Compressive Line Spectrum Estimation with Clustering and Interpolation

Dian Mo

Department of Electrical and Computer Engineering
University of Massachusetts
Amherst, MA, 01003
mo@umass.edu

Marco F. Duarte

Department of Electrical and Computer Engineering
University of Massachusetts
Amherst, MA, 01003
mduarte@ecs.umass.edu

Abstract—**We consider the standard line spectral estimation problem when the number of observed samples is significantly lower than that prescribed by the Nyquist rate. Two families of sparsity-based methods have recently been proposed for this problem. The first one uses an atomic norm minimization algorithm where the atoms correspond to complex exponentials of varying frequencies. The second one defines the sparse coefficient vectors for the signals of interest by designing parametric dictionaries that can be leveraged by sparse approximation algorithms involving clustering and interpolation. This paper compares the performance of these two algorithm families. Experiments show their advantages and disadvantages in terms of precision and complexity.**

Index Terms—**compressive sensing, line spectral estimation, parametric dictionary, K-median clustering**

I. INTRODUCTION

Atomic norm minimization has been recently proposed to apply sparsity concepts in line spectral estimation for fully sampled and subsampled signals [1, 2]. Simultaneously, the authors and other researchers have propose algorithms for compressive sensing of spectrally sparse signals that rely on sparsity dictionaries [3–7]. In particular, both atomic norm methods [1, 2] and certain greedy methods [7] have been shown to be successful if there is sufficient spacing between the unknown frequencies present in the signal. This paper evaluates the performance of these new sparsity-based algorithms for the line spectral estimation problem in comparison with atomic norm minimization methods.

Both types of methods exploit the sparsity structure of the signals of interest in line spectral estimation to estimate the underlying frequencies. By defining a proper penalty function to measure the sparsity of the signal, the estimation problem can be solved via an optimization method. Algorithms based on ℓ_1-norm minimization are widely used in parameter estimation when the signal has sparse coefficients in some basis suitable for estimation purposes [8]. In contrast, the atomic norm limits the search to the signals that can be expressed as superpositions of a small number of signal atoms [1, 2], and provides a direct estimate of the spectrally sparse signal. The unknown frequencies can then be obtained from the recovered signal using standard line spectrum estimation algorithms. One of the drawbacks of atomic norm minimization methods is the complexity of the semidefinite programming solver involved.

As an alternative, one can introduce parametric dictionaries (PDs), whose elements are the complex exponentials corresponding to frequencies sampled from a fixed frequency range; the signals of interest can then be expressed as a linear combination of a small number of dictionary elements. The sparse dictionary coefficient vector can be recovery by standard CS algorithms, and the support of the recovered sparse vector encodes the values of the frequencies present in the signal. These algorithms, however, are limited by the increasing coherence between dictionary elements as the number of sampled frequencies contained increases. Several recently proposed approaches aim to mitigate the impact of the coherence on the performance of signal recovery [3–5].

In previous work, we incorporated K-median clustering within a sparse approximation algorithm and applied it to compressive parameter estimation to ameliorate the coherence handicap while providing better measures and guarantees for the frequency estimation error [6, 7]. Furthermore, using more sophisticated models on top of sparsity, such as manifold models, we can refine the estimates of the frequency values so that they are not restricted to those captured by the dictionary elements. In this paper, we will show that our algorithms are applicable beyond the standard compressive sensing setup and can be extended easily to the line spectral estimation problem to which atomic norm minimization has been recently applied. We then perform an experimental comparison for the performance of both types of algorithms.

II. BACKGROUND

A. Signal Model in Line Spectral Estimation

Line spectral estimation refers to the problem of estimating the frequencies of a signal composed as a linear combination of a small number of complex sinusoids. To be precise, the spectra of the signal of interest x consists of K spikes located at unknown positions $f_1, f_2, \ldots f_K$ in the frequency range $\Theta = [0, 1)$. Thus, each time the samples of interest signal can be expressed as

$$x[n] = \sum_{k=1}^{K} c_k e^{j2\pi f_k n}, n = 0, 1, 2, \ldots, N-1. \quad (1)$$

Rather than collecting all N time samples at time set $D = \{0, 1, \ldots, N-1\}$, we sample the signal only at a certain

time subset $T = \{t_1, t_2, \ldots, t_M\} \subset D$. The signal can be subsampled in two ways: random subsampling, where each t_i is selected uniformly at random from the set D; and uniform subsampling, where $t_i = i/\kappa - 1$, and $\kappa = M/N$ is the subsampling ratio. In this paper, we will focus on random subsampling since the latter is a straightforward adaption of the former. Additionally, each coefficient c_k is a complex number with phases drawn independently from the uniform distribution $\mathcal{U}[0, 2\pi]$. The goal of line spectral estimation is to estimate those unknown frequencies $f_1, f_2, \ldots, f_K$ from the subsampled observations $y = x_T$, when no noise is involved, or $y = x_T + w$ when some additive noise is present; in both cases, x_T is the vector that contains all samples of signal on the subset the T.

B. Atomic Norm Minimization

Recently, a new penalty known as the atomic norm has been introduced for purposes of convex programming-based frequency estimation [1, 2]. More specifically, a set of atoms $a(f, \phi) \in \mathbb{C}^N$ are defined as

$$a(f, \phi)[n] = e^{j(2\pi f n + \phi)}, n \in D, \tag{2}$$

where $f \in \Theta$, and $\phi \in [0, 2\pi)$ [9]. Then the signal of interest has the form

$$x[n] = \sum_{k=1}^{K} c_k e^{j 2\pi f_k n} = \sum_{k=1}^{K} |c_k| a(f_k, \phi_k), \tag{3}$$

where ϕ_k is the phase for the coefficient c_k. The set of atoms $\mathcal{A} = \{a(f, \phi) : f \in \Theta, \phi \in [0, 2\pi)\}$ provides an infinite-size dictionary for the signals of interest x, which can be expressed as a linear combination of a finite number of elements of $\mathcal{A}$, in a manner similar to the finite-dimensional sparsity models.

The atomic norm is defined in terms of the most compact expansion of the signal x in terms of the elements of $\mathcal{A}$:

$$\|x\|_{\mathcal{A}} = \inf \left\{ \sum_{k=1}^{K} |c_k| : x = \sum_{k=1}^{K} |c_k| a(f_k, \phi_k) \right\}. \tag{4}$$

It is easy to see that when the set of elements in $\mathcal{A}$ are the canonical vectors, the atomic norm is exactly the standard ℓ_1 norm used in sparse signal recovery. For the set of atoms $\mathcal{A}$ detailed above and noiseless observations, the atomic norm allows us to write the line spectral estimation as the optimization problem

$$\begin{aligned} \min_{x} \quad & \|x\|_{\mathcal{A}} \\ \text{s.t} \quad & x_T = y, \end{aligned} \tag{5}$$

where the atoms involved in $\|x\|_{\mathcal{A}}$ identify the component frequencies. The atomic norm has an equivalent semidefinite programming (SDP) formulation as

$$\begin{aligned} \|x\|_{\mathcal{A}} = \min_{x, u, t} \quad & \frac{1}{2N} \text{trace} \left(\text{Toep}(u) \right) + \frac{1}{2} t \\ \text{s.t} \quad & \begin{bmatrix} \text{Toep}(u) & x \\ x^* & t \end{bmatrix} \succeq 0 \end{aligned} \tag{6}$$

where $\text{Toep}(u)$ is a Hermitian Toeplitz matrix defined by the vector u. Although the problem (5) is a semidefinite program

for which many solvers exist, it will be more convenient to solve its dual problem, which has the form

$$\begin{aligned} \max_{q} \quad & \langle q_T, y \rangle_{\mathbb{R}} \\ \text{s.t} \quad & \|q\|_{\mathcal{A}}^* \leq 1 \\ & q_{T^C} = 0, \end{aligned} \tag{7}$$

where $\langle x, y \rangle_{\mathbb{R}}$ denotes the real part of inner product between x and y, and $\|q\|_{\mathcal{A}^*}$ is dual atomic norm given by

$$\|q\|_{\mathcal{A}}^* = \sup_{\|x\|_{\mathcal{A}} \leq 1} \langle q, x \rangle_{\mathbb{R}} = \sup_{f \in [0, 1]} |\langle q, a(f, 0) \rangle|. \tag{8}$$

Also, the dual problem (7) has an equivalent SDP form [10] as

$$\begin{aligned} \hat{q} = \max_{q} \quad & \langle q_T, y \rangle_{\mathbb{R}} \\ \text{s.t} \quad & \begin{bmatrix} H & q \\ q^* & 1 \end{bmatrix} \succeq 0 \\ & q_{T^C} = 0, \end{aligned} \tag{9}$$

where H is an Hermitian matrix defined by

$$\sum_{i=1}^{N-j} H_{i, i+j} = \begin{cases} 1, & j = 0 \\ 0, & j = 1, 2, \ldots, N-1 \end{cases}. \tag{10}$$

The dual solution $\hat{q}$ to the dual problem provides a way to determine the unknown frequencies due to the fact that the polynomial $Q(f) := \langle \hat{q}, a(f, 0) \rangle$ achieves a unit absolute value if and only if $f \in \{f_1, f_2, \ldots, f_k\}$.

The atomic norm minimization method can estimate any arbitrary unknown frequencies if they are well separated [11]. However, the precision and complexity of estimation will be highly depend on the choice of semidefinite program solver used. In practice, the amount of computation required for accurate estimation can be large.

C. Compressive Parameter Estimation via Clustering

An alternative approach to compressive parameter estimation proceeds by designing parametric dictionaries (PDs). Let $\Omega = \{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_l\} \subset \Theta$ denote a set of parameters (e.g., frequencies) sampled with a sampling step Δ. The PD for parameter estimation collects all parametric signals corresponding to the sampled parameters $\Psi = [\psi(\tilde{f}_1), \psi(\tilde{f}_2), \ldots, \psi(\tilde{f}_l)]$, where $\psi(\cdot)$ is the signal function, e.g., $\psi(\tilde{f}) = a(\tilde{f}, 0)$. Thus, the signal of interest can be expressed as a linear combination of the dictionary elements $x = \Psi c$, $c \in \mathbb{C}^l$, when all the unknown parameters are contained in the sampling set Ω. Therefore, finding the unknown parameters from CS measurements $y = \Phi x$, where Φ is the measurement matrix, reduces to finding all dictionaries appearing in the signal representation or, equivalently, finding the nonzero entries of the coefficient vector c.

Following the convention of greedy algorithm in CS, a proxy of coefficient vectors is obtained via the correlation of the observations with PD, i.e., $v = \Psi^* \Phi^* y = (\Phi \Psi)^* (\Phi \Psi) c$. The entries of v achieve a local maximum when they corresponds to the nonzero entries of c. Thus, the goal of parameter estimation reduces to locate the entries of v that are local

maxima, which is usually achieved via a hard thresholding operator that returns the sparse approximation with smallest ℓ_2 distance to the proxy.

Such PD-based parameter estimation can be perfect only if the sampling set Ω contains all unknown frequencies. Nonetheless, one can attempt to reduce the sparse approximation error by reducing the parameter step size Δ. However, highly dense PD sampling increases the similarity between adjacent PD elements, resulting also in a highly coherent PD [12]. The coherence of a dictionary is measured by the maximum inner product of normalized dictionaries, i.e.,

$$\mu(\Omega) = \max_{0 \leq i \neq j \leq l} \frac{\left|\psi(\tilde{f}_i)^* \psi(\tilde{f}_j)\right|}{\|\psi(\tilde{f}_i)\|_2 \|\psi(\tilde{f}_2)\|_2}. \tag{11}$$

The high coherence makes it difficult for hard thresholding to identify the local maxima since it will unavoidably select entries around the largest local maxima. Furthermore, an additional issue in PD-based frequency estimation is that the recovery algorithms in CS can only guarantee small ℓ_2-norm error in the coefficient vector estimate. Such guarantees result from the core hard thresholding operator. However, these guarantees are meaningless for parameter estimation because the ℓ_2 norm cannot provide a precise measure of the parameter estimation error.

We recently introduced an approach to compressive parameter estimation that relies on the earth mover's distance (EMD) to measure the error in the coefficient vector (i.e., the distance between the estimated and the true coefficient vectors) in terms of similarity between the values and locations of their nonzero entries. The EMD between two vectors is obtained by optimizing the flow of mass among the entries of one vector to make it match with the other [13, 14]. Let p and q be two K-sparse coefficient vectors with nonzero entries, and let $I, J \subset 0, 1, \ldots, l - 1$ be their respective supports. The EMD between p and q will be obtained by the following linear program:

$$\text{EMD}(p, q) = \min_{g_{ij}} \sum_{i,j} g_{ij} d_{ij}$$
$$\text{s.t.} \sum_{j \in J} g_{ij} = p_i$$
$$\sum_{i \in I} g_{ij} = q_j$$
$$g_{ij} \geq 0, i \in I, j \in J, \tag{12}$$

where d_{ij} is the distance between the entry i and entry j [15]. When the parameter samples are sorted, d_{ij} is proportional to the difference between the sampled parameters $\tilde{f}_i$ and $\tilde{f}_j$, and thus the EMD between coefficient vectors is proportional to the corresponding parameter error [7]. This inspires us to find the EMD-optimal sparse approximation, which returns the closest sparse approximation to the proxy in terms of EMD, in order to provide EMD-based guarantees in compressive parameter estimation.

It can be shown that K-median clustering finds such sparse approximation: the sparse vector with nonzero entry corresponding to the centroids obtained from K-median clustering performed on the proxy vector entries has the smallest EMD to the proxy. By replacing the hard thresholding in standard CS recovery systems with K-median clustering, one can easily formulate a EMD-optimal compressive parameter estimation algorithm.

In addition, when the sampling step Δ is sufficiently small, it is possible to interpolate the functional map $\psi(\cdot)$ between the samples contained in the dictionary [5–7].

III. CLUSTERING METHOD FOR FREQUENCY ESTIMATION

There is a similarity in atomic norm minimization and PD-based parameter estimation in that both methods exploit the concept of sparsity for signals of interest in terms of the set of atoms or the PD, respectively. However, in atomic norm minimization, the sparsity is enforced by minimizing the atomic norm, while in PD-based parameter estimation, a sparse approximation algorithm is used to recover the sparse coefficient vector.

The atomic norm minimization is highly depend on the equivalent semidefinite programming (SDP) problem, which enables the optimization to be solved precisely. However, it is not easy for such method to be extended to the case that the observations cannot be obtained by subsampling. The equivalent SDP form of dual problem in (9) will not exist when the observations are obtained in the general case $y = \Phi x$.

On the contrary, it is quite straightforward to extend the clustering method to frequency estimation from few samples. The linear process of subsampling a signal on the time set T has the matrix representation $\Phi \in \mathbb{R}^{M \times N}$ containing only the rows of the identity matrix corresponding to T, i.e.,

$$\Phi_{ij} = \begin{cases} 1, & j = t_i \\ 0, & \text{otherwise} \end{cases} \quad i = 1, 2, \ldots, M, t_i \in T. \tag{13}$$

Using this notation, the observations from the subsampling process can be expressed in the standard CS form $y = \Phi x$. It is possible to show that such random subsampling measurements are suitable for recovery of spectral sparse signals [16].

A simple modified clustering subspace pursuit (CSP), which incorporates the clustering method into the subspace pursuit (SP) sparse recovery algorithm [17], is shown in Algorithm 1. CSP repeatedly uses the K-median clustering $S = \mathbb{C}(v, \Omega, K)$ on a set of points with weights corresponding to the entry values of the proxy v, and locations given by the sampled frequencies Ω. The centroids S resulting from the clustering process indicate the locations of the nonzero entries of the sparse vector approximation to v, which will be used to refine the previous estimates after another clustering. In the CSP algorithm, the proxy is preprocessed by a thresholding step, cf. step 5, to increase the rate of decay of the autocorrelation function and to improve the performance of clustering in the subsequent step. Furthermore, we refine the estimates of the component frequencies by using polar interpolation between elements of the PD, which allows us to obtain

Algorithm 1 *Clustering Subspace Pursuit (CSP)*

Input: measurement vector y, subsampling set T, frequency number K, set of sampled frequencies Ω, signal function $\psi(\cdot)$, threshold t

Output: estimated signal $\widehat{x}$, estimated frequencies $\widehat{f}$

1: **Initialize:** $\widehat{x} = 0$, generate PD $\Psi = \psi(\Omega)$, generate sensing matrix Φ from T.
2: **repeat**
3:      $y_r = y - \Phi\widehat{x}$ {Compute residual}
4:      $v = (\Phi\Psi)^* y_r$ {Obtain proxy from residual}
5:      $v(|v| \leq t) = 0$ {Threshold proxy}
6:      $\widehat{f} = \widehat{f} \cup \mathbb{C}(v, \Omega, K)$ {Augment estimates}
7:      $\Psi_S = \psi\left(\widehat{f}\right)$ {Obtain PD subset for estiamtes}
8:      $c = (\Phi\Psi_S)^+ y$ {Obtain proxy of estimates}
9:      $\widehat{f} = \mathbb{C}(c, \widehat{f}, K)$ {Refine estimates}
10:     $(\widehat{x}, \widehat{f}) = \mathbb{P}(y, \widehat{f}, \psi(\cdot))$ {Interpolate estimates}
11: **until** a convergence criterion is met

---

frequency estimates of arbitrary resolution in order to bypass the restriction on the parameter values encoded by the PD, as will be described in the next section.

IV. POLAR INTERPOLATION

A recently proposed alternative to improve the estimation performance of PDs when the unknown frequencies are not all contained in the sampling set of frequency range is to use interpolation in the PD elements. The motivation behind such approaches is that the low-dimensional manifold that expresses the relationship between frequencies and signals can be approximated well in small neighborhoods by a closed-form expression that integrates as much knowledge of the manifold characteristics as possible while remaining computationally feasible. Therefore, the observation for a frequency value outside of the sampling set Ω can be accurately estimated by its surrounding sampled frequencies using interpolation. Although Taylor series interpolation has been used in this case, certain applications like frequency estimation feature parametric invariance of the norm and distances between signals, and are better suited to a polar interpolation scheme [18, 19].

For frequency estimation, the signals corresponding to normalized complex exponentials at various frequency values share the same unit magnitude and can be characterized by a manifold contained in the surface of a high-dimensional hyper-sphere in $\mathbb{C}^N$. A small segment of this manifold can therefore be approximated by an arc of a circle which is uniquely determined by a triplet of PD elements corresponding to the sampled frequencies to be contained in the segment. It is possible to find a basis for the span of the triplet of elements that provides a trigonometric map from the angle between the middle element and the observed signal to the differential of the frequency values for these two signals.

More specifically, assume that the frequency range is sampled with a step size Δ. The unknown frequency f_k is linked to the closest value $\tilde{f}_l$ within the sampled set Ω; therefore, the

observed signal $\psi(f_k)$ lies on the segment of the manifold $\{\psi(f) | \tilde{f}_l - \frac{\Delta}{2} \leq f \leq f_l + \frac{\Delta}{2}\}$. This segment is to be approximated by the unique circular arc that contains the triplet of dictionary elements $\{\psi(f_l - \frac{\Delta}{2}), \psi(f_l), \psi(f_l + \frac{\Delta}{2})\}$. The polar approximation is obtained as

$$
\psi(f_k) \approx d(\tilde{f}_l) + r\cos\left(\frac{2\left(f_k - \tilde{f}_l\right)}{\Delta}\sigma\right) u(\tilde{f}_l)
$$
$$
+ r\sin\left(\frac{2\left(f_k - \tilde{f}_l\right)}{\Delta}\sigma\right) v(\tilde{f}_l), \qquad (14)
$$

where $d(\tilde{f}_l)$, $u(\tilde{f}_l)$, and $v(\tilde{f}_l)$ are a basis for the circle corresponding to its center and trigonometric coordinates and the constants r and σ represent the radius and the half-angle of the relevant circular arc. The approximation basis elements can be computed in closed form using the formula

$$
\left[d(\tilde{f}_l), u(\tilde{f}_l), v(\tilde{f}_l)\right] = \left[\psi\left(\tilde{f}_l - \frac{\Delta}{2}\right), \psi\left(\tilde{f}_l\right), \psi\left(\tilde{f}_l + \frac{\Delta}{2}\right)\right]
$$
$$
\times \begin{bmatrix} 1 & 1 & 1 \\ r\cos(\sigma) & r & r\cos(\sigma) \\ -r\sin(\sigma) & 0 & r\sin(\sigma) \end{bmatrix}^{-1},
$$
$$(15)$$

which intuitively provides the mapping between the angles $\{-\sigma, 0, \sigma\}$ and the PD element triplet. When multiple frequencies are observed simultaneously, we collect the estimation basis elements $d(\tilde{f}_l)$, $u(\tilde{f}_l)$ and $v(\tilde{f}_l)$ into the matrices D, U, and V so that the observed signal x can be expressed as as

$$
x = \sum_{k=1}^{K} c_k \psi(\theta_k) \approx Dc + U\alpha + V\beta, \qquad (16)
$$

where c, α and β collect the trigonometric coefficients from the individual approximations (14). The resulting coefficients to the equation (16) α and β yield an estimate of the frequencies via the bijective relation

$$
f_k = \tilde{f}_l + \frac{\Delta}{2\sigma}\arctan\left(\frac{\beta_k}{\alpha_k}\right). \qquad (17)
$$

V. NUMERICAL EXPERIMENTS

To compare the performance between the two types of line spectral estimation methods, we conduct a series of numerical experiments. In all experiments, the signals are sampled at integer times $D = 0, 1, \ldots, N-1$. The subsampling set T contains M elements randomly selected from D. All $K = 4$ unknown frequencies are chosen uniformly at random from the frequency range Θ, while the sampled frequencies used in the design of Φ are uniformly sampled from Θ with sampling step $\Delta = 0.01$Hz.

Our experiments compare the frequency estimation performance of clustering subspace pursuit (CSP) and atomic norm minimization solved by semidefinite programming (SDP). Additionally, our experiments consider band-exclusion subspace pursuit (BSP), which uses band exclusion to prevent the highly
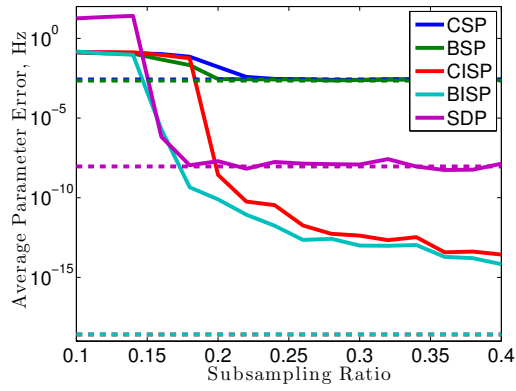
Fig. 1. Average frequency estimation error as a function of subsampling ratio when noiseless samples are observed. Dashed lines represents the average error when the subsampling ratio $\kappa = 1$
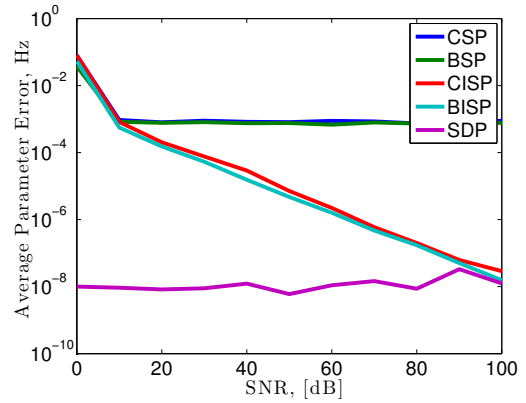

Fig. 2. Average frequency estimation error as a function of the signal to noise ratio when the subsampling ratio $\kappa = 0.4$.

Fig. 3. Average estimation time as a function of signal length when noiseless samples are observed and subsampling ratio $\kappa = 0.4$

coherent dictionaries appearing simultaneously in signal representation to address the coherence issue in compressive parameter estimation [5, 20]. The two subspace pursuit algorithms are also implemented with polar interpolation to improve the estimation performance, cf. step 10 of Algorithm 1, and are denoted CISP and BISP, respectively. Both CSP and CISP use a threshold level $t = 0.15$, while the maximum allowed coherence level for BSP and BISP is set to $\nu = 0.2$. The SDP problem is solved via the CVX toolbox with the solver SDPT3 [21, 22].

The first experiment considers the frequency estimation performance as a function of the subsampling ratio $\kappa = M/N$. As a reference, dotted lines mark the performance of the different algorithms when $\kappa = 1$, i.e., in the full sampling case. Figure 1 shows the average frequency estimation error over 100 randomly drawn spectrally sparse signals for each subsampling ratio. The parameter error refers to the minimum match cost between the true and estimated frequencies as obtained by the Hungarian matching algorithm [23]. When the numbers of true and estimated frequencies are not matched, the smaller set is supplemented with an additional cost equal to the frequency range N. When the subsampling ratio is small ($\kappa \leq 0.2$), all methods fail to accurately estimate the frequencies. The poor performance of SDP in low subsampling ratios is due to the fact that SDP fails to estimate all the frequencies present. As the subsampling ratio increases, the performance of all methods improves rapidly. Although both subspace pursuit algorithms without polar interpolation cannot exceed the limit of sampling step Δ, their interpolated versions greatly decrease the average error, even improving on the performance of SDP. Note, however, that it may still be possible for SDP to have improved performance by selecting more sophisticated solvers; nonetheless, this is likely to involve an additional computational cost penalty.

The second experiment considers the frequency estimation performance as a function of the signal to noise ratio (SNR), where the observation noise is additive white Gaussian. Figure 2 shows the average frequency estimation error over 100
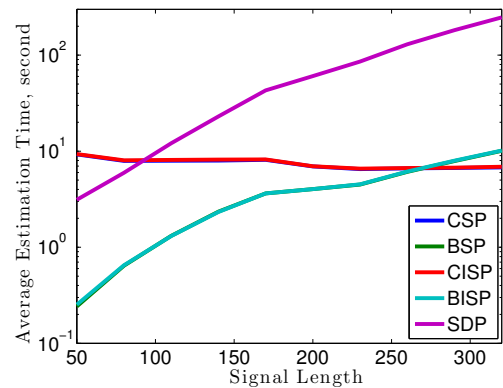
randomly drawn spectrally sparse signals per SNR, where the subsampling ratio is set to $\kappa = 0.4$. Although the SDP approach shows better performance over different noise levels than all subspace pursuit methods, the advantage over the interpolation-aided subspace pursuit algorithms is reduced as the SNR improves.

The third experiment considers the estimation time as a function of signal length in noiseless case. Figure 3 shows the average estimation time over 100 randomly drawn spectrally sparse signals per signal length when the subsampling ratio is set to $\kappa = 0.4$. The time required for SDP approach increases rapidly as the signal length increases and will be unacceptable when the signal length is beyond 350 samples. On the contrary, our subspace pursuit method implemented with K-median clustering shows its advantage in computational cost stability.

VI. CONCLUSION

Both atomic norm minimization and PD-based compressive parameter estimation exploit the sparsity of signals of interest in a chosen basis or atom set to estimate the underlying frequencies. Atomic norm minimization uses the atomic norm to form a semidefinite program that solves line spectrum estima-

tion. In contrast, the PD-based parameter estimation methods convert the spectrum estimation into a sparse coefficient vector recovery problem. Although atomic norm minimization is able to obtain estimates precisely and stably with or without noise, it is not easy to apply the method in the case that signal observations are not obtained via (sub)sampling. In contrast, the proposed PD-based parameter estimation algorithms can easily perform frequency estimation from subsampled observations. Furthermore, the PD-based parameter estimation algorithms can be more precise and less time-consuming when aided by polar interpolation.

Due to the flexibility of PD-based parameter estimation, we believe that the proposed methods can also be used in the super-resolution problem, which is the time-frequency dual problem of line spectrum estimation. In super-resolution, low-frequency coefficients of a signal that consists of a super-positions of point sources are observed, and the goal is to resolve the signals at high frequencies [24]. Our future work will extend the PD-based parameter estimation to the super-resolution problem.

REFERENCES

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

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

[3] A. C. Fannjiang and W. Liao, "Coherence pattern-guided compressive sensing with unresolved grids," *SIAM J. Imageing Sci.*, vol. 5, no. 1, pp. 179 – 202, Feb. 2012.

[4] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Appl. and Comput. Harmonic Anal.*, vol. 35, no. 1, pp. 111 – 129, Jan. 2013.

[5] K. Fyhn, H. Dadkhahi, and M. F. Duarte, *Spectral Compressive Sensing with Polar Interpolation*. IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP), May 2013, pp. 6225 – 6229.

[6] D. Mo and M. F. Duarte, "Compressive parameter estimation with earth mover's distance via k-median clustering," in *Proc. SPIE Wavelets and Sparsity XV*, vol. 8858, San Diego, CA, Aug. 2013.

[7] ——, "Performance of compressive parameter estimation with earth mover's distance via k-median clustering," Dec. 2014, submitted. Available at http://arxiv.org/pdf/1412.6724.

[8] S. Shaobing, C. D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Computing*, vol. 43, no. 1, pp. 129 –159, Jan. 2001.

[9] V. Chandrasekaranm, B. Rechtw, P. A. Parrilom, and A. S. Willskym, "The convex geometry of linear inverse problems," *Found. Comput. Math.*, vol. 12, no. 6, pp. 805 – 849, Dec. 2012.

[10] B. A. Dumitrescu, *Positve Trigonometric Polynomials and Signal Processing Applicaitons*. Springer Netherlands, 2007.

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

[12] D. L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via $\ell 1$ minimization," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, no. 5, pp. 2197 – 2202, Mar. 2003.

[13] P. Indyk and E. Price, "k-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance," in *ACM Symp. Theory of Computing (STOC)*, San Diego, CA, Jun 2011, pp. 627 – 636.

[14] R. Gupta, P. Indyk, and E. Price, "Sparse recovery for earth mover distance," in *48th Ann. Allerton Conf. Commun., Control, and Computing*, Monticello, IL, Sep. 2010, pp. 1742 – 1744.

[15] O. Pele and M. Werman, "Fast and robust earth mover's distance," in *Int. Conf. Comput. Vision (ICCV)*, Kyoto, Japan, Sep. 2009, pp. 460 – 467.

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

[17] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, 2009.

[18] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Proc.*, vol. 59, no. 10, pp. 4735 – 4744, Oct. 2011.

[19] K. Fyhn, M. F. Duarte, and S. H. Jensen, "Compressive parameter estimation for sparse translation-invariant signals using polar interpolation," *IEEE Trans. Signal Proc.*, vol. 64, no. 4, pp. 870–881, Feb. 2015.

[20] ——, "Compressive time delay estimation using interpolation," in *IEEE Global Conf. on Signal and Inf. Proc. (GlobalSIP)*, Austin, TX, Dec. 2013, pp. 624 – 624.

[21] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[22] R. H. Tütüncü, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming*, vol. 95, no. 2, pp. 189 – 217, Oct. 2002.

[23] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Society for Industrial and Applied Math.*, vol. 5, no. 1, pp. 32 – 38, Mar. 1957.

[24] E. Candès and C. Fernandez-Granda, "Super-resolution from noisy data," *J. Fourier Analysis and Applicat.*, vol. 19, no. 6, pp. 1229 – 1254, Dec. 2013.