Adaptive Interference Removal for Un-coordinated Radar/Communication Co-existence

Le Zheng, Member, IEEE, Marco Lops, Senior Member and Xiaodong Wang, Fellow, IEEE

Abstract

Most existing approaches to co-existing communication/radar systems assume that the radar and communication systems are coordinated, i.e., they are aware of the existence of each other, and share information such as position and channel information. In this paper, we consider an un-coordinated scenario where a communication receiver is to operate in the presence of a number of radars, of which only a sub-set may be active, which poses the problem of estimating the interfering waveforms and canceling them prior to demodulation. Two algorithms are proposed for such a joint waveform estimation/data demodulation problem, both exploiting sparsity of a proper representation of the interference and of the vector containing the errors of the data block, so as to implement an iterative joint interference removal/data demodulation process. The former algorithm is based on classical on-grid compressed sensing (CS), while the latter forces an atomic norm (AN) constraint: in both cases the radar parameters and the communication demodulation errors can be estimated by solving a convex problem. We also propose a way to improve the efficiency of the AN-based algorithm. The performance of these algorithms are demonstrated through extensive simulations, taking into account a variety of conditions concerning both the interferers and the respective channel states.

Index Terms

Radar/communication co-existence, atomic norm, compressed sensing, off-grid, sparsity.

I. INTRODUCTION

The ever increasing demand for spectrum and the consequent shortage of available bandwidths pave the way to communication/radar co-existing system architectures [1]–[4]. This inevitably produces inter-system interference that degrades the performance of both systems: in particular, the radar transmit power may be large enough to

Le Zheng and Xiaodong Wang are with Electrical Engineering Department, Columbia University, New York, USA, 10027, e-mail: le.zheng.cn@gmail.com, wangx@ee.columbia.edu.

Marco Lops is with the DIEI, Universita degli Studi di Cassino e del Lazio Meridionale, Cassino 03043, Italy (e-mail: lops@unicas.it, e.grossi@unicas.it).
significantly degrade the performance of the communication system. Techniques such as interference mitigation [5], pre-coding or spatial separation [6]–[8], waveform design [9]–[12] allow both radar and communications to share the spectrum and co-exist: for example, in [6], [8], [12], the radar interference is eliminated by forcing the radar waveforms to live in the null space of the interference channel between the radar transmitters and the communication receiver. Motivated by the cooperative methods in cognitive radio networks, later works [13]–[15] exploit some prior knowledge to jointly design the radar waveform and communication code-book by minimizing a measure of the mutual interference under certain constraints.

Most existing approaches assume that the radar and communication systems are aware of the existence of each other, and share information: for example, ad hoc design of radar waveforms or beam-formers is proposed in [5]–[10] to reduce the mutual interference, while [11]–[15] rely on making channel information available to the communication and radar through the transmission of pilot training. Otherwise stated, these approaches rely on a centralized architecture, namely on a strict coordination between the active players in order to allow co-existence. The situation we refer to in this contribution is one wherein a communication system should share its spectrum with an ensemble of potential interferers, i.e., a set of radar/sensing systems. When observed on a conveniently long time interval, this situation is akin to a highly non-stationary environment, wherein the sources of interference may vary over time, and so do the corresponding waveforms, timing and channels. While the interference produced by the (unique) communication system on the active radars may be neglected, due to both the order of magnitude of the powers in play and some specific countermeasure that can be taken (for example, the active radars may use suitable beam-forming techniques to get rid of interference from a known location [16]), the interference produced onto the communication system may be highly detrimental, and must be dealt with.

In this un-coordinated scenario, the only information the communication system can rely upon is that the interfering waveforms live in the subspace of a dictionary, and that they impinge on the communication receiver (RX) with unknown, possibly time-varying delays: as a consequence, the communication RX must be made adaptive, in order to accomplish jointly the two tasks of interference estimation/removal and data demodulation.

The approach we propose here focuses on guaranteeing the performance of the communication system and relies on the size of the dictionary from which the radar waveforms are picked up: we first show that, adopting a suitable representation domain, the interfering signals hitting the communication RX are sparse. On a parallel track, if iterative demodulation/re-modulation algorithms are implemented, the vector containing the demodulation errors of a data block should be itself sparse (and become sparser and sparser as the iterations go), whereby a joint interference removal/data demodulation process can take great advantage of existing algorithms forcing sparsity constraints.

Unfortunately, however, the scenario we consider in this paper is not ensured to lend itself to direct application
of compressed sensing (CS) theory [17], which relies on the fact that signals can be sparsely represented by a finite discrete dictionary [18]–[21]: the presence of relevant continuous parameters, such as delays, could indeed lead to remarkable degradations from model mismatch, should a simple discretization of the parameter space be implemented [22]. We thus also explore the applicability of the recently developed mathematical theory of continuous sparse recovery for super-resolution [23]–[25] and in particular of Atomic-Norm (AN) minimization techniques successfully employed for continuous frequency recovery from incomplete data [25], [26], direction-of-arrival estimation [27], channel estimation [28] and line spectral estimation [29].

Given the above framework, we thus propose two algorithms for joint waveform estimation and data demodulation in the overlaid radar/communication architecture, the former based on classical on-grid CS, the latter forcing an AN constraint: in both cases the radar parameters and the communication demodulation errors can be estimated by solving a convex problem. We also propose a way to improve the efficiency of the AN-based algorithm. The merits of these algorithms are demonstrated through extensive simulations, taking into account a variety of conditions concerning both the interferers and the respective channel states.

The remainder of the paper is organized as follows. In Section II, we present the signal model of the co-existed radar and communication system. In Section III, we develop the proposed CS-based algorithms using both the atomic norm and the $\ell_1$-norm. In Section IV, an accelerated algorithm for solving the atomic norm-based algorithm is proposed. Simulation results are presented in Section V. Section VI, finally, contains concluding remarks.

II. PROBLEM FORMULATION

We consider a situation with one communication system and $J$ active radars. Suppose the $j$-th radar transmits the coded waveform

$$s_j(t) = \sum_{n=1}^{N} g_j(n)\xi(t - (n - 1)T),$$

(1)

where $g_j(n)$ is the $n$-th code, $N$ is the number of codes, $\xi(t)$ satisfies the Nyquist criterion with respect to $T$ and the bandwidth is $1/T$. Let $\bar{g}_j = [\bar{g}_j(1), \bar{g}_j(2), ..., \bar{g}_j(N)]^T \in \mathbb{C}^{N \times 1}$ be the Discrete Fourier Transform (DFT) of $g_j = [g_j(1), g_j(2), ..., g_j(N)]^T \in \mathbb{C}^{N \times 1}$, i.e., $\bar{g}_j = Fg_j$ with $F = [f_1, f_2, ..., f_N]^H \in \mathbb{C}^{N \times N}$ denoting the DFT matrix. We assume that $\bar{g}_j$ lives in a low-dimensional subspace spanned by the columns of a known $N \times K$ matrix $\bar{D} = [\bar{d}_1, \bar{d}_2, ..., \bar{d}_N]^H \in \mathbb{C}^{N \times K}$ with $\bar{d}_n \in \mathbb{C}^{K \times 1}$ and $K \ll N$, i.e., $\bar{g}_j = \bar{D}h_j$ for some unknown $h_j \in \mathbb{C}^{K \times 1}$.

Specifically, in the case that the radar waveform is known a priori, we have $K = 1$ and $\bar{D}$ becomes an $N \times 1$ vector.

Without loss of generality, we assume that $\|h_j\|_2 = 1$ for $j = 1, 2, ..., J$. For simplicity, we also assume that there is one path between the radar TX and the communication RX. This assumption is true for narrow-band radar systems [14] or as the interference is dominated by the direct path between the radar TX and the communication
RX. The interference produced by \( J \) active radars - with \( J \) possibly unknown - onto the communication RX can be expressed as

\[
y_I(t) = \sum_{j=1}^{J} \sum_{n=1}^{N} c_j g_j(n) \xi(t - (n-1)T - \tau_j),
\]

where \( \tau_j \) and \( c_j \) denote the delay and path gain of the \( j \)-th radar, respectively.

The communication TX transmits data symbols \( b = [b(1), b(2), ..., b(M)]^T \in \mathbb{C}^{M \times 1} \) with \( M \leq N \). We assume that \( b \in B \) where \( B \) denotes the set of possible \( b \) values. Defining the data received at the RX side as \( x = [x(1), x(2), ..., x(N)]^T \in \mathbb{C}^{N \times 1} \), we have

\[
x = H Ab,
\]

where \( A \in \mathbb{C}^{N \times M} \), and \( H \in \mathbb{C}^{N \times N} \) is the channel matrix. This model subsumes a number of communication systems. For example, in a Code-Division Multiple Access (CDMA) system [30], the elements in \( b \) are the symbols transmitted by \( M \) active users: The columns of \( A \) are the signatures of the users, and \( H \) is a diagonal matrix representing the channel gains. Another example is an OFDM system, in which \( A \) is the IDFT matrix with \( M = N \), \( H \) is the channel matrix and \( x \) is the received data in time domain. In practice, \( A \) is known and the channel matrix \( H \) can be obtained through the transmission of pilot signals. Let \( \bar{x} = [\bar{x}(1), \bar{x}(2), ..., \bar{x}(N)]^T \in \mathbb{C}^{N \times 1} \) be the DFT of \( x \), i.e., \( \bar{x} = Fx \).

The received communication signal is given by

\[
y_C(t) = \sum_{n=1}^{N} x(n) \xi(t - (n-1)T - \tau_C),
\]

where \( \tau_C \) denotes the overall delay of the communication transmission. As the communication TX and RX are synchronized, \( \tau_C \) is assumed to be known.

At the communication RX, the signal contains both the communication signal and the radar interference, i.e.,

\[
y(t) = y_I(t) + y_C(t) + \bar{w}(t),
\]

\[
y(t) = \sum_{j=1}^{J} \sum_{n=1}^{N} c_j g_j(n) \xi(t - (n-1)T - \tau_j) + \sum_{n=1}^{N} x(n) \xi(t - (n-1)T - \tau_C) + \bar{w}(t),
\]

where \( \bar{w}(t) \) is the measurement noise. Projecting \( y(t) \) onto \( \xi(t - t' - \tau_C) \) results in

\[
r(t') = \langle y(t), \xi(t - t' - \tau_C) \rangle
\]

\[
r(t') = \sum_{n=1}^{N} x(m) R_\xi(t' - (n-1)T) + \sum_{j=1}^{J} \sum_{n=1}^{N} c_j g_j(n) R_\xi(t' - (n-1)T - \tau_j + \tau_C) + w(t'),
\]

where \( R_\xi(\cdot) \) is the auto-correlation function of \( \xi(\cdot) \), i.e., \( R_\xi(\tau) = \langle \xi(t), \xi(t - \tau) \rangle \) with \( \langle \cdot \rangle \) denoting inner product, \( w(t') = \langle \bar{w}(t), \xi(t - t' - \tau_C) \rangle \). The auto-correlation function \( R_\xi(t) \) is considered substantially time-limited in a finite interval, \([−T', T']\) say. Letting \( \tau_{\text{min}} = \min_j \{\tau_j\} \) and \( \tau_{\text{max}} = \max_j \{\tau_j\} \) be respectively the minimum and the
maximum delays, tied to the corresponding minimal and maximum distances of all of the potential radar systems from the receiver, we define

\[ \tilde{r}(k) = \int_{\tau_{\min} - \tau_C}^{\tau_{\max} - \tau_C + (N-1)T + T'} r(t')e^{-i2\pi(k-1)T'} dt' \]

\[ = \sum_{n=1}^{N} x(n) \int_{\tau_{\min} - \tau_C}^{\tau_{\max} - \tau_C + (N-1)T + T'} R_\xi(t' - (n-1)T)e^{-i2\pi(k-1)T'} dt' + \sum_{j=1}^{J} N \sum_{n=1}^{N} c_j g_j(n) \int_{\tau_{\min} - \tau_C}^{\tau_{\max} - \tau_C + (N-1)T + T'} R_\xi(t' - (n-1)T - \tau_j + \tau_C)e^{-i2\pi(k-1)T'} dt' \]

\[ \approx \sum_{n=1}^{N} x(n) \int_{-\infty}^{\infty} R_\xi(t' - (n-1)T)e^{-i2\pi(k-1)T'} dt' + \sum_{j=1}^{J} e^{-i2\pi(k-1)T'} \sum_{n=1}^{N} c_j g_j(n) \int_{-\infty}^{\infty} R_\xi(t' - (n-1)T)e^{-i2\pi(k-1)T'} dt' \]

(7)

for \( k = 1, 2, \ldots, N \), where \( \tau_j' = \frac{\tau_j - \tau_C}{T} \). For simplicity, we define \( \tilde{w}(k) = \int_{\tau_{\min} - \tau_C}^{\tau_{\max} - \tau_C + (N-1)T + T'} w(t')e^{-i2\pi(k-1)T'} dt' \).

It is assumed that \( \tilde{w}(k) \) is a complex Gaussian variable with zero mean and variance \( \sigma_w^2 \), i.e., \( \tilde{w}(k) \sim CN(0, \sigma_w^2) \).

Suppose that \( R_\xi(\cdot) \) satisfies \( \int_{-\infty}^{\infty} R_\xi(t)e^{-i2\pi(k-1)T'} dt \approx 1 \) for \( k = 1, 2, \ldots, N \). The above condition is rigorously true in the relevant case that \( R_\xi(\cdot) \) is a sinc function (i.e., \( \beta = 0 \)), as it happens if the communication system employs an OFDM format, and is approximately true for any other format fitting the model (3) as far as bandwidth efficiency is pursued by forcing a small zero excess bandwidth factor as compared to \( 1/T \), e.g., if \( \xi(\cdot) \) is a Square Root Raised Cosine (SQRRT) with small roll-off factor \( \beta \). Then we have

\[ \int_{-\infty}^{\infty} R_\xi(t - (n-1)T)e^{-i2\pi(k-1)T'} dt = e^{-i2\pi(n-1)(k-1)T}N \].

(8)

Plugging (8) into (7), we have

\[ \tilde{r}(k) = \sum_{n=1}^{N} x(m)e^{-i2\pi(n-1)(k-1)N} + \tilde{w}(k) + \sum_{j=1}^{J} e^{-i2\pi(k-1)T'} \sum_{n=1}^{N} c_j g_j(n)e^{-i2\pi(n-1)(k-1)N} \]

\[ = \tilde{x}(k) + \tilde{w}(k) + \sum_{j=1}^{J} c_j \tilde{g}_j(k)e^{-i2\pi(k-1)T'} \]

(9)

where \( \tilde{x}(k) \) and \( \tilde{g}_j(k) \) are the \( k \)-th element of \( \tilde{x} \) and \( \tilde{g}_j \), respectively.

We define \( \tilde{r} = [\tilde{r}(1), \tilde{r}(2), \ldots, \tilde{r}(N)]^T \in \mathbb{C}^{N \times 1} \). As outlined in the introduction, the communication RX has to remove the radar interference from the measurement with no knowledge of the number, the delays and the waveforms of the active radars, i.e., under uncertainty concerning \( J \), \( \{\tau_j\}_{1 \leq j \leq J} \) and \( \{\tilde{g}_j\}_{1 \leq j \leq J} \). In principle, demodulation may be undertaken simply ignoring the presence of interference, i.e., through the operation \( \hat{b} = \Psi(\tilde{r}) \), with \( \Psi(\cdot) \) being the decoding function operating on the received signal, which would obviously lead to an uncontrolled symbol-error-rate (SER). The approach we take here instead relies on a joint interference-estimation symbol-demodulation
process. In particular, defining \( \hat{x} = [\hat{x}(1), \hat{x}(2), \ldots, \hat{x}(N)]^T = H A \hat{b} \in \mathbb{C}^{N \times 1} \) as the “estimated communication signal”, the presence of errors in the decision process results into a non-zero difference vector \( \hat{x} - F \hat{x} = F H A v \) where \( v = b - \hat{b} \). Obviously, the \( k \)-th element of \( z = [z(1), z(2), \ldots, z(N)]^T = \bar{r} - \hat{x} \) is given by

\[
   z(k) = \langle H A v, f_k \rangle + \sum_{j=1}^J c_j g_j(k) e^{-\frac{(2\pi k \tau_j) e}{N}} + \bar{w}(k),
\]

\[
   = \langle H A v, f_k \rangle + \sum_{j=1}^J c_j a(\tau_j^H) e_k h_j + \bar{w}(k),
\]

\[
   = \langle H A v, f_k \rangle + \langle X, \hat{d}_k e_k^H \rangle + \bar{w}(k), \tag{10}
\]

where we have defined \( \langle X, Y \rangle = \text{Tr}(Y^H X) \), \( X = \sum_{j=1}^J c_j h_j a(\tau_j^H) \) with \( a(\tau) = [1, e^{\frac{i 2\pi \tau}{N}}, \ldots, e^{\frac{i 2\pi (N-1) \tau}{N}}]^T \), \( e_k \) is the \( k \)-th column of the \( N \times N \) identity matrix \( I_N \). Once estimates of \( X \) and \( v \) are available, the radar interference can be obtained and canceled from the measurements and the symbols re-demodulated. Hence, the main problem is to estimate \( X \) and \( v \) from the noisy measurements \( z \). Note that \( z(k) \) contains both the radar interference and the residual of communication signal caused by the mis-demodulations. The mixing of both signals causes great difficulties for the estimation, which inspires us to exploit some structural information about the desired solution, and in particular their sparsity, as detailed in the next section.

### III. Proposed Algorithms

Equation (10) highlights that data demodulation and interference mitigation are coupled, in the sense that they should be accomplished jointly and that poor performance in estimating either one has detrimental effects on the estimate of the other. In fact, in order to remove the radar interference, we need to estimate the matrix \( X \) from the observations (10), but this would obviously require also estimating the error vector \( v \), which boils down to correctly demodulating the data block. To this end, we design iterative algorithms, exploiting structural information on the desired solution. In the \( l \)-th iteration, the demodulated symbols are denoted as \( \hat{b}(l) \), and \( z(l) = \bar{r} - F \hat{x}(l-1) \) where \( \hat{x}(l-1) = H A \hat{b}(l-1) \) is the estimate of \( x \). The following two types of sparsity are exploited in the problem:

1) The signal \( X \) is a combination of \( J \) complex exponentials \( a(\tau_j^H) \) with unknown modulation \( c_j h_j \), and the number of complex exponentials is much smaller than the number of measurements, i.e., \( J \ll N \).

2) Ideally, the vector \( v \) should be an all-zero vector. As a consequence, denoting \( v(l) = b - \hat{b}(l-1) \) the result of the \( l \)-th iteration, we want \( \|v(l\|_0 = L_I \) to be as small as possible, and in any case we want to force the condition \( L_I \ll N \).
A. Joint Waveform Estimation and Demodulation Based on On-grid CS Algorithm

In the first iteration, \( \hat{\mathbf{b}}^{(0)} = \Psi(\bar{\mathbf{r}}) \) so \( \mathbf{v}^{(1)} = \mathbf{b} - \hat{\mathbf{b}}^{(0)} \). According to (10), jointly estimating \( c_j \mathbf{h}_j, \mathbf{a}(\tau_j^{'}) \) and \( \mathbf{v}^{(1)} \) is a non-linear problem. However, it can be linearized by using an overcomplete dictionary matrix

\[
\hat{\mathbf{A}} = [\mathbf{a}(\tau_1^{'},),\mathbf{a}(\tau_2^{'},),...,\mathbf{a}(\tau_{\tilde{J}}^{'})] \in \mathbb{C}^{N \times \tilde{J}},
\]

(11)

with \( \{\tau_j^{'}\}_{j=1,2,...,\tilde{J}} \) denoting sets of uniformly spaced points of the radar delay. Define \( \tilde{J} \) as the number of columns of \( \hat{\mathbf{A}} \) where \( \tilde{J} \geq N \). For sufficiently large \( \tilde{J} \), the delay is densely sampled. Let \( \mathbf{\alpha} = [\tilde{c}_1 \mathbf{h}_1^T, \tilde{c}_2 \mathbf{h}_2^T, ... , \tilde{c}_j \mathbf{h}_j^T]^T \in \mathbb{C}^{JK \times 1} \) be the sparse vector whose non-zero elements correspond to \( c_j \mathbf{h}_j \) in (10).

As usual, forcing a constraint onto the \( \ell_0 \)-norm is impractical, since it results in an NP-hard non-convex optimization problem, and \( \ell_1 \)-norm regularization is used instead, i.e., \( \|\mathbf{\alpha}\|_1 = \sum_{k=1}^{JK} |\alpha(k)| \) and \( \|\mathbf{v}\|_1 = \sum_{k=1}^{N} |v(k)| \). Define \( \mathcal{V} \) as the set of all possible differences \( \mathbf{b} - \hat{\mathbf{b}}^{(0)} \) when the two vectors both belong to \( \mathcal{B} \). We notice that the constraint \( \mathbf{v}^{(1)} \in \mathcal{V} \) results in a non-convex problem, and would cause much difficulty in solving the optimization problem. This constraint is removed and the non-linear joint estimation problem is thus reduced to a linear parameter estimation problem, i.e., the estimation of the linear amplitude vectors \( \mathbf{\alpha} \) and \( \mathbf{v}^{(1)} \), under a sparsity constraint:

\[
(\tilde{\mathbf{\alpha}}^{(1)}, \tilde{\mathbf{v}}^{(1)}) = \arg \min_{\mathbf{\alpha} \in \mathbb{C}^{JK \times 1}, \mathbf{v}^{(1)} \in \mathbb{C}^{N \times 1}} \frac{1}{2} \| \mathbf{z} - \Phi \mathbf{v}^{(1)} - \mathbf{Y} \|_2^2 + \tilde{\lambda} \|\mathbf{\alpha}\|_1 + \tilde{\gamma} \|\mathbf{v}^{(1)}\|_1,
\]

(12)

where

\[
\Phi = \begin{bmatrix}
\mathbf{f}_1 H \mathbf{A} \\
\mathbf{f}_2 H \mathbf{A} \\
\vdots \\
\mathbf{f}_{\tilde{J}} H \mathbf{A}
\end{bmatrix},
\]

(13)

\[
\mathbf{Y} = \begin{bmatrix}
\mathbf{a}(\tau_1^{'}) H \mathbf{e}_1 \mathbf{d}_1^H & \mathbf{a}(\tau_2^{'}) H \mathbf{e}_1 \mathbf{d}_1^H & ... & \mathbf{a}(\tau_{\tilde{J}}^{'}) H \mathbf{e}_1 \mathbf{d}_1^H \\
\mathbf{a}(\tau_1^{'}) H \mathbf{e}_2 \mathbf{d}_2^H & \mathbf{a}(\tau_2^{'}) H \mathbf{e}_2 \mathbf{d}_2^H & ... & \mathbf{a}(\tau_{\tilde{J}}^{'}) H \mathbf{e}_2 \mathbf{d}_2^H \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{a}(\tau_1^{'}) H \mathbf{e}_N \mathbf{d}_N^H & \mathbf{a}(\tau_2^{'}) H \mathbf{e}_N \mathbf{d}_N^H & ... & \mathbf{a}(\tau_{\tilde{J}}^{'}) H \mathbf{e}_N \mathbf{d}_N^H
\end{bmatrix},
\]

(14)

\( \tilde{\lambda} \) and \( \tilde{\gamma} \) are weighting parameters determining the sparsity of the reconstruction. In practice, we set \( \tilde{\lambda}, \tilde{\gamma} = \sigma_w \sqrt{2 \log(\tilde{J}K)} \). As (12) is convex, it can be solved with standard convex solvers.

By solving (12), we obtain the estimates \( \tilde{\mathbf{\alpha}}^{(1)} \) and \( \tilde{\mathbf{v}}^{(1)} \), whereby the symbols can be corrected and re-demodulated as:

\[
\hat{\mathbf{b}}^{(1)} = \arg \min_{\mathbf{b} \in \mathcal{B}} \| \mathbf{b} - \hat{\mathbf{b}}^{(0)} - \tilde{\mathbf{v}}^{(1)} \|_2.
\]

(15)

Note that the re-demodulation process makes use of the structural information of the communication symbols, i.e., \( \mathbf{b} \in \mathcal{B} \). After the re-demodulation, the demodulated symbols belong to the constellation alphabet, whereby \( \mathbf{v}^{(2)} \in \mathcal{V} \).
When the estimation of the symbols in the first iteration is very accurate and the measurement noise is small, i.e., \( \|v^{(1)}\|_0 \) and \( \sigma_w \) are both small, then all the mistakenly demodulated symbols can be corrected by applying (15). In many cases, however, the interference from the radars is strong, and \( \hat{b}^{(0)} \) contains many demodulation errors. Thus, we need to iterate the joint interference removal/data demodulation process. Specifically, in the \( l \)-th iteration \( (l \geq 2) \), \( \hat{\alpha}^{(l)} \) and \( \hat{v}^{(l)} \) are estimated as

\[
(\hat{\alpha}^{(l)}, \hat{v}^{(l)}) = \arg \min_{\alpha \in \mathbb{C}^{K \times 1}, \nu^{(l)} \in \mathbb{C}^{N \times 1}} \frac{1}{2} \|z^{(l)} - \Phi \nu^{(l)} - \Upsilon \alpha\|_2^2 + \lambda \|\alpha\|_1 + \gamma \|v^{(l)}\|_1.
\]  

(16)

Then re-demodulation is undertaken by (15) with \( \hat{b}^{(0)} \) replaced by \( \hat{b}^{(l-1)} \) and \( \hat{v}^{(1)} \) replaced by \( \hat{v}^{(l)} \). As the iteration goes, some wrong symbols are corrected, and the demodulation error \( v^{(l)} \) becomes sparser as \( l \) increases. The proposed algorithm iterates until \( \hat{b}^{(l-1)} = \hat{b}^{(l)} \) or the maximum number of iteration is reached.

Let \( \hat{\alpha} \) be the estimate of \( \alpha \) when the algorithm terminates. The radar delays can be identified by locating the non-zero entries of \( \hat{\alpha} \). If the solution \([\hat{\alpha}_{(j-1)K+1}, \hat{\alpha}_{(j-1)K+2}, \ldots, \hat{\alpha}_{(j-1)K+K}]^T \) is not a zero vector, i.e., \( \|\hat{c}_j \hat{h}_j\|_2 \neq 0 \), then a radar interference exists at delay \( \tau'_j T + \tau_C \). Note that one cannot resolve the inherent scaling ambiguity between each \( \hat{c}_j \) and the corresponding \( \hat{h}_j \), which is in any case not essential since \( c_j g_j \) is of interest for radar interference removal purposes. The estimated time domain radar waveform is then given by

\[
\hat{c}_j \hat{g}_j = F^H D \hat{c}_j \hat{h}_j.
\]  

(17)

The CS algorithm based on \( \ell_1 \)-minimization (CS-L1) is capable of super-resolving the spectrum of the sparse signal under certain conditions of the matrices \( \Phi \) and \( \Upsilon \) [31]. However, the spectrum of interest is discretized into a number of grids, and the radars may not exactly reside on the grids. The off-grid radar position can lead to mismatches in the model and deteriorate the performance, as detailed later on in the section.

\section*{B. Joint Waveform Estimation and Demodulation Based on Off-grid CS Algorithm}

As anticipated, (10) reveals that \( X \) is a linear combination of modulated complex exponentials with arbitrary phases, where the frequencies do not fall onto discrete grids. In this subsection, we use the atomic norm to build a sparse representation which does not suffer from the off-grid problem. We define the atomic norm [32] associated to \( X \) as

\[
\|X\|_\mathcal{A} = \inf \{ \mu > 0 : X \in \mu \text{conv}(\mathcal{A}) \}
\]

\[
= \inf_{c_j, \tau'_j, \|h_j\|_2 = 1} \left\{ \sum_j |c_j| : X = \sum_j c_j h_j a(\tau'_j)^H \right\},
\]  

(18)

where \( \text{conv}(\cdot) \) denotes the convex hull of the input atom set, and the set of atoms is defined as

\[
\mathcal{A} = \{ h a(\tau')^H : \tau' \in [0, 1), \|h\|_2 = 1, h \in \mathbb{C}^{K \times 1} \}.
\]  

(19)
For future developments, we introduce also the following equivalent form of the atomic norm for the atom set $A$ [32]:

$$\|X\|_A = \inf_{\mathbf{u}, T} \left\{ \frac{1}{2N} \text{Tr}(\text{Toep}(\mathbf{u})) + \frac{1}{2} \text{Tr}(T), \right. \\
\left. \text{s.t.} \begin{bmatrix} \text{Toep}(\mathbf{u}) & X^H \\ X & T \end{bmatrix} \succeq 0 \right\},$$

where $\mathbf{u} \in \mathbb{C}^{N \times 1}$ is a complex vector whose first entry is real, Toep($\mathbf{u}$) denotes the $N \times N$ Hermitian Toeplitz matrix whose first column is $\mathbf{u}$, and $T$ is a Hermitian $K \times K$ matrix.

The atomic norm can be exploited to enforce sparsity in the atom set $A$ [25], [32]. Based on (10), and paralleling the arguments outlined in the previous sub-section, the $l$-th iteration achieves estimates, $\hat{X}^{(l)}$ and $\hat{v}^{(l)}$ say, of $X$ and $v^{(l)}$ by processing $z^{(l)}$ and solving the optimization problem:

$$\begin{align*}
(\hat{X}^{(l)}, \hat{v}^{(l)}) &= \min_{X, v^{(l)}} \sum_{k=1}^{N} \frac{1}{2} |z^{(l)}(k)| - \left\{ \left( f_k, HAv^{(l)} \right) - \left( X, \tilde{d}_k e_k^H \right) \right\}^2 + \lambda \|X\|_A + \gamma \|v^{(l)}\|_1,
\end{align*}$$

where $\lambda > 0$ and $\gamma > 0$ are the weight factors. In practice, we set $\lambda \simeq \sigma_w \sqrt{KN \log(KN)}$ and $\gamma \simeq \sigma_w \sqrt{K \log(KN)}$.

In light of (20), the above can be transformed into the following SDP:

$$\begin{align*}
(\hat{X}^{(l)}, \hat{v}^{(l)}) &= \arg \min_{X, T, u^{(l)}} \sum_{k=1}^{N} \frac{1}{2} |z^{(l)}(k)| - \left\{ \left( f_k, HAv^{(l)} \right) - \left( X, \tilde{d}_k e_k^H \right) \right\}^2 \\
&\phantom{=} + \frac{\lambda}{2N} \text{Tr}(\text{Toep}(\mathbf{u})) + \frac{\lambda \text{Tr}(T)}{2} + \gamma \|v^{(l)}\|_1,
\end{align*}$$

where Toep(·) denotes the Toeplitz matrix whose first column is the input vector. The above problem is convex, and can be solved by using a convex solver. We name the algorithm based on solving (21) as the CS algorithm based on atomic norm (CS-AN). Similar to the CS-L1 algorithm, CS-AN iterates with (22) and (15) until $\hat{b}^{(l-1)} = \hat{b}^{(l)}$.

From now on, $\hat{X}$ denotes the estimate of $X$ when the algorithm terminates.

Solving (22) does not directly provide the estimates of the delays of the active radars, $\{\tau_j'\}_{j=1}^J$. Notice however that each row of $\hat{X}$ is a linear combination of several sinusoids, in that, denoting $\hat{X}_{k,1:N} \in \mathbb{C}^1 \times N$ the $k$-th row of $\hat{X}$, we have $\hat{X}_{k,1:N} = \sum_{j=1}^{J} \hat{c}_j h_j(k) a(\tau_j')^H$. Hence, $\hat{\beta}_j(k) = \hat{c}_j h_j(k)$ and $\hat{\tau}_j'$ can be obtained by MUSIC [33] or prony’s method [34] with $\hat{X}_{k,1:N}$ as input. Denoting MUSIC(·) the operation of the MUSIC algorithm, it outputs $\hat{J}_k$ components with different amplitudes and delays\(^1\), i.e.,

$$T_k = \{ \hat{\beta}_j(k), \hat{\tau}_j'(k) \}_{j=1,2,...,\hat{J}_k} = \text{MUSIC}(\hat{X}_{k,1:N})$$

\(^1\)Here $\hat{J}_k$ is not necessarily the same as $\hat{J}$ because $\beta_j(k)$ can be zero while $\beta_j(l) \neq 0$ for some $l \neq k$. For example, the radar has $K$ candidate waveforms, and select one for transmission. In such case, $h_j$ is a $K \times 1$ vector having $K-1$ zero elements and one element with magnitude 1. As a result, the number of output delays should satisfy $\hat{J}_k \leq \hat{J}$.}

\[^1\]
for \( k = 1, 2, \ldots, K \). An association is then needed to combine \( \hat{\tau}_j(k) \)s of different \( k \) and estimate the delays, \( \hat{\tau}_j \) say, of the radar signals. In practice, the calculation of \( \hat{\tau}_j \) may not be accurate due to the computational error. Hence, if the estimated delay for different \( k \) is smaller than a small threshold \( \delta \), then they are regarded as the signal from the same radar, and the corresponding \( \hat{\tau}_j(k) \)s are combined. For clarity, we summarize the association process in Algorithm 1. \( \mathcal{S} = \{ (\hat{\tau}_j(k), \hat{c}_j \hat{h}_j) \}_{j=1,2,\ldots,J} \) is the set of the estimated radar delays and waveform parameters. In the algorithm, for \( \{ \hat{\beta}_l(k), \hat{\tau}_l(k) \} \), if there exists \( (\hat{\tau}_m, \hat{c}_m \hat{h}_m) \in \mathcal{S} \) such that \( |\hat{\tau}_m - \hat{\tau}_l(k)| \leq \delta \) and \( \hat{c}_m \hat{h}_m(k) = 0 \), then both components belong to the same radar and the radar delay is updated by the weighted summation of \( \hat{\tau}_l(k) \) and \( \hat{\tau}_m \). Otherwise, an additional radar with the delay and amplitude information is added to the set \( \mathcal{S} \). Notice that, once again, the inherent scaling ambiguity between each \( \hat{c}_j \) and the corresponding \( \hat{h}_j \). Hence, we only estimate \( c_j g_j \) via (17) with \( \hat{c}_j \hat{h}_j \) replaced by \( \hat{c}_j \hat{h}_j \).

**Algorithm 1** Radar delay and path gain estimation

Input \( T_k \) for \( k = 1, 2, \ldots, K, \delta \).

1. \( \hat{J} = 0, \mathcal{S} = \{ \} \).

For \( k = 1, 2, \ldots, K \)

   For \( l = 1, 2, \ldots, \hat{J} \)

   If there exists \( (\hat{\tau}_m, \hat{c}_m \hat{h}_m) \in \mathcal{S} \) such that \( |\hat{\tau}_m - \hat{\tau}_l(k)| \leq \delta \) and \( \hat{c}_m \hat{h}_m(k) = 0 \)

      \( \hat{\tau}_m = \frac{\sum_{j=1}^{K} |c_j \hat{h}_m(j) + \hat{\beta}_l(k) \hat{\tau}_l(k)|}{\sum_{j=1}^{K} |c_j \hat{h}_m(j) + \hat{\beta}_l(k)|} \).

      \( \hat{c}_m \hat{h}_m(k) = \hat{\beta}_l(k) \).

   Else

      \( \mathcal{S} = \{ \mathcal{S}, (\hat{\tau}_m, \hat{c}_m \hat{h}_m) \} \) where \( \hat{c}_j \hat{h}_j(k) = \hat{\beta}_l(k) \) and \( \hat{c}_j \hat{h}_j(m) = 0 \) for \( m \neq k \).

      \( \hat{J} = \hat{J} + 1 \).

   End If

End For

End For

Return \( \mathcal{S}, \hat{J} \).

An alternative approach to delay estimation consists in solving the dual problem of (21), which is given by

\[
\begin{align*}
\max_{\nu} & \quad \langle \nu, z^{(l)} \rangle_{\mathbb{R}} - \frac{1}{2} \|\nu\|_2^2, \\
\text{s.t.} & \quad \|\mathcal{D}(\nu)\|_A^* \leq \lambda, \\
& \quad \left\| \sum_{k=1}^{N} \nu_k A^H H_H f_k \right\|_\infty \leq \gamma,
\end{align*}
\]

where \( \nu \in \mathbb{C}^{N \times 1} \) is the dual variable, and \( \|\mathcal{D}(\nu)\|_A^* = \sup_{\|X\|_A \leq 1} \langle \mathcal{D}(\nu), X \rangle_{\mathbb{R}} \) is the dual norm with \( \mathcal{D}(\nu) = \).
\[ \sum_n \nu(n) \tilde{d}_n e_n^H \in \mathbb{C}^{K \times N}, \quad \langle D(\nu), X \rangle_{\mathbb{R}} = \text{Re}(\text{Tr}(X^H D(\nu))). \]

Following the derivation in [32], [35], the following lemma can be obtained:

**Lemma 1:** Suppose \( \hat{X} = \sum_{j=1}^{J} \hat{\varepsilon}_j \hat{h}_j a(\hat{\tau}'_j)^H \) and \( \hat{\nu} \) are the primal solution, then the dual polynomial \( q(\tau') = D(\nu)a(\tau') \) satisfies

\[
q(\tau') = \lambda \frac{\hat{\varepsilon}_j}{|\hat{\varepsilon}_j|} \hat{h}_j, \quad j = 1, 2, ..., \hat{J},
\]

where \( \hat{J} \) is the number of estimated delays, \((\cdot)_j\) denotes the \( j \)-th element of the input vector.

Based on (25), the delays of the active radars can be obtained by identifying points where the dual polynomial has modulus \( \lambda \), i.e., \( ||q(\tau')||_2 = \lambda \). Moreover, the dual solution provides another way to detect the mistaken demodulation: in places where mistaken demodulation occurs, the magnitude of \( \sum_{k=1}^{N} \nu_k A^H H^H f_k \) equals \( \gamma \).

**C. Example**

The previous discussion highlights that the inherent coupling of interference estimation and data demodulation has a deep impact on the performance of the proposed approach. In order to illustrate further this point and to highlight the rationale behind the AN criterion, we consider a simple scenario wherein an OFDM system with \( N = 64 \) and \( FHA = I_N \) is to co-exist with \( J = 2 \) radar systems with \( K = 3 \). We focus on the results of the first implementation and of the first iteration of (12) and (22) in order to assess the ability of the CS-L1 and the CS-AN algorithms in detecting, identifying and ranging the active transmitters. The simulations have been performed by generating data according to (10), while the \( h_j \)'s are uniformly generated with \( ||h_j||_2 = 1 \) for \( j = 1, 2, ..., J \). Due to the coupling of interference estimation and data demodulation, the initial symbol error rate (SER) plays a key role, and we assume the two initial values of 0.1 and 0.3. In the example, the communication uses binary phase-shift keying (BPSK) and the wrong symbols are randomly placed in \( \hat{b}^{(0)} \) based on the SER. The grid parameter of the CS algorithm has been set as \( \tilde{J} = 4N \). For both algorithms, the weighting parameters have been optimized so that the mean-squared-error (MSE) of the estimation is minimized.

Fig. 1 and Fig. 2 give the results when the SER is 0.1 and 0.3, respectively. The basis mismatch inherent in the CS-L1 algorithm returns more false alarms already in the initial iteration, but in both cases the number of false detections at \( t = 0 \) is definitely unacceptable. Not surprisingly, the first iteration dramatically cleans the environment, showing that for both algorithms the interference picture becomes much clearer and much closer to the reality: the CS-AN algorithm, however, is definitely superior to the CS-L1 algorithm under both considered values of the SER, which confirms the importance of a proper basis matching in the interference identification-estimation phase.
Fig. 1. Comparison between (a)(b) CS-L1 algorithm and (c)(d) CS-AN algorithm.

Fig. 2. Comparison between (a)(b) CS-L1 algorithm and (c)(d) CS-AN algorithm.
IV. Fast Algorithm Based on Non-convex Factorization

The problem in (21), which can be solved using off-the-shelf solvers such as SeDuMi [36] and SDPT3 [37] once the formulation (22) is adopted. Unfortunately, these solvers tend to be slow, especially for large-dimensional problems. A possible alternative, in these circumstances, could be the Alternating Direction Method of Multipliers (ADMM) [38], which requires an eigenvalue decomposition at each iteration, still entailing a computational complexity $O((N + K)^3)$, again posing a complexity issue for large-scale problems.

In this section, we derive a fast method for solving this SDP via the non-convex factorization proposed by Burer and Monteiro [39]. For brevity, the superscript $l$ of the variables $z$ and $v$ are omitted in what follows. Defining

$$Z = \begin{bmatrix} U & X^H \\ X & T \end{bmatrix} \in \mathbb{C}^{(N+K)\times(N+K)},$$

we rewrite (22) as

$$\begin{array}{ll}
\min_{Z,v} & \sum_{k=1}^N \frac{1}{2} |z(k) - \langle f_k, HA_v \rangle - \langle X, \bar{d}_k e_k^H \rangle|^2 + \frac{\lambda}{2N} \text{Tr}(U) + \frac{\lambda \text{Tr}(T)}{2} + \gamma \|v\|_1,
\text{s.t.} & \mathcal{P}_{\text{Toep}}(U) = U, Z \succeq 0,
\end{array}$$

where $\mathcal{P}_{\text{Toep}}(\cdot)$ denotes the projection of the input matrix onto a Toeplitz matrix, i.e.,

$$\mathcal{P}_{\text{Toep}}(U) = \text{Toep}(\mathcal{G}(U))$$

with $\mathcal{G}(U)$ outputing an $N$-dimensional vector whose $(k+1)$-th element is the mean value of the $k$-th subdiagonal elements of the input matrix.

As will be shown later, the algorithm can be accelerated if the solution to (28) is of low-rank. Note that the number of active radars is usually small in practice, which leads to small $\hat{J}$. We show in the following lemma that under some conditions the rank of the solution to $Z$ equals to $\hat{J}$, which enables us to accelerate the algorithm through non-convex factorization. The proof is given in Appendix A.

**Lemma 2:** Suppose $\hat{X} = \sum_{j=1}^{\hat{J}} \hat{e}_j \hat{H}_j \mathbf{a}(\hat{\tau}_j)^H$ is the solution to (21). If $N \geq 257$ and $\Delta_{\hat{\tau}} \geq \frac{1}{N-1}$, where

$$\Delta_{\hat{\tau}} = \inf_{\hat{\tau}_1', \hat{\tau}_2' \in [0,1], \hat{\tau}_1' \neq \hat{\tau}_2'} \min \left\{ |\hat{\tau}_1' - \hat{\tau}_2'|, 1 - |\hat{\tau}_1' - \hat{\tau}_2'| \right\},$$

then there exists $\hat{Z}$ as a solution to (28) that satisfies $\text{rank}(\hat{Z}) = \hat{J}$.

If an upper bound on the number of active radars, say $\bar{J}$, is known in advance, then we can introduce the extra constraint $\text{rank}(Z) \leq \bar{J}$ in (28), whereby restricting the search space to matrices of rank at most $\bar{J}$ does not change

\[ \text{The condition } N \geq 257 \text{ is more like a technical requirement but not an obstacle in practice.} \]
the globally optimal value. Additionally, we relax the constraint $P_{\text{Toep}}(U) = U$, replacing it with the penalty term $\frac{\theta}{2} \| P_{\text{Toep}}(U) - U \|^2_F$, so that the problem is recast as

$$
\min_{Z,v} \quad \zeta(Z,v) = \sum_{k=0}^{N-1} \frac{1}{2} \left| z(k) - \langle f_k, HA v \rangle - \langle X, d_k e_k^H \rangle \right|^2 + \frac{\lambda}{2N} \text{Tr}(U)
+ \frac{\lambda \text{Tr}(T)}{2} + \frac{\theta}{2} \| P_{\text{Toep}}(U) - U \|^2_F + \gamma \| v \|_1,
$$

s.t. $Z \succeq 0$, rank$(Z) \leq \tilde{J}$.

Setting $Z = VV^H$ where $V \in \mathbb{C}^{(N+K) \times \tilde{J}}$, then (31) becomes an unconstrained optimization of $\min_{V,v} \zeta(VV^H, v)$. Though this unconstrained problem is non-convex, its dimension is lower than that of the original problem in (28) and has no conic constraint, which leads to reduced computational complexity. A very effective means to undertake the desired minimization of $\zeta(VV^H, v)$ is to resort to a Conjugate Gradient (CG) algorithm [40], which is a fast first-order algorithm. The algorithm requires the objective function to be smooth, therefore we approximate $\| \cdot \|_1$ in $\zeta(VV^H, v)$ with the convex, differentiable function:

$$
\| v \|_1 \approx \psi_\mu(v) = \mu \sum_{m=1}^M \log \left( \frac{\exp(|u_m|/\mu) + \exp(-|u_m|/\mu)}{2} \right) = \mu \sum_{m=1}^M \log \cosh(|u_m|/\mu)
$$

where $\mu$ controls the smoothing level $^3$. Hence, the problem becomes

$$
\min_{V,v} \quad \tilde{\zeta}(VV^H, v),
$$

where

$$
\tilde{\zeta}(Z,v) = \sum_{k=0}^{N-1} \frac{1}{2} \left| z(k) - \langle f_k, HA v \rangle - \langle X, d_k e_k^H \rangle \right|^2 + \frac{\lambda}{2N} \text{Tr}(U)
+ \frac{\lambda \text{Tr}(T)}{2} + \frac{\theta}{2} \| P_{\text{Toep}}(U) - U \|^2_F + \gamma \psi_\mu(v).
$$

The minimization problem (33) can be effectively solved using the CG algorithm, undertaking the iteration

$$
V^{(k)} = V^{(k-1)} + \varsigma_k P_k,
$$

$$
v^{(k)} = v^{(k-1)} + \varsigma_k p_k,
$$

where $\varsigma_k$ is the step size, $p_k$ and $P_k$ are the search directions at step $k$, evaluated as the weighted sum of the gradient at present iteration and the search direction used at the previous one. Specifically, if $\nabla_v \tilde{\zeta}(V^{(k-1)}(V^{(k-1)})^H, v^{(k-1)})$ and $\nabla_v \tilde{\zeta}(V^{(k-1)}(V^{(k-1)})^H, v^{(k-1)})$ denote the gradients of $\tilde{\zeta}(VV^H, v)$ at the $k$-th iteration, then we have

$$
P_k = -\nabla_v \tilde{\zeta}(V^{(k-1)}(V^{(k-1)})^H, v^{(k-1)}) + \omega_k P_{k-1},
$$

$$
p_k = -\nabla_v \tilde{\zeta}(V^{(k-1)}(V^{(k-1)})^H, v^{(k-1)}) + \omega_k p_{k-1},
$$

$^3$Similar approach has also been used in [41] for the smoothing of the $\ell_1$-norm in the objective function. In fact, there is nothing special about this choice and we believe that some other twice continuously differentiable approximation to $\| \cdot \|_1$ would work and yield qualitatively similar results.
where
\[
\omega_k = \frac{\langle \nabla \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)}) \rangle}{\langle p_{k-1}, Q_{k-1} \rangle} + \frac{\langle \nabla u \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)}, q_{k-1} \rangle}{\langle p_{k-1}, q_{k-1} \rangle},
\]
(39) with \(Q_k = \nabla \tilde{\phi} (V^{(k)} (V^{(k)})^H, v^{(k)}) - \nabla \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)})\) and \(q_k = \nabla \tilde{\phi} (V^{(k)} (V^{(k)})^H, v^{(k)}) - \nabla v \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)})\). The expressions of the gradients are derived in Appendix B.

For clarity, we summarize the proposed non-convex solver in Algorithm 2. To guarantee that the objective function does not increase with \(k\), the CG utilizes the Armijo line search [42] (line 6 of Algorithm 2), so that the algorithm converges to a stationary point of the surrogate problem, namely, the point where the smoothed objective function (34) has vanishing gradient.

**Algorithm 2 Conjugate gradient algorithm**

1. Initialize \(V^{(0)}, v^{(0)}, k = 0\).

Do
2. \(k = k + 1\).
3. Compute \(\nabla \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)})\) and \(\nabla u \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)})\).
4. \(P_k = -\nabla \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)}), p_k = -\nabla v \tilde{\phi} (V^{(k-1)} (V^{(k-1)})^H, v^{(k-1)}),\)
5. Calculate \(P_k\) and \(p_k\) according to (37) and (38) where \(\omega_k\) is obtained by (39).
6. Update \(V^{(k)}\) and \(v^{(k)}\) according to (35) and (36) where \(\zeta_k\) is obtained via Armijo line search.
While \(\| \nabla \tilde{\phi} (V^{(k)} (V^{(k)})^H) \|_2 \leq \epsilon\).
7. Obtain \(\hat{X}\) according to (27) with \(\hat{Z} = V^{(k)} (V^{(k)})^H\).

Return \(\hat{v} = v^{(k)}, \hat{X}\).

The computational complexity of the proposed algorithm at each iteration is mainly determined by the calculation of \(V V^H\), whose complexity is \(O((N + K)^2 \tilde{J})\). As \(\tilde{J}\) is much smaller than \((N + K)\), the complexity per-iteration is much smaller than that of a classical eigenvalue decomposition, whereby, for large dimensional problems, the proposed non-convex approach can be faster than those based on the first-order methods such as ADMM and projected gradient descent. We illustrate this fact through a simulation example, whose results are reported in Fig. 3. The dimension of the signal is \(N = 256\), while the other parameters are \(J = 2, L = 5\) and \(K = 5\). The variance of the noise is \(\sigma_w^2 = 0.01\). The non-convex solver is implemented by solving (33) with the CG algorithm. We compare the MSE of the proposed algorithm with that given by solving (22) with CVX [43] and ADMM solver.
As can be seen from Fig. 3, the result of the proposed algorithm is close to the solution given by the CVX after 1000 iterations: since the proposed algorithm is also much faster than the ADMM, it appears much more suitable for real-time implementation.

V. SIMULATION RESULTS

A. Simulation Setup

In order to demonstrate the performance of the proposed algorithms, we simulate a scenario with multiple radars and one communication receiver. The communication system uses OFDM signal with frequency spacing between adjacent subcarrier of 10kHz, $N = 128$ subcarrier frequencies and total bandwidth 1.28MHz. The symbol length is $T = 100\mu s$ and Quadrature Phase-Shift Keying (QPSK) modulation is used. The channel matrix is generated as $F^H \Gamma F$ where $\Gamma$ is an identity matrix whose diagonal elements are complex random variables whose modulus follows a Ricean distribution: as a consequence, we model the ray impinging on the receiver as the superposition of a non-fading component, with power $\rho^2$ say, and a fluctuating (zero-mean) component, with mean square value $\sigma_h^2$, so that the Rician factor is $\frac{\rho^2}{\sigma_h^2}$ [44, p. 79]. As to the active radars, they are modeled as point sources in our simulations. For simplicity, the amplitudes of the path gains $c_j$ are generated with fixed magnitude and random phase, and the magnitude is controlled by the power of the paths. We define the signal-to-interference ratio (SIR) of the communication RX as the power ratio of the communication signal and the radar interference. Specifically, the SNR and SIR are defined as

\[
\text{SNR} = \frac{(\rho^2 + \sigma_h^2)\epsilon_b}{\sigma_w^2},
\]
\[
\text{SIR} = \frac{N\epsilon_b(\rho^2 + \sigma_h^2)}{\left\| \sum_{j=1}^J c_j g_j \right\|^2_2},
\]

Fig. 3. Convergence behavior of the proposed non-convex solver. The non-convex solver takes 35.6 seconds with 1000 iterations. The ADMM solver takes 108.9 seconds with 1000 iterations. The CVX solver takes 237.7 seconds.
where $\varepsilon_b$ is the power of each symbol in $b$. Notice that, for fixed SNR, lower values of $\rho$ account for larger fluctuations, up to the limit of $\rho = 0$, $\text{SNR} = \frac{\sigma_w^2}{\sigma_b^2}$, which corresponds to a Rayleigh-fading channel. In what follows we use $\frac{\rho}{\sigma_b} = 3$, unless otherwise specified and then study the impact of the channel fluctuation on the achievable performance.

In keeping with the model of Section II, $\bar{g}_j$ lives in a low-dimensional subspace spanned by the columns of $\bar{D}$ matrix, i.e., $\bar{g}_j = \bar{D}h_j$. We use the setting that $\bar{D} = FD$ where $D = [d_1, d_2, ..., d_K] \in \mathbb{C}^{N \times K}$ with $d_k \in \mathbb{C}^{N \times 1}$ and $K = 5$. In our simulation, the radars use pulse waveforms and each pulse uses Gaussian random code with length $N' = 32$. Specifically, $d_k$ satisfies $d_k(m) \sim \mathcal{CN}(0, 1/N')$ for $1 \leq m \leq N'$ and $d_k(m) = 0$ for $N' + 1 \leq m \leq N$. The columns of $\bar{D}$ can be obtained by taking the DFT of $d_k$. Fig. 4 gives an example of the signal at the communication RX when $J = 1$. The SIR of the example is set as 0dB: The figure clearly demonstrates how dramatic the effect of even a single co-existing radar can be.

Some other parameters of the simulations are given as follows.

1. In our simulation, we set $\tau_j > \tau_C$, while $\tau_j - \tau_C$ is randomly generated between $10 \mu s$ and $70 \mu s$ for $j = 1, 2, ..., J$.

2. For comparison purposes, we also show the performance of $\hat{b} = \Psi(\bar{r})$, i.e. of a demodulator operating on the raw data: this is named “original demodulation”, since its result is the initial point to be provided to the iterative algorithms.

3. For the proposed CS-L1 algorithm, we discretize the continuous parameter space to a finite set of grid points of cardinality $\bar{J} = 4N$. The weighting parameters for CS-L1 are $\bar{\lambda} = \sigma_w \kappa \sqrt{2 \log(\bar{J}K)}$ and $\bar{\gamma} = 0.5 \sigma_w \sqrt{2 \log(\bar{J}K)}$, where $\kappa$ is the average norm of the column vectors of matrix $\Upsilon$. 

---

Fig. 4. Plots of radar interference, communication signal and the received signal of communication RX in (a) time domain and (b) frequency domain. In (a), the magnitude of the signal is plotted against time. In (b), the real part of the signal is plotted versus frequency sample.
4. For the proposed CS-AN algorithm, the weighting parameters are set as $\lambda = \sigma_w \sqrt{KN \log(KN)}$ and $\gamma = \frac{\lambda}{\sqrt{N}}$. We use the proposed non-convex algorithm to solve (33) where $\bar{J} = 10$, $\mu = 0.01$ and $\varrho = 5$. The algorithm stops as the norm of the gradient is smaller than 0.01.

5. We evaluate the root-mean-squared-error (RMSE) of the radar delay estimation and the relative mean-squared-error (MSE) of the estimated waveform for the proposed CS-L1 and CS-AN algorithms. Specifically, the delay RMSE and relative waveform MSE are calculated as

$$\text{RMSE}_\tau = \sqrt{\frac{1}{MCJ} \sum_{m=1}^{MC} \sum_{j=1}^{J} (\tau_j^{(m)} - \hat{\tau}_j^{(m)})^2},$$

$$\text{MSE}_{cg} = \frac{1}{MCJ} \sum_{m=1}^{MC} \sum_{j=1}^{J} \left\| \frac{c_j^{(m)} g_j^{(m)} - \hat{c}_j^{(m)} \hat{g}_j^{(m)}}{\left\| c_j^{(m)} g_j^{(m)} \right\|^2} \right\|^2,$$

respectively, where $MC$ is the number of runs; $\tau_j^{(m)}$, $c_j^{(m)}$ and $g_j^{(m)}$ are the delay, path gain and waveform of the $j$-th radar in the $m$-th run, respectively, while $\hat{\tau}_j^{(m)}$ and $\hat{c}_j^{(m)} \hat{g}_j^{(m)}$ are the respective estimates.

### B. Performance

We firstly compare the SER performance of the proposed algorithms. The number of active radars is set as $J = 2$ in the simulation. As can be seen from Fig. 5, where the effect of the SIR is studied: both the proposed CS-L1 and CS-AN algorithms provide better SER performance than the original demodulation algorithm. The CS-AN algorithm also outperforms the CS-L1 algorithm in all situations. Notice that, as the SNR is 8dB, the performance gain brought by both CS algorithms is limited: this is obviously due to the fact that, under fading channel, the performance of even an interference-free OFDM/QPSK would be poor, and the coupling between data demodulation and interference removal explains the poor performance at low SIR. For example, the considered Ricean channel with Rice factor 9 is approximately equivalent to a Nakagami-m fading with parameter $m \simeq 5.3$ [44, p.79], whereby the error probability for an isolated OFDM/QPSK at SNR = 8 dB would be slightly larger than $10^{-2}$, a value which is approximately restored as the SIR becomes increasingly large. As the SNR increases to 18dB, the CS algorithms provides significant improvement over the original demodulation, but is significantly outperformed by the CS-AN especially in the low SIR region. This is due to the fact that, in the low SIR region, basis mismatch prevents correcting the demodulation errors in the first iteration $\hat{b}^{(0)}$ through a CS-L1 algorithm, while the CS-AN algorithm, much more accurate in detecting and ranging the interference sources, allows a much more effective error correction. Needless to say, both algorithms restore the original OFDM/QPSK performance for increasingly large SIR in a much faster way than the original demodulation: it might thus be inferred that, even at SIR as low as -5dB, the SER achieved through the proposed CS-AN algorithm - in the order of $10^{-2}$ - would in principle

---

4 We use the relative MSE rather than the RMSE to evaluate the accuracy because it reflects the loss in energy.
allow communication to be sustained once forward error correction (FEC) decoding [45] is undertaken, while no communication could take place if either of the other two algorithms were adopted.

We then verify the accuracy of radar waveform and delay estimation of the proposed CS-L1 and CS-AN algorithms. Note that the received signal contains both communication and radar signal, and the estimation accuracy depends on both the power of the radar signals to be estimated and the performance of the demodulator: as a consequence, the estimation accuracy may be not necessarily increasing with the SIR, especially for “intermediate” SIR values where radar signals are not strong enough to prevail on the communication signal, but still produce significant demodulation error: this intuition is confirmed by the plots in Fig. 6, where the relative MSE of waveform estimation and the RMSE of the delay estimation is represented as a function of the SIR. As expected, the CS-AN algorithm provides much better accuracy than the CS-L1 algorithm. Notice also the apparently contradictory effects of the SNR. Indeed, large SNR’s guarantee good demodulation performance, with a beneficial effect on the interference estimation due to the coupling, but also represent larger amount of disturbance the estimator of the radar waveforms is confronted to: under this point of view, the advantage of the CS-AN over the CS-L1 is visible, since the latter seems to take much greater advantage of the more reliable demodulation process granted by larger values of the SNR. On the other hand, the results also demonstrate that, as the SNR is low, the SIR has paradoxically a detrimental effect on the estimation accuracy: this is because, for large SIR and low SNR, the demodulation process is completely unreliable, which explains the trends of the curves in Fig. 6.

The effects of $J$ and $K$ are elicited in Fig. 7. The simulations are run with an SNR of 18dB and an SIR of 0dB. In Fig. 7a, we set $\rho/\sigma_h = 3$ and plot the SER against the number of active radars: As $J$ increases, the sparsity of the problem is reduced, and the sources of interference - with the respective unknown parameters to be estimated - increase, which obviously results in a visible performance degradation for both algorithms, and not
Fig. 6. Accuracy of (a) the waveform estimation and (b) radar delay estimation for different SIRs.

Fig. 7. Plots of SER against (a) $J$, and (b) $K$.

Fig. 8. Plots of SER against $\rho/\sigma_h$. 
even an CS-AN algorithm is effective if \( J \geq 5 \). In Fig. 7b, the number of active radars is set as \( J = 2 \) and we examine the SER behavior for varying \( K \), i.e. for increasing dictionary size. Although a performance degradation is evident, the robustness of the CS-AN algorithm to \( K \) is a definitely appealing advantage over the other two algorithms. Finally, we investigate the effect of the channel model on the performance in Fig. 8, assuming \( J = 2 \), SIR = 0dB and SNR = 18dB: on the horizontal axis we represent the Ricean factor, whereby decreasing values of \( \rho/\sigma_h \) represent weaker and weaker direct paths, up to the case \( \rho = 0 \) where no direct path is present. Notice that only the CS-AN algorithm guarantees, in such a severe scenario, an acceptable performance and is able to restore an error probability in the order of \( 10^{-2} \) as \( \rho/\sigma_h > 2 \).

VI. CONCLUSIONS

In this paper, we have proposed two algorithms for joint waveform estimation and demodulation in the overlaid communication and radar systems. One of them is based on the on-grid compressed sensing (CS) technique and uses \( \ell_1 \)-norm to exploit the sparsity of the radar signal components and the sparsity of the demodulation error. The other one is a CS-based algorithm using both the atomic norm and the \( \ell_1 \)-norm to exploit the sparsity of the radar signal components and the sparsity of the demodulation error, respectively. We have derived an fast algorithm to compute the solution to the formulated CS-AN problem. Simulation results show that the proposed algorithms provide better SER compared to the original demodulation.

APPENDIX

A. Proof of Lemma 2

Suppose the solution to (21) is \( \hat{X} = \sum_{j=1}^{J} \hat{c}_j \hat{h}_j a(\hat{\tau}_j) H \), the following lemma states the condition of the unique atomic decomposition:

Lemma 3: [21] \( \hat{X} \) is the unique atomic decomposition satisfying that \( \| \hat{X} \|_A = \sum_{j=1}^{J} |\hat{c}_j| \) if \( N \geq 257 \) and \( \Delta_{\tau'} \geq \frac{1}{(N-1)/4} \).

The lemma above gives the value of \( \| \hat{X} \|_A \). Since (21) and (28) are equivalent, we obtain

\[
\| \hat{X} \|_A = \inf_{\tilde{U}, \tilde{T}} \left\{ \frac{1}{2N} \text{Tr}(\tilde{U}) + \frac{1}{2} \text{Tr}(\tilde{T}), \text{s.t.,} \mathcal{P}_{\text{Toep}}(\tilde{U}) = \tilde{U}, \tilde{Z} \succeq 0 \right\},
\]

where the relation of \( \tilde{Z}, \tilde{U} \) and \( \tilde{T} \) are given in (27). Hence, \( \tilde{Z} \) is the solution to (28) once the equality holds. Then we need to prove that there exist Toeplitz matrix \( \tilde{U} \) and matrix \( \tilde{T} \) such that \( \frac{1}{2N} \text{Tr}(\tilde{U}) + \frac{\text{Tr}(\tilde{T})}{2} = \| \hat{X} \|_A \) with rank(\( \tilde{Z} \)) = \( \hat{J} \) and \( \tilde{Z} \succeq 0 \). Let \( \tilde{u} = \sum_{j=1}^{\hat{J}} \hat{c}_j a(\hat{\tau}_j)' H \). Following the Caratheodory-Toeplitz lemma [25], [46], we have

\[
\hat{U} = \text{Toep}(\tilde{u}) = \sum_{j=1}^{\hat{J}} |\hat{c}_j| a(\hat{\tau}_j)' a(\hat{\tau}_j)' H.
\]
In such case, the matrix
\[
\tilde{Z} = \sum_{j=1}^{j} |\hat{c}_j| \begin{bmatrix}
a(\hat{\tau}'_j) \\
\hat{h}_j
\end{bmatrix} \begin{bmatrix}
a(\hat{\tau}'_j)^H & \hat{h}_j^H
\end{bmatrix} \succeq 0
\] (46)
is rank-$\hat{J}$. Note that \(\frac{1}{2N} \text{Tr} \left( \hat{U} \right) + \frac{\text{Tr}(\hat{T})}{2} = \sum_{j=1}^{j} |\hat{c}_j| = ||\hat{X}||_A, \hat{Z} \) is the solution to (28), which accomplishes the proof.

B. The calculation of \(\nabla_v \tilde{\zeta}(VV^H, v)\) and \(\nabla_V \tilde{\zeta}(VV^H, v)\)

The gradient \(\nabla_v \tilde{\zeta}(VV^H, v)\) can be directly calculated as
\[
\nabla_v \tilde{\zeta}(VV^H, v) = \sum_{k=1}^{N} \left( \langle f_k, HA \rangle + \langle X, \tilde{d}_ke_k^H \rangle - z(k) \right) A^H H^H f_k + \gamma \nabla_v \psi_\mu(v),
\] (47)
where the \(m\)-th element of \(\nabla_v \psi_\mu(v) \in \mathbb{C}^{M \times 1}\) is
\[
\nabla_v \psi_\mu(v) = \frac{\sinh(v_m/\mu)}{\cosh(v_m/\mu)} v_m.
\] (48)

Then we derive the gradient with respect to \(V\). Following the chain rule, we have
\[
\nabla_V \tilde{\zeta}(VV^H, v) = \left[ \nabla_Z \tilde{\zeta}(Z, v) \right]_{Z=VV^H} V,
\] (49)
so the problem becomes calculating the gradient \(\nabla_Z \tilde{\zeta}(Z, v)\). For the convenience of our calculation, \(\tilde{\zeta}(Z, v)\) is rewritten as
\[
\tilde{\zeta}(Z, v) = \frac{\lambda \text{Tr}(U)}{2N} + \frac{\lambda \text{Tr}(T)}{2} + \sum_{k=0}^{N-1} \frac{1}{2} \left| z(k) - \langle f_k, HA \rangle - \langle X, \tilde{d}_ke_k^H \rangle \right|^2 + \gamma \psi_\mu(v)
\]
\[
+ \frac{\rho}{2} \left[ \sum_{k=1}^{N} \left( m_k^H m_k - \frac{1}{N-k} m_k^H I_{N-|k|} m_k \right) \phi_m(m_k) \right],
\] (50)
where \(U, X, T\) are submatrices of \(Z\) whose structure is given in (27), \(m_k \in \mathbb{C}^{(N-|k|) \times 1}\) is the \(k\)-th subdiagonal elements of \(U\), \(I_k = [1, 1, ..., 1]^T\) is an \(k\) dimensional all one vector. After manipulation, \(\tilde{\zeta}(Z, v)\) can be rewritten in a quadratic form:
\[
\tilde{\zeta}(Z, v) = \langle Z, Q(Z) \rangle / 2 + \langle C, Z \rangle + \tilde{\zeta}(v),
\] (51)
where \(\tilde{\zeta}(v)\) is a function that depends on \(v\); \(C\) and \(Q(Z)\) can be respectively computed by
\[
C = \frac{1}{2} \begin{bmatrix}
\frac{\lambda}{N} E_N & -\bar{Y}^H \\
-\bar{Y} & \lambda E_K
\end{bmatrix},
\] (52)
\[
Q(Z) = \begin{bmatrix}
\mathcal{E}(U) & \sum_{k=1}^{N} \langle e_k^H \tilde{d}_k, X \rangle e_k d_k^H \\
\sum_{k=1}^{N} \langle d_k e_k^H, X \rangle & 0_{K,K}
\end{bmatrix}.
\] (53)
Here $0_{K,K} \in \mathbb{C}^{K \times K}$ is a zero matrix; $E_N$ is an identity matrix with dimension $N$;

$$
\bar{Y} = \sum_{k=1}^{N} e_k \bar{a}^H_k(z(k) - \langle f_k, HAv \rangle), \\
(54)
$$

$$
\Xi(\bar{Z}) = \frac{P}{2} \sum_{k=1-N}^{N-1} \text{diag}(\nabla \phi_k(m_k), k), \\
(55)
$$

with

$$
\nabla \phi_k(m_k) = 2 \left( m_k - \frac{1}{N - |k|} I_{N-|k|} I_{N-|k|}^H m_k \right), \\
(56)
$$

and $\text{diag}(m, k)$ outputs an $N \times N$ matrix whose $k$-th sub-diagonal is the input vector $m$, and the rest of the elements are zero. The derivative of $\tilde{\zeta}(Z, v)$ is

$$
\nabla_Z \tilde{\zeta}(Z, v) = Q(Z) + C. \\
(57)
$$

Plugging (57) into (49), $\nabla_V \tilde{\zeta}(VV^H, v)$ can be obtained.

REFERENCES

[1] A. R. Chiriyath, B. Paul, G. M. Jacyna, and D. W. Bliss, “Inner bounds on performance of radar and communications co-existence,” IEEE Transactions on Signal Processing, vol. 64, no. 2, pp. 464–474, 2016.
[2] H. Griffiths and S. Blunt, “T09-Spectrum engineering and waveform diversity,” in Radar Conference. IEEE, 2014, pp. 36–36.
[3] H. Griffiths, L. Cohen, S. Watts, E. Mokole, C. Baker, M. Wicks, and S. Blunt, “Radar spectrum engineering and management: technical and regulatory issues,” Proceedings of the IEEE, vol. 103, no. 1, pp. 85–102, 2015.
[4] F. Hessar and S. Roy, “Spectrum sharing between a surveillance radar and secondary Wi-Fi networks,” IEEE Transactions on Aerospace and Electronic Systems, vol. 52, no. 3, pp. 1434 – 1448, 2016.
[5] H. Deng and B. Himed, “Interference mitigation processing for spectrum-sharing between radar and wireless communications systems,” IEEE Transactions on Aerospace and Electronic Systems, vol. 49, no. 3, pp. 1911–1919, 2013.
[6] A. Babaei, W. H. Tranter, and T. Bose, “A practical precoding approach for radar/communications spectrum sharing,” in Cognitive Radio Oriented Wireless Networks (CROWNCOM), 2013 8th International Conference on. IEEE, 2013, pp. 13–18.
[7] A. Aubry, A. De Maio, M. Piezzo, M. M. Naghsh, M. Soltanalian, and P. Stoica, “Cognitive radar waveform design for spectral coexistence in signal-dependent interference,” in Radar Conference, 2014 IEEE. IEEE, 2014, pp. 0474–0478.
[8] A. Khawar, A. Abdel-Hadi, and T. C. Clancy, “Spectrum sharing between s-band radar and lte cellular system: A spatial approach,” in Dynamic Spectrum Access Networks (DYSSPAN), 2014 IEEE International Symposium on. IEEE, 2014, pp. 7–14.
[9] A. Aubry, A. De Maio, M. Piezzo, and A. Farina, “Radar waveform design in a spectrally crowded environment via nonconvex quadratic optimization,” IEEE Transactions on Aerospace and Electronic Systems, vol. 50, no. 2, pp. 1138–1152, 2014.
[10] K.-W. Huang, M. Biçä, U. Mitra, and V. Koivunen, “Radar waveform design in spectrum sharing environment: Coexistence and cognition,” in Radar Conference (RadarCon), 2015 IEEE. IEEE, 2015, pp. 1698–1703.
[11] L. Zheng, M. Lops, X. Wang, and E. Grossi, “Joint design of overlaid communication systems and pulsed radars,” arXiv preprint arXiv:1703.10184, 2017.
[12] S. Sodagari, A. Khawar, T. C. Clancy, and R. McGwier, “A projection based approach for radar and telecommunication systems coexistence,” in Global Communications Conference (GLOBECOM), 2012 IEEE. IEEE, 2012, pp. 5010–5014.
[13] B. Li, A. Petropulu, and W. Trappe, “Optimum co-design for spectrum sharing between matrix completion based MIMO radars and a MIMO communication system,” IEEE Transactions on Signal Processing, vol. 64, no. 17, pp. 4562–4575, 2016.
[14] B. Li and A. Petropulu, “MIMO radar and communication spectrum sharing with clutter mitigation,” in Radar Conference (RadarConf), 2016 IEEE. IEEE, 2016, pp. 1–6.
[15] A. Turlapaty and Y. Jin, “A joint design of transmit waveforms for radar and communications systems in coexistence,” in Radar Conference. IEEE, 2014, pp. 0315–0319.
[16] J. Liu, H. Li, and B. Himed, “Joint optimization of transmit and receive beamforming in active arrays,” IEEE Signal Processing Letters, vol. 21, no. 1, pp. 39–42, 2014.
[17] D. L. Donoho, “Compressed sensing,” IEEE Transactions on information theory, vol. 52, no. 4, pp. 1289–1306, 2006.
[18] L. Stanković, I. Orović, S. Stanković, and M. Amin, “Compressive sensing based separation of nonstationary and stationary signals overlapping in time-frequency,” IEEE Transactions on Signal Processing, vol. 61, no. 18, pp. 4562–4572, 2013.
[19] B. Jokanovic and M. Amin, “Reduced interference sparse time-frequency distributions for compressed observations,” IEEE Transactions on Signal Processing, vol. 63, no. 24, pp. 6698–6709, 2015.
[20] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, “Recovery of sparsely corrupted signals,” IEEE Transactions on Information Theory, vol. 58, no. 5, pp. 3115–3130, 2012.
[21] Z. Yang and L. Xie, “Exact joint sparse frequency recovery via optimization methods,” IEEE Transactions on Signal Processing, vol. 64, no. 19, pp. 5145–5157, 2014.
[22] 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.
[23] E. J. Candès and C. Fernandez-Granda, “Super-resolution from noisy data,” Journal of Fourier Analysis and Applications, vol. 19, no. 6, pp. 1229–1254, 2013.
[24] ——, “Towards a mathematical theory of super-resolution,” Communications on Pure and Applied Mathematics, vol. 67, no. 6, pp. 906–956, 2014.
[25] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, “Compressed sensing off the grid,” IEEE transactions on information theory, vol. 59, no. 11, pp. 7465–7490, 2013.
[26] B. N. Bhaskar, G. Tang, and B. Recht, “Atomic norm denoising with applications to line spectral estimation,” IEEE Transactions on Signal Processing, vol. 61, no. 23, pp. 5987–5999, 2013.
[27] Z. Tan, Y. C. Eldar, and A. Nehorai, “Direction of arrival estimation using co-prime arrays: A super resolution viewpoint,” IEEE Transactions on Signal Processing, vol. 62, no. 21, pp. 5565–5576, 2014.
[28] Y. Chi and Y. Chen, “Compressive two-dimensional harmonic retrieval via atomic norm minimization,” IEEE Transactions on Signal Processing, vol. 63, no. 4, pp. 1030–1042, 2015.
[29] G. Tang, P. Shah, B. N. Bhaskar, and B. Recht, “Robust line spectral estimation,” in 2014 48th Asilomar Conference on Signals, Systems and Computers. IEEE, 2014, pp. 301–305.
[30] D. Guo and S. Verdú, “Randomly spread cdma: Asymptotics via statistical physics,” IEEE Transactions on Information Theory, vol. 51, no. 6, pp. 1983–2010, 2005.
[31] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a structured union of subspaces,” IEEE Transactions on Information Theory, vol. 55, no. 11, pp. 5302–5316, 2009.
[32] D. Yang, G. Tang, and M. B. Wakin, “Super-resolution of complex exponentials from modulations with unknown waveforms,” IEEE Transactions on Information Theory, vol. 62, no. 10, pp. 5809–5830, 2016.
[33] A. Naha, A. K. Samanta, A. Routray, and A. K. Deb, “Determining autocorrelation matrix size and sampling frequency for MUSIC algorithm,” IEEE Signal Processing Letters, vol. 22, no. 8, pp. 1016–1020, 2015.
[34] P. Stoica and R. L. Moses, *Introduction to spectral analysis*. Prentice hall Upper Saddle River, 1997, vol. 1.

[35] C. Fernandez-Granda, G. Tang, X. Wang, and L. Zheng, “Demixing sines and spikes: Robust spectral super-resolution in the presence of outliers,” *arXiv preprint arXiv:1609.02247*, 2016.

[36] J. F. Sturm, “Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones,” *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[37] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, “SDPT3-a MATLAB software package for semidefinite programming, version 1.3,” *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 545–581, 1999.

[38] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[39] S. Burer and R. D. Monteiro, “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization,” *Mathematical Programming*, vol. 95, no. 2, pp. 329–357, 2003.

[40] Y.-H. Dai and C.-X. Kou, “A nonlinear conjugate gradient algorithm with an optimal property and an improved wolfe line search,” *SIAM Journal on Optimization*, vol. 23, no. 1, pp. 296–320, 2013.

[41] J. Sun, Q. Qu, and J. Wright, “Complete dictionary recovery over the sphere i: Overview and the geometric picture,” *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 853–884, 2017.

[42] N. Boumal and P.-A. Absil, “Low-rank matrix completion via preconditioned optimization on the grassmann manifold,” *Linear Algebra and its Applications*, vol. 475, pp. 200–239, 2015.

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

[44] A. Godsmith, *Wireless Communications*. Cambridge University Press, 2005.

[45] T. Ji and W. Stark, “Rate-adaptive transmission over correlated fading channels,” *IEEE transactions on communications*, vol. 53, no. 10, pp. 1663–1670, 2005.

[46] 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.