda_p(\delta)$ (as in (52) except that the set from which $\tilde{\boldsymbol{\omega}}$ is chosen has changed). We show that $M = O\left(\epsilon^{-2}(2K) \log\left(N(2K)\epsilon^{-1}\delta^{-1}\right)\right)$ measurements suffice to provide such a guarantee with high probability.

Theorem 4. *Suppose that $\boldsymbol{A}$ is an $M \times N$ measurement matrix whose entries are drawn i.i.d. from* Uniform $\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$. *Let* $\mathbf{X}(\boldsymbol{\omega}) = [\mathbf{x}(\omega_1)\ \mathbf{x}(\omega_2)\ \ldots\ \mathbf{x}(\omega_K)]$ *denote an $N \times K$ matrix of sinusoids (46) with $\boldsymbol{\omega} = (\omega_1 \ldots \omega_K) \in \mathbb{R}^K$. Let $\Lambda_p(\delta) = \{\boldsymbol{\omega} : $ smallest singular value of $\mathbf{X}(\boldsymbol{\omega})$ is greater than or equal to $\delta\}$. For any $\epsilon > 0$ and $\delta > 0$, we have*

$$1 - \epsilon \leq \sqrt{\frac{N}{M}} \frac{\|\boldsymbol{A}\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|}{\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|} \leq 1 + \epsilon, \forall \boldsymbol{\omega} \in \Lambda_p(\delta), \mathbf{g} \in \mathbb{C}^K, \quad (53)$$

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

Remarks:

• Returning to the problem posed in (52), suppose that the smallest singular value of $\mathbf{X}(\tilde{\omega})$, further minimized over all values of $\tilde{\omega} \in \tilde{\Theta}$ is $\sigma_{min} > 0$. Then, $\tilde{\Theta}$ is contained in $\Lambda_p(\sigma_{min})$ and using Theorem 4, $M = O\left(\epsilon^{-2}(2K) \log\left(N(2K)\epsilon^{-1}\sigma_{min}^{-1}\right)\right)$ measurements suffice to guarantee the required ϵ-isometry.

• While the singular values of $\mathbf{X}(\tilde{\omega})$ depend only on frequency differences, we leave the question of quantifying σ_{min} (e.g., in terms of the minimum pairwise separation $\Delta\omega$ of frequencies for $\boldsymbol{\omega} \in \Theta$) and μ (the radius of the ball around each $\boldsymbol{\omega} \in \Theta$) as an open problem. The problem of lower bounding the singular values of the Fourier matrix ($\mathbf{X}(\boldsymbol{\omega})$, when choosing $\{h_n\}$ in (46) as the all-ones sequence) as a function of minimum frequency separation has been investigated in [29] (using Gershgorin-type bounds). Similar ideas may be useful in our present context as well, but again, these are fundamental and difficult questions regarding the original frequency estimation problem (with a full set of measurements) that are beyond our scope here. We are, however, able to provide an explicit characterization for the special case of a single sinusoid in Appendix D.

• The previous remark also explains why we choose to restrict ω' to $\Theta'(\omega) = \Theta \backslash B(\boldsymbol{\omega}, \mu)$. The singular value of $\mathbf{X}(\tilde{\omega})$ when $\boldsymbol{\omega}, \omega' \in \Theta$ can be made arbitrarily small by allowing $\omega' \to \boldsymbol{\omega}$. Thus, in this case, we cannot directly use Theorem 4 to quantify the number of measurements required. However, this does not necessarily mean that an isometry cannot be provided for closely spaced sinusoids. Indeed, we show in Appendix D that, for $K = 1$, it is possible to provide an isometry no matter how close ω and ω' get.

Proof of Theorems 3 and 4: We give a proof of Theorem 4 along the lines of the proof in [6], where the authors extend the JL lemma (which gives the number of compressive measurements needed to preserve the geometry of a discrete point cloud) to a manifold by sampling the manifold and exploiting its continuity. Details of the proof can be found in Appendix B. A similar proof can be given for Theorem 3, which we briefly sketch in Appendix C.

A. Pairwise isometry for frequency estimation of a single sinusoid

In the preceding discussions, we quantify the number of measurements needed to give pairwise isometries for a mixture of K sinusoids in two distinct regimes: when the frequencies (ω, ω') are "far apart" and in the limit of $\omega' \to \omega$ (tangent plane isometries). We now consider a single sinusoid ($K = 1$) and provide pairwise isometries for *all* frequency pairs. In order to do this, we consider two regimes of frequency pairs (ω_1, ω_2): closely spaced and well-separated. For the set of well-separated frequencies, say $\{(\omega_1, \omega_2) : |\omega_1 - \omega_2| > \psi\}$, we obtain a bound on the smallest singular value of $\mathbf{X}(\omega) = [\mathbf{x}(\omega_1)\ \mathbf{x}(\omega_2)]$ and use it in Theorem 4 to immediately infer the number of measurements needed to guarantee pairwise ϵ-isometry for sinusoids from this set. The challenge then is in providing a similar result for sinusoids whose frequencies are separated by less than ψ. We solve this problem in two stages: first, we use Theorem 3 to infer the number of measurements needed to guarantee tangent plane ϵ-isometries for all frequencies (loosely, pairwise isometries for $\omega_1 \to \omega_2$). We then use the continuity of the sinusoidal manifold to extend these tangent plane ϵ-isometries to a pairwise 2ϵ-isometry for closely-spaced frequencies $\{(\omega_1, \omega_2) : |\omega_1 - \omega_2| < \psi\}$.

Theorem 5. *Suppose that $\boldsymbol{A}$ is an $M \times N$ measurement matrix whose entries are drawn i.i.d. from* Uniform $\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$. *Let $\mathbf{x}(\omega)$ denote a sinusoid (46) of frequency ω with weights $\{h_n\}$ such that $\sum |h_n|^2 = 1$. Let $H(\omega) = \sum_{n=1}^{n=N} |h_n|^2 e^{j\omega(n-(N+1)/2)}$ be such that (i) the maxima of $|H(\omega)|^2$ that occur at frequencies other than $\omega = 0$ (side-lobes) are smaller than some constant $D < 1$ (independent of N) and (ii) $|H(\omega)|^2$ is non-increasing in $(0, \pi/(2N))$. Then, for any $\epsilon > 0$,*

$$1 - \epsilon \leq \sqrt{\frac{N}{M}} \frac{\|g_1\boldsymbol{A}\mathbf{x}(\omega_1) - g_2\boldsymbol{A}\mathbf{x}(\omega_2)\|}{\|g_1\mathbf{x}(\omega_1) - g_2\mathbf{x}(\omega_2)\|} \leq 1 + \epsilon, \ \forall g_1, g_2, \omega_1, \omega_2$$
$$(54)$$

with high probability when $M = O(\epsilon^{-2} \log(N\epsilon^{-1}(1 - \tau\chi)^{-1}\zeta^{-1}\alpha^{-1}))$ where $\tau = 1/\|d\mathbf{x}(\omega)/d\omega\|$, $\chi = |dH(0)/d\omega|$, $\alpha = 1/(N\tau)$ and $\zeta = -\frac{N^{-2}}{2}\frac{d^2|H(0)|^2}{d\omega^2}$ are parameters of the windowing sequence $\{|h_n|^2\}$.

We give the proof of Theorem 5 in Appendix D. The condition that (i) $|H(\omega)|^2$ is monotonic in $(0, \pi/(2N))$ and (ii) all side-lobes peaks of $|H(\omega)|^2$ are smaller than an absolute constant $D < 1$ (the main-lobe peak $|H(0)|^2 = 1$ since $\sum |h_n|^2 = 1$) are mild. These conditions are satisfied by windowing sequences $\{|h_n|^2\}$ commonly used for spectral estimation, such as the all-ones, Hamming, Hanning, Triangular and Blackman sequences.

Remark: We state and prove Theorems 3, 4 and 5 for compressive measurements with projection weights (elements of the matrix $\boldsymbol{A}$) taken from $X \sim$ Uniform$\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$. In addition to the concentration results on $\|\boldsymbol{A}\mathbf{v}\|^2$ of the form (8) which we need to preserve the geometry of a discrete point cloud, we use the fact that the Frobenius norm $\|\boldsymbol{A}\|_F = \sqrt{M}$ w.p. 1 for this choice of distribution X. When the elements of $\boldsymbol{A}$ are drawn from other distributions such as the gaussian distribution for which these concentration results on $\|\boldsymbol{A}\mathbf{v}\|^2$ are also available [22], $\|\boldsymbol{A}\|_F^2$, which is the sum of the square of all elements of $\boldsymbol{A}$, can be shown to fall within $M(1 \pm \delta)$ w.h.p. Therefore, the conclusions of Theorems 3, 4 and 5 also apply when the elements of $\boldsymbol{A}$ are drawn from these distributions.

VIII. Conclusions and Future Work

For parameter estimation in AWGN, we have identified isometry conditions under which the only effect of making compressive measurements is an SNR penalty equal to the dimensionality reduction factor. We prove this by establishing a connection between the isometry conditions and the CRB/ZZB. For a mixture of K sinusoids of length N, we show that $O(K \log N K \delta^{-1})$ measurements suffice to provide such isometries, where δ is the smallest singular value of appropriate matrices (stronger results are obtained for $K = 1$). Based on the connection between the ZZB and CRB, we also observe that, in order to avoid large estimation errors, the compressive measurements must not only preserve the geometry, but the SNR after the dimension reduction penalty must also be above a threshold. We illustrate this by showing that, for frequency estimation for a single sinusoid, the convergence of the ZZB to the CRB can be used to tightly predict the number of measurements needed to avoid error floors.

We leave open the issue of establishing the relationship between the smallest singular value δ and the minimum spacing between sinusoids, and whether the stronger isometry results established for a single sinusoid can be extended to $K > 1$. Another interesting topic for future work is the development of an analytical understanding of multi-dimensional sinusoid estimation, motivated by practical applications such as large 2D arrays for mm-wave communication [11] and imaging. Finally, investigation of compressive parameter estimation in non-Gaussian settings is an interesting problem with few known results.

Appendix A
(Extended) Ziv-Zakai Bound Review[24]

Consider the problem of estimating a parameter $\boldsymbol{\theta}$ from measurements

$$\mathbf{y} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{z}, \ \boldsymbol{\theta} \in \Theta, \mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbb{I}). \qquad (55)$$

For an estimator $\hat{\boldsymbol{\theta}}(\mathbf{y})$, let $\boldsymbol{\epsilon} = \hat{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}$ denote the estimation error. The ZZB lower bounds the error $\mathbb{E}|\mathbf{a}^T \boldsymbol{\epsilon}|^2$ for any $\mathbf{a} \in \mathbb{R}^K$ by relating it to the probabilities of error in a sequence of detection problems. We begin by describing one of the detection problems.

Consider a simplified version of the preceding model, in which the parameter $\boldsymbol{\theta}$ takes only two values $\boldsymbol{\phi}$ and $\boldsymbol{\phi} + \boldsymbol{\delta}$, occurring with probabilities $p(\boldsymbol{\phi})/(p(\boldsymbol{\phi}) + p(\boldsymbol{\phi} + \boldsymbol{\delta}))$ and $p(\boldsymbol{\phi} + \boldsymbol{\delta})/(p(\boldsymbol{\phi}) + p(\boldsymbol{\phi} + \boldsymbol{\delta}))$, respectively. There are two possible ways to estimate $\boldsymbol{\theta}$:
- *Optimal detection-theoretic approach:* Compute the Bayesian posterior probabilities $p(\boldsymbol{\phi}|\mathbf{y})$ and $p(\boldsymbol{\phi} + \boldsymbol{\delta}|\mathbf{y})$. Choose $\boldsymbol{\phi}$ if $p(\boldsymbol{\phi}|\mathbf{y}) > p(\boldsymbol{\phi}+\boldsymbol{\delta}|\mathbf{y})$ and $\boldsymbol{\phi}+\boldsymbol{\delta}$ otherwise. Denote the probability of error with this approach by $f(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta})$.
- *Heuristic approach using the estimate $\hat{\boldsymbol{\theta}}(\mathbf{y})$:* Form the estimate $\hat{\boldsymbol{\theta}}(\mathbf{y})$; this could take any value in Θ, and is not restricted to $\{\boldsymbol{\phi}, \boldsymbol{\phi}+\boldsymbol{\delta}\}$. Classify based on the following rule: if $\mathbf{a}^T \hat{\boldsymbol{\theta}}(\mathbf{y}) < \mathbf{a}^T \boldsymbol{\phi} + (h/2)$, where $h = \mathbf{a}^T \boldsymbol{\delta}$, choose $\boldsymbol{\phi}$ to have occurred; else, choose $\boldsymbol{\phi} + \boldsymbol{\delta}$. Denote the probability of error with this scheme by $P_{sub}(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta})$.

Since the Bayesian detection rule is optimal, we have $f(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta}) \leq P_{sub}(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta})$.

In order to use this observation to bound $\mathbb{E}|\mathbf{a}^T \boldsymbol{\epsilon}|^2$, we begin with the identity

$$\mathbb{E}\left|\mathbf{a}^T \boldsymbol{\epsilon}\right|^2 = \frac{1}{2} \int_0^\infty \Pr\left(\left|\mathbf{a}^T \boldsymbol{\epsilon}\right| \geq h/2\right) h \ dh, \qquad (56)$$

and relate $\Pr\left(\left|\mathbf{a}^T \boldsymbol{\epsilon}\right| \geq h/2\right)$ to the probability of error with the heuristic rule $P_{sub}(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta})$ as follows:

$$\Pr\left(\left|\mathbf{a}^T \boldsymbol{\epsilon}\right| \geq h/2\right) = \int_{\mathbb{R}^K} (p(\boldsymbol{\phi}) + p(\boldsymbol{\phi} + \boldsymbol{\delta}))$$
$$P_{sub}(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta}) \ d\boldsymbol{\phi}, \quad (57)$$

where $\boldsymbol{\delta}$ is *any* vector satisfying $\mathbf{a}^T \boldsymbol{\delta} = h$. We now use the lower bound $P_{sub}(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta}) \geq f(\boldsymbol{\phi}, \boldsymbol{\phi} + \boldsymbol{\delta})$ in (57) and substituting back in (56), we get the basic version of the ZZB.

We can further tighten the bound in two ways: (a) by choosing $\boldsymbol{\delta}$ appropriately and (b) by exploiting the fact that $\Pr\left(\left|\mathbf{a}^T \boldsymbol{\epsilon}\right| \geq h/2\right)$ is non-increasing using the valley filling operation $\mathcal{V}\{\ \}$, defined as $\mathcal{V}\{q(h)\} = \max_{r \geq 0} q(h + r)$ (refer [24] for details). This gives us the ZZB in (16).

Appendix B
Proof of Theorem 4

Let $\boldsymbol{\omega} = [\omega_1 \ \cdots \ \omega_K]^T$, $\mathbf{g} = [g_1 \ \cdots \ g_K]^T$ and $\mathbf{X}(\boldsymbol{\omega}) = [\mathbf{x}(\omega_1) \ \cdots \ \mathbf{x}(\omega_K)]$. We note that an ϵ-isometry for all vectors of the form $\mathbf{X}(\boldsymbol{\omega})\mathbf{g}$ such that $\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\| > \delta$ and $\|\mathbf{g}\| = 1$ is equivalent to (53). We discretize the frequencies $[0, 2\pi]$ uniformly into R points (R is specified later) and obtain the set F. We first prove a $2\epsilon_0$ isometry for all vectors in the span of $\mathbf{X}(\mathbf{q})$ for all frequency tuplets $\mathbf{q} \in F^K$ (i.e., $\mathbf{q} = [q_1 \ \cdots \ q_K]^T$ with $q_l \in F$). We then extend this to a $3.5\epsilon_0$ isometry for vectors $\mathbf{X}(\boldsymbol{\omega})\mathbf{g}$ such that $\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\| > \delta$ and $\|\mathbf{g}\| = 1$ by: (a) approximating them to nearby points in the span of $\mathbf{X}(\mathbf{q})$, (b) choosing $R = O(N^{1.5} K^{0.5} \delta^{-1} \epsilon_0^{-1})$ so that the approximation is good.

Sampling: For any tuplet of sampled frequencies $\mathbf{q} \in F^K$, if $\mathbf{A}$ preserves the norm of $\left(6\epsilon_0^{-1}\right)^{2K}$ well-chosen samples in the span of $\mathbf{X}(\mathbf{q})$ up to $\epsilon_0 < 2/5$, it can be shown that $\mathbf{A}$ will preserve the norms of all vectors in the span of $\mathbf{X}(\mathbf{q})$ up to $2\epsilon_0$ [4] (since we are concerned only with ℓ_2 distances from sampled points, we map the unit ball in $\mathbb{C}^K$ to the unit ball in $\mathbb{R}^{2K}$ using the map $f(\mathbf{z}) = [\Re\{\mathbf{z}^T\} \ \Im\{\mathbf{z}^T\}]^T$ and use corresponding covering arguments in [4]. The other argument used in [4], which is closure w.r.t. to addition is satisfied in $\mathbb{C}^K$ as well). Since there are R^K sampled frequency tuplets $\mathbf{q} \in F^K$, by demanding that $\mathbf{A}$ preserves the norm of $R^K \left(6\epsilon_0^{-1}\right)^{2K}$ samples, we can provide a $2\epsilon_0$ isometry for the span of $\mathbf{X}(\mathbf{q}) \ \forall \mathbf{q} \in F^K$.

Isometry for mixtures of arbitrary frequencies: We now extend this to an $3.5\epsilon_0$ isometry result for vectors of the form $\mathbf{X}(\boldsymbol{\omega})\mathbf{g}$ such that $\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\| > \delta$ and $\|\mathbf{g}\| = 1$ by choosing R appropriately.

Let $\mathbf{q}$ be a tuplet in F^K that is close to $\boldsymbol{\omega}$ satisfying $\max_l |q_l - \omega_l| \leq \pi/R$. We let $\mathbf{e}_l = \mathbf{x}(\omega_l) - \mathbf{x}(q_l)$ and bound the absolute value of each term of $\mathbf{e}_l$ using the mean value theorem to get $\|\mathbf{e}_l\| \leq \pi N/(\sqrt{2}R)$. We use this to calculate a bound on the difference between a vector $\mathbf{X}(\boldsymbol{\omega})\mathbf{g}$ and its

approximated version $\mathbf{X}(\mathbf{q})\mathbf{g}$. Using the definition of $\mathbf{e}_l$, we obtain

$$\mathbf{X}(\boldsymbol{\omega})\mathbf{g} = \mathbf{X}(\mathbf{q})\mathbf{g} + \sum_{l=1}^{l=K} g_l \mathbf{e}_l. \tag{58}$$

Using the triangle inequality and the fact that $\sum_l |g_l| \le \sqrt{2K}$ (since $\|\mathbf{g}\|=1$), we have

$$\frac{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|}{\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|} \vartriangleleft 1 \pm \frac{\pi\sqrt{K}NR^{-1}}{\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|}. \tag{59}$$

where $x \vartriangleleft y \pm z$ denotes $y - z \le x \le y + z$.

Next, we bound the difference between the vectors $\boldsymbol{A}\mathbf{X}(\boldsymbol{\omega})\mathbf{g}$ and $\boldsymbol{A}\mathbf{X}(\mathbf{q})\mathbf{g}$. We see that $\|\boldsymbol{A}\|_F = \sqrt{M}$ and, therefore, have $\|\boldsymbol{A}\mathbf{e}_k\| \le \sqrt{M}\|\mathbf{e}_k\|$. Furthermore, since $\boldsymbol{A}$ preserves the norms of all vectors of the form $\mathbf{X}(\mathbf{q})\mathbf{g}$, where $\mathbf{q} \in F^K$ up to an isometry constant $2\epsilon_0$ (and scale factor of $\sqrt{M/N}$), we get

$$\sqrt{\frac{N}{M}}\frac{\|\boldsymbol{A}\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|}{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|} \vartriangleleft 1 \pm \left(2\epsilon_0 + \frac{\pi N\sqrt{NK}R^{-1}}{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|}\right). \tag{60}$$

Before we proceed to give the isometry result, we need to characterize how small $\|\mathbf{X}(\mathbf{q})\mathbf{g}\|$ can be in (60). Since $\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\| > \delta$, from (59) we have the following:

$$\frac{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|}{\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|} \vartriangleleft 1 \pm \pi\sqrt{K}N(R\delta)^{-1}. \tag{61}$$

Choosing $R = (4\pi)N\sqrt{NK}\epsilon_0^{-1}\delta^{-1}$, we have that

$$\frac{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|}{\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|} \vartriangleleft 1 \pm 0.25\epsilon_0. \tag{62}$$

For this choice of R, from (60), we see that

$$\sqrt{\frac{N}{M}}\frac{\|\boldsymbol{A}\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|}{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|} \vartriangleleft 1 \pm \left(2\epsilon_0 + \frac{0.25\epsilon_0\delta}{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|}\right). \tag{63}$$

Using the lower bound from (62), $\|\mathbf{X}(\mathbf{q})\mathbf{g}\| \ge (1 - 0.25\epsilon_0)\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\| \ge \delta(1 - 0.25\epsilon_0)$,

$$\sqrt{\frac{N}{M}}\frac{\|\boldsymbol{A}\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|}{\|\mathbf{X}(\mathbf{q})\mathbf{g}\|} \vartriangleleft 1 \pm 2.5\epsilon_0. \tag{64}$$

Substituting the bounds for $\|\mathbf{X}(\mathbf{q})\mathbf{g}\|$ in terms of $\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|$ from (62), we have that

$$\sqrt{\frac{N}{M}}\frac{\|\boldsymbol{A}\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|}{\|\mathbf{X}(\boldsymbol{\omega})\mathbf{g}\|} \vartriangleleft 1 \pm 3.5\epsilon_0. \tag{65}$$

Number of measurements: It only remains to specify the number of measurements M required to preserve the norms of the $R^K\left(6\epsilon_0^{-1}\right)^{2K}$ samples up to ϵ_0. Using the value for R just obtained, and setting $\epsilon = 3.5\epsilon_0$, we see that we must preserve the norms of $(18 \times 7^3\pi N^{1.5}K^{0.5}\epsilon^{-3}\delta^{-1})^K$ vectors (samples) up to $2\epsilon/7$ w.h.p. We relate the probability of preserving these norms to the number of measurements M via the concentration results (8) for Uniform$\{\pm 1/\sqrt{N}, \pm j/\sqrt{N}\}$ (setting δ in (8) and (9) to $32\epsilon/49$ – here we have used the fact that when $\epsilon < 1$, $\max\{(1 + 2\epsilon/7)^2 - 1, 1 - (1 - 2\epsilon/7)^2\} < 32\epsilon/49$). We employ the union bound and (8) to compute the probability that the norm of atleast one sample is not preserved. This probability becomes vanishingly small for $M = O\left(\epsilon^{-2}K\log\left(NK\epsilon^{-1}\delta^{-1}\right)\right)$ measurements, which concludes the proof.

APPENDIX C
PROOF OF THEOREM 3

For the matrix $\mathbf{T}(\boldsymbol{\omega})$, K of the columns are of the form $\tau d\mathbf{x}(\omega)/d\omega$, while the remaining K are of the form $\mathbf{x}(\omega)$. When $\tau d\mathbf{x}(\omega)/d\omega$ is approximated by $\tau d\mathbf{x}(q)/d\omega$, where q is the frequency on an uniformly spaced frequency grid with R points that is the closest to ω, the norm of the approximation error is upper bounded by $\pi N/(\sqrt{2}R)$. The upper bound on the norm of the error in approximating $\mathbf{x}(\omega)$ by $\mathbf{x}(q)$ used in theorem 4 is also $\pi N/(\sqrt{2}R)$. Therefore, by following the proof of theorem 4 with K set to $2K$ (because number of columns of $\mathbf{X}(\boldsymbol{\omega})$ is only K), we obtain the proof for theorem 3.

APPENDIX D
PROOF OF THEOREM 5

We present the results for closely spaced frequencies first (tangent plane isometries), and then move to the well-separated setting.

Tangent plane isometry: For a single sinusoid, the tangent plane matrix at ω is given by $\mathbf{T}(\omega) = [\mathbf{x}(\omega)\ \tau d\mathbf{x}(\omega)/d\omega]$ where $\tau = 1/\|d\mathbf{x}(\omega)/d\omega\|$. The smallest singular value of $\mathbf{T}(\omega)$, denoted by σ_{tangent}, satisfies

$$\sigma_{\text{tangent}}^2 = 1 - \tau\left|\langle\mathbf{x}(\omega), d\mathbf{x}(\omega)/d\omega\rangle\right| \tag{66}$$

$$= 1 - \tau\left|\sum_{n=1}^{n=N}|h_n|^2 j\omega\left(n - (N+1/2)\right)\right| \tag{67}$$

where the second equality is obtained from the definition of the sinusoid (46) by noting that the nth entry of $\mathbf{x}(\omega)$ is $h_n e^{j\omega(n-(N+1)/2)}$. From the definition of $H(\omega)$, we see that,

$$\sigma_{\text{tangent}}^2 = 1 - \tau\left|dH(0)/d\omega\right|, \tag{68}$$

and therefore $\sigma_{\text{tangent}} = \sqrt{1 - \tau\chi}$ where $\chi = |dH(0)/d\omega|$. By Jensen's inequality, we see that $\chi^2 < 1/\tau^2$ when the weight sequence $\{h_n\}$ has more than one non-zero tap. Thus, $\tau\chi < 1$ and therefore σ_{tangent} is strictly positive. Setting $\delta = \sqrt{1 - \tau\chi}$ in Theorem 3, we can provide tangent plane ϵ-isometries for a single sinusoid with $M = O\left(\epsilon^{-2}\log\left(N\epsilon^{-1}(1 - \tau\chi)^{-1}\right)\right)$ measurements.

Extending tangent plane isometry to pairwise isometry for frequencies separated by at most $1/N^{1.5}$: We now extend ϵ-isometry of the tangent planes to a pairwise 2ϵ-isometry for any two frequencies ω_1, ω_2 whose separation $\Delta = \omega_2 - \omega_1$ is "small" (we quantify how small later) by exploiting continuity. Let $q = (\omega_1 + \omega_2)/2$ be the average of the two frequencies. For small values of $|\Delta|$, a first-order Taylor series expansion for $\mathbf{x}(\omega_1)$ and $\mathbf{x}(\omega_2)$ around $\mathbf{x}(q)$ will have small errors. Such an expansion gives us

$$\mathbf{x}(\omega_1) = \mathbf{x}(q) - (\Delta/2)(d\mathbf{x}(q)/d\omega) + \mathbf{e}_1, \tag{69}$$
$$\mathbf{x}(\omega_2) = \mathbf{x}(q) + (\Delta/2)(d\mathbf{x}(q)/d\omega) + \mathbf{e}_2, \tag{70}$$

where $\mathbf{e}_1, \mathbf{e}_2$ are the approximation errors. Consider a linear combination $\mathbf{X}(\boldsymbol{\omega})\mathbf{g}$ where $\mathbf{X}(\boldsymbol{\omega}) = [\mathbf{x}(\omega_1)\ \mathbf{x}(\omega_2)]$ and $\mathbf{g} = [g_1\ g_2]$. This can be written as

$$\mathbf{X}(\boldsymbol{\omega})\mathbf{g} = \mathbf{v} + \mathbf{e} \tag{71}$$

where $\mathbf{e} = g_1\mathbf{e}_1 + g_2\mathbf{e}_2$ and

$$\mathbf{v} = (g_1 + g_2)\,\mathbf{x}(q) + (\Delta/2)\,(g_2 - g_1)\,(d\mathbf{x}(q)/d\omega), \quad (72)$$

lies in the span of $\mathbf{T}(q) = [\mathbf{x}(q)\ \tau(d\mathbf{x}(q)/d\omega)]$, the tangent plane at $\omega = q$.

Since $\mathbf{A}$ guarantees ϵ-isometries for tangent planes at all frequencies, for any vector $\mathbf{T}(q)\mathbf{h}$ in the tangent plane at q, the quantity $\|\mathbf{AT}(q)\mathbf{h}\|$ is bounded within $(1 \pm \epsilon)\sqrt{M/N}\|\mathbf{T}(q)\mathbf{h}\|$. Expanding out $\|\mathbf{AX}(\omega)\mathbf{g}\|/\|\mathbf{X}(\omega)\mathbf{g}\|$ in terms of $\mathbf{v}$ and $\mathbf{e}$ and applying the tangent plane isometry condition to $\|\mathbf{Av}\|/\|\mathbf{v}\|$, we can show that

$$\sqrt{\frac{N}{M}}\frac{\|\mathbf{AX}(\omega)\mathbf{g}\|}{\|\mathbf{X}(\omega)\mathbf{g}\|} \lhd 1 \ \pm\ \left(\epsilon + \frac{5\sqrt{N}\|\mathbf{e}\|}{\|\mathbf{v}\|}\right). \quad (73)$$

where $x \lhd y \pm z$ denotes $y - z \leq x \leq y + z$. Next, we get bounds on $\|\mathbf{e}\|$ and $\|\mathbf{v}\|$ as follows. First, we use the mean value theorem to show that the error is bounded as $\|\mathbf{e}\| \leq N^2\Delta^2/(4\sqrt{2})$. Next, since $\mathbf{v}$ lies in the span of $\mathbf{T}(q)$, we can use the bound on the minimum singular value of $\mathbf{T}(q)$ to get $\|\mathbf{v}\| \geq \sqrt{1 - \tau\chi}|\Delta|/(\sqrt{2}\tau)$. The details are given in Appendix E. Substituting these bounds in the above equation, we obtain

$$\sqrt{\frac{N}{M}}\frac{\|\mathbf{AX}(\omega)\mathbf{g}\|}{\|\mathbf{X}(\omega)\mathbf{g}\|} \lhd 1 \ \pm\ \left(\epsilon + \frac{5\tau|\Delta|N^{2.5}}{4\sqrt{1 - \tau\chi}}\right). \quad (74)$$

We note that $\tau = 1/\|d\mathbf{x}(\omega)/d\omega\|$ scales as $1/N$. Therefore, defining a scale-invariant constant $\alpha = 1/(N\tau)$, we see that, as long as the frequency separation $|\Delta| \leq (4\alpha\epsilon\sqrt{(1 - \tau\chi)/5})/N^{1.5}$, we can get a 2ϵ isometry

$$\sqrt{\frac{N}{M}}\frac{\|\mathbf{AX}(\omega)\mathbf{g}\|}{\|\mathbf{X}(\omega)\mathbf{g}\|} \lhd 1 \ \pm\ 2\epsilon. \quad (75)$$

Thus, if $\mathbf{A}$ provides an $\epsilon/2$ tangent plane isometry for all frequencies (which can be achieved with $M = O\left(\epsilon^{-2}\log\left(N\epsilon^{-1}(1 - \tau\chi)^{-1}\right)\right)$ measurements), we can extend it to a pairwise ϵ-isometry for the set of frequencies ω_1, ω_2 whose separation $|\omega_1 - \omega_2| \leq (4\alpha(\epsilon/2)\sqrt{(1 - \tau\chi)}/5)/N^{1.5}$.

Pairwise isometry for frequencies separated by more than $1/N^{1.5}$: We now use Theorem 4 to quantify the number of measurements necessary to guarantee pairwise ϵ-isometry for two frequencies that are separated by more than $\mu/N^{1.5}$, where $\mu = (4\alpha(\epsilon/2)\sqrt{(1 - \tau\chi)}/5)$.

First, we obtain a bound on the smallest singular value of $\mathbf{X}(\omega_1, \omega_2) = [\mathbf{x}(\omega_1)\ \mathbf{x}(\omega_2)]$. Denoting the smallest singular value by $\sigma^2_{\mathrm{signal}}$, we can show that it satisfies

$$\sigma^2_{\mathrm{signal}} = 1 - |\langle\mathbf{x}(\omega_1), \mathbf{x}(\omega_2)\rangle|. \quad (76)$$

Furthermore, we can show that $|\langle\mathbf{x}(\omega_1), \mathbf{x}(\omega_2)\rangle| = |H(\omega_1 - \omega_2)|$, where $H(\omega) = \sum_{n=1}^{n=N}|h_n|^2 e^{j\omega(n - (N+1)/2)}$. Thus, we have $\sigma^2_{\mathrm{signal}} = 1 - |H(\omega_1 - \omega_2)|$.

Suppose now that $|\omega_1 - \omega_2| > \mu/N^{1.5}$. For large values of N, the smallest singular value of $\mathbf{X}(\omega_1, \omega_2)$ is bounded as

$$\sigma_{\mathrm{signal}} > \sqrt{\frac{0.4\zeta\mu^2}{N}}, \ \text{ where } \zeta = -\frac{N^{-2}}{2!}\left.\frac{d^2|H(\omega)|^2}{d\omega^2}\right|_{\omega=0}. \quad (77)$$

The details are given in Appendix F.

We now apply Theorem 4 with $\delta = \sqrt{0.4\zeta\mu^2/N}$. The set of all frequencies $|\omega_1 - \omega_2| > \mu/N^{1.5}$ is

contained in $\Lambda_p(\sqrt{0.4\zeta\mu^2/N})$ and thus, we can guarantee pairwise ϵ-isometry for this set with $M = O\left(\epsilon^{-2}\log\left(N\epsilon^{-1}(1 - \tau\chi)^{-1}\zeta^{-1}\alpha^{-1}\right)\right)$ measurements.

Combining the isometries in the regimes $|\omega_1 - \omega_2| \leq \mu/N^{1.5}$ and $|\omega_1 - \omega_2| \geq \mu/N^{1.5}$ completes the proof of Theorem 5.

APPENDIX E
EXTENDING TANGENT PLANE ISOMETRY

We first derive a bound on $\|\mathbf{e}\|$. Applying the triangle inequality to $\mathbf{e}$, we obtain $\|\mathbf{e}\| \leq |g_1|\|\mathbf{e}_1\| + |g_2|\|\mathbf{e}_2\|$. Since the quantity we wish to bound $\|\mathbf{AX}(\omega)\mathbf{g}\|/\|\mathbf{X}(\omega)\mathbf{g}\|$ does not depend on $\|\mathbf{g}\|$, we can, without loss of generality, restrict attention to $\|\mathbf{g}\| = 1$. Thus, we have $|g_i| \leq 1$. We use the mean value theorem to obtain bounds on $\|\mathbf{e}_i\|$ $i = 1, 2$ (the mean value theorem relates $\mathbf{e}_i$ to $d^2\mathbf{x}(\omega_i')/d\omega^2$ for some $\omega_i' \in [\omega_1, \omega_2]$) and ultimately get $\|\mathbf{e}\| \leq N^2\Delta^2/(4\sqrt{2})$.

In order to obtain a lower bound for $\|\mathbf{v}\|$, we rewrite $\mathbf{v}$ as

$$\mathbf{v} = \mathbf{T}(q)\begin{bmatrix} \sqrt{2} & 0 \\ 0 & \frac{\Delta}{\sqrt{2}\tau} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} g_1 \\ g_2 \end{bmatrix}. \quad (78)$$

We now recall that the minimum singular value of the product of two matrices is at least as large as the product of their minimum singular values. The minimum singular value of $\mathbf{T}(q)$ is $\sigma_{\mathrm{tangent}} = \sqrt{1 - \tau\chi}$ and the corresponding value for the other two matrices are $|\Delta|/\sqrt{2}\tau$ and 1 respectively. Thus, the minimum singular value of the product of the three matrices is greater than $\sqrt{1 - \tau\chi} \times \frac{|\Delta|}{\sqrt{2}\tau}$. Since $\|\mathbf{g}\| = 1$, we immediately get the desired bound on $\|\mathbf{v}\|$.

APPENDIX F
SMALLEST SINGULAR VALUE FOR WELL-SEPARATED FREQUENCIES

We wish to obtain a lower bound for the smallest singular value $\sigma^2_{\mathrm{signal}}$ of the matrix $[\mathbf{X}(\omega_1)\ \mathbf{X}(\omega_2)]$ when the frequencies satisfy $|\omega_1 - \omega_2| > \mu/N^{1.5}$. First, we note that this is equivalent to upper-bounding $|H(\omega_1 - \omega_2)|$ since $\sigma^2_{\mathrm{signal}} = 1 - |H(\omega_1 - \omega_2)|$. Since $|H(\omega)|$ is not necessarily monotonic (imagine that $|h_n|^2$ is the Hamming window; $|H(\omega)|$, being the magnitude of the Fourier transform of $|h_n|^2$, has sidelobes), it is not true in general that that the maximum of $|H(\omega)|, |\omega| > \mu/N^{1.5}$ occurs at $\omega = \mu/N^{1.5}$. However, we now make two observations that allow us to analyze the behavior of $|H(\omega)|$ only at the minimum separation $\mu/N^{1.5}$.

First, if there were no restrictions on the frequencies (ω_1, ω_2), $|H(\omega_1 - \omega_2)|$ has a maximum $(= 1)$ when $\omega_1 = \omega_2$. Second, because (i) the set of the frequencies we are excluding $|\omega_1 - \omega_2| < \mu/N^{1.5}$ is very small (it is smaller than $\pi/(2N)$ for large enough N) and (ii) we restrict ourselves to sequences $\{h_n\}$ such that $|H(\omega)|$ is monotone in $(0, \pi/(2N))$, the maximum of $|H(\omega_1 - \omega_2)|, \pi/(2N) > |\omega_1 - \omega_2| > \mu/N^{1.5}$ is guaranteed to occur when $\omega_2 = \omega_1 \pm \mu/N^{1.5}$.

For small values of $|\omega_1 - \omega_2|$, we can expand $|H(\omega)|^2$ around $\omega = 0$ to get

$$|H(\omega_1 - \omega_2)|^2 = 1 - \zeta N^2(\omega_1 - \omega_2)^2 \pm O\left(N^4(\omega_1 - \omega_2)^4\right), \quad (79)$$

where $\zeta = -(N^{-2}/2!) \, d^2|H(\omega)|^2/d\omega^2\big|_{\omega=0}$. For $|\omega_1 - \omega_2| = \mu/N^{1.5}$, we have

$$\left|H\left(\mu/N^{1.5}\right)\right|^2 = 1 - \zeta\mu^2/N \pm O\left(1/N^2\right). \quad (80)$$

we see that $|H(\mu/N^{1.5})|$ approaches 1 with increasing N. Since we assume that all side-lobes are smaller than $D < 1$, there exists some N beyond which the maximum of $|H(\omega_1 - \omega_2)|, |\omega_1 - \omega_2| > \pi/(2N)$ is guaranteed to be smaller than $|H(\mu/N^{1.5})|$. Therefore, for sufficiently large N, the maximum of $|H(\omega_1 - \omega_2)|$ for all $|\omega_1 - \omega_2| > \mu/N^{1.5}$ occurs at $|\omega_1 - \omega_2| = \mu/N^{1.5}$. Plugging the expression for $|H(\omega_1 - \omega_2)|$ in σ_{signal}^2 for this frequency separation, we have that,

$$N\sigma_{\text{signal}}^2 \geq 0.5\zeta\mu^2 \pm O\left(1/N\right). \quad (81)$$

To arrive at the above expression we have used the following: $|H(\omega)| \leq 1 \ \forall \omega$ (since $\sum|h_n|^2 = 1$). This gives us

$$1 - |H(\omega)| \geq (1 + |H(\omega)|)(1 - |H(\omega)|)/2 \ \forall \omega. \quad (82)$$

Therefore, $\sigma_{\text{signal}}^2 \geq (1 - |H(\mu/N^{1.5})|^2)/2$, which yields (81). The first term in (81) given by $0.5\zeta\mu^2$, is a constant and the second term decays to zero (as $1/N$). Therefore, for large values of N, the second term is much smaller than the first and σ_{signal}^2 is bounded away from zero. In particular, for large enough N (how large it needs to be depends on μ and the behavior of $|H(\omega)|^2$ at $\omega = 0$), we have $\sigma_{\text{signal}} > \sqrt{0.4\zeta\mu^2/N}$.

REFERENCES

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

[2] 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.

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

[4] 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.

[5] D. Achlioptas, "Database-friendly random projections," ser. PODS '01. New York, NY, USA: ACM, 2001, p. 274281.

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

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

[8] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, 2012.

[9] A. Fannjiang and W. Liao, "Coherence pattern-guided compressive sensing with unresolved grids," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 179–202, 2012.

[10] D. Ramasamy, S. Venkateswaran, and U. Madhow, "Compressive estimation in AWGN: General observations and a case study," in *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, 2012, pp. 953–957.

[11] ——, "Compressive tracking with 1000-element arrays: A framework for multi-Gbps mm wave cellular downlinks," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, 2012, pp. 690–697.

[12] G. Tang, B. N. Bhaskar, and B. Recht, "Near Minimax Line Spectral Estimation," *CoRR*, vol. abs/1303.4348, 2013.

[13] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *CoRR*, vol. abs/1204.0562, 2012.

[14] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed Sensing off the Grid," *CoRR*, vol. abs/1207.6053, 2012.

[15] E. Arias-Castro and Y. Eldar, "Noise Folding in Compressed Sensing," *Signal Processing Letters, IEEE*, vol. 18, no. 8, pp. 478–481, 2011.

[16] J. Haupt and R. Nowak, "Compressive sampling for signal detection," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 3, april 2007, pp. III–1509 –III–1512.

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

[18] I. Maravic and M. Vetterli, "Sampling and reconstruction of signals with finite rate of innovation in the presence of noise," *Signal Processing, IEEE Transactions on*, vol. 53, no. 8, pp. 2788–2805, 2005.

[19] Z. Ben-Haim, T. Michaeli, and Y. Eldar, "Performance Bounds and Design Criteria for Estimating Finite Rate of Innovation Signals," *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 4993–5015, 2012.

[20] M. Mishali and Y. C. Eldar, "Xampling: Compressed Sensing for Analog Signals," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2012.

[21] H. L. Van Trees, K. L. Bell, and Z. Tian, *Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2013.

[22] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," arXiv e-print, Nov. 2010, chapter 5 of: Compressed Sensing, Theory and Applications. Edited by Y. Eldar and G. Kutyniok. Cambridge University Press, 2012.

[23] H. Van Trees and K. Bell, "Bayesian bounds for parameter estimation and nonlinear filtering and tracking," 2007.

[24] K. Bell, Y. Steinberg, Y. Ephraim, and H. Van Trees, "Extended Ziv-Zakai lower bound for vector parameter estimation," *Information Theory, IEEE Transactions on*, vol. 43, no. 2, pp. 624 –637, mar 1997.

[25] 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, 1974.

[26] S. Basu and Y. Bresler, "A global lower bound on parameter estimation error with periodic distortion functions," *Information Theory, IEEE Transactions on*, vol. 46, no. 3, pp. 1145–1150, 2000.

[27] E. Candes and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *CoRR*, vol. abs/1203.5871, 2012.

[28] A. Weiss and B. Friedlander, "Preprocessing for direction finding with minimal variance degradation," *Signal Processing, IEEE Transactions on*, vol. 42, no. 6, pp. 1478–1485, 1994.

[29] P. Ferreira, "Super-resolution, the recovery of missing samples and vandermonde matrices on the unit circle," in *Proceedings of the Workshop on Sampling Theory and Applications, Loen, Norway*, 1999.