

OFF-THE-GRID SPECTRAL COMPRESSED SENSING WITH PRIOR INFORMATION

Kumar Vijay Mishra, Myung Cho, Anton Kruger and Weiyu Xu

Department of Electrical and Computer Engineering
The University of Iowa, Iowa City IA 52242

ABSTRACT

Recent research in off-the-grid compressed sensing (CS) has demonstrated that, under certain conditions, one can successfully recover a spectrally sparse signal from a few time-domain samples even though the dictionary is continuous. In this paper, we extend off-the-grid CS to applications where some prior information about spectrally sparse signal is known. We specifically consider cases where a few contributing frequencies or poles, but not their amplitudes or phases, are known *a priori*. Our results show that equipping off-the-grid CS with the known-poles algorithm can increase the probability of recovering all the frequency components.

Index Terms— compressed sensing, spectral estimation, basis mismatch, matrix completion, known poles

1. INTRODUCTION

Compressed sensing (CS) is a new sampling paradigm which postulates uniting two critical steps involved in processing a signal: digital data acquisition and its compression [1] [2]. To recover a signal from fewer random measurements, CS algorithms harness the inherent sparsity of the signal under some appropriate basis or dictionary.

For example, consider a frequency-sparse signal $x[l]$ represented as a sum of s complex exponentials,

$$x[l] = \sum_{j=1}^s c_j e^{i2\pi f_j l} = \sum_{j=1}^s |c_j| a(f_j, \phi_j)[l], \quad l \in \mathcal{N} \quad (1.1)$$

where $c_j = |c_j| e^{i\phi_j}$ ($i = \sqrt{-1}$) represents the complex coefficient of the frequency $f_j \in [0, 1]$, with amplitude $|c_j| > 0$, phase $\phi_j \in [0, 2\pi)$, and frequency-atom $a(f_j, \phi_j)[l] = e^{i(2\pi f_j l + \phi_j)}$. We use the index set $\mathcal{N} = \{l \mid 0 \leq l \leq n-1\}$, where $|\mathcal{N}| = n, n \in \mathbb{N}$, to represent the time samples of the signal. It is customary to label only the frequency information - either the exponentials $e^{i2\pi f_j l}$ or just f_j - as *poles* [3] [4].

When f_j takes values only on a *discrete frequency grid*, the Discrete Fourier Transform (DFT) matrix can be used as an appropriate finite discrete dictionary for the sparse representation of $x[l]$. However, it is quite possible for the true frequencies to be anywhere in the *continuous domain* $[0, 1]$. Since the true continuous-domain frequencies may lie off the center of the DFT bins, the DFT representation in this case would destroy the sparsity of the signal and result in the so-called “basis mismatch” [5]. This can be mitigated to a certain extent by finer discretization of the DFT grid. But that could lead to higher correlation of the sensing matrix and, thus, computationally infeasible or expensive signal recovery [6].

These numerical problems associated with the spectral spill-over in the Dirichlet kernel have recently been addressed by the *off-the-grid compressed sensing* approach [7] [8]. This method relies on atomic norm minimization and guarantees recovery of frequencies

lying anywhere in the continuous domain $[0, 1]$ from a limited number of random observations, provided the line spectrum satisfies nominal resolution conditions. A two-dimensional generalization to this method involves Hankel matrix completion and guarantees robustness against corruption of data [9].

These approaches to off-the-grid frequency recovery assume that, other than the sparse nature of signal frequencies, little is known about the signal *a priori*. However, in many applications such as radar [10], acoustics [11], and power systems [12], some information about the signal may already be known through previous measurements or known electrical properties of the signal source. For example, a radar engineer may be aware of the Doppler signature of standard targets such as a fighter jet or a commercial airplane [10]. Similarly, in a weather radar scenario, the range of Doppler frequencies and the spectrum widths for certain types of precipitation events (storms or tornadoes) are often known from previous observations [13].

In this work, we consider a frequency-sparse signal composed of multiple frequencies, of which a few frequencies are known *a priori*. However, the amplitudes and phases of these frequencies are not known. We then propose the *known-poles* algorithm based on the *conditional* atomic norm minimization to first recover the complex coefficients of these known frequencies. Once the known frequencies are filtered from the original signal, the problem then simplifies to a search for the remaining frequency components in the signal.

1.1. Main Results

Suppose p out of s frequencies contributing to $x[l]$ are known *a priori*. Suppose that f_j 's, $s - p + 1 \leq j \leq s$, are known frequencies. Then,

$$x[l] = \sum_{j=1}^{s-p} c_j e^{i2\pi f_j l} + \sum_{j=s-p+1}^s d_j e^{i2\pi f_j l}, \quad l \in \mathcal{N} \quad (1.2)$$

where c_j and d_j are *unknown* complex coefficients of these unknown and known frequencies respectively. Our main result in this work demonstrates that if the known frequency information is included in the off-the-grid atomic norm minimization approach, then all the remaining frequencies can be exactly recovered with a higher probability, even though all the frequencies are *continuous-domain* frequencies. Further, the recovery of the entire unknown spectral content using *known-poles* algorithm is possible with a smaller number of random observations. When the frequencies do not satisfy any minimum resolution conditions, our algorithm suffers lesser degradation in the spectral recovery performance compared to the algorithm which does not use any prior information.

1.2. Relation To Prior Work

Estimation of the parameters of a frequency-sparse signal has been a long-pursued problem [14]. A description of several common tools and methods dealing with the line spectral estimation of regularly sampled signals in the presence or absence of noise can be found in [15] and [16]. These methods have often been accompanied by an undercurrent of research on the spectral analysis with available prior information. For example, the classical Prony's method can be easily modified to account for known frequencies [11]. Variants of the subspace-based frequency estimation methods such as MUSIC and ESPRIT have also been formulated [17] [18], where prior knowledge can be incorporated for parameter estimation. For applications wherein only approximate knowledge of the frequencies is available, the spectral estimation described in [19] applies circular von Mises probability distribution on the spectrum.

For irregularly spaced samples, sparse signal recovery methods which leverage on prior information have recently gained attention for general applications [20] [21] [22] as well as, specifically, spectral estimation [23]. To the best knowledge of the authors, the CS-based spectral estimation methods - aforementioned and others - have not considered spectral components lying off the given frequency grid.

Our model and methods for recovery of the frequency-sparse signal in the continuous dictionary are inspired by [8] and [7]. We additionally show that, if some poles are known, the performance of the recovery can be improved in terms of the probability of successful recovery and the number of random samples needed.

2. SYSTEM MODEL

The signal in (1.1) can be modeled as a positive linear combination of the unit-norm frequency-atoms $a(f_j, \phi_j)[l] \in \mathcal{A} \subset \mathbb{C}^n$ where $\mathcal{A}$ is the set of all frequency-atoms. These frequency atoms are basic units for synthesizing the frequency-sparse signal. Further, suppose the signal in (1.1) is observed on the index set $\mathcal{M} \subset \mathcal{N}$, $|\mathcal{M}| = m \ll n$ where m observations are chosen uniformly at random. Then, to estimate the remaining $\mathcal{N} \setminus \mathcal{M}$ samples of the signal x , [24] suggests minimizing the atomic norm $\|\hat{x}\|_{\mathcal{A}}$ - a sparsity-enforcing analog of ℓ_1 norm for a general atomic set $\mathcal{A}$ - among all vectors $\hat{x}$ leading to the same observed samples as x . The atomic norm is given by,

$$\|\hat{x}\|_{\mathcal{A}} = \inf_{c_j, f_j} \left\{ \sum_{j=1}^s |c_j| : \hat{x}[l] = \sum_{j=1}^s c_j e^{i2\pi f_j l}, l \in \mathcal{M} \right\} \quad (2.1)$$

The semidefinite formulation of $\|\hat{x}\|_{\mathcal{A}}$ is defined as follows:

Definition 2.1. [8] Let T_n denote the $n \times n$ positive semidefinite Toeplitz matrix, $t \in \mathbb{R}^+$, $\text{Tr}(\cdot)$ denote the trace operator and $(\cdot)^*$ denote the complex conjugate. Then,

$$\|\hat{x}\|_{\mathcal{A}} = \inf_{T_n, t} \left\{ \frac{1}{2|\mathcal{N}|} \text{Tr}(T_n) + \frac{1}{2}t : \begin{bmatrix} T_n & \hat{x} \\ \hat{x}^* & t \end{bmatrix} \succeq 0 \right\} \quad (2.2)$$

The positive semidefinite Toeplitz matrix T_n is related to the frequency atoms through the following result by Carathéodory [25]:

$$T_n = URU^* \quad (2.3)$$

$$\text{where } U_{l', j'} = a(f_{j'}, \phi_{j'})[l'], \quad (2.4)$$

$$R = \text{diag}([b_1, \dots, b_{r'}]) \quad (2.5)$$

The diagonal elements of R are real and positive, and $r' = \text{rank}(T_n)$.

Consistent with this definition, the atomic norm minimization problem for the frequency-sparse signal recovery can now be formulated in a semidefinite program (SDP) with m affine equality constraints:

$$\begin{aligned} & \text{minimize}_{T_n, \hat{x}, t} \quad \frac{1}{2|\mathcal{N}|} \text{Tr}(T_n) + \frac{1}{2}t \\ & \text{subject to} \quad \begin{bmatrix} T_n & \hat{x} \\ \hat{x}^* & t \end{bmatrix} \succeq 0 \\ & \quad \hat{x}[l] = x[l], l \in \mathcal{M} \end{aligned} \quad (2.6)$$

3. KNOWN-POLES ALGORITHM

We now consider the case when some frequency components are known *a priori* but their corresponding amplitudes and phases are not. A common approach to harness the prior information about the sparse signal is to replace the classical ℓ_1 norm with the weighted ℓ_1 norm [21] [22]. However, the *known poles* signal does not lead to a trivial application of the weighted ℓ_1 approach, since all the frequencies have a continuous domain. Therefore, we propose the *conditional* atomic norm minimization as the *known-poles* algorithm.

Let the index set of all the frequencies be $\mathcal{S}$, $|\mathcal{S}| = s$. Let $\mathcal{P}$ be the index set of all the known frequencies, and $|\mathcal{P}| = p$. Namely, we assume that the signal x contains some known frequencies f_j , $j \in \mathcal{P} \subseteq \mathcal{S}$, $|\mathcal{P}| = p$. For unknown frequencies, let us denote their complex coefficients as d_j and their *phaseless frequency atoms* as $\alpha_j[l] = a(f_j, 0)[l] = e^{i2\pi f_j l}$. We define the *conditional atomic norm* $\|\hat{x}\|_{\mathcal{A}|\mathcal{P}}$ for the *known poles* as follows:

$$\|\hat{x}\|_{\mathcal{A}|\mathcal{P}} = \inf_{c_j, d_j, f_j} \left\{ \sum_{j=1}^{s-p} |c_j| : \hat{x}[l] = \sum_{j=1}^{s-p} c_j e^{i2\pi f_j l} + \sum_{j=s-p+1}^s d_j e^{i2\pi f_j l}, l \in \mathcal{M} \right\} \quad (3.1)$$

The semidefinite formulation for $\|\hat{x}\|_{\mathcal{A}|\mathcal{P}}$ is given as follows (proof omitted due to space limitations):

Definition 3.1. The conditional atomic norm for a vector $\hat{x}$ is given by

$$\|\hat{x}\|_{\mathcal{A}|\mathcal{P}} = \inf_{T_n, \hat{x}, t, d_j} \left\{ \frac{1}{2|\mathcal{N}|} \text{Tr}(T_n) + \frac{1}{2}t : \begin{bmatrix} T_n & \hat{x} \\ \hat{x}^* & t \end{bmatrix} \succeq 0 \right\} \quad (3.2)$$

where $\hat{x}[l] = \hat{x}[l] - \sum_{j \in \mathcal{P}} \alpha_j[l] d_j$ represents the positive combination of complex sinusoids with unknown poles.

The conditional atomic norm minimization problem can be posed as the following semidefinite formulation in a similar way as in (2.6):

$$\begin{aligned} & \text{minimize}_{T_n, \hat{x}, \hat{x}, t, d_j} \quad \frac{1}{2|\mathcal{N}|} \text{Tr}(T_n) + \frac{1}{2}t \\ & \text{subject to} \quad \begin{bmatrix} T_n & \hat{x} \\ \hat{x}^* & t \end{bmatrix} \succeq 0 \\ & \quad \hat{x}[l] = x[l], l \in \mathcal{M} \\ & \quad \hat{x}[l] = \hat{x}[l] + \sum_{j \in \mathcal{P}} \alpha_j[l] d_j, l \in \mathcal{M} \end{aligned} \quad (3.3)$$

$\tilde{x}$ can be viewed as the signal filtered of the *known poles*. The remaining unknown frequencies in $\tilde{x}$ can be identified by the frequency localization approach [8] based on computing the dual polynomial, which we restate for $\tilde{x}$ in Algorithm 3.1.

Algorithm 3.1. *Known-poles algorithm*

- 1: Solve the semidefinite program (3.3) for *conditional* atomic norm minimization to obtain $\tilde{x}$.
- 2: Solve the following dual problem to obtain the optimum solution q^*

$$\begin{aligned} & \underset{q}{\text{maximize}} \quad \langle q, \tilde{x} \rangle_{\mathbb{R}} \\ & \text{subject to} \quad \|q\|_{\mathcal{A}} \leq 1 \\ & \quad q[l] = 0, \quad l \in \mathcal{N} \setminus \mathcal{M} \end{aligned} \quad (3.4)$$

where $\|\cdot\|$ denotes the dual-norm.

- 3: The unknown frequencies f_j , identify as $|\langle q^*, \alpha_j \rangle| = 1$ where $j \in \mathcal{P}$. For $j \notin \mathcal{S} \setminus \mathcal{P}$, $|\langle q^*, \alpha_j \rangle| < 1$.
 - 4: All the unknown frequencies identified, hereafter their corresponding complex coefficients can be recovered by solving a system of simultaneous linear equations $\tilde{x}[l] - \sum_{j \in \mathcal{S} \setminus \mathcal{P}} c_j \alpha_j[l] = 0$.
-

4. PERFORMANCE ANALYSIS

In this section, we set out to give necessary and sufficient conditions under which (3.3) recovers existing frequencies f_j , $1 \leq j \leq s$ and their complex coefficients, regardless of what amplitudes these s frequency components take. We first need to define a notion of “feasible matrix” $A \in \mathbb{C}^{m \times (n+s)}$. Recall that f_r , $1 \leq r \leq s$, are existing frequencies, and f_r , $s-p+1 \leq r \leq s$, are known frequencies. For a matrix $A \in \mathbb{C}^{m \times (n+s)}$ and an index set $K \subseteq \{1, 2, \dots, n+s\}$, we define $\mathcal{D}(A_K)$ as the dictionary (set) of columns in A corresponding to the index set K . We further define a few index sets as follows:

$$\begin{aligned} \mathcal{S}^* &= \{r | n+1 \leq r \leq n+s\} \\ \mathcal{P}^* &= \{r | n+s-p+1 \leq r \leq n+s\} \\ \mathcal{N}^* &= \{r | 1 \leq r \leq n\} \\ \mathcal{N}^* \cup (\mathcal{S}^* \setminus \mathcal{P}^*) &= \{r | 1 \leq r \leq n+s-p\} \end{aligned}$$

In the following definition of a “feasible matrix” $A \in \mathbb{C}^{m \times (n+s)}$, we will use $A_{\mathcal{S}^*}$ to accommodate the basis vectors for existing frequencies, and $A_{\mathcal{P}^*}$ to accommodate the basis vectors for known frequencies.

Definition 4.1. A matrix $A \in \mathbb{C}^{m \times (n+s)}$ is called a feasible matrix if

- $A_r = a(f_{r-n}, \phi_{r-n})[\mathcal{M}]$, $n+1 \leq r \leq n+p+s$, where $\mathcal{M}$ is the sampling time index set;
- $A_r = a(\tilde{f}_r, \tilde{\phi}_r)[\mathcal{M}]$, $1 \leq r \leq n$, where $\tilde{f}_r$, $1 \leq r \leq n$, are n distinct frequencies in $[0, 1]$; and $\tilde{\phi}_r \in [0, 2\pi)$, $1 \leq r \leq n$;
- $f_j \notin \{\tilde{f}_r | 1 \leq r \leq n\}$ when $s-p+1 \leq j \leq s$.

We use $\mathbf{N}(A)$ to denote the null space of A . We are now ready to give a statement of the conditions for successful recovery.

Theorem 4.2. Let $x[l] = \sum_{j=1}^s c_j e^{i2\pi f_j l} = \sum_{j=1}^s |c_j| a(f_j, \phi_j)[l]$, $l \geq 0$. Assume that $\{f_j | s-p+1 \leq j \leq s\}$ are p known frequencies (but their amplitudes and phases are still unknown). Then, for all possible amplitudes $|c_j| > 0$, $1 \leq j \leq s$, the known-poles

algorithm uniquely recovers all the frequencies, their corresponding phases and amplitudes, if and only if the following holds.

For every feasible matrix $A \in \mathbb{C}^{m \times (n+s)}$ and for every $h \in \{h | h \in \mathbf{N}(A), h \neq 0, h_{\mathcal{N}^* \cup (\mathcal{S}^* \setminus \mathcal{P}^*)} \in \mathbb{R}^{n+s-p}, h_{\mathcal{S}^* \setminus \mathcal{P}^*} < 0, h_{\mathcal{N}^*} \geq 0\}$, $\|h_{\mathcal{N}^*}\|_1 > \|h_{\mathcal{S}^* \setminus \mathcal{P}^*}\|_1$ unless the following three statements hold true simultaneously.

- (1) $a(f_j, \phi_j) \in \mathcal{D}(A_{\mathcal{N}^*})$, when $1 \leq j \leq s-p$;
- (2) $h_{r_1} = -h_{r_2}$, whenever $r_1 \in \mathcal{N}^*$, $r_2 \in \mathcal{S}^* \setminus \mathcal{P}^*$, and $A_{r_2} = A_{r_1}$;
- (3) $h_{r_3} = 0$, if $r_3 \in \mathcal{N}^* \cup \mathcal{S}^*$ and $A_{r_3} \notin \mathcal{D}(A_{\mathcal{S}^* \setminus \mathcal{P}^*})$.

Due to space limitations, we omit the proof in this paper. We remark that, intuitively, when the size of known frequencies (namely $|\mathcal{P}^*|$) increases, it may be easier to satisfy the key condition $\|h_{\mathcal{N}^*}\|_1 > \|h_{\mathcal{S}^* \setminus \mathcal{P}^*}\|_1$ in Theorem 4.2 since, roughly speaking, $\|h_{\mathcal{S}^* \setminus \mathcal{P}^*}\|_1$ decreases when $|\mathcal{P}^*|$ increases. So we expect that when more prior information is known, it becomes easier to recover all the frequencies.

5. NUMERICAL SIMULATIONS

We evaluated the *known-poles* algorithm through a number of simulations using SDPT3 [26] to solve the semidefinite program. In all our experiments, the s frequencies of the artificially generated signal were drawn at random in the band $[0, 1]$. Except for Experiment 4, the sampled frequencies were also constrained to have the minimum modulo spacing of $\Delta f = 1/\lfloor (n-1)/4 \rfloor$ between the adjacent frequencies. This is the theoretical resolution condition for the results in [8], although numerical experiments suggested that frequencies could be closer, i.e., Δf could be $1/(n-1)$. While working with the *known poles*, we draw the first known frequency uniformly at random from the set of s frequencies. As the number p of *known poles* increases, we retain the previously drawn known frequencies and draw the next known frequency uniformly at random from the remaining set of existing signal frequencies.

The phases of the signal frequencies were sampled uniformly at random in $[0, 2\pi)$. The amplitudes $|c_j|$, $j = 1, \dots, s$ were drawn randomly from the distribution $0.5 + \chi_1^2$ where χ_1^2 represents the chi-squared distribution with 1 degree of freedom.

Experiment 1. We simulated a low-dimensional model with the triple $(n, m, s) = (32, 9, 4)$ and first solved the semidefinite program (2.6) which does not use any prior information, i.e., $p = 0$. For the same realization of the signal, we then successively increase p up to $s-1$, and solve the optimization (3.3) of the *known-poles* algorithm. At every instance of solving an SDP, we record the number k of successfully recovered frequencies along with their complex coefficients. This number also includes the known frequencies if the recovery process returns exact values of their complex coefficients. $k = s$ corresponds to *complete success*, i.e., recovering all of the unknown spectral content. $k = 0$ is *complete failure*, including the case when the complex coefficients of the known frequencies could not be recovered. Figure 1 shows the probability P of recovering k frequencies over 1000 trials. We observe that even though the complex coefficients of the known frequencies are unknown, the *known-poles* algorithm increases the probability of accurately recovering all or some of the unknown spectral content.

Experiment 2. We repeat the first experiment for the higher-dimensional pair $(n, m) = (256, 40)$ and vary s . The probability P

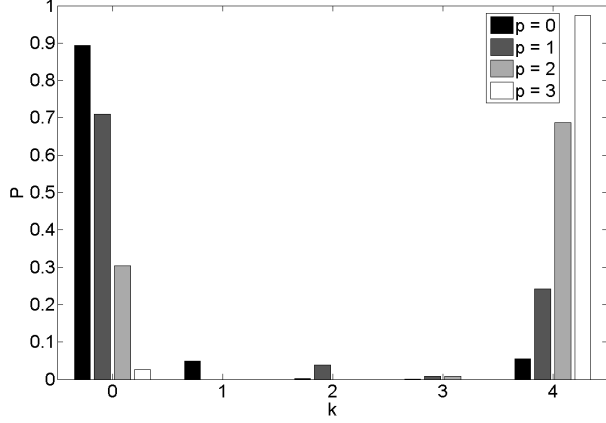


Fig. 1: The probability P of recovering the unknown spectral content. The probability is computed for 1000 random realizations of the signal for the triple $(n, m, s) = (32, 9, 4)$. (For $k > 0$, $k \leq p$ being the invalid cases, the corresponding bars have been omitted.)

over 100 random realizations of the signal is shown in Figure 2 for selected values of s . We observe that the probability of successfully recovering all the frequencies using the *known-poles* Algorithm 3.1 increases with p .

Experiment 3. Figure 3 shows the probability P of *complete success*, (i.e. $k = s$), as a function of m over 100 trials for the twin $(n, s) = (80, 6)$. We note that, the *known-poles* algorithm achieves the same recovery probability when compared to (2.6) with a smaller number of random observations.

Experiment 4. We now consider these two cases: (a) when $\Delta f = 1/(n-1)$, the resolution limit for the numerical experiments in [8], and (b) when the frequencies are drawn uniformly at random and do not adhere to any minimum resolution conditions. Figure 4 shows the probability P of recovering k frequencies over 1000 trials for the triple $(n, m, s) = (40, 15, 7)$. We note that the probability of *complete success* with *known poles* suffers relatively little degradation for the random frequency resolutions. These trials include instances when the minimum resolution condition does not hold, formulation in (2.6) shows *complete failure* but the *known-poles* algorithm recovers the unknown spectral content with *complete success*.

6. SUMMARY

Our off-the-grid CS formulation for the *known-poles* algorithm suggests that using the known frequencies of the signal, the signal recovery performance can be improved leading to a higher probability of recovering some or all the residual spectral information - frequency, amplitude and phase - with smaller number of random observations. In the future, it would be interesting to further investigate theoretical performance limits of off-the-grid compressed sensing with prior information.

7. ACKNOWLEDGEMENT

The authors would like to thank Gongguo Tang and Benjamin Recht of University of Wisconsin at Madison for helpful e-mail discussions related to their work in [8].

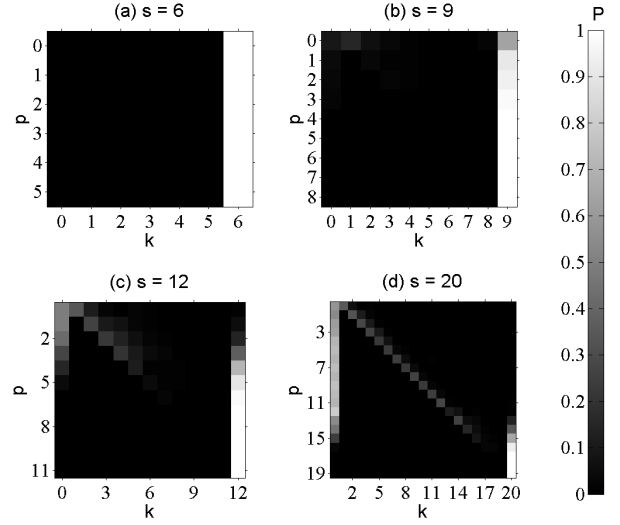


Fig. 2: The probability P of recovering the unknown spectral content for selected values of s . The probability is computed for 100 random realizations of the signal with $(n, m) = (256, 40)$. (The lower diagonal cases when $k > 0$, $k \leq p$ are invalid, and do not contribute to the result.)

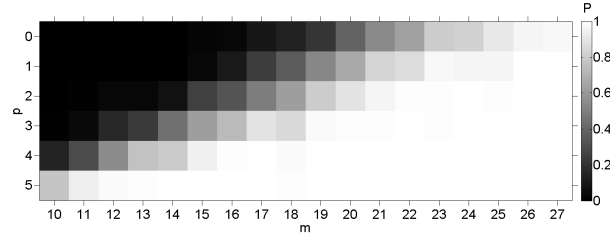


Fig. 3: A higher probability P of recovering all the unknown frequency content can be achieved with a smaller number m of random observations using the *known-poles* algorithm. The probability is computed for 100 random realizations with $(n, s) = (80, 6)$.

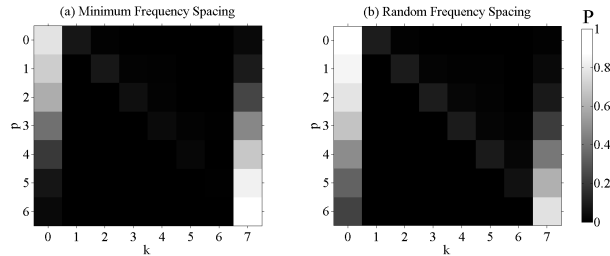


Fig. 4: Performance of the *known-poles* algorithm when the frequencies do not satisfy any nominal resolution conditions. The probability P of successfully recovering k frequencies is computed for 1000 realizations of the signal with dimensions $(n, m, s) = (40, 15, 7)$. (a) $\Delta f = 1/(n-1)$ (b) Frequencies are selected uniformly at random in the band $[0, 1]$.

8. REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] G. Kutyniok, "Theory and applications of compressed sensing," *arXiv preprint arXiv:1203.3815*, 2012.
- [3] G. Majda and M. Wei, "A simple procedure to eliminate known poles from a time series," *IEEE Transactions on Antennas and Propagation*, vol. 37, no. 10, pp. 1343–1344, 1989.
- [4] M. Wei and G. Majda, "A new theoretical approach for Prony's method," *Linear Algebra and its Applications*, vol. 136, pp. 119–132, 1990.
- [5] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.
- [6] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," *Applied and Computational Harmonic Analysis*, vol. 35, no. 1, pp. 111–129, 2013.
- [7] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, 2013, to appear.
- [8] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *arXiv preprint arXiv:1207.6053*, 2012.
- [9] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *arXiv preprint arXiv:1304.8126*, 2013.
- [10] M. I. Skolnik, *Radar handbook*, McGraw-Hill, third edition, 2008.
- [11] D. H. Trivett and A. Z. Robinson, "Modified Prony method approach to echo-reduction measurements," *The Journal of the Acoustical Society of America*, vol. 70, pp. 1166, 1981.
- [12] J. Zygarlicki and J. Mroczka, "Prony's method used for testing harmonics and interharmonics in electrical power systems," *Metrology and Measurement Systems*, vol. 19, no. 4, pp. 659–672, 2012.
- [13] R. J. Doviak and D. S. Zrnić, *Doppler radar and weather observations*, pp. 87–118, Dover Publications, second edition, 2006.
- [14] B. G. C. F. M. R. de Prony, "Essai expérimental 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, 1795, in French.
- [15] S. L. Marple Jr, *Digital spectral analysis with applications*, Englewood Cliffs, NJ, Prentice-Hall, Inc., 1987.
- [16] P. Stoica and R. L. Moses, *Spectral analysis of signals*, Prentice Hall, 2005.
- [17] D. A. Linebarger, R. D. DeGroat, E. M. Dowling, P. Stoica, and G. L. Fudge, "Incorporating a priori information into MUSIC-algorithms and analysis," *Signal Processing*, vol. 46, no. 1, pp. 85–104, 1995.
- [18] P. Wirfält, G. Bouleux, M. Jansson, and P. Stoica, "Subspace-based frequency estimation utilizing prior information," in *IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 533–536.
- [19] D. Zachariah, P. Wirfält, M. Jansson, and S. Chatterjee, "Line spectrum estimation with probabilistic priors," *Signal Processing*, vol. 93, no. 11, pp. 2969 – 2974, 2013.
- [20] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [21] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *IEEE International Symposium on Information Theory*, 2009, pp. 483–487.
- [22] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [23] S. Bourguignon, H. Carfantan, and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 575–585, 2007.
- [24] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.
- [25] C. C. Carathéodory, "Über den variabilitätsbereich der fourierschen konstanten von positiven harmonischen funktionen," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 32, no. 1, pp. 193–217, 1911, in German.
- [26] 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, 2003.