Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Compressive sensing off the grid

Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht
University of Wisconsin-Madison

***Abstract*— We consider the problem of estimating the frequency components of a mixture of s complex sinusoids from a random subset of n regularly spaced samples. Unlike previous work in compressive sensing, the frequencies are not assumed to lie on a grid, but can assume any values in the normalized frequency domain $[0, 1]$. We propose an atomic norm minimization approach to exactly recover the unobserved samples, which is then followed by any linear prediction method to identify the frequency components. We reformulate the atomic norm minimization as an exact semidefinite program. By constructing a dual certificate polynomial using random kernels, we show that roughly $s \log s \log n$ random samples are sufficient to guarantee the exact frequency estimation with high probability, provided the frequencies are well separated. Extensive numerical experiments are performed to illustrate the effectiveness of the proposed method. Our approach avoids the basis mismatch issue arising from discretization by working directly on the continuous parameter space. Potential impact on both compressive sensing and line spectral estimation, in particular implications in sub-Nyquist sampling and super-resolution, are discussed.**

I. Introduction

In many modern signal processing systems, acquiring a signal by a sampling mechanism in an efficient and cost-effective manner is a major challenge. For computational, cost and storage reasons it is often desirable to not only acquire, but also to subsequently compress the acquired signal. A fundamental idea that has the potential to overcome the somewhat wasteful process of acquiring a signal using expensive hi-fidelity sensors, followed by compression and a subsequent loss of fidelity, is the possibility of compressive sensing: i.e. the realization that it is often possible to combine the signal acquisition and the compression step by sampling the signal in a novel way [1]–[4]. Compressive sensing explores different mechanisms that allow one to acquire a succinct representation of the underlying system while simultaneously achieving compression.

Despite the tremendous impact of compressive sensing on signal processing theory and practice, its development thus far has focused on signals with sparse representation in finite discrete dictionaries. However, signals we encounter in the real world are usually specified by continuous parameters, especially those in radar, sonar, sensor array, communication, seismology, remote sensing. Wideband analog signal with sparse spectrum is another example that is closely tied to sampling theory [5]–[7]. In order to apply the theory of CS, researchers in these fields adopt a discretization procedure to reduce the continuous parameter space to a finite set of grid points [8]–[14]. While this seemingly simple strategy gives superior performance for many problems provided that the true parameters do fall into the grid set, the discretization has several significant drawbacks.

Indeed, one major weakness of the discretization approach is *basis mismatch*, where the true signal cannot even be sparsely represented by the assumed discrete dictionary [13], [15], [16]. One might attempt to remedy basis mismatch by using a finer discretization or adding extra basis elements. However, increasing the size of the dictionary will also increase its coherence. Common wisdom in compressive sensing suggests that higher coherence will also degrade performance casting doubt as to whether over-discretization is beneficial to solving the problems. Finer gridding also results in higher computational complexity and numerical instability, further diminishing any advantage it might have in sparse recovery.

We overcome the issues arising from discretization by working directly on the continuous parameter space for a specific problem, where we estimate the continuous frequencies and amplitudes of a mixture of complex sinusoids from *randomly* observed time samples. This problem captures all of the essential ingredients of applying compressive sensing to problems with continuous dictionaries. In particular, the frequencies are not assumed to lie on a grid, and can instead take arbitrary values across the bandwidth of the signal. With a time-frequency exchange, our model is exactly the same as the one in Candès, Romberg, and Tao's foundational work on compressive sensing [1], except that we do not assume the frequencies lie on an equispaced grid. This major difference presents a significant technical challenge as the resulting dictionary is no longer an orthonormal Fourier basis, but is an infinite dictionary with continuously many atoms and arbitrarily high coherence.

In this paper we consider signals whose spectra consist of spike trains with unknown locations in a *continuous* normalized interval $[0, 1]$ and whose amplitudes are random but unknown. Rather than sampling the signal at all times $t = 0, \ldots, n-1$ we randomly sample the signal at times $t_1, \ldots t_m$ where each $t_j \in \{0, \ldots, n-1\}$. Our main contributions in this paper are the following:

1) Provided the original signal has a resolvable spectrum (in a sense that we make precise later), we show that the original signal can be reconstructed *exactly* from the random samples .
2) Our reconstruction algorithm is formulated as the solution to an atomic norm [17] minimization problem. We show that this convex optimization problem can be exactly reformulated as a semidefinite program (hence our methodology is computationally tractable)

978-1-4673-4539-2/12/$31.00 ©2012 IEEE

by exploiting a well-known result in systems theory called the bounded real lemma.

3) Our proof technique requires the demonstration of an explicit dual certificate that satisfies certain interpolation conditions. The production of this dual certificate requires us to consider certain random polynomial kernels, and derive concentration inequalities for these kernels. These results may be of independent interest to the reader.
4) We validate our theory by extensive numerical experiments that confirm random under-sampling as a means of compression, followed by atomic norm minimization as a means of recovery are a viable, superior alternative to discretization techniques.

This paper is organized as follows. In Section II we introduce the class of signals under consideration, the sampling procedure, and the recovery algorithm formally. Theorem II.2 is the main result of this paper. We outline connections to prior art and the foundations that we build upon in Section III. In Section IV we present the main proof idea (though we omit the detailed proofs due to space limitations). In Section V we present some supporting numerical experiments. In Section VI we make some concluding remarks and mention future directions.

II. Model and Main Results

We start with introducing the signal model and present the main results.

A. Problem Setup

Suppose we have a signal

$$x_j^\star = \frac{1}{\sqrt{|J|}} \sum_{k=1}^{s} c_k e^{i2\pi f_k j}, j \in J, \tag{1}$$

composed of complex sinusoids with s distinct frequencies $\Omega = \{f_1, \ldots, f_s\} \subset [0,1]$. Here J is an index set and $|J|$ is the size of J. In this paper, J is either $\{0, \ldots, n-1\}$ or $\{-2M, \ldots, 2M\}$ for some positive integers n and M. Any mixture of sinusoids with frequencies bandlimited to $[-W, W]$, after appropriate normalization, can be assumed to have frequencies in $[0,1]$. Consequently, a bandlimited signal of such a form leads to samples of the form (1).

We emphasize that, unlike previous work in compressive sensing where the frequencies are assumed to lie on a finite set of discrete points [1], [13], [14], [18], the frequencies in this work could lie anywhere in the normalized continuous domain $[0,1]$.

Instead of observing $x_j^\star$ for all $j \in J$, we observe only entries in a subset $T \subset J$. The goal is to recover the missing entries from the observed entries by exploiting the sparsity of frequencies in the continuous domain $[0,1]$. Once the missing entries are recovered exactly, the frequencies can be identified by Prony's method, the matrix pencil approach, or other linear prediction methods. After identifying the frequencies, the coefficients $\{c_k\}_1^s$ can be obtained by solving a linear system.

B. Atomic Norms

Define atoms $a(f) \in \mathbb{C}^{|J|}, f \in [0,1]$ via

$$[a(f)]_j = \frac{1}{\sqrt{|J|}} e^{i2\pi f j}, j \in J \tag{2}$$

and rewrite the signal model (1) in matrix-vector form:

$$x^\star = \sum_{k=1}^{s} c_k a(f_k) = \sum_{k=1}^{s} |c_k| e^{i\phi(c_k)} a(f_k) \tag{3}$$

where $\phi(\cdot) : \mathbb{C} \mapsto [0, 2\pi)$ extracts the phase of a complex number. The set of atoms $\mathcal{A} = \{e^{i\phi} a(f), f \in [0,1], \phi \in [0, 2\pi)\}$ are building blocks of the signal $x^\star$, the same way that canonical basis vectors are building blocks for sparse signals, and unit-norm rank one matrices are building blocks for low-rank matrices. In sparsity recovery and matrix completion, the unit balls of the sparsity-enforcing norms, e.g., the ℓ_1 norm and the nuclear norm, are exactly the convex hulls of their corresponding building blocks. In a similar spirit, we define an atomic norm $\|\cdot\|_{\mathcal{A}}$ by identifying its unit ball with the convex hull of $\mathcal{A}$:

$$\begin{aligned} \|x\|_{\mathcal{A}} &= \inf\{t > 0 : x \in t \operatorname{conv}(\mathcal{A})\} \\ &= \inf_{\substack{c_k \geq 0,\ \phi_k \in [0,2\pi) \\ f \in [0,1]}} \Big\{ \sum_k c_k : x = \sum_k c_k e^{i\phi_k} a(f_k) \Big\}. \end{aligned} \tag{4}$$

Roughly speaking, the atomic norm $\|\cdot\|_{\mathcal{A}}$ can enforce sparsity in $\mathcal{A}$ because low-dimensional faces of $\operatorname{conv}(\mathcal{A})$ correspond to signals involving only a few atoms. The idea of using atomic norm to enforce sparsity for a general set of atoms was first proposed and analyzed in [17].

Equation (4) indeed defines a norm if the set of atoms is bounded, centrally symmetric, and absorbent, which are satisfied by our choice of $\mathcal{A}$. The dual norm of $\|\cdot\|_{\mathcal{A}}$ is

$$\|z\|_{\mathcal{A}}^* = \sup_{\|x\|_{\mathcal{A}} \leq 1} \langle z, x \rangle_{\mathbb{R}} = \sup_{\phi \in [0,2\pi), f \in [0,1]} \langle z, e^{i\phi} a(f) \rangle_{\mathbb{R}} \tag{5}$$

In this paper, for complex column vectors x and z, the complex and real inner products are defined as

$$\langle z, x \rangle = x^* z, \langle z, x \rangle_{\mathbb{R}} = \operatorname{Re}(x^* z) \tag{6}$$

respectively, where the superscript * represents conjugate transpose.

C. Computational Method

With the atomic norm $\|\cdot\|_{\mathcal{A}}$, we recover the missing entries of $x^\star$ by solving the following convex optimization problem:

$$\underset{x}{\text{minimize}} \ \|x\|_{\mathcal{A}} \ \text{subject to} \ x_j = x_j^\star, j \in T. \tag{7}$$

At this point, it is not clear at all that solving (7) is computationally feasible, despite that it is a convex optimization (norm minimization subject to linear constraint, more explicitly). In this subsection, we present an exact computational method based on semidefinite programming to solve (7). When $J = \{0, \cdots, n-1\}$ or $\{-2M, \cdots, 2M\}$,

the atomic norm $\|x\|_{\mathcal{A}}$ defined in (4) is equal to the optimal value of the following semidefinite program:

$$\underset{X,u,t}{\text{minimize}} \ \frac{1}{2}\operatorname{tr}(X) \text{ subject to } X = \begin{bmatrix} \operatorname{Toep}(u) & x \\ x^* & t \end{bmatrix} \succeq 0. \tag{8}$$

Here the linear Toeplitz operator constructs a Hermitian Teoplitz matrix $\operatorname{Toep}(u)$ from a complex vector $u \in \mathbb{C}^n$ whose first element u_0 is real. The case for $J = \{0, \ldots, n-1\}$ was first shown in [19] via the bounded real lemma [20, Section 4.3]. The other case can be proved in a similar manner.

The semidefinite programming characterization (8) of the atomic norm allows us to reformulate the optimization (7) as a semidefinite program:

$$\begin{aligned} \underset{X,x,u,t}{\text{minimize}} \ & \frac{1}{2}\operatorname{tr}(X) \\ \text{subject to} \ & X = \begin{bmatrix} \operatorname{Toep}(u) & x \\ x^* & t \end{bmatrix} \succcurlyeq 0 \\ & x_j = x_j^\star, j \in T. \end{aligned} \tag{9}$$

D. Random Model

To quantify the performance of (7), we adopt a Bernoulli observation model and a random sign signal model.

Bernoulli Observation Model: We observe entries in J independently with probability p. Let $\delta_j = 1$ or 0 indicate whether we observe the jth entry. Then $\{\delta_j\}_{j \in J}$ are i.i.d. Bernoulli random variables such that

$$\mathbb{P}(\delta_j = 1) = p. \tag{10}$$

Therefore, on average we observe $p|J|$ entries,

Random Sign Model: We also assume that the complex signs for the coefficients c_k are uniformly random, namely, $\{\operatorname{sign}(c_k) = c_k/|c_k|, k = 1, \ldots, s\}$ are independent and identically distributed according to the uniform distribution on the complex unit circle.

E. Main Results

Our results depend on a condition on the minimum separation between the frequencies defined as:

$$\Delta_f = \min_{k \neq j} |f_k - f_j| \tag{11}$$

where the distance $|f_k - f_j|$ is understood as the wrap-around distance on the unit circle. The lower bound on Δ_f required by our theorems determines the maximum frequency resolution.

Our major result concerns the symmetric case $J = \{-2M, \ldots, 2M\}$, where we use a Bernoulli observation model with

$$\mathbb{P}(\delta_j = 1) = p = m/M, j = -2M, \cdots, 2M. \tag{12}$$

Theorem II.1. *Suppose we observe the time samples of*

$$x_j = \frac{1}{\sqrt{4M+1}} \sum_{k=1}^{s} c_k e^{i2\pi f_k j} \tag{13}$$

on the index set $T \subset J = \{-2M, \ldots, 2M\}$ according to the Bernoulli observation model (12) *and* $\operatorname{sign}(c_k)$*s follow i.i.d. uniform distribution on the complex unit circle. If $\Delta_f \geq \frac{1}{M}$, then there exists a numerical constant C such that*

$$m \geq C \max\left\{\log^2 \frac{M}{\delta}, s \log \frac{s}{\delta} \log \frac{M}{\delta}\right\}, \tag{14}$$

is sufficient to guarantee that with probability at least $1-\delta$, $x^\star$ is the unique optimizer to (7).

It is not difficult to translate this result to the case $J = \{0, \ldots, n-1\}$.

Corollary II.2. *Suppose we observe the time samples of*

$$x_j = \frac{1}{\sqrt{n}} \sum_{k=1}^{s} c_k e^{i2\pi f_k j} \tag{15}$$

on the index set $T \subset J = \{0, \ldots, n-1\}$ according to the Bernoulli observation model with the average number of samples $m = \mathbb{E}|T|$ and $\operatorname{sign}(c_k)$*s follow i.i.d. uniform distribution on the complex unit circle. If $\Delta_f \geq \frac{1}{\lfloor (n-1)/4 \rfloor}$, then there exists a numerical constant C such that*

$$m \geq C \max\left\{\log^2 \frac{n}{\delta}, s \log \frac{s}{\delta} \log \frac{n}{\delta}\right\}, \tag{16}$$

is sufficient to guarantee that with probability at least $1-\delta$, $x^\star$ is the unique optimizer to (7).

Remark II.3. **(Sub-Nyquist Sampling)** The normalization in our model is such that we sample above the Nyquist frequency when we observe all entries. To see this, suppose the frequencies lie in $[-W, W]$, and $x^\star(t)$ is a continuous signal of the form:

$$x^\star(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{s} c_k e^{i2\pi w_k t}, w_k \in [-W, W], t \in \mathbb{R}. \tag{17}$$

By taking regular spaced Nyquist samples at $t \in \{0/2W, 1/2W, \ldots, (n-1)/2W\}$, we observe

$$\begin{aligned} x_j^\star := x^\star(j/2W) &= \frac{1}{\sqrt{n}} \sum_{k=1}^{s} c_k e^{i2\pi \frac{w_k}{W} j} \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^{s} c_k e^{i2\pi f_k j} \text{ with } f_k = \frac{w_k}{2W} \in \left[-\frac{1}{2}, \frac{1}{2}\right], \end{aligned} \tag{18}$$

exactly the same as our model (1) after a trivial translation of the frequency domain.

However, if we use random sampling studied in this paper, we take exponentially less samples, namely, reducing the number of samples from n to $O(s \log s \log n)$. Therefore, by exploiting sparsity in the signal spectrum, we could sample beyond Nyquist, exponentially. Compressed sensing has shown that such scaling is possible when the frequencies are constrained to lie on a discrete grid [1], [5]–[7]. The main innovation in this work is that we do not assume such a grid and allow the frequencies to take a continuum of values.

Remark II.4. **(Resolution)** The number of measurements required is roughly at the order of $O(s \log s \log n)$. The resolution we get is roughly $\frac{4}{n}$. Therefore, by using random sampling, we increase the resolution from $\frac{4}{s \log s \log n}$, which is what we get using equispaced sampling [21], to $\frac{4}{n}$, i.e.,

an exponential improvement. We comment that numerical simulations in Section V suggest that the critical separation is actually $\frac{1}{n}$. We leave tightening our bounds by the extra constant of 4 to future work.

F. Basis Mismatch Conundrum

Our result resolves the basis mismatch conundrum of the discretization method. First, we note that we can approximately solve the optimization problem (7) is described below by discretizing the frequency domain $[0, 1]$ into N grid points $\{j/N\}_0^{N-1} \subset [0, 1]$ and considering the set of atoms:

$$\mathcal{A}_N = \left\{ e^{i\phi} a\left(j/N\right) : \phi \in [0, 2\pi), 0 \le j \le N-1 \right\} \subset \mathcal{A}. \tag{19}$$

The set of atoms $\mathcal{A}_N$ also defines an atomic norm $\|\cdot\|_{\mathcal{A}_N}$. The authors of [19] showed that $\|\cdot\|_{\mathcal{A}_N}$ approximates $\|\cdot\|_{\mathcal{A}}$ for reasonably fine discretization. More specifically, for any $x \in \mathbb{C}^{|J|}$, we have

$$\left(1 - 2\pi|J|/N\right)\|x\|_{\mathcal{A}_N} \le \|x\|_{\mathcal{A}} \le \|x\|_{\mathcal{A}_N}. \tag{20}$$

The computation of $\|\cdot\|_{\mathcal{A}_N}$ involves solving a second-order cone problem, or more specifically, a basis pursuit problem:

$$\|x\|_{\mathcal{A}_N} = \min\{\|c\|_1 : x = Ac\}, \tag{21}$$

where $A \in \mathbb{C}^{|J|\times N}$ is the matrix whose jth column is $a\left(j/N\right)$. If the coefficients are real, then this problem can be solved by linear programming. If we replace $\|x\|_{\mathcal{A}}$ with $\|x\|_{\mathcal{A}_N}$ in (7), the resulting optimization is also a basis pursuit problem:

$$\underset{c}{\text{minimize}} \ \|c\|_1 \text{ subject to } x_j^\star = (Ac)_j, j \in T. \tag{22}$$

However, as mentioned in the introduction, when the frequencies do not fall onto the grid, the signal $x^\star$ is not sparse or even compressible [13], [15], potentially degrading the performance of sparsity recovery. If one tries to remedy the issue by increasing the level of discretization, not only does one need to solve larger convex programs, but the coherence of the resulting dictionary will also increase. Hence, there seems to be a dilemma in sparse representation and incoherence regarding whether to use finer grid.

In our atomic norm formulation, the dictionary can be seen as having continuously many elements and is arbitrarily coherent globally. Therefore, Corollary II.2 shows that the global coherence of the frame is not an obstacle to recovery. What matters more is the local coherence characterized by the separation between frequencies in the true signal. Although we do observe performance improvement by increasing the discretization level, overly fine grid usually leads to dictionaries that are poorly conditioned, yielding unstable numerical procedures, and a substantial increase in computational complexity. In this regard, the continuous formulation (7) and the corresponding semidefinite program (9) seem to be a more desirable choice. We validate this claim empirically in Section V.

III. Prior Art and Inspirations

Accurate frequency estimation from time samples is a fundamental task in signal processing. When the frequency domain is composed of line spectra, the so called *line spectral estimation* is critical to many applications including sensor array networks, array signal processing, NMR, radar target detection and sonar [22]. Problems involving signals composed of shifted versions of a common mother function [23], such as those in neural spike train detection [24] and LIDAR [25], can also be formulated as a line spectral estimation using Fourier transform.

Frequency estimation is extensively studied and techniques for estimating sinusoidal frequencies from time samples goes back to Prony [26]. Many linear prediction algorithms based on Prony's method were proposed to estimate the frequencies from *regularly spaced* time samples. A survey of these methods can be found in [27] and an extensive list of references is given in [28]. With equispaced samples, these root-finding based procedures deal with the problem directly on the continuous frequency domain, and can recover frequencies as long as the number of samples is at least twice of the number of frequencies, no matter how closely these frequencies are located [22], [26]–[28]. In recent work [21], Candès and Fernandez-Granda studied this problem from the point of view of convex relaxations and proposed a total-variation norm minimization formulation that provably recovers the spectrum exactly. However, the convex relaxation requires the frequencies to be well separated. In [19], the authors proposed using atomic norm to denoise a line spectral signal corrupted with gaussian noise, and reformulated the resulting atomic norm minimization as a semidefinite program using bounded real lemma [20]. Denosing is important to frequency estimation since the frequencies in a line spectral signal corrupted with moderate noise can be identified by linear prediction algorithms. Since the atomic norm framework in [19] is essentially the same as the total-variation norm framework of [21], the same semidefinite program can also be applied to total-variation norm minimization.

What is common to all aforementioned approaches, including linear prediction methods, is the reliance on observing uniform or equispaced time samples. In sharp contrast, we show that nonuniform sampling is not only a viable option, and that the original spectrum can be recovered exactly in the continuous domain, but in fact is a means of *compressive* sampling. Indeed non-uniform sampling allows us to effectively sample the signal at a sub-Nyquist rate. For array signal processing applications, this corresponds to the reduction in the number of sensors required for exact recovery, since each sensor obtains one spatial sample of the field. An extensive justification of the necessity of using randomly located sensor arrays can be found in [29]. To the best of our knowledge, little is known about exact line-spectrum recovery with non-uniform sampling, except sporadic work using ℓ_2-norm minimization to recover the missing samples [30], or based on nonlinear least square data fitting [31].

An interesting feature related to using convex optimization based methods for estimation such as [21] is a particular resolvability condition: the separation between frequencies is required to be greater than $\frac{4}{n}$ where n is the number of measurements. Linear prediction methods do not have a resolvability limitation, but it is known that in practice the numerical stability of root finding limits how close the frequencies can be. Theorem II.1 can be viewed as an extension of the theory to nonuniform samples. Note that our approach gives an exact semidefinite characterization using the bounded real lemma, and is hence computationally tractable. We believe our results have potential impact on two related areas: extending compressive sensing to continuous dictionaries, and extending line spectral estimation to nonuniform sampling, thus providing new insight in sub-Nyquist sampling and super-resolution.

IV. Proof Ideas

In this section we briefly describe the proof architecture. The interested reader can refer to the technical report for more details. The following proposition indicates that the key is to construct a dual polynomial to certify the unique optimality of $x^\star$.

Proposition IV.1. *The original signal $x^\star$ is the unique optimizer to (7) if there exists a dual polynomial*

$$Q(f) = \frac{1}{\sqrt{|J|}} \sum_{j \in J} q_j e^{-i2\pi j f} \tag{23}$$

satisfying

$$Q(f_k) = \operatorname{sign}(c_k), \forall f_k \in \Omega \tag{24}$$

$$|Q(f)| < 1, \forall f \notin \Omega \tag{25}$$

$$q_j = 0, \forall j \notin T. \tag{26}$$

In the equispaced case where $T = J$, Candès and Fernandez-Granda [21] exhibited an explicit dual certificate composed of the squared Fejér kernel (denoted by $K(\cdot)$) as follows: Write

$$Q(f) = \sum_{k=1}^{s} \alpha_k K(f - f_k) + \sum \beta_k K'(f - f_k).$$

One then selects the coefficients α and β such that the interpolation conditions (24) are satisfied, and in addition such that at the f_ks, $Q(f)$ attains a local maximum (so that (25) is satisfied). Note that the condition (26) is absent in their setting since the set T^c is empty. The squared Fejer kernel is a good candidate kernel because it attains the value of 1 at its peak, and rapidly decays to zero. Provided a separation conditions is satisfied by the original signal, a suitable set of coefficients α and β can always be found.

The main difference in our setting is that the demands of our polynomial $Q(f)$ are much stricter as manifested by (26). Our polynomial is required to have *most* of its coefficients be zero. The main proof technique then is to modify the kernel $K(\cdot)$ to its random counterpart $K_{\mathrm{r}}(\cdot)$, which has nonzero coefficients only on the random subset T, with the property that $\mathbb{E}K_{\mathrm{r}} = K$ and such that K_{r} concentrates tightly around K.

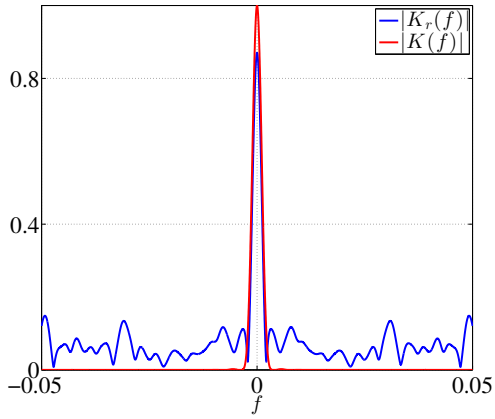


Fig. 1: Plot of kernels: $|K_{\mathrm{r}}(f)|$ and $|K(f)|$

The rest of the proof consists of three steps:

1) Show the construction is meaningful with high probability by establishing the invertibility of certain linear system using matrix Bernstein inequality [32];
2) Show the random perturbation introduced to the dual polynomial by the random kernel is small on a set of finite grid points with high probability, implying the random dual polynomial satisfies the constraints in Proposition IV.1 on the grid; This step is proved using a modification of the idea in [33].
3) Extend the result to $[0, 1]$ using Bernstein's polynomial inequality [34].

V. Numerical Experiments

We conducted a series of numerical experiments to test the performance of (7) under various parameter settings (see Table I). We use $J = \{0, \ldots, n-1\}$ for all numerical experiments.

We compared the performance of two algorithms: the semidefinite program (9) solved by the SDPT3 solver [35] and the basis pursuit (22) obtained through discretization, which was solved using CVX [36] coupled with SDPT3. All parameters of the SDPT3 solver were set to default values and CVX precision was set to 'high'. In the following, we use SDP and BP to label the semidefinite program algorithm and the basis pursuit algorithm, respectively. For the BP, we used three levels of discretization, which are 4, 16, and 64 multiples of the signal dimension.

To generate our instances of form (3), we sampled $s = \rho_s n$ normalized frequencies from $[0, 1]$, either *randomly*, or *equispaced*, with an additional constraint on the minimal separation Δ_f. The signal coefficient magnitudes $|c_1|, \cdots, |c_s|$ are either *unit*, i.e., equal to 1, or *fading*, i.e., equal to $.5+w^2$ with w a zero mean unit variance Gaussian random variable. The signs $\{e^{i\phi_k}, k = 1, \cdots, s\}$ follow either Bernoulli ± 1 distribution, labeled as *real*, or uniform distribution on the complex unit circle, labeled as *complex*. A length n signal

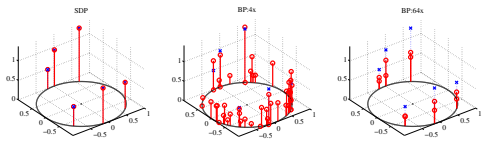


Fig. 2: Frequency Estimation: Blue represents the true frequencies, while red represents the estimated ones.

TABLE I: Parameter configurations

n	64, 128, 256
ρ_s	1/16, 1/32, 1/64
ρ_m/ρ_s	5, 10, 20
$\lvert c_k \rvert$	unit, fading
frequency	random, equispaced
sign	real, complex

was then formed according to model (3). As a final step, we uniformly sample $\rho_m n$ entries of the resulting signal.

We tested the algorithms on four sets of experiments. In the first experiment, by running the algorithms on a randomly generated instance with $n = 256, s = 6$ and 40 uniform samples, we compare SDP and BP's ability of frequency estimation and visually illustrate the the effect of discretization. We see from Figure 2 that SDP recovery followed by matrix pencil approach to retrieve the frequencies gives the most accurate result. We also observe that increasing the level of discretization can increase BP's accuracy in locating the frequencies.

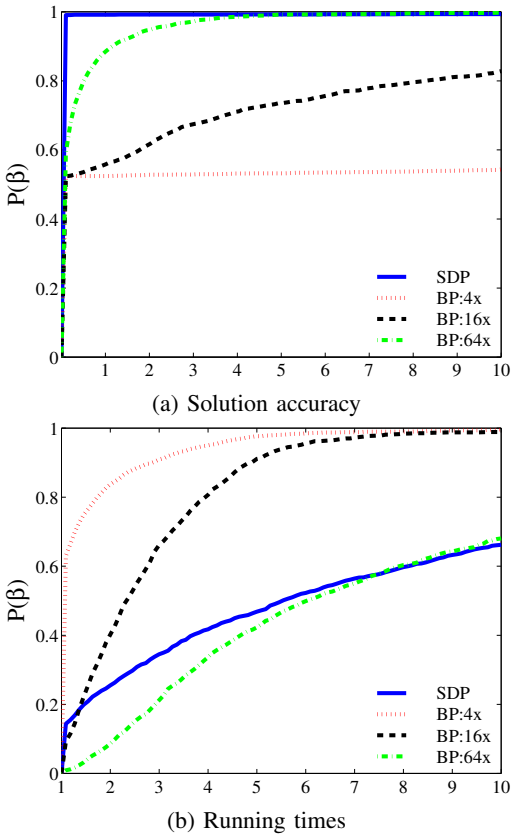
In the second set of experiments, we compare the performance of SDP and BP with three levels of discretization in terms of solution accuracy and running time. The parameter configurations are summarized in Table I. Each configuration was repeated 10 times, resulting a total of 2160 valid experiments.

We use the performance profile as a convenient way to compare the performance of different algorithms. The performance profile proposed in [37] visually presents the performance of a set of algorithms under a variety of experimental conditions. More specifically, let $\mathcal{P}$ be the set of experiments and $\mathcal{M}_a(p)$ specify the performance of algorithm a on experiment p for some metric $\mathcal{M}$ (the smaller the better), e.g., running time and solution accurary. Then the performance profile $\mathcal{P}_a(\beta)$ is defined as

$$\mathcal{P}_a(\beta) = \frac{\#\{p \in \mathcal{P} : \mathcal{M}_a(p) \leq \beta \min_a \mathcal{M}_a(p)}{\#(\mathcal{P})}, \beta \geq 1. \tag{27}$$

Roughly speaking, $\mathcal{P}_a(\beta)$ is the fraction of experiments such that the performance of algorithm a is within a factor β of that of the best performed one.

We show the performance profiles for numerical accuracy and running times in Figure 3a and 3b, respectively. We can see that SDP outperforms BP for all tested discretization



(a) Solution accuracy

(b) Running times

Fig. 3: Performance profiles

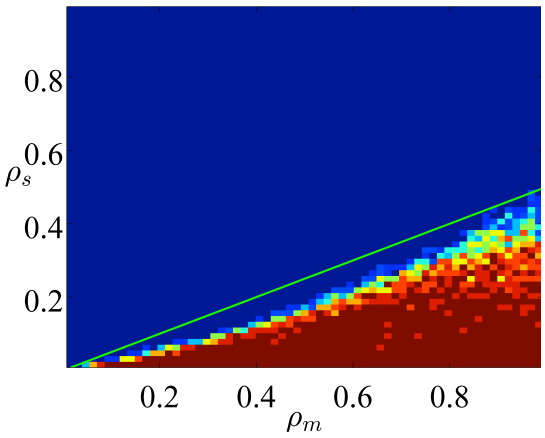
levels in terms of numerical accuracy. But SDP is also the slowest.

To give the reader a better idea of the numerical accuracy and the running time, we present their medians for the four algorithms in Table II. As one would expect, the running time increases as the discretization level increases. We also observe that SDP is very accurate, with an median error at the order of 10^{-9}. Increasing the level of discretization can increase the accuracy of BP. However, with discretization level $N = 64n$, we get a median accuracy at the order of 10^{-6}, but the median running time already exceeds that of SDP.

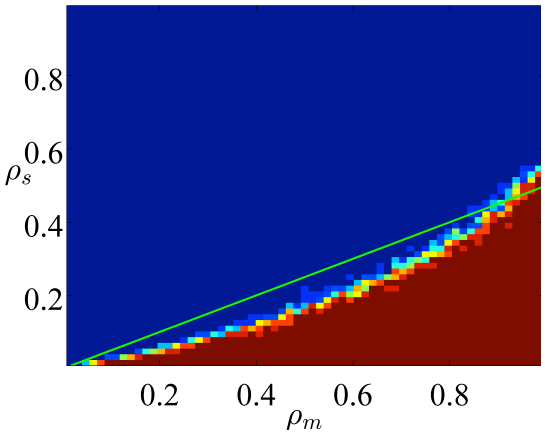
TABLE II: Medians for solution accuracy and running time

	SDP	BP: 4x	BP: 16x	BP: 64x
Running Time (s)	42.48	11.40	21.48	54.87
Solution Accuracy	9.50e-10	5.44e-04	2.84e-05	2.46e-06

In the third set of experiments, we compiled two phase transition plots. To prepare the Figure 4a, we pick $n = 128$ and vary $\rho_s = \frac{2}{n} : \frac{2}{n} : \frac{100}{n}$ and $\rho_m = \frac{2}{n} : \frac{2}{n} : \frac{126}{n}$. For each fixed (ρ_m, ρ_s), we randomly generate $s = n\rho_s$ frequencies while maintaining a frequency separation $\Delta_f \geq \frac{1}{n}$. The coefficients are generated with random magnitudes and random phases, and the entries are observed uniform randomly. We then run the SDPT3-SDP algorithm to recover the missing entries. The recovery is considered successful if the relative error $\|\hat{x} - x^\star\|_2 / \|x^\star\|_2 \leq 10^{-6}$. This process

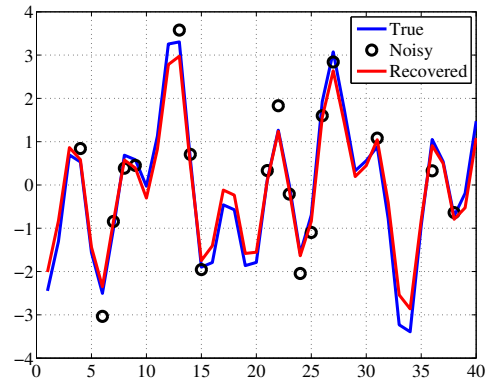


(a) Phase transition with $\Delta_f \geq \frac{1}{n}$.

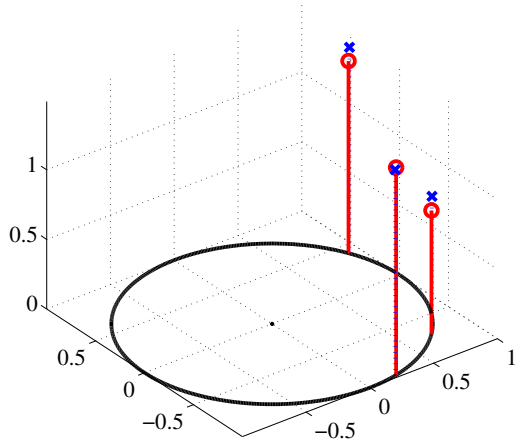


(b) Phase transition with $\Delta_f \geq \frac{1.5}{n}$.

Fig. 4: Phase transition: The plots are on a color scale where red represents success while blue represents failure.



(a) Real part of true, noisy, and recovered signals



(b) True frequencies (blue) and recovered frequencies (red)

Fig. 5: Noisy frequency recovery

was repeated 10 times and the rate of success was recorded. Figure 4a shows the phase transition results. The x-axis indicates the fraction of observed entries ρ_m, while the y-axis is $\rho_s = \frac{s}{n}$. The color represents the rate of success with red corresponding to perfect recovery and blue corresponding to complete failure.

We also plot the line $\rho_s = \rho_m/2$. Since a signal of s frequencies has $2s$ degrees of freedom, including s frequency locations and s amplitudes, this line serves as the boundary above which any algorithm should have a chance to fail. In particular, Prony's method requires $2s$ consecutive samples in order to recover the frequencies and the magnitudes.

From Figure 4a, we see that there is a transition from perfect recovery to complete failure. However, the transition boundary is not very sharp. In particular, we notice failures below the boundary of the transition where complete success should happen. Examination of the failures show that they correspond to instances with minimal frequency separations marginally exceeding $\frac{1}{n}$. We expect to get cleaner phase transitions if the frequency separation is increased.

To prepare Figure 4b, we repeated the same process in preparing Figure 4a except that the frequency separation was increased from $\frac{1}{n}$ to $\frac{1.5}{n}$. In addition, to respect the minimal separation, we reduced the range of possible sparsity levels to $\{2, 4, \ldots, 70\}$. We now see a much sharper phase transition. The boundary is actually very close to the $\rho_s = \rho_m/2$ line. When ρ_m is close to 1, we even observe successful recovery above the line.

In the last set of experiment, we use a simple example to illustrate the noise robustness of the proposed method. The signal was generated with $n = 40$, $s = 3$, random frequencies, fading amplitudes, and random phases. A total number of 18 uniform samples indexed by T were taken. The noisy observations y was generated by adding complex noise w with bounded ℓ_2 norm $\varepsilon = 2$ to $x_T^\star$. We denoised and recovered the signal by solving the following optimization:

$$\underset{x}{\text{minimize}} \ \|x\|_{\mathcal{A}} \text{ subject to } \|y - x_T\|_2 \leq \varepsilon, \qquad (28)$$

which clearly is equivalent to a semidefinite program. Matrix pencil approach was then applied to the recovered x to retrieve the frequencies. Figure 5 illustrates (28)'s ability in approximate frequency recovery in presence of noise.

VI. Conclusion and Future Work

By leveraging the framework of atomic norm minimization, we were able to resolve the basis mismatch problem

in compressed sensing of line spectra. For signals with well-separated frequencies, we show the number of samples needed is roughly propositional to the number of frequencies, up to polylogarithmic factors. This recovery is possible even though our continuous dictionary is not incoherent at all and does not satisfy any sort of restricted isometry conditions.

There are several interesting future directions to be explored to further expand the scope of this work. First, it would be useful to understand what happens in the presence of noise. We cannot expect exact support recovery in this case, as our dictionary is continuous and any noise will make the exact frequencies un-identifiable. In a similar vein, techniques like those used in [21] that still rely on discretization are not applicable for our current setting. However, since our numerical method is rather stable, we are encouraged that a theoretical stability result is possible.

Second, we are interested in exploring the class of signals that are semidefinite characterizable in hopes of understanding which signals can be exactly recovered. Models involving image manifolds may fall into this category [38]. Fully exploring the space of signals that can be acquired with just a few specially coded samples provides a fruitful and exciting program of future work.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Thy.*, vol. 52, no. 2, pp. 489–509, 2006.

[2] E. J. Candès and M. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Thy.*, vol. 52, no. 4, pp. 1289–1306.

[4] R. Baraniuk, "Compressive Sensing [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121.

[5] J. A. Tropp, J. N. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient Sampling of Sparse Bandlimited Signals," *IEEE Trans. Inf. Thy.*, vol. 56, no. 1, pp. 520–544, 2010.

[6] M. Mishali, Y. C. Eldar, and A. J. Elron, "Xampling: Signal Acquisition and Processing in Union of Subspaces," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4719–4734.

[7] M. Mishali and Y. C. Eldar, "From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 375–391.

[8] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, 2005.

[9] M. A. Herman and T. Strohmer, "High-Resolution Radar via Compressed Sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2275–2284.

[10] A. C. Fannjiang, T. Strohmer, and P. Yan, "Compressed Remote Sensing of Sparse Objects," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 595–618, Jan. 2010.

[11] R. Baraniuk and P. Steeghs, "Compressive Radar Imaging," in *Radar Conference, 2007 IEEE*, 2007, pp. 128–133.

[12] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076.

[13] M. Duarte and R. Baraniuk, "Spectral compressive sensing," *dsp.rice.edu*, Jul. 2011.

[14] H. Rauhut, "Random sampling of sparse trigonometric polynomials," *Applied and Computational Harmonic Analysis*, vol. 22, no. 1, pp. 16–42, Jan. 2007.

[15] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, 2011.

[16] M. A. Herman and T. Strohmer, "General Deviants: An Analysis of Perturbations in Compressed Sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349.

[17] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems," *arXiv.org*, vol. math.OC, Dec. 2010.

[18] E. J. Candès and J. Romberg, "Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions," *Foundations of Computational Mathematics*, vol. 6, no. 2, Apr. 2006.

[19] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic norm denoising with applications to line spectral estimation," *arXiv.org*, vol. cs.IT, Apr. 2012.

[20] B. Dumitrescu, *Positive Trigonometric Polynomials and Signal Processing Applications*. Springer Verlag, Feb. 2007.

[21] E. J. Candès and C. Fernandez-Granda, "Towards a Mathematical Theory of Super-Resolution," *arXiv.org*, vol. cs.IT, Mar. 2012.

[22] P. Stoica and R. L. Moses, *Spectral analysis of signals*, 1st ed. Upper Saddle River, New Jersey: Prentice Hall, 2005.

[23] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of Sparse Translation-Invariant Signals With Continuous Basis Pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, 2011.

[24] ——, "Neural spike identification with continuous basis pursuit," in *Computational and Systems Neuroscience (CoSyNe)*, Salt Lake City, Utah, Feb. 2011.

[25] C. E. Parrish and R. D. Nowak, "Improved Approach to LIDAR Airport Obstruction Surveying Using Full-Waveform Data," *Journal of Surveying Engineering*, vol. 135, no. 2, pp. 72–82, May 2009.

[26] B. G. R. de Prony, "Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures," *Journal de l'école Polytechnique*, vol. 1, no. 22, pp. 24–76.

[27] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse Sampling of Signal Innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, 2008.

[28] P. Stoica, "List of references on spectral line analysis," *Signal Processing*, vol. 31, no. 3, pp. 329–340, 1993.

[29] L. Carin, D. Liu, and B. Guo, "Coherence, Compressive Sensing, and Random Sensor Arrays," *IEEE Antennas and Propagation Magazine*, vol. 53, no. 4, pp. 28–39.

[30] E. Dowski, C. Whitmore, and S. Avery, "Estimation of randomly sampled sinusoids in additive noise," *Acoustics, Speech and IEEE Trans. Signal Process.*, vol. 36, no. 12, pp. 1906–1908, 1988.

[31] P. Stoica, J. Li, and H. He, "Spectral Analysis of Nonuniformly Sampled Data: A New Approach Versus the Periodogram," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 843–858.

[32] J. A. Tropp, "User-Friendly Tail Bounds for Sums of Random Matrices," *Foundations of Computational Mathematics*, Aug. 2011.

[33] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, Apr. 2007.

[34] A. Schaeffer, "Inequalities of A. Markoff and S. Bernstein for polynomials and related functions," *Bull Amer Math Soc*, 1941.

[35] K. C. Toh, M. Todd, and R. H. Tutuncu. SDPT3: a MATLAB software for semidefinite-quadratic-linear programmin. [Online]. Available: http://www.math.nus.edu.sg/~mattohkc/sdpt3.html

[36] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming (web page and software)," *Available at http://cvxr.com/cvx*.

[37] E. D. Dolan and J. J. Moré, "Benchmarking Optimization Software with Performance Profiles," *Mathematical Programming, Series A*, vol. 91, no. 2, pp. 201–203, Mar. 2002.

[38] M. Wakin, "A manifold lifting algorithm for multi-view compressive imaging," in *Picture Coding Symposium, 2009. PCS 2009*. IEEE, 2009, pp. 1–4.