Distributed MIMO Radar Based on Sparse Sensing: Analysis and Efficient Implementation

Bo Li, *Student member, IEEE*, and Athina P. Petropulu, *Fellow, IEEE*

Abstract

In sparse sensing based distributed MIMO radars, the problem of target estimation is formulated as a sparse vector recovery problem, where the vector to be recovered is block sparse, or equivalently, the sensing matrix is block-diagonal and the sparse vector consists of equal-length blocks that have the same sparsity profile. This paper derives the theoretical requirements and performance guarantees for the application of sparse recovery techniques to this problem. The obtained theoretical results confirm previous, simulations based observations, that exploiting the block sparsity of the target vector can further reduce the amount of measurements needed for successful target estimation. For signal recovery, two low-complexity approaches are proposed. The first one is an ADMM-based sparse signal recovery algorithm, which in addition to significantly reducing computations is also amenable to a parallel and semi-distributed implementation. The second approach decouples the location and speed estimation into two separate stages, with each stage addressing a sparse recovery problem of lower dimension while maintains high estimation accuracy.

Index Terms

Distributed MIMO radars, sparse sensing, restricted isometry property, ADMM, decoupled estimation, computation performance.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) radars [1]–[4] have received considerable attention in recent years due to their improved performance over traditional phase arrays. Depending on the placement of antennas, MIMO radars can be classified into collocated [1], [2] and widely separated [3], [4]. Collocated MIMO radars exploit phase differences in target returns induced by transmit and receive antennas, to effectively increase the array aperture and achieve high resolution. Distributed MIMO radars enjoy spatial diversity, introduced by the multiple independents paths between the targets and the transmit/receive antennas, and thus achieve improved target estimation performance.

By exploiting the sparsity of targets in the radar scene, sparse sensing [5]–[7] has been studied in the context of both collocated [8]–[12], and distributed MIMO radars [13], [14]. In [13], [14], the problem of target location and

speed estimation in distributed MIMO radars is investigated as a block sparse signal recovery problem. The target vector contains the attenuation coefficients for all paths between the grid points and the transmit/receive antenna pairs. If there is no target present at a certain grid point, the corresponding entries in the target vector are zero. Since the number of targets is much smaller than the number of grid points, the target vector is sparse. The block sparsity in the target vector arises by grouping together entries corresponding to paths between a given grid point and all transmit/receive antenna pairs. Block matching pursuit (BMP) is applied in [13] for signal support recovery. Simulations in [13] show that BMP outperforms the basis pursuit method, which ignores the block sparsity. The advantage of block sparsity was also studied in [14], where a group Lasso approach was used to exploit the block sparsity. Again, simulations in [14] show that exploiting block sparsity results in significant detection performance gains over methods which just consider unstructured sparsity. To the best of our knowledge, there are no theoretical works on the performance of sparse sensing based distributed MIMO radars. Although there are theoretical works on sparse sensing based collocated MIMO radars [8]–[11], those results cannot be extended to the distributed MIMO radar scenario.

Employing sparse signal recovery techniques in radar systems, on one hand, relieves the volume of data that needs top be collected, but on the other hand, introduces significant computational complexity. In [14] a group Lasso with proximal gradient algorithm (GLasso-PGA) was used, and in [15], a mixed ℓ_1/ℓ_2 norm optimization with interior point method (L-OPT-IPM) was used. GLasso-PGA and L-OPT-IPM achieve better estimation performance than BMP but involve higher computational complexity and require careful tuning of manually chosen parameters. The computation becomes prohibitive as the dimension of the sparse target vector increases. The approach of [14] exploited the block diagonal structure of the sensing matrix to propose a decomposition of the original problem into smaller size problems, thus reducing complexity. However, the scheme of [14] did not exploit all available structural information, such as the identical sparsity profile of the sub-vectors in the target vector.

The contribution of this paper is two-fold: (i) it provides performance guarantees for the target location and speed estimation in sparse sensing based distributed MIMO radars, and (ii) it proposes two low-complexity approaches for target estimation. Regarding the performance guarantees, by permuting the columns of the measurement matrix we reformulate the block-sparse signal recovery problem into a problem in which the measurement matrix, $\boldsymbol{\Psi}$, is block diagonal (BD) and the sparse target vector, $\mathbf{s}$, contains equal-sized blocks that have the same sparsity profile. This reformulation facilitates restricted isometry property (RIP)-based performance analysis. Once the RIP of $\boldsymbol{\Psi}$ holds with respect to sparse signals with the aforementioned structure, the vector $\mathbf{s}$ can be obtained as the solution to a mixed ℓ_2/ℓ_1-optimization program (L-OPT) [16]. Our theoretical results confirm that the BD structure in $\boldsymbol{\Psi}$ and the sparsity structure in $\mathbf{s}$ reduce the number of measurements needed for target estimation. Further, our RIP-based analysis provides a uniform recovery guarantee, which means that once $\boldsymbol{\Psi}$ satisfies the RIP, target estimation can be achieved with high probability even in the worst case. In relation to the literature, the proposed RIP analysis is related to that for a Toeplitz matrix, presented in [17], except that our BD measurement matrix contains additional complex exponential factors introduced by the moving targets.

Regarding our low complexity contribution, we first propose a fast algorithm to solve the L-OPT problem based

on the alternating direction method of multipliers (ADMM). This ADMM based approach is amenable to parallel implementation, which allows for reduction of running time. A semi-distributed implementation of the solution is also discussed, in which the computations are distributed among all the receive nodes, thus obviating the need of a powerful fusion center. Simulations validate the efficiency of the proposed algorithm and show that the proposed algorithm is robust over a wide range of a manually chosen parameter. Another approach to lower complexity is to reduce the dimension of the L-OPT problem. The joint location-speed space is a Cartesian product of the location space and the speed space, which produces a high dimensional sparse target vector $\mathbf{s}$. A decoupled two-stage model is proposed to lower the problem dimension. In the first stage, we derive the sparse model for the location space by absorbing the unknown target Doppler effect into the sparse target vector. The target locations are estimated via sparse recovery algorithms with much lower dimension. In the second stage, the location estimates are used to greatly reduce the dimension of the sparse model for speed estimation. The ADMM based approach can also be integrated into both stages of the decoupled estimation framework. It is shown via simulation that the decoupled scheme reduces both the computation and the required number of measurements, while it maintains good performance. A related decoupled scheme was proposed in [18] for compressive sensing based step-frequency MIMO radar with collocated antennas. The matched filtering method was used to provide an initial estimate to reduce the space that needs to be discretized. However, while matched filtering requires large amount of measurements for high resolution and reliable estimation, our decoupled approach, provides a high resolution initial estimate with much fewer measurements.

The paper is organized as follows. Section II provides some background and introduces notation. The sparse model for distributed MIMO radar system is presented in Section III. In Section IV, we derive the $\mathcal{A}$-RIP of the measurement matrix, which is used to provide the performance of L-OPT in Section V. In Section VI, an efficient algorithm based on ADMM is proposed for the target estimation. Parallel and semi-decentralized implementation schemes are discussed in Section VI-B. Section VII presents the decoupled location and speed estimation together with discussions on the computation complexity and required number of measurements. Simulation results are given in Section VIII, and conclusions are presented in Section IX.

II. BACKGROUND ON BLOCK SPARSITY

In the context of compressive sensing, the focus is to exploit the structure in the sparse signal and the measurement matrix for improving the sparse signal recovery [19]. Block sparsity in the sparse signal was investigated in [16], [20], [21], where the elements in the sparse signal vector appear in blocks.

Let us consider a block sparse vector $\mathbf{s} \in \mathbb{C}^{MN}$ with at most K nonzero blocks out of N equal-sized blocks, *i.e.*, $M \triangleq |\mathcal{I}_n|, \forall n \in \mathbb{N}_N^+$, where $\mathcal{I}_n$ is the index set for the n-th block. Let us denote by $\mathcal{A}_B^K$ the space in which the block sparse vectors lie.

Given the noisy measurement vector $\mathbf{z} = \mathbf{\Psi s} + \mathbf{n}$ with $\mathbf{\Psi} \in \mathbb{C}^{L \times NM}$ as the measurement matrix and $\mathbf{n} \in \mathbb{C}^L$ as

the additive noise vector, the recovery of $\mathbf{s} \in \mathcal{A}_B^K$ is achieved via the following convex optimization problem

$$\min_{\mathbf{s}} \sum_{n=1}^{N} \|\mathbf{s}[\mathcal{I}_n]\|_2 \ \ s.t. \ \|\mathbf{z} - \mathbf{\Psi s}\|_2 \leq \epsilon. \tag{1}$$

which is referred to as mixed ℓ_2/ℓ_1-optimization program (L-OPT) [16]. The effectiveness of using L-OPT relies on the restricted isometry property (RIP) of $\mathbf{\Psi}$ with respect to vectors in $\mathcal{A}_B^{2K}$.

Definition 1 ([22]): For a union of certain subspaces denoted by $\mathcal{A}$, $\mathbf{\Psi}$ is said to satisfy the $\mathcal{A}$-restricted isometry property with constant $\delta \in (0, 1)$, in short, $\mathcal{A}$-RIP(K, δ), if δ is the smallest value such that

$$(1 - \delta)\|\mathbf{s}\|_2^2 \leq \|\mathbf{\Psi s}\|_2^2 \leq (1 + \delta)\|\mathbf{s}\|_2^2 \tag{2}$$

holds for all $\mathbf{s} \in \mathcal{A}$.

The above definition is for general union of subspaces. If $\mathbf{\Psi}$ satisfies the RIP over $\mathcal{A}_B^{2K}$, or equivalently, if $\mathbf{\Psi}$ satisfies the $\mathcal{A}_B$-RIP$(2K, \delta_{2K})$, then the next lemma shows that the solution of (1), *i.e.*, $\hat{\mathbf{s}}$, is a good approximation of $\mathbf{s}$.

Lemma 1 (Theorem 2 in [16]): If $\mathbf{\Psi}$ satisfies the $\mathcal{A}_B$-RIP$(2K, \delta_{2K})$ with $\delta_{2K} < \sqrt{2} - 1$, then for the solution of (1), $\hat{\mathbf{s}}$, it holds that

$$\|\hat{\mathbf{s}} - \mathbf{s}\|_2 \leq \frac{4\sqrt{1 + \delta_{2K}}}{1 - (1 + \sqrt{2})\delta_{2K}} \epsilon \triangleq g(\epsilon). \tag{3}$$

It is shown in [16] that Gaussian measurement matrices require fewer measurements to satisfy the $\mathcal{A}_B$-RIP$(2K, \delta_{2K})$ as compared to the number of measurements needed to satisfy the RIP$(2K, \delta_{2K})$. Therefore, exploiting block sparsity in $\mathbf{s}$ reduces the required number of measurements for sparse recovery. In the following, we put the sparse sensing-based MIMO radar problems into the framework of block-sparse signal recovery, and derive the $\mathcal{A}_B$-RIP of the corresponding measurement matrix.

III. SIGNAL MODEL

We consider a MIMO radar system with M_t transmit nodes (TX) and M_r receive nodes (RX), which are widely separated. Let (x_i^t, y_i^t) and (x_i^r, y_i^r) denote the locations of the i-th transmit and receive antenna in cartesian coordinates, respectively. The i-th TX antenna transmits repeated pulses with pulse repetition interval T. Each pulse contains the modulated waveform $w_i(t)e^{j2\pi f_i t}$, where f_i is the carrier frequency, and $w_i(t)$ is the continuous-time baseband waveform. We assume that transmit waveforms are jointly Gaussian with zero mean and variance σ_0^2. Let us assume that there are K moving targets present. For simplicity, we consider a clutter-free environment [8]–[14], [18]. In practice, preprocessing techniques can be employed to suppress the clutter. For example, if the covariance matrix of the clutter is known, beamforming can be used to suppress the clutter [32]. Also, if the clutter is static while the target is moving, Doppler filters [30] and the technique of change detection can be used to remove the clutter [33], [34].

The location-speed space is discretized on grid $\mathbf{\Theta} \triangleq \mathbf{\Theta}_{loc} \times \mathbf{\Theta}_{spd}$, where the location grid is $\mathbf{\Theta}_{loc} \triangleq \{(x_n, y_n), n = 1, \ldots, N_1\}$, $N_1 \triangleq N_x \times N_y$, and the speed grid is $\mathbf{\Theta}_{spd} \triangleq \{(v_x^n, v_y^n), n = 1, \ldots, N_2\}$, $N_2 \triangleq N_{vx} \times N_{vy}$. Denoting the cardinality of $\mathbf{\Theta}$ as N, it holds that $N = N_1 \times N_2$. It is assumed that the targets fall on grid points. Let us

denote by Ξ the set of all different transmit and receive antenna pairs. It is clear that $|\Xi| = M_t M_r$. In the sequel, the subscript $(ij) \in \Xi$ with $i \in \mathbb{N}_{M_t}^+$ and $j \in \mathbb{N}_{M_r}^+$ denotes the pair of the i-th transmit antenna and the j-th receive antenna. Suppose that the j-th receive antenna obtains L T_s-spaced samples from each pulse transmitted by antenna i. On stacking the samples from P pulses into vector $\mathbf{z}_{ij}$ it holds that [14]

$$\mathbf{z}_{ij} = \boldsymbol{\Psi}_{ij}\mathbf{s}_{ij} + \mathbf{n}_{ij}, \; \forall(ij) \in \Xi, \tag{4}$$

where $\mathbf{s}_{ij} = \left[s_{ij}^1, \ldots, s_{ij}^N\right]^T$, with s_{ij}^n being non-zero only if there is a target at the n-th grid point (here n refers to a particular ordering of grid points of the 4-dimensional space into a vector of length N); $\mathbf{n}_{ij}$ denotes the additive noise; $\boldsymbol{\Psi}_{ij}$ is defined in term of its columns, $\boldsymbol{\Psi}_{ij}^n$, *i.e.*,

$$\boldsymbol{\Psi}_{ij}^n = \mathbf{D}(f_{ij}^n) \otimes \mathbf{w}_{i,\tau_{ij}^n}, \; \forall n \in \mathbb{N}_N^+, \tag{5}$$

where $\otimes$ denotes Kronecker product,

$$\mathbf{D}(f_{ij}^n) \triangleq \left[1, e^{j2\pi f_{ij}^n T}, \ldots, e^{j2\pi f_{ij}^n T(P-1)}\right]^T, \tag{6a}$$

$$\mathbf{w}_{i,\tau_{ij}^n} \triangleq \left[w_i[\tau_{ij}^n], \ldots, w_i[(L-1)T_s + \tau_{ij}^n]\right]^T \tag{6b}$$

with $w_i[\tau_{ij}^n]$ denoting the sample of the i-th transmit waveform at time index τ_{ij}^n. τ_{ij}^n and f_{ij}^n respectively denote the propagation time and Doppler frequency associated with the n-th grid and the (ij)-th TX/RX antenna pair. It holds that

$$f_{ij}^n = \frac{\left\langle (v_x^n, v_y^n), \mathbf{d}_{in}^t \right\rangle}{\lambda_i \|\mathbf{d}_{in}^t\|_2} + \frac{\left\langle (v_x^n, v_y^n), \mathbf{d}_{jn}^r \right\rangle}{\lambda_i \|\mathbf{d}_{jn}^r\|_2}, \tag{7}$$

where $\mathbf{d}_{in}^{t/r} \triangleq ((x_i^{t/r}, y_i^{t/r}) - (x_n, y_n))$ denotes the vector from the n-th grid to the i-th TX/RX antenna, and λ_i is the carrier wavelength of the i-th transmitter. In the model based on (5), we assume that the targets are moving relatively slowly so that the Doppler effect can be approximated as constant during one pulse, *i.e.*, $f_{ij}^n LT_s \ll 1$. In the literature, it is common to make such an assumption for pulse Doppler processing [12], [13], [30], [31]. Actually, we can show that our model works for wide range of target moving speeds by using different system parameters. Substituting the expression of f_{ij}^n in (7) into $f_{ij}^n LT_s \ll 1$ gives

$$\frac{\left\langle (v_x^n, v_y^n), \mathbf{d}_{in}^t \right\rangle}{\|\mathbf{d}_{in}^t\|_2} + \frac{\left\langle (v_x^n, v_y^n), \mathbf{d}_{jn}^r \right\rangle}{\|\mathbf{d}_{jn}^r\|_2} \ll \frac{c}{LT_s f_i},$$

Via the Cauchy-Schwarz inequality, we have $\|\mathbf{v}^n\|_2 \triangleq \sqrt{(v_x^n)^2 + (v_y^n)^2} \ll \frac{c}{2LT_s f_i}$ for all $n \in \mathbb{N}_N^+$. For carrier frequency $f_i = 5GHz$, waveform bandwidth $10MHz$, sampling frequency $f_s = 1/T_s = 20MHz$ and $L = 20$, the model based on (5) is valid for target speed much smaller than $3 \times 10^4 m/s$, which could cover speeds as high as transonic. By increasing the waveform bandwidth and the sampling frequency, the model would be valid even for supersonically moving targets. In Section VIII, we choose waveform bandwidth $25MHz$ and sampling frequency $f_s = 50MHz$, thus the model would be valid if the target speed is much smaller than $7.5 \times 10^4 m/s$.

The fusion center collects the sample vectors from all TX/RX antenna pairs and stack them into a column vector $\mathbf{z}$ of length $LPM_t M_r$, *i.e.*,

$$\mathbf{z} = \left[(\mathbf{z}_{11})^T, \ldots, (\mathbf{z}_{M_t M_r})^T\right]^T = \boldsymbol{\Psi}\mathbf{s} + \mathbf{n}, \tag{8}$$

where $\mathbf{s} = \left[(\mathbf{s}_{11})^T, \ldots, (\mathbf{s}_{M_t M_r})^T\right]^T$, $\mathbf{n} = \left[(\mathbf{n}_{11})^T, \ldots, (\mathbf{n}_{M_t M_r})^T\right]^T$, and $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\Psi}_{11}, \ldots, \boldsymbol{\Psi}_{M_t M_r})$.

Note that each vector $\mathbf{s}_{ij}$ contains zero entries except the entries corresponding to grid points occupied by targets. Thus, the vector $\mathbf{s}$ is a concatenation of $M_t M_r$ sub-vectors that share the same sparsity profile, and have exactly K nonzero entries each. We can see that $\mathbf{s}$ lies in $\mathcal{A}_0^K \subset \mathbb{C}^{N M_t M_r}$, defined as

$$
\begin{aligned}
\mathcal{A}_0^K \triangleq \big\{ \mathbf{y} \in \mathbb{C}^{N M_t M_r} \big| \; & \mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_{M_t M_r}^T]^T, \\
& \mathbf{y}_j \in \mathbb{C}^N, \text{supp}(\mathbf{y}_i) = \text{supp}(\mathbf{y}_j), \\
& |\text{supp}(\mathbf{y}_j)| \le K, \; \forall i, j \in \mathbb{N}_{M_t M_r}^+ \big\},
\end{aligned}
\tag{9}
$$

where $\mathbf{y}_j$'s are uniformly partitioned blocks of $\mathbf{y}$, $\text{supp}(\cdot)$ denotes the index set of nonzero entries of a vector, *i.e.*, the support of a vector, and $|\cdot|$ denotes the cardinality of a set.

In the following section we provide the $\mathcal{A}$-RIP analysis of the BD measurement matrix $\boldsymbol{\Psi}$.

IV. THE $\mathcal{A}$-RIP OF THE MEASUREMENT MATRIX

Let us first state two lemmas which will be used later.

Lemma 2: Let $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{y} \in \mathbb{C}^N$ be vectors with i.i.d complex Gaussian entries with zero mean and variance σ^2. Then for every $0 < t < 4\sigma^2 N$, it holds that

$$
\Pr\left(\|\mathbf{x}\|_2^2 - \mathbb{E}\{\|\mathbf{x}\|_2^2\} \ge t\right) \le e^{-\frac{t^2}{16 N \sigma^4}},
\tag{10a}
$$

$$
\Pr\left(\left|\|\mathbf{x}\|_2^2 - \mathbb{E}\{\|\mathbf{x}\|_2^2\}\right| \ge t\right) \le 2 e^{-\frac{t^2}{16 N \sigma^4}}.
\tag{10b}
$$

For every $t > 0$ it holds that

$$
\Pr\left(|\langle \mathbf{x}, \mathbf{y} \rangle| \ge t\right) \le 2 e^{-\frac{t^2}{4\sigma^2 (N \sigma^2 + t/2)}}.
\tag{11}
$$

where $\langle \mathbf{x}, \mathbf{y} \rangle \triangleq \mathbf{x}^H \mathbf{y}$, and $(\cdot)^H$ denotes Hermitian transpose.

Proof: Lemma 2 is derived based on Lemma 5 and 6 in [17]. By substituting k and t in [17, Lemma 5] respectively by N and τ, we have $\Pr\left(\|\mathbf{x}\|_2^2 - \mathbb{E}\{\|\mathbf{x}\|_2^2\} \ge 2\sigma^2 \sqrt{N\tau} + 2\sigma^2 \tau\right) \le e^{-\tau}$. If $\tau < N$, then it holds that $4\sigma^2 \sqrt{N\tau} \ge 2\sigma^2 \sqrt{N\tau} + 2\sigma^2 \tau$. We have

$$
\begin{aligned}
& \Pr\left(\|\mathbf{x}\|_2^2 - \mathbb{E}\{\|\mathbf{x}\|_2^2\} \ge 4\sigma^2 \sqrt{N\tau}\right) \\
\le & \Pr\left(\|\mathbf{x}\|_2^2 - \mathbb{E}\{\|\mathbf{x}\|_2^2\} \ge 2\sigma^2 \sqrt{N\tau} + 2\sigma^2 \tau\right) \le e^{-\tau}.
\end{aligned}
$$

Inequality (10a) is readily proved by denoting $t \triangleq 4\sigma^2 \sqrt{N\tau} < 4\sigma^2 N$. Similarly, we can prove inequality (10b). Lastly, it is clear to see that (11) holds by directly substituting k in Lemma 6 in [17] by N. ∎

Lemma 3: Let $\{x_i\}$ and $\{y_i\}$, $i = 1, \ldots, Q$ be sequences of identically distributed, zero-mean, Gaussian variables with variance σ^2. All variables are independent except that the last I ($I \in [1, Q)$) variables of $\{x_i\}$ are the first I variables of $\{y_i\}$, *i.e.*, $x_{i+Q-I} = y_i$ for any $i \in [1, I]$. Then

$$
\Pr\left(\left|\sum_{i=1}^Q x_i y_i\right| \ge t\right) \le 4 \exp\left(-\frac{(Q-1)t^2}{8 Q \sigma^2 (Q \sigma^2 + t/2)}\right).
$$

We know that $\{x_i y_i\}_{i=1}^Q$ are not mutually independent. Lemma 3 can be proven by a splitting trick, as in [17].

A. Observations on The Gram of The Normalized $\boldsymbol{\Psi}$

Note that $\mathbb{E}\{\|\boldsymbol{\Psi}_{ij}^n\|_2^2\} = LP\sigma_0^2$. Since in the compressive sensing literature measurement matrices with normalized columns are typically considered, we will provide observations on the normalized measurement matrix $\bar{\boldsymbol{\Psi}} = \boldsymbol{\Psi}/\sqrt{LP\sigma_0^2}$. The Gram of $\bar{\boldsymbol{\Psi}}$, denoted here by $\mathbf{G}$, is also block-diagonal, *i.e.*, $\mathbf{G} = \text{diag}(\mathbf{G}_{11}, \ldots, \mathbf{G}_{M_tM_r})$ where $\mathbf{G}_{ij} = \bar{\boldsymbol{\Psi}}_{ij}^H \bar{\boldsymbol{\Psi}}_{ij}$ and $\bar{\boldsymbol{\Psi}}_{ij} \triangleq \boldsymbol{\Psi}_{ij}/\sqrt{LP\sigma_0^2}$.

Consider the (n,l)-th entry in $\mathbf{G}_{ij}$. It holds that

$$
\begin{aligned}
\mathbf{G}_{ij}(n,l) &\equiv \frac{1}{LP\sigma_0^2}\langle \boldsymbol{\Psi}_{ij}^n, \boldsymbol{\Psi}_{ij}^l\rangle \\
&= \frac{1}{LP\sigma_0^2}\langle \mathbf{D}(f_{ij}^n), \mathbf{D}(f_{ij}^l)\rangle\langle \mathbf{w}_{i,\tau_{ij}^n}, \mathbf{w}_{i,\tau_{ij}^n}\rangle.
\end{aligned}
\tag{12}
$$

The following three cases are analyzed:

Case (i) For $n = l$, *i.e.*, the diagonal entries, it holds that $\langle \mathbf{D}(f_{ij}^n), \mathbf{D}(f_{ij}^n)\rangle = P$. Now, $G_{ij}(n,n) = \frac{1}{L\sigma_0^2}\mathbf{w}_{i,\tau_{ij}^n}^T\mathbf{w}_{i,\tau_{ij}^n}$, which is the sum of squares of i.i.d Gaussian variables with $\mathbb{E}\{G_{ij}(n,n)\} = 1$. Applying (10b) in Lemma 2, it holds that

$$
\Pr(|G_{ij}(n,n) - 1| > t) \leq 2\exp\left(-\frac{Lt^2}{16}\right).
\tag{13}
$$

Case(ii) the n-th and l-th grid points have different propagation delay, *i.e.*, $\tau_{ij}^n \neq \tau_{ij}^l$. From (12), it holds that $\mathbb{E}\{G_{ij}(n,l)\} = 0$ and

$$
|G_{ij}(n,l)| = \frac{1}{LP\sigma_0^2}\left|\mathbf{w}_{i,\tau_{ij}^n}^T\mathbf{w}_{i,\tau_{ij}^l}\right|\phi_{f_{ij}^n, f_{ij}^l}(P),
\tag{14}
$$

where

$$
\phi_{f_{ij}^n, f_{ij}^l}(P) \triangleq |\langle \mathbf{D}(f_{ij}^n), \mathbf{D}(f_{ij}^l)\rangle| \in [0, P].
\tag{15}
$$

A probabilistic bound on $|G_{ij}(n,l)|$ can be found as

$$
\Pr(|G_{ij}(n,l)| > t) \leq \Pr\left(\frac{1}{L\sigma_0^2}\left|\mathbf{w}_{i,\tau_{ij}^n}^T\mathbf{w}_{i,\tau_{ij}^l}\right| > t\right).
\tag{16}
$$

Now, we only need to provide the bound on the inner product of $\mathbf{w}_{i,\tau_{ij}^n}$ and $\mathbf{w}_{i,\tau_{ij}^l}$. Note that $\mathbf{w}_{i,\tau_{ij}^n}$ and $\mathbf{w}_{i,\tau_{ij}^l}$ are both sampled from the i-th waveform, and may share some common entries. The general bound (11) in Lemma 2 referring to two distinct i.i.d random vectors cannot be applied directly. Applying Lemma 3 for (16) gives

$$
\Pr(|G_{ij}(n,l)| > t) \leq 4\exp\left(-\frac{(L-1)t^2}{8(1+t/2)}\right).
\tag{17}
$$

Case(iii) the n-th and l-th grid points introduce the same propagation delay ($\tau_{ij}^n = \tau_{ij}^l$) but have different Doppler frequencies ($f_{ij}^n \neq f_{ij}^l$). Consider the absolute value

$$
|G_{ij}(n,l)| = \frac{1}{LP\sigma_0^2}\mathbf{w}_{i,\tau_{ij}^n}^T\mathbf{w}_{i,\tau_{ij}^n}\phi_{f_{ij}^n, f_{ij}^l}(P),
\tag{18}
$$

which can be viewed as the squared norm of random vector $\sqrt{\frac{1}{LP\sigma_0^2}\phi_{f_{ij}^n, f_{ij}^l}(P)}\mathbf{w}_{i,\tau_{ij}^n}$ with i.i.d zero-mean Gaussian entries with variance $\sigma_1^2 = \frac{1}{LP}\phi_{f_{ij}^n, f_{ij}^l}(P)$. Applying the unilateral bound (10a) in Lemma 2 , we have

$$
\begin{aligned}
\Pr(|G_{ij}(n,l)| > t) &\leq \exp\left(-\frac{1}{L}\left(\frac{t - L\sigma_1^2}{4\sigma_1^2}\right)^2\right) \\
&= \exp\left(-\frac{L}{16}\left(\frac{Pt}{\phi_{f_{ij}^n, f_{ij}^l}(P)} - 1\right)^2\right) \leq \exp\left(-\frac{Lt^2}{16}\right),
\end{aligned}
\tag{19}
$$

where the last inequality holds if

$$P \geq (1/t + 1)\phi_{f_{ij}^n, f_{ij}^l}(P). \tag{20}$$

B. $\mathcal{A}$-RIP of The Normalized Measurement Matrix

Equipped with the above observations, we will first establish the RIP of $\bar{\boldsymbol{\Psi}}$ with respect to sparse vectors in $\mathcal{A}_1^K \subset \mathbb{C}^{NM_t M_r}$, which is defined as

$$\mathcal{A}_1^K \triangleq \big\{ \mathbf{y} \in \mathbb{C}^{NM_t M_r} \,\big|\, \mathbf{y} = [\mathbf{y}_1^T, \ldots, \mathbf{y}_{M_t M_r}^T]^T, \mathbf{y}_j \in \mathbb{C}^N,$$

$$|\text{supp}(\mathbf{y}_i)| = |\text{supp}(\mathbf{y}_j)| \leq K, \forall i, j \in \mathbb{N}_{M_t M_r}^+ \big\}$$

where $\mathbf{y}_j$'s are uniformly partitioned blocks of $\mathbf{y}$. It holds that $\mathcal{A}_0^K \subset \mathcal{A}_1^K$.

Theorem 1: For any $\delta_K \in (0, 1)$, $\bar{\boldsymbol{\Psi}}$ satisfies $\mathcal{A}_1$-RIP(K, δ_K) with probability exceeding $(1 - 4(N\sqrt{M_t M_r})^{-1})$ whenever

$$L \geq 48\delta_K^{-1} K^2 \log(N\sqrt{M_t M_r}) + 1, \tag{21a}$$

$$P \geq (\delta_K^{-1} K + 1)\beta(P), \tag{21b}$$

where

$$\beta(P) \triangleq \sup_{(ij) \in \Xi} \phi_{ij}(P) \triangleq \sup_{\substack{f_{ij}^n \neq f_{ij}^l, \\ (ij) \in \Xi}} \phi_{f_{ij}^n, f_{ij}^l}(P). \tag{22}$$

Proof: See Appendix A ∎

Note that the technique used to prove Theorem 1 can only exploit the structure characterized by $\mathcal{A}_1^K$, and not the additional structure characterized by $\mathcal{A}_0^K$. Vectors in $\mathcal{A}_0^K$ consist of sub-vectors that have the same support. However, only the support cardinality of the sub-vectors matters in the proof of $\mathcal{A}$-RIP presented in Appendix A. The positions of the nonzero entries would introduce no difference to the bound for the off-diagonal entries in (44). In the next section, the $\mathcal{A}_1$-RIP of $\bar{\boldsymbol{\Psi}}$ in Theorem 1 will be relaxed to the $\mathcal{A}_0$-RIP, which is further used to guarantee the effectiveness of applying L-OPT.

Remark 1: In the proof of Theorem 1, the sparsity structures in $\bar{\boldsymbol{\Psi}}$ and $\mathbf{s}$ is exploited to reduce the required number of measurements. To emphasize the advantage of the block-sparse structure in our scenario, we compare to a scenario in which the block-structure is ignored, and the recovery is based on a full Toeplitz matrix of size $LM_t M_r \times NM_t M_r$ and a sparse vector with $KM_t M_r$ nonzero entries at arbitrary locations. From [17], a full Toeplitz matrix satisfies the RIP if L is of the order of $\mathcal{O}(K^2 M_t M_r \log(N\sqrt{M_t M_r}))$, which is $M_t M_r$ times larger than the bound in (21a). Comparing that to (21a) suggests that exploiting the block sparsity reduces the number of samples needed. This validates previous simulation-based observations [14] and results in Section VIII-B of this paper, suggesting that exploiting the structure in both $\boldsymbol{\Psi}$ and $\mathbf{s}$ allows for reduction of the number of samples, L, needed for target estimation.

Remark 2: From (21b), we know that the number of required pulses is determined by $\beta(P)$, which is the maximum of $\phi_{f_{ij}^n, f_{ij}^l}(P)$ over the Doppler grid set for all TX/RX pairs. From the definitions in (6a) and (15), it

holds that

$$\beta(P) \triangleq \sup_{\substack{f_{ij}^n \neq f_{ij}^l, \\ (ij) \in \Xi}} \left| \frac{\sin\left(\pi P(f_{ij}^n - f_{ij}^l)T\right)}{\sin\left(\pi(f_{ij}^n - f_{ij}^l)T\right)} \right|.$$

The quantity $\beta(P)$ is determined by the pulse repetition interval T, the number of pulses, P, and the Doppler grid set. From the definition in (7), the Doppler grid set further depends on the speed grid set Θ_2, the antennas position and even the target location grid set Θ_1. Therefore, it is rather difficult to analytically characterize $\beta(P)$. Generally speaking, in order to increase the speed resolution, we can increase either the number of pulses, or the pulse repetition interval. For a given MIMO radar configuration and target space discretization, we can use numerical methods to find the minimum P that satisfies (21b). In Section VIII-A, we present an example to show how P is chosen.

V. Performance of Distributed MIMO Radars Using Sparse Sensing

To apply the L-OPT for the sparse model in (8), we permute the columns of $\boldsymbol{\Psi}$ and correspondingly permute the entries of $\mathbf{s}$ to generate block sparsity in the target vector. Then, $\mathbf{s}$ is recovered by solving the problem

$$\min \sum_{n=1}^{N} \|\mathbf{s}[\mathcal{I}_n]\|_2 \quad s.t. \ \|\mathbf{z} - \mathcal{P}_M(\boldsymbol{\Psi})\mathcal{P}_v(\mathbf{s})\|_2 \leq \epsilon_0. \tag{23}$$

where $\mathcal{P}_M$ is the column permutation matrix applied on $\boldsymbol{\Psi}$ and $\mathcal{P}_v : \mathcal{A}_0^K \to \mathcal{A}_B^K$ is the corresponding permutation operator applied on $\mathbf{s}$; $\mathcal{I}_n$, $\forall n \in \mathbb{N}_N^+$, is the set with cardinality $M_t M_r$ containing the indices of the n-th entries from all sub-vectors $\mathbf{s}_{ij}$; ϵ_0 is a manually chosen parameter related to the norm of vector $\mathbf{n}$. In the above, $\mathcal{P}_v(\mathbf{s}) = [\mathbf{s}[\mathcal{I}_1], \ldots, \mathbf{s}[\mathcal{I}_N]]^T$ is block sparse.

The reconstructed target location-speed scene $\hat{\mathbf{s}}$ contains location and speed and target complex Radar Cross Section (RCS) information on all K targets. Let us use as performance metric the error $\|\hat{\mathbf{s}} - \mathbf{s}\|_2$. As shown in Lemma 1, the effectiveness of (23) is guaranteed if $\mathcal{P}_M(\boldsymbol{\Psi})$ satisfies the $\mathcal{A}_B$-RIP$(2K, \delta_{2K})$. Combining with the $\mathcal{A}$-RIP analysis of Section IV, the following proposition provides an error bound when applying L-OPT for the recovery of $\mathbf{s}$, along with the requirements on the number of measurements and pulses.

Proposition 1: Consider the signal model in (8). For any $\delta_{2K} < \sqrt{2} - 1$, if L and P satisfy that

$$L \geq 192\delta_{2K}^{-1}K^2 \log(N\sqrt{M_t M_r}) + 1, \tag{24a}$$

$$P \geq (2\delta_{2K}^{-1}K + 1)\beta(P), \tag{24b}$$

then for any $\mathbf{s} \in \mathcal{A}_0^K$, the error in the solution of the L-OPT problem of (23) is bounded as

$$\|\hat{\mathbf{s}} - \mathbf{s}\|_2 \leq g\left(\frac{\epsilon_0}{\sqrt{LP\sigma_0^2}}\right) \tag{25}$$

with probability exceeding $(1 - 4(N\sqrt{M_t M_r})^{-1})$.

Proof: According to Theorem 1, $\bar{\boldsymbol{\Psi}}$ satisfies the $\mathcal{A}_1$-RIP$(2K, \delta_{2K})$ with probability exceeding $(1 - 4(N\sqrt{M_t M_r})^{-1})$ under conditions:

$$L \geq 48\delta_{2K}^{-1}(2K)^2 \log(N\sqrt{M_t M_r}) + 1,$$

$$P \geq (\delta_{2K}^{-1}(2K) + 1)\beta(P),$$

which are obtained by substituting K and δ_K in (21) respectively by $2K$ and δ_{2K}. As we can see, the above two conditions are equivalent to those in (24). Since $\mathcal{A}_0^{2K} \subset \mathcal{A}_1^{2K}$, we know that $\bar{\bar{\Psi}}$ also satisfies the $\mathcal{A}_0$-RIP$(2K, \delta_{2K}^0)$ with $\delta_{2K}^0 \le \delta_{2K}$. In [23, Proposition 1], we have shown that the $\mathcal{A}_B$-RIP of $\mathcal{P}_M(\bar{\bar{\Psi}})$ is equivalent to the $\mathcal{A}_0$-RIP of $\bar{\bar{\Psi}}$. Applying Lemma 1 to (23) for the normalized measurement matrix $\bar{\bar{\Psi}}$ proves the claims of the proposition. ∎

Remark 3: The significance of Proposition 1 is that it guarantees theoretically that sparse modeling and block sparse recovery algorithms can be effectively applied to distributed MIMO radars. If there is no additive noise, *i.e.*, $\epsilon_0 = 0$, based on (25), $\mathbf{s}$ can be recovered exactly. When noise is present, the performance is stable in the sense that the estimation error is bounded for any $\mathbf{s} \in \mathcal{A}_0^K$. In the proof of Proposition 1, the $\mathcal{A}_B$-RIP of $\mathcal{P}_M(\bar{\bar{\Psi}})$ is established via its equivalence to the $\mathcal{A}_0$-RIP of $\bar{\bar{\Psi}}$. The direct block-RIP analysis for $\mathcal{P}_M(\bar{\bar{\Psi}})$ is difficult, because $\mathcal{P}_M(\bar{\bar{\Psi}})$ has a complicated structure. Also, the $\mathcal{A}_0$-RIP of $\bar{\bar{\Psi}}$ is established indirectly via the $\mathcal{A}_1$-RIP analysis of $\bar{\bar{\Psi}}$. Since $\mathcal{A}_0^{2K}$ is only a small subset of $\mathcal{A}_1^{2K}$, the $\mathcal{A}_0$-RIP of $\bar{\bar{\Psi}}$ may be satisfied with much weaker conditions on L and P which are required by the $\mathcal{A}_1$-RIP of $\bar{\bar{\Psi}}$. That is to say that the L-OPT in (23) may perform well with smaller L and P than those in (24).

Remark 4: In the compressive sensing literature, there are two kinds of sparse recovery guarantees: the uniform and non-uniform [29]. A uniform guarantee means that once Ψ satisfies the $\mathcal{A}_0$-RIP, target estimation can be achieved with high probability for any $\mathbf{s} \in \mathcal{A}_0^K$. A uniform recovery guarantee attracts a lot of research interest in the compressive sensing literature [6], [16], [17], [22], [29] and applications in collocated MIMO radars [10], [11]. Proposition 1 provides bounds on L and P for the uniform recovery guarantee including the worst case. On the other hand, the simulation gives the average performance for given L and P. This explains why much smaller L and P perform well in the simulation in Section VIII. As one can see both in Theorem 1 and Preposition 1, the bound on L scales quadratically with the sparsity level K. The quadratically scaled bound is the proved tightest bound for many structured measurement matrices [10], [11], [17], [29]. To the best of our knowledge, Theorem 1 is the first result on the $\mathcal{A}$-RIP of the measurement matrix Ψ with block diagonal structure in sparse sensing based distributed MIMO radars, modeled via (8). Although there might be the possibility to break the quadratic bottleneck on L, that would call for complete different techniques and it is out of the scope of this paper.

Remark 5: The L-OPT problem is convex and can be solved directly using the interior point method with complexity of $\mathcal{O}((NM_rM_t)^3)$. This means that the computational cost may be prohibitive if the dimension- NM_tM_r is large. In the following two sections, we tackle the computation issue in two ways, namely, we propose an ADMM-based algorithm with lower complexity (see Section VI), and we propose decoupling the location and speed estimation, which effectively lowers the dimensionality of the problem (see Section VII).

VI. Fast Signal Recovery based on ADMM

In this section, we present an ADMM-based approach for solving the L-OPT problem for the general case of moving targets. Preliminary results of our work, for the case of stationary targets can be found in [24].

Note that $\mathbf{z}, \boldsymbol{\Psi}, \mathbf{s}$ and $\mathbf{n}$ are all complex. The majority of ADMM literature deals with real variables. However, we can easily reformulate our problem with real variables as follows.

$$\underbrace{(\mathbf{z} \otimes \mathcal{F}_v)}_{\triangleq \tilde{\mathbf{z}}} = \underbrace{(\boldsymbol{\Psi} \otimes \mathcal{F}_M)}_{\triangleq \tilde{\boldsymbol{\Psi}}} \underbrace{(\mathbf{s} \otimes \mathcal{F}_v)}_{\triangleq \tilde{\mathbf{s}}} + \underbrace{(\mathbf{n} \otimes \mathcal{F}_v)}_{\triangleq \tilde{\mathbf{n}}}, \tag{26}$$

where the operators $\mathcal{F}_v$ and $\mathcal{F}_M$ are defined in terms of the real and imaginary parts of $z \in \mathbb{C}$, *i.e.*, respectively, $\Re\{z\}$ and $\Im\{z\}$, as follows.

$$\mathcal{F}_v(z) \triangleq [\Re\{z\}, \Im\{z\}]^T, \ \mathcal{F}_M(z) \triangleq \begin{bmatrix} \Re\{z\} & -\Im\{z\} \\ \Im\{z\} & \Re\{z\} \end{bmatrix}.$$

For any vector $\mathbf{v}$, $\mathbf{v} \otimes \mathcal{F}_v$ applies the operator $\mathcal{F}_v$ on all the entries of $\mathbf{v}$. Similarly, for any matrix $\mathbf{M}$, $\mathbf{M} \otimes \mathcal{F}_M$ applies the operator $\mathcal{F}_M$ on all the entries of $\mathbf{M}$. It is clear that $\tilde{\boldsymbol{\Psi}}$ is still block diagonal with $\tilde{\boldsymbol{\Psi}}_{ij} \triangleq (\boldsymbol{\Psi}_{ij} \otimes \mathcal{F}_M) \in \mathbb{R}^{(2LP) \times (2N)}$ as is its (ij)-th diagonal block; and $\tilde{\mathbf{s}}$ is composed by $M_t M_r$ sub-vectors $\tilde{\mathbf{s}}_{ij} \triangleq (\mathbf{s}_{ij} \otimes \mathcal{F}_v) \in \mathbb{R}^{2N}$ that share the same sparsity profile and have exactly $2K$ nonzero real entries.

The L-OPT problem corresponding to (26) is given by

$$\min \sum_{n=1}^{N} \|\tilde{\mathbf{s}}[\mathcal{I}_n]\|_2 \ \ s.t. \ \|\tilde{\mathbf{z}} - \tilde{\boldsymbol{\Psi}}\tilde{\mathbf{s}}\|_2 \le \epsilon_0, \tag{27}$$

where the set $\mathcal{I}_n, \forall n \in \mathbb{N}_N^+$, with cardinality $2M_t M_r$, contains the indices of the $(2n-1)$-th and $(2n)$-th entries from all equal-length sub-vectors $\tilde{\mathbf{s}}_{ij}, \forall (ij) \in \Xi$. The equivalent unconstrained problem, known as group Lasso, is as follows:

$$\min \ \frac{1}{2} \|\tilde{\mathbf{z}} - \tilde{\boldsymbol{\Psi}}\tilde{\mathbf{s}}\|_2^2 + \lambda \sum_{n=1}^{N} \|\tilde{\mathbf{s}}[\mathcal{I}_n]\|_2 \tag{28}$$

where λ is the regularization parameter. The second term enforce the solution to be group sparse. If prior information on the sparse target vector exists, it can be incorporated by introducing constraints, *i.e.*,

$$\min \ \frac{1}{2} \|\tilde{\mathbf{z}} - \tilde{\boldsymbol{\Psi}}\tilde{\mathbf{s}}\|_2^2 + \lambda \sum_{n=1}^{N} \|\tilde{\mathbf{s}}[\mathcal{I}_n]\|_2$$

$$s.t. \ \tilde{\mathbf{s}} \in \Omega^{2NM_t M_r}, \tag{29}$$

where $\Omega^{2NM_t M_r}$ can be any general convex set, determined by the prior. In this paper, we consider complex attenuation factors with magnitude less than ω_0, which means $|s_{ij}^n| \in [0, \omega_0] \triangleq \Omega$. Such prior can be obtained, for example, based on the distance between the region of interest and the TX/RX pairs. Thus, the constraint $\tilde{\mathbf{s}} \in \Omega^{2NM_t M_r}$ is satisfied if $\|[\tilde{\mathbf{s}}[2i-1], \tilde{\mathbf{s}}[2i]]\|_2 \in \Omega, \ \forall i \in \mathbb{N}_{NM_t M_r}^+$, where $\tilde{\mathbf{s}}[i]$ denotes the i-th entry of $\tilde{\mathbf{s}}$.

In the following, we use the alternating direction method of multipliers (ADMM) [25] to solve the problem.

A. A Fast Algorithm Based on the ADMM

We introduce auxiliary variables $\mathbf{y}$ and $\mathbf{x}$ and rewrite (29) as

$$\min \frac{1}{2}\|\tilde{\mathbf{z}} - \tilde{\mathbf{\Psi}}\tilde{\mathbf{s}}\|_2^2 + \sum_{n=1}^{N}\lambda\|\mathbf{y}_n\|_2$$

$$s.t. \ \mathbf{y}_n = \mathbf{D}_n\tilde{\mathbf{s}}, \ \forall n \in \mathbb{N}_N^+, \tag{30}$$

$$\mathbf{x} = \tilde{\mathbf{s}}, \mathbf{x} \in \Omega^{2NM_tM_r},$$

where $\mathbf{D}_n$ is the matrix of dimension $(2M_tM_r) \times (2NM_tM_r)$ that selects the entries in $\tilde{\mathbf{s}}$ indexed by $\mathcal{I}_n$; the vector $\mathbf{y}$ is defined as $[\mathbf{y}_1^T, \ldots, \mathbf{y}_N^T]^T$. We have $\mathbf{y} = \mathbf{D}\tilde{\mathbf{s}}$ where $\mathbf{D} = [\mathbf{D}_1^T, \ldots, \mathbf{D}_N^T]$ permutates $\tilde{\mathbf{s}}$ into $\mathbf{y}$. The auxiliary variable $\mathbf{y}$ is to isolate $\tilde{\mathbf{s}}$ from the group sparsity-inducing term $\sum\|\cdot\|_2$; the magnitude constraint is now imposed on $\mathbf{x}$ instead of $\tilde{\mathbf{s}}$.

Let us now apply the ADMM after grouping the variables into two blocks, *i.e.*, $(\mathbf{y}, \mathbf{x})$ and $\tilde{\mathbf{s}}$. The augmented Lagrangian of the above optimization problem can be written as follows

$$\mathcal{L}(\tilde{\mathbf{s}}, \mathbf{y}, \mathbf{x}; \mu, \nu) = \frac{1}{2}\|\tilde{\mathbf{z}} - \tilde{\mathbf{\Psi}}\tilde{\mathbf{s}}\|_2^2 + \nu^T(\mathbf{x} - \tilde{\mathbf{s}}) + \frac{\rho_2}{2}\|\mathbf{x} - \tilde{\mathbf{s}}\|_2^2$$

$$+ \sum_{n=1}^{N}\left(\lambda\|\mathbf{y}_n\|_2 + \mu_n^T(\mathbf{y}_n - \mathbf{D}_n\tilde{\mathbf{s}}) + \frac{\rho_1}{2}\|\mathbf{y}_n - \mathbf{D}_n\tilde{\mathbf{s}}\|_2^2\right), \tag{31}$$

where $\rho_1, \rho_2 > 0$ and $\mu \triangleq [\mu_1^T, \ldots, \mu_N^T]^T \in \mathbb{R}^{2NM_tM_r}$ and $\nu \in \mathbb{R}^{2NM_tM_r}$ are the Lagrangian multipliers.

Based on the framework of ADMM, we can solve (30) by alternatively iterating over $\mathbf{y}, \mathbf{x}$ and $\tilde{\mathbf{s}}$. The $\mathbf{y}$-subproblem is well studied in the literature [26] and its solution is given explicitly by the shrinkage operator

$$\mathbf{y}_n^{k+1} = \max\left\{\|\bar{\mathbf{s}}_n^k\|_2 - \frac{\lambda}{\rho_1}, 0\right\}\frac{\bar{\mathbf{s}}_n^k}{\|\bar{\mathbf{s}}_n^k\|_2}, \ \forall n \in \mathbb{N}_N^+, \tag{32}$$

where $\bar{\mathbf{s}}_n^k \triangleq \mathbf{D}_n\tilde{\mathbf{s}}^k - \mu_n^k/\rho_1$. In total, the computation cost of (32) scales as $\mathcal{O}(NM_tM_r)$.

For the $\mathbf{x}$-subproblem, we have

$$\mathbf{x}^{k+1} = \mathcal{P}_\Omega\left(\tilde{\mathbf{s}}^{k+1} - \frac{\nu^k}{\rho_2}\right), \tag{33}$$

where $\mathcal{P}_\Omega(\mathbf{x})$ projects $(\mathbf{x}[2i-1], \mathbf{x}[2i])$ onto the region $\{(x,y)|x^2 + y^2 \le \omega_0\}$, $\forall i \in \mathbb{N}_{NM_tM_r}^+$. The overall computation of (33) involves $\mathcal{O}(NM_tM_r)$ operations.

The $\tilde{\mathbf{s}}$-subproblem is a least squares problem. The minimizer is attained if

$$0 = \frac{\partial}{\partial\tilde{\mathbf{s}}}\mathcal{L}(\tilde{\mathbf{s}}, \mathbf{y}^{k+1}, \mathbf{x}^{k+1}; \mu^k, \nu^k) = \mathbf{A}\tilde{\mathbf{s}} - \mathbf{b}^k, \tag{34}$$

where $\mathbf{A} = \tilde{\mathbf{\Psi}}^T\tilde{\mathbf{\Psi}} + (\rho_1 + \rho_2)\mathbf{I}_{2NM_tM_r}$ and $\mathbf{b}^k = \tilde{\mathbf{\Psi}}^T\tilde{\mathbf{z}} + \mathbf{D}^T\mu^k + \rho_1\mathbf{D}^T\mathbf{y}^{k+1} + \nu^k + \rho_2\mathbf{x}^{k+1}$, and $\mathbf{I}_N$ denotes the identity matrix of dimension $N \times N$. The solution can be obtained by solving the following system of linear equations

$$\mathbf{A}\tilde{\mathbf{s}}^{k+1} = \mathbf{b}^k. \tag{35}$$

Given the signal model, $\mathbf{A}$ is fixed for all iterations. The computational effort for $\mathbf{b}^k$ in each iteration only involves permutation and addition of vectors; this is because $\tilde{\mathbf{\Psi}}^T\tilde{\mathbf{z}}$ is also fixed. In addition, $\mathbf{A}$ is block diagonal because

$\tilde{\boldsymbol{\Psi}}^T \tilde{\boldsymbol{\Psi}}$ is block diagonal. The system of (35) can be written into a set of subsystems of linear equations as follows

$$\mathbf{A}_m \tilde{\mathbf{s}}_m^{k+1} = \mathbf{b}_m^k, \ \forall m \in \mathbb{N}_{M_t M_r}^+, \tag{36}$$

where $\mathbf{A}_m$ denotes the m-th diagonal block of matrix $\mathbf{A}$; $\mathbf{v}_m$ denotes the m-th uniformly partitioned block of vector $\mathbf{v}$. From the definition of $\mathbf{A}$, we know that

$$\mathbf{A}_m = \tilde{\boldsymbol{\Psi}}_{ij}^T \tilde{\boldsymbol{\Psi}}_{ij} + (\rho_1 + \rho_2)\mathbf{I}_{2N}, \ \forall m \in \mathbb{N}_{M_t M_r}^+, \tag{37}$$

where $j = \lfloor \frac{m-1}{M_t} \rfloor + 1$ and $i = m - (j-1)M_t$. $\lfloor a \rfloor$ denotes the largest integer that is smaller than a. From the definition of $\mathbf{A}_m$, it is easy to show that $\mathbf{A}_m$ is symmetric and positive definite for any $\rho_1, \rho_2 > 0$. Therefore, each system in (36) can be solved efficiently using iterative methods, such as the Preconditioned Conjugate Gradient (PCG) method, with cost about $\mathcal{O}(N^2)$ operations. The total number of operations to solve (35) is of the order of $\mathcal{O}(N^2 M_t M_r)$.

Finally, the update for multipliers μ and ν can be carried out as

$$\nu^{k+1} = \nu^k + \rho_2(\mathbf{x}^{k+1} - \tilde{\mathbf{s}}^{k+1}), \tag{38a}$$

$$\mu^{k+1} = \mu^k + \rho_1(\mathbf{y}^{k+1} - \mathbf{D}\tilde{\mathbf{s}}^{k+1}), \tag{38b}$$

with linear complexity $\mathcal{O}(N M_t M_r)$.

The convergence of the above iterations is guaranteed by results in the ADMM literature [25]. The iterations stop when the decrease of the objective value in (29) drops below certain threshold or the number of iterations exceeds certain value.

Remark 6: The bounds on L and P of Proposition 1 apply to (27) and (28) exactly in the same way as to (23). The problems in (27) and (23) are identical because the transformation only involves the separation of real and imaginary parts. Also, (28) is equivalent to (27) because (28) and (27) are the dual problems of each other. It can be shown using convex analysis techniques [28] that for any $\lambda > 0$ the solution of (28) is a minimizer of (27) for certain ϵ_0. The same bounds on L and P of Proposition 1 also guarantee to solve (29), which result in a smaller recovery error than (23). The bounded constraint in (29) reduces the feasible set in (23) by incorporating prior information on $\mathbf{s}$. This means that the solution of (29) is at least as accurate as that of (23) with the same conditions on L and P of Proposition 1. As shown in the simulations, the additional constraint indeed improves the accuracy of the solution over other methods that also use the same L and P.

Remark 7: The advantages of the proposed algorithm can be summarized as follows. First, the computational cost is low. As we know, solving (30) using an interior point method would involve $\mathcal{O}((N M_t M_r)^3)$ operations [27]. For the proposed algorithm, the computational cost in each iteration is dominated by solving the system of linear equations (35), which is $\mathcal{O}(N^2 M_t M_r)$. The reduction of computations is more significant as the number of antennas increases. Second, the estimation accuracy of $\mathbf{s}$ is improved by introducing the amplitude constraints on the sparse target vector. Also, the performance is robust over wide range of regularization parameter λ. This is validated via simulations in Section VIII. Lastly, due to the block diagonal structure in $\boldsymbol{\Psi}$, the update of $\tilde{\mathbf{s}}^{k+1}$

in (35) can be achieved by updating independent sub-vectors in $\tilde{\mathbf{s}}^{k+1}$. The good separability in the update of all variables affords a parallel and decentralized implementation, as discussed in the next section.

B. Parallel and Semi-distributed Implementation

1) Parallel Implementation: In the $(k+1)$th iteration, it is clear that all pairs $(\mathbf{x}^k[2i-1], \mathbf{x}^k[2i])$ in $\mathbf{x}^k$ are updated independent of the others, thus, the computations can be done in parallel. A similar parallel scheme applies to μ^k and ν^k, and the update of $\mathbf{y}_n^k$. The subsystems in (36) can also be solved in parallel. Assuming that there are multiple computing units available at the fusion center, the target estimation running time can be significantly reduced. The parallel implementation here is different from the decoupled Lasso of [14], because here, the identical sparsity profile in the sub-vectors of the target vector is utilized via the auxiliary variable $\mathbf{y}$.

2) Fusion Center Aided Semi-Distributed Implementation: The ADMM based approach described in Section VI-A requires a fusion center to perform all the computations. However, a semi-distributed implementation is also possible. For each iteration, $\mathbf{x}$ (respectively for $\mathbf{s}$ and ν) can be divided into blocks, each of which can be updated locally at the receive antenna. However, the update of $\mathbf{y}$ and μ cannot be done locally. A fusion center performs the update of $\mathbf{y}$ and μ.

The fusion center aided semi-distributed scheme is summarized in Algorithm 1. In the implementation of Algorithm 1, $\tilde{\mathbf{s}}_m^{k+1} \in \mathbb{R}^{2N}$ (respectively for $\nu_m^{k+1}, \mathbf{x}_m^{k+1}$), $\forall m \in \mathbb{N}_{M_t M_r}^+$, denotes the m-th block of uniformly partitioned $\tilde{\mathbf{s}}^{k+1}$. $\mathbf{y}_m^{k+1} \in \mathbb{R}^N$ denotes the m-th block of uniformly partitioned $\mathbf{D}^T \mathbf{y}^{k+1}$. The receive node j updates $\mathbf{x}_m^{k+1}, \nu_m^{k+1}$ and $\mathbf{s}_m^{k+1}$ for all $m \in \mathcal{T}_j \triangleq \{(j-1)M_t + i | i \in \mathbb{N}_{M_t}^+\}$. The fusion center updates $\mathbf{y}$ and μ. Thus, the computation cost is $\mathcal{O}(N^2 M_t)$ at each node and $\mathcal{O}(N M_t M_r)$ at the fusion center. One can see that the computations are distributed among all receive nodes. The computation and memory required by the fusion center is only linear in the dimension of sparse vector $\mathbf{y}$, which is significantly lower than that of the fusion center of Section VI-A. Thus, even one of the receivers can be assigned to serve as the fusion center, and perform the update of $\mathbf{y}$ and μ. In each iteration, each node communicates $\tilde{\mathbf{s}}_m^{k+1}$ and $\mathbf{y}_m^{k+1}$ to the fusion center. However, the communication load decreases in a few iterations, because the nonzero entries in $\tilde{\mathbf{s}}_m^{k+1}$ and $\mathbf{y}_m^{k+1}$ would be on the order of $\mathcal{O}(K)$.

A fully distributed scheme would also be possible, but would require consensus. However, consensus-based implementations converge slowly, which would be a problem in target estimation and tracking applications.

VII. DECOUPLED LOCATION AND SPEED ESTIMATION

Instead of jointly estimating the target location-speed in the discretized location-speed space $\boldsymbol{\Theta} = \boldsymbol{\Theta}_{loc} \times \boldsymbol{\Theta}_{spd}$ (dimension $N = N_1 \times N_2$) we can decouple the estimation into target location estimation and the speed estimation. As it will be shown, such decoupling lowers complexity and required fewer measurements.

A. The Decoupled Signal Model

First, we describe the sparse model in the discretized target location space, $\boldsymbol{\Theta}_{loc}$, of dimension N_1, where $N_1 \ll N$. For target location estimation, it suffices to collect the measurements from one pulse only. During the

---

Algorithm 1 Semi-Distributed Implementation

One peer receive node is chosen as the fusion center.

Input $\tilde{\boldsymbol{\Psi}}, \tilde{\mathbf{z}}, \lambda, \rho_1, \rho_2$

Initialization $\tilde{\mathbf{s}}^{(0)} = \mathbf{x}^{(0)} = \mathbf{y}^{(0)} = 0$, $\mu^{(0)} = \nu^{(0)} = 0$

Iteration

Fusion Center:

    compute $\mathbf{y}_n^{k+1}$, $\forall n \in \mathbb{N}_N^+$ by (32);

Node $j \in \mathbb{N}_{M_r}^+$: for all $m \in \mathcal{T}_j$

    download $\mathbf{y}_m^{k+1}$ from the fusion center;

    compute $\mathbf{x}_m^{k+1} = \mathcal{P}_\Omega(\tilde{\mathbf{s}}_m^{k+1} - \nu_m^k/\rho_2)$;

    compute $\tilde{\mathbf{s}}_m^{k+1}$ by solving (36);

    compute $\nu_m^{k+1} = \nu_m^k + \rho_2(\mathbf{x}_m^{k+1} - \tilde{\mathbf{s}}_m^{k+1})$;

    upload $\tilde{\mathbf{s}}_m^{k+1}$ to the fusion center;

Fusion Center:

    compute μ^{k+1} by (38b);

---

p-th pulse (for some fixed p), the sample vector $\mathbf{z}_{ij}^p \in \mathbb{C}^L$, corresponding to the (ij)-th TX/RX antenna pair node pair, equals

$$\mathbf{z}_{ij}^p = \boldsymbol{\Psi}_{ij}^p \mathbf{s}_{ij}^p + \mathbf{n}_{ij}^p, \ \forall (ij) \in \Xi, \tag{39}$$

where $\mathbf{n}_{ij}^p \in \mathbb{C}^L$ denotes the additive noise. The matrix $\boldsymbol{\Psi}_{ij}^p \in \mathbb{R}^{L \times N_1}$ has $\mathbf{w}_{i,\tau_{ij}^n}$ as its n-th column. The vector $\mathbf{s}_{ij}^p \in \mathbb{C}^{N_1}$ is K-sparse and its n-th entry equals $\mathbf{s}_{ij}^p(n) = \beta_{ij}^k e^{2\pi f_{ij}^k T(p-1)}$, if there is a target at the n-th grid point (here β_{ij}^k is the corresponding target reflectivity); otherwise it equals 0. For slowly moving targets, the Doppler effect can be approximated as constant during one pulse, thus, the Doppler effect here is absorbed into $\mathbf{s}_{ij}^p$. At the fusion center, the sample vector corresponding to the p-th pulse, formed based on all TX/RX pairs, $\mathbf{z}^p \in \mathbb{C}^{LM_t M_r}$, equals

$$\mathbf{z}^p = \left[(\mathbf{z}_{11}^p)^T, \ldots, (\mathbf{z}_{M_t M_r}^p)^T \right]^T = \boldsymbol{\Psi}^p \mathbf{s}^p + \mathbf{n}^p, \tag{40}$$

where $\mathbf{s}^p = \left[(\mathbf{s}_{11}^p)^T, \ldots, (\mathbf{s}_{M_t M_r}^p)^T \right]^T$, $\mathbf{n}^p = [(\mathbf{n}_{11}^p)^T, \ldots, (\mathbf{n}_{M_t M_r}^p)^T]^T$, and $\boldsymbol{\Psi}^p = \text{diag}(\boldsymbol{\Psi}_{11}^p, \ldots, \boldsymbol{\Psi}_{M_t M_r}^p)$.

The location vector $\mathbf{s}^p$ can be recovered by applying the L-OPT method of Section II. The recovery performance is given in the following proposition.

Proposition 2: Consider the location estimation model in (40). For any $\delta_{2K} < \sqrt{2} - 1$, if L is such that

$$L \geq 192 \delta_{2K}^{-1} K^2 \log(N_1 \sqrt{M_t M_r}) + 1, \tag{41}$$

then the error of the L-OPT solution, $\hat{\mathbf{s}}^p$, is bounded as $\|\hat{\mathbf{s}}^p - \mathbf{s}^p\|_2 \leq g\left(\epsilon_p / \sqrt{L\sigma_0^2}\right)$ with probability exceeding $(1 - 4(N_1 \sqrt{M_t M_r})^{-1})$.

The proof of Proposition 2 follows the same spirit as that of Proposition 1. The key is to show that $\boldsymbol{\Psi}^p$ satisfies the $\mathcal{A}_1$-RIP$(2K, \delta_{2K})$ given the conditions on L, which can be proven along the lines of Section IV.

Once we obtain the target locations from $\hat{\mathbf{s}}^p$, we can use them to reduce the dimension of the speed estimation problem. Let $\boldsymbol{\Theta}_{loc,\mathcal{I}}$ be the pruned target location space. If the number of targets, K, is known, then $\boldsymbol{\Theta}_{loc,\mathcal{I}}$ consists of the grid points corresponding to the K largest entries of the recovered $\mathbf{s}_{ij}^p$. If K is unknown, we keep a slightly larger portion of the location grid points in $\boldsymbol{\Theta}_{loc,\mathcal{I}}$. The sparse model in (8) can be applied for $\boldsymbol{\Theta}_{loc,\mathcal{I}} \times \boldsymbol{\Theta}_{spd}$ instead of the entire space $\boldsymbol{\Theta}$. On denoting $\tilde{K} \triangleq |\boldsymbol{\Theta}_{loc,\mathcal{I}}|$, it holds that $\tilde{K} \sim \mathcal{O}(K) \ll N_1$. The number of measurements can also be reduced due to the lower dimension of the location-speed space $\boldsymbol{\Theta}_{loc,\mathcal{I}} \times \boldsymbol{\Theta}_{spd}$. During each pulse, the (ij)-th TX/RX pair will only use $\tilde{L} \ll L$ measurements.

Let us denote by $\mathcal{I}_r$ the index set that selects $\tilde{L}P$ measurements during all P pulses, and by $\mathcal{I}_c$ the index set that selects $\boldsymbol{\Theta}_{loc,\mathcal{I}} \times \boldsymbol{\Theta}_{spd}$ out of $\boldsymbol{\Theta}$. The sparse model for speed estimation becomes

$$\mathbf{z}_{\mathcal{I}_r} = \boldsymbol{\Psi}_{\mathcal{I}} \mathbf{s}_{\mathcal{I}_c} + \mathbf{n}_{\mathcal{I}_r}, \tag{42}$$

where $\mathbf{z}_{\mathcal{I}_r} \in \mathbb{C}^{\tilde{L}PM_tM_r}$ consists of entries of $\{\mathbf{z}_{ij}\}_{(ij)\in\Xi}$ indexed by $\mathcal{I}_r$, and similarly for $\mathbf{s}_{\mathcal{I}_c} \in \mathbb{C}^{\tilde{K}N_2M_tM_r}$ and $\mathbf{n}_{\mathcal{I}_r} \in \mathbb{C}^{\tilde{L}PM_tM_r}$; $\boldsymbol{\Psi}_{\mathcal{I}} \in \mathbb{C}^{\tilde{L}PM_tM_r \times \tilde{K}N_2M_tM_r}$ is the corresponding block diagonal matrix. The diagonal blocks of $\boldsymbol{\Psi}_{\mathcal{I}}$ consist of rows and columns of $\{\boldsymbol{\Psi}_{ij}\}_{(ij)\in\Xi}$ respectively indexed by $\mathcal{I}_r$ and $\mathcal{I}_c$. Note that $\mathbf{s}_{\mathcal{I}_c}$ can be uniformly partitioned into M_tM_r K-sparse sub-vectors, which share the same sparse profile. The final location-speed estimation can be achieved based on $\mathbf{s}_{\mathcal{I}_c}$, by solving an L-OPT problem (see (42)).

B. Complexity and Discussion

The decoupled location and speed estimation scheme needs to solve two sparse recovery problems. If the interior point method is adopted to solve the L-OPT problems, the total computation cost would be $\mathcal{O}((N_1M_tM_r)^3 + (\tilde{K}N_2M_tM_r)^3)$. Recall that solving (8) using the interior point method requires cost of $\mathcal{O}((N_1N_2M_tM_r)^3)$. The computation saving comes from the lower dimensions of the decoupled scheme than that of the original problem in (8). Moreover, in (40), the measurement matrix $\boldsymbol{\Psi}^p$ is block diagonal, and the sparse vector $\mathbf{s}^p$ has group sparsity. Thus, the ADMM-based algorithm in Section VI can be used to recover $\mathbf{s}^p$. It is easy to show that the Gram matrix of $\boldsymbol{\Psi}^p$ is block diagonal, symmetric and positive semidefinite. The computation of the ADMM-based algorithm will be $\mathcal{O}(N_1^2M_tM_r)$. Similarly for (42), the ADMM-based algorithm can also be used to recover $\mathbf{s}_{\mathcal{I}_c}$ with cost $\mathcal{O}(\tilde{K}^2N_2^2M_tM_r)$. The decoupled location and speed estimation using the ADMM further reduces the computational complexity. The parallel technique (or semi-distributed implementation) can also be applied here to distribute the computations among multiple processors (or receiver nodes).

The decoupled scheme requires LM_tM_r measurements from the p-th pulse for location estimation and $(P-1)\tilde{L}M_tM_r$ from the rest $(P-1)$ pulses for the final speed estimation. Thus, the receivers can operate at sampling frequency $1/T_s$ during the first pulse, and reduce the sampling frequency to $\frac{\tilde{L}}{LT_s}$ thereafter. Compared to the joint estimation scheme, less measurements are needed.

Given fixed amount of measurements, the estimation performance is better for smaller location-speed space $\boldsymbol{\Theta}_{loc,\mathcal{I}} \times \boldsymbol{\Theta}_{spd}$ once it contains the location grid points possessed by targets. Since the dimension of $\boldsymbol{\Theta}_{loc,\mathcal{I}} \times \boldsymbol{\Theta}_{spd}$ is controlled by $\tilde{K}$, it is expected that the estimation error achieves the minimum when $\tilde{K} = K$ if the locations of K targets are estimated correctly in the first stage. We will show in the simulation that this happens even for reasonable low SNRs. It is also shown that the estimation error of the decouple scheme with partial measurements may even be lower than that of the joint scheme using all the measurements.

VIII. NUMERICAL RESULTS

We consider a MIMO radar system with M_t TX and M_r RX antennas, distributed uniformly on a circle of radius of $6,000m$ and $3,000m$, respectively. Each TX radar transmits pulses with pulse repetition interval 0.125 ms and $5GHz$ carrier frequency. The variance of Gaussian waveform is $\sigma_0^2 = 1$. Each RX radar works with sampling frequency of 50MHz on the received baseband signal, which are corrupted by zero-mean Gaussian noise with variance σ_n^2. The signal-to-noise ratio (SNR) is defined as $10\log_{10}(\sigma_0^2/\sigma_n^2)$.

The probing space is discretized on a $N_x \times N_y$ grid, starting from point $[8000m, 8000m]$ with grid spacing equal to $10m$. The velocity space in default is fixed as a uniform 4×1 grid on $v_x \in [100, 130]m/s$, $v_y = 100m/s$ *i.e.*, $N_{vx} = 4, N_{vy} = 1$, unless otherwise is stated. We randomly generate K targets on the grid. The magnitude of the complex reflection coefficients for each target in each trial is randomly generated from uniform distribution in the range of $[0.1, 0.8]$.

All the simulations are carried out on a PC with Intel Core i7 CPU and 8GB memory. The number of independent trials is 100 unless otherwise stated.

A. On The Number of Pulses P

We first illustrate the choice of the number of pulses P via the inequality $\beta(P)/P \leq \delta_{2K}/(2K+\delta_{2K})|_{\delta_{2K}=\sqrt{2}-1} \triangleq \gamma_0$ when only a single target is considered, *i.e.*, $K = 1$. We consider $M_t = 2, M_r = 4$. For the case of $N_x = 25, N_y = 4$, Fig. 1 shows values of $\phi_{ij}(P)/P$ for all TX/RX pairs under different values of P. We choose the smallest P such that the maximum of $\phi_{ij}(P)/P$, $\forall (ij) \in \Xi$, is smaller than γ_0, *i.e.*, $\beta(P)/P \leq \gamma_0$. Based on Proposition 1, this value guarantees the performance under the worst cases. In the following simulation, we will show that even a smaller P works well.

B. The Advantage of Exploiting Group Sparsity

It can be seen from Theorem 1 that exploiting the sparsity in the target vector reduces the required measurements, or equivalently, it improves the performance with the same amount of measurements. In this simulation, we evaluate the advantage of exploiting the sparsity structure in the target vector. We consider $M_t = 2, M_r = 4, N_x = 25$ and $N_y = 4$. For the proposed L-OPT based method in (23), we use the interior point method with $\epsilon_0 = 2\sqrt{LM_tM_r}\sigma_n$ [15]. For comparison, we implement BPDN which just minimizes the ℓ_1-norm of $\mathbf{s}$ and ignores the sparsity structure in $\mathbf{s}$. The constraint in BPDN is exactly the same as that for L-OPT method. The parameter ϵ_0 for BPDN is chosen
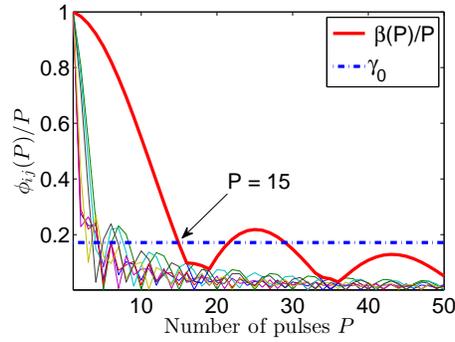
Fig. 1: Results on the choice of the number of pulses, P.
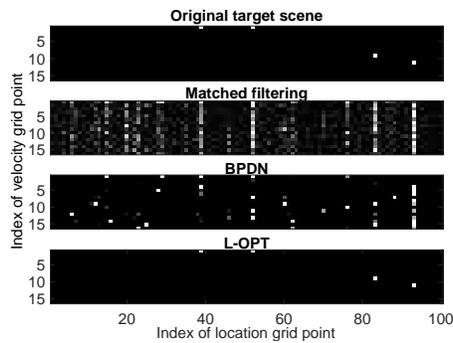
Fig. 2: An illustration of target scene estimation. The MIMO radar system has $M_t = 2$ receive and $M_r = 4$ transmit antennas. We sample $L = 50$ samples per pulse from $P = 6$ pulses. There are $K = 4$ targets. The target space of interest is with parameters $N_x = 25, N_y = 4, N_{vx} = 4$ and $N_{vy} = 4$.

(a)

(b)
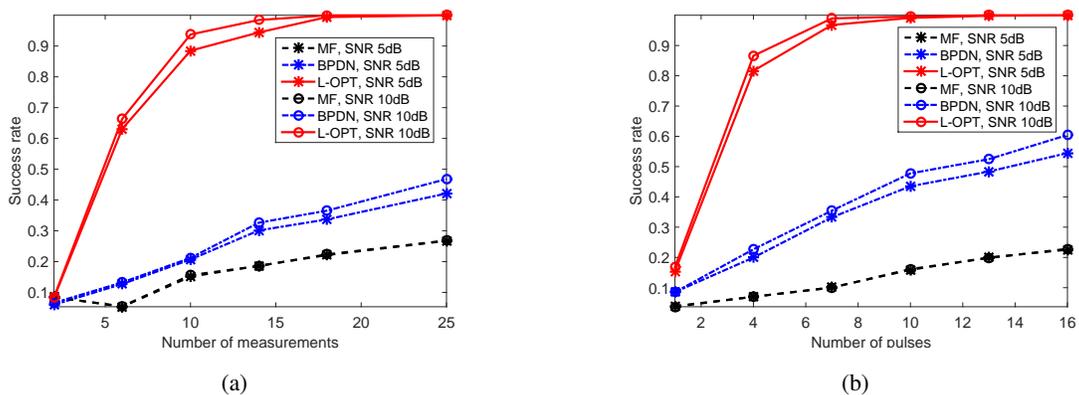
Fig. 3: Performance for the Matched Filtering (MF), the BPDN and the L-OPT methods. (a) Results under different number of measurements and SNRs, $K = 20, P = 3$; (b) Results under different number of pulses and SNRs, $K = 20, L = 6$.

the same as that for L-OPT. Also, the traditional matched filtering (MF) approach [30] is also implemented for comparison. The MF method uses the same L and P as the sparsity-based methods. The received signal is correlated with the transmitted signal distorted by different Doppler shifts and time delays.

An illustration of target scene estimation is presented in Fig. 2. The MIMO radar system has $M_t = 2$ receive and $M_r = 4$ transmit antennas. We sample $L = 50$ samples per pulse from $P = 6$ pulses. There are $K = 4$ targets. The target space of interest is with parameters $N_x = 25, N_y = 4, N_{vx} = 4$ and $N_{vy} = 4$. The traditional MF method has complexity of $\mathcal{O}(LPNM_tM_r)$, which is much lower compared to that of the BPDN and the L-OPT methods. However, the target scene estimate by the MF method is corrupted by lots of false peaks. The BPDN method achieves a relative cleaner target scene estimate, but the strong false peaks may still degrade the target estimation results. The L-OPT method utilizes the group sparsity in $\mathbf{s}$, which helps to remove the false peaks in the estimate. It can be seen from Fig. 2 that the L-OPT method indeed achieves the best target scene estimation.

In Fig. 3(a), we plot the successful recovery rates by the MF, the BPDN and the L-OPT methods under different number of measurements, L, and SNRs with $K = 20, P = 3$. The recovery rate of the L-OPT method drops dramatically if L is smaller than 10. This observation validates the claim in Proposition 1 that L should be larger than certain value to maintain a high probability of target location and speed estimation. Based on (24a) in Proposition 1, the bound on L is 5×10^5, which is much larger than the values of L here. As discussed in Remark 3, the L-OPT method performs well with much smaller L and P than those in (24). In Fig. 3(b), we plot the successful recovery rates under different number of pulses, P, with $K = 20, L = 6$. Similar observations can be made for the number of pulses. From both Fig. 3(a) and (b), the successful recovery rates of the L-OPT method are higher than those of the BPDN method under all the settings. As implied by Theorem 1, L-OPT outperforms BPDN because it exploits the sparsity structure in $\mathbf{s}$. In addition, the sparsity based methods, both the BPDN and the L-OPT methods, outperform the traditional MF method in terms of the success rate of target estimation.

C. Efficient Algorithm Based on The ADMM

In this section, we evaluate the efficient algorithm based on the ADMM proposed in Section VI-A. We consider the same simulation setting in Section VIII-B except that the SNR is set to be 5 dB. The dimension of the target vector in (26) is 6400×1 with $16 \times K$ nonzero entries. The BOMP [13] method, the GLasso-PGA method [14] and the L-OPT method using the interior point method (L-OPT-IPM) are implemented for comparison. For GLasso-PGA, we choose $\lambda = 0.02$ for the best performance. For L-OPT-IPM, we set $\epsilon_0 = 2\sqrt{LPM_tM_r}\sigma_n$ with knowledge of $\sigma_n^2 = \sigma_0^2 10^{10/SNR}$. For the proposed ADMM based method, preconditioned conjugate gradient is used to solve the system of (35). The estimation error $\|\hat{\mathbf{s}} - \mathbf{s}\|_2$ and the CPU running time are used as the performance metrics. All results are averaged over 100 independent trials.

We first fix $K = 10, P = 3$ and evaluate the root of total squared error $\|\hat{\mathbf{s}} - \mathbf{s}\|_2$ under different number of measurements L. The results are plotted in Fig. 4. The proposed algorithm, labeled as ADMM-based in the legend, achieves lower estimation errors with less CPU run time as compared to GLasso-PGA and L-OPT-IPM under all L's. The CPU run time of the proposed algorithm remains less than $10s$, while the run time of L-OPT-IPM grows
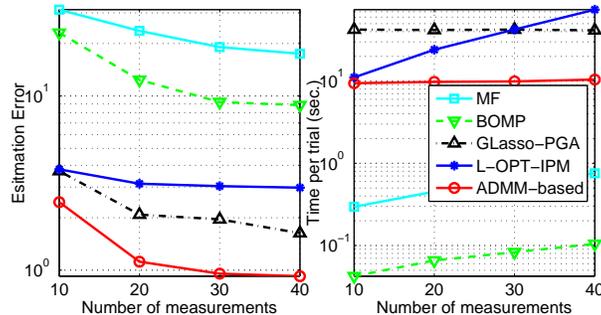
Fig. 4: Performance under different number of measurements with $K = 10$ and $P = 3$; for GLasso-PGA $\lambda = 0.02$; and for the proposed ADMM-based method $\lambda = 4$, $\rho_1 = \rho_2 = 1$.

superlinearly with L. We should note that the L-OPT-IPM solves the problem of (23), while the ADMM method solves the problem of (29). Compared to the problem of (23), there is an additional bounded constraint in (29), which changes the optimization problem. The additional constraint in (29) imposes a smaller feasible set, which may introduce the denoising effect because the noise-corrupted candidates outside the feasible set are excluded. In contrast, the solution of (23) obtained by the L-OPT-IPM method may fall out of the feasible set in (29). As shown in Fig. 4, the additional constraint in (29) improves the accuracy of the solution with the same L and P over the L-OPT-IPM method.

Next, we consider the performance of the proposed scheme for different number of targets and fixed number of measurements $L = 20$ and pulses $P = 3$. The results are plotted in Fig. 5. For all values of K, the proposed ADMM-based algorithm achieves lower estimation errors than the GLasso-PGA method using around one quarter CPU running time.

Fig. 6 shows the performance of the proposed scheme for different number of pulses and fixed number of measurements $L = 10$ and targets $K = 10$. The ADMM-based method achieves the smallest estimation errors, and takes much less CPU runtime than the GLasso-PGA method does.

It is noted that the MF method takes much less CPU runtime compared to the sparsity based methods because its complexity is $\mathcal{O}(LPNM_tM_r)$. However, the sparsity based methods achieve lower estimation errors and higher success rates at the cost of complexity increase.

In the above simulation, L-OPT-IPM requires knowledge of the noise variance σ_n^2. The regularization parameter λ in GLasso-PGA and the proposed algorithm also need to be manually tuned. In fact, the choice of such parameters are critical for the estimation performance. In Fig. 7, we plot the estimation errors for a wide range of λ. We observe that the estimation error of the proposed ADMM-based algorithm remains very small for a wide range of λ's, while for GLasso-PGA, the range of good λ's is very narrow. The robustness to λ makes the proposed algorithm good candidate for real world applications.
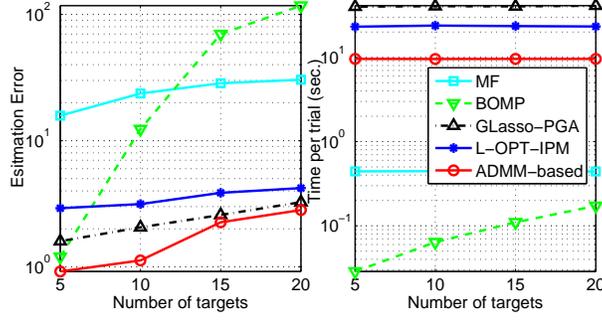
Fig. 5: Performance under different number of targets with $L = 20$ and $P = 3$; for GLasso-PGA $\lambda = 0.02$; and for the proposed ADMM-based method $\lambda = 4$, $\rho_1 = \rho_2 = 1$.
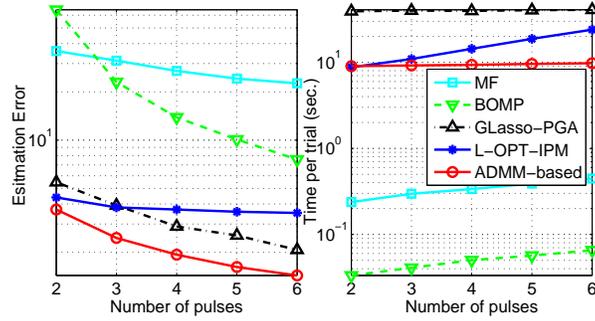
Fig. 6: Performance under different number of pulses with $K = 10$ and $L = 10$; for GLasso-PGA $\lambda = 0.02$; and for the proposed ADMM-based method $\lambda = 4$, $\rho_1 = \rho_2 = 1$.

D. The Performance of the Decoupled Scheme

In this section, we evaluate the performance of the decouple scheme proposed in Section VII under different values of $\tilde{L}$ and $\tilde{K}$. Recall that $\tilde{K}$ is the number of location grid points kept after the location estimation stage, and $\tilde{L}$ is the number of measurements used to achieve final location and speed estimation in the second stage. We aim to illustrate the extent to which $\tilde{K}$ and $\tilde{L}$ can be reduced. For comparison, the joint location and speed estimation
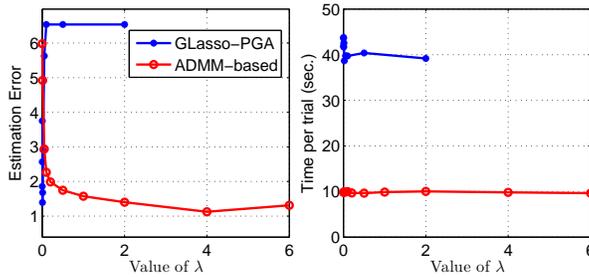
Fig. 7: Performance under different values of λ with $K = 10, L = 20$ and $P = 3$.

is implemented using the ADMM-based algorithm in Section VI, which is referred to as the joint scheme. We consider $M_t = 2, M_r = 4, N_x = 25$ and $N_y = 4$. The number of targets, measurements and pulses are fixed as $K = 10, L = 50$, and $P = 3$. The performance metrics used in this subsection are the successful recovery rate, root of the total squared error, and the CPU running time.

Fig. 8 shows the performance under different values of $\tilde{L}$. The performance of the joint scheme remains constant and serves as the reference because it is not affected by $\tilde{L}$. As for the decoupled scheme, the successful recovery rate goes up if more number of measurements are used. Also, the estimation accuracy of **s** increases when more measurements are used. The success rate and the estimation error of the decoupled scheme become very close to that of the joint scheme when only $\tilde{L} = 20$ out of $L = 50$ measurements (40%) are used. The location and speed estimation is correct with high probability even when we only keep 20% the measurements. The running time of the decoupled scheme is less than one tenth of that of the joint scheme. Note that the decoupled scheme estimation error becomes even smaller than that of the joint scheme when $\tilde{L}$ is large enough. It is because that the dimension of the location-speed space in the second stage of the decoupled scheme is much smaller than that used in the joint scheme, *i.e.*, $\tilde{K}N_2 \ll N_1N_2$.

The performance results under different values of $\tilde{K}$ are plotted in Fig. 9. The success rate results in Fig. 9(a) show that the success rate of the decouple scheme is 100% even if $\tilde{K} = K$. This means that the location estimation is accurate and stable in noise (5dB and 0dB are shown in the figure). The estimation error results in Fig. 9(b) indicate that the error is large if $\tilde{K}$ is smaller that K. If $\tilde{K}$ is set properly larger than K, the estimation error of the decoupled scheme is lower than that of the joint scheme. Also, note that once the correct location grid points are kept, the smaller the $\tilde{K}$ is, the smaller the dimension of the pruned space in the decoupled scheme is, and thus the smaller the estimation error is.

Based on the simulations in the subsection, we conclude that the decoupled scheme can reduce both the computation and the required number of measurements, while maintaining high estimation accuracy in practice.
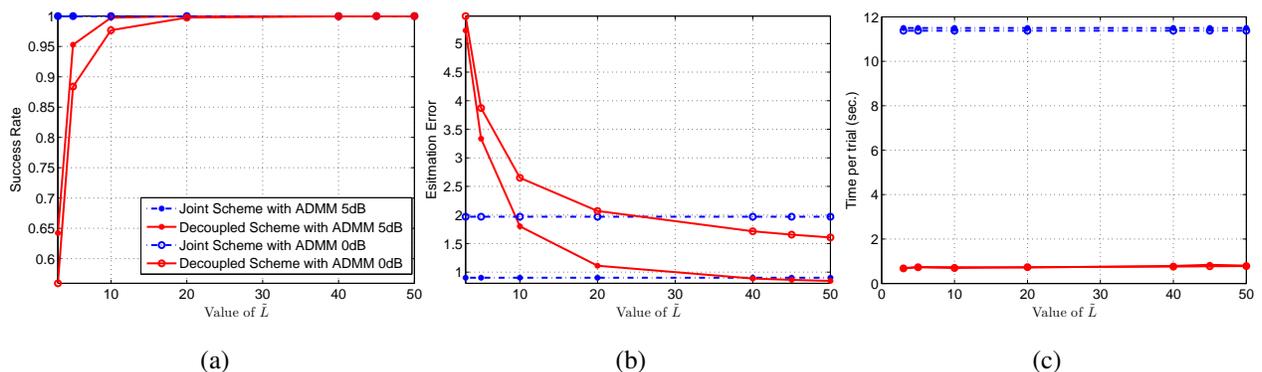
Fig. 8: Performance of the decoupled scheme under different values of $\tilde{L}$ with $\tilde{K} = 2K$. (a) the successful recovery rate, (b) the estimation error, and (c) the CPU running time per trial in seconds.

(a)   (b)   (c)
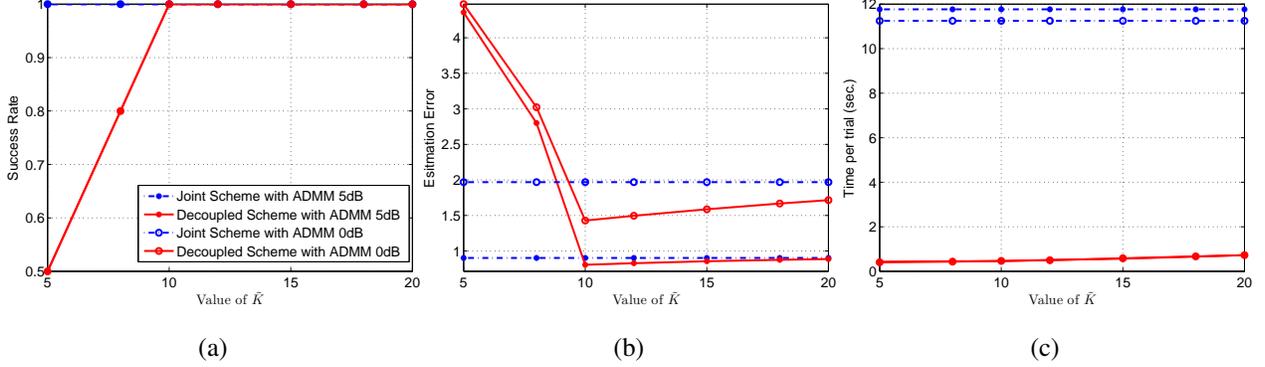
Fig. 9: Performance of the decouple scheme under different values of $\tilde{K}$ with $\tilde{L} = 0.8L$. (a) the successful recovery rate, (b) the estimation error, and (c) the CPU running time per trial in seconds.

E. Off-grid targets and grid refinement

In all the previous simulations, all targets are assumed to be on the grid. In this section, we consider the case in which the targets may be between the grid points. For a certain target space, a dense grid would generate a problem with large dimension, the solution of which would demand high computational cost. On the other hand, a coarse grid would introduce large estimation errors. One could use the grid refinement scheme [27] to reduce the complexity while maintaining estimation performance. Let us consider the location and velocity estimation of $K = 4$ moving targets. The true target parameters are given in the second row of Table I. The target space of interest is $x \in [8000, 8200], y \in [8000, 8100], v_x \in [100, 180]$ and $v_y \in [100, 140]$. In a dense uniform grid $\boldsymbol{\Theta}_d$ with $N_x = 41, N_y = 21, N_{vx} = 9$ and $N_{vy} = 5$, the location and velocity grid spacings would be $5m$ and $10m/s$, respectively, for which all targets fall on the grid. However, the corresponding total number of grid points and the dimension of the sparse target vector $\mathbf{s}$ are $N = 38745$ and $N M_t M_r = 309960$, respectively. It is too time demanding to estimate targets using $\boldsymbol{\Theta}_d$. We instead use a three-level grid refinement scheme. An initial coarse grid $\boldsymbol{\Theta}_c$ with $N_x = 11, N_y = 6, N_{vx} = 5$ and $N_{vy} = 3$ is used to discretize the target space, which reduces the total number of grid points to 990. The location and velocity grid spacings are $20m$ and $20m/s$, respectively. Note that for this grid, all four targets fall between grid points. The first round estimate $\hat{\mathbf{s}}_1$ is obtained from the reconstructed target vector using the L-OPT method and $\boldsymbol{\Theta}_c$. A refined grid $\boldsymbol{\Theta}_r$ is generated in the neighborhood of the dominant peaks in $\hat{\mathbf{s}}_1$ with location grid spacing $10m$ and velocity grid spacing $10m/s$. The total number of grid points in the refined grid is 2153, which is still very small compared to that of $\boldsymbol{\Theta}_d$. Note that all four targets still fall off the refined grid. The second round estimate $\hat{\mathbf{s}}_2$ is obtained and the grid refinement procedure is repeated for a second time. Finally, we obtain the third round estimate $\hat{\mathbf{s}}_3$ using a further refined grid $\boldsymbol{\Theta}_{r'}$ with 2924 grid points, $5m$ location grid spacing and $10m/s$ velocity grid spacing. The true and estimated target scenes in the location space are shown in Fig. 10, where the vectors $\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2$ and $\hat{\mathbf{s}}_3$ are mapped onto grid $\boldsymbol{\Theta}_d$ by interpolating with zeroes. The estimated target locations and velocities are given in Table I. We observe that the grid refinement scheme can effectively reduce the computational complexity and achieve accurate target estimation. Also, in the first two
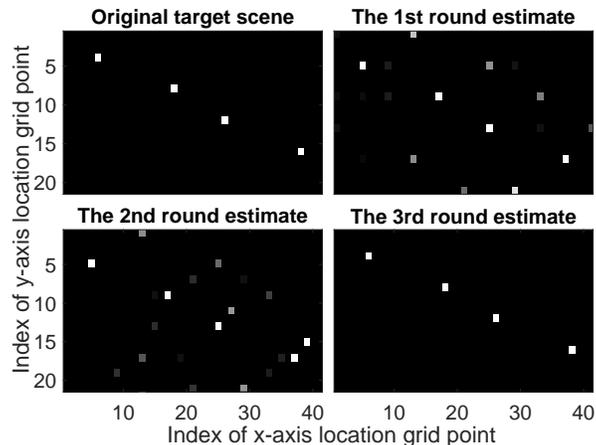
Fig. 10: The estimated target scene in location space by a three-level grid refinement scheme.

TABLE I: The estimated target locations and velocities by a three-level grid refinement scheme. The results are given in the form of $[x, y, v_x, v_y]$ with metrics m and m/s respectively for location and velocity.

	Target 1	Target 2	Target 3	Target 4
True	[8025,8015,120,100]	[8085,8035,140,120]	[8125,8055,160,120]	[8185,8075,180,140]
1st estimate	[8020,8020,120,100]	[8080,8040,140,120]	[8120,8060,160,120]	[8180,8080,180,140]
2nd estimate	[8020,8020,120,100]	[8080,8040,140,120]	[8120,8060,160,120]	[8180,8080,180,140]
3rd estimate	[8025,8015,120,100]	[8085,8035,140,120]	[8125,8055,160,120]	[8185,8075,180,140]

rounds, the off-grid targets are captured by the closest grid points. This indicates that the proposed methods are robust to off-grid targets.

IX. CONCLUSIONS

We have considered moving target estimation using distributed, sparsity based MIMO radars. We have provided the uniform recovery guarantee by analyzing the $\mathcal{A}$-RIP of the block diagonal measurement matrix. The proposed theoretical results validate that the structures in both $\mathbf{\Psi}$ and $\mathbf{s}$ result in reduction of the number of measurements needed, or result in improved target estimation for the same L.

Two low-complexity approaches have been proposed to reduce the computation while maintaining the estimation performance. The first approach was an ADMM-based sparse signal recovery algorithm. Simulation results have indicated that this approach significantly lowers the computational complexity for target estimation with improved accuracy as compared to the approaches using proximal gradient algorithm and interior point method. The second approach decouples the location and speed estimation into two separate stages. The location estimation obtained in the first stage is used to prune the target location-speed space in the speed estimation stage. Simulations have indicated that the decoupled scheme can reduce both the computation and the required number of measurements, while maintaining high estimation accuracy.

APPENDIX A

PROOF OF THEOREM 1

Proof: Here we only focuss on the bounds for the off-diagonal entries in the Gram of $\bar{\Psi}$, $\mathbf{G} = \bar{\Psi}^H \bar{\Psi}$. For the diagonal entries, *i.e.*, $n = l$ as in case (i), the union bound can be easily obtained based on (13).

The off-diagonal entries may be from either case (ii) or case (iii). In order to arrive at a uniform union bound, we need to unify the bounds in (17) and (19) for these two cases. Inequality (17) for case (ii) can be relaxed as

$$\Pr\left(|G_{ij}(n,l)| > t\right) \leq 4\exp\left(-\frac{L-1}{16}t^2\right). \tag{43}$$

Under condition (20), the probabilistic bound in (19) for case (iii) can be relaxed to the same as that in (43) for case (ii).

Following the classical procedure of RIP analysis in [17], we need to evaluate the radii of the Gergošin's discs for the sub-matrix of $\mathbf{G}$ constructed based on the support of $\mathbf{s}$. Recall that $\bar{\Psi}$ is block diagonal and there are only K nonzero entries in each sub-vector $\mathbf{s}_{ij}$ in $\mathbf{s}$. Therefore, there are only $(K-1)$ instead of $(K-1)M_tM_r$ off-diagonal entries contributing to the radii of the Gergošin's discs. This reduction comes from the BD structure of $\tilde{\Psi}$ and the sparsity profile of $\mathbf{s}$ characterized by $\mathcal{A}_1^K$. Here we choose $\delta_d \triangleq \frac{\delta_K}{K}$ and $\delta_o \triangleq \frac{(K-1)\delta_K}{K}$. Substituting t with $\delta_o/(K-1)$ in (43) gives the unified bound for any of the off-diagonal entries, *i.e.*,

$$\Pr\left(|G_{ij}(n,l)| > \frac{\delta_K}{K}\right) \leq 4\exp\left(-\frac{(L-1)\delta_K^2}{16K^2}\right). \tag{44}$$

under the condition of (21b), which is derived by substituting $t = \delta_K/K$ into (20). Following the steps of [17] the proof follows. ∎

REFERENCES

[1] Li, J., and Stoica, P., "MIMO radar with colocated antennas," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 106–114, 2007.

[2] Li, J., and Stoica, P., Xu, L., and Roberts, W., "On parameter identifiability of MIMO radar," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 968–971, 2007.

[3] Haimovich, A. M., Blum, R. S., and Cimini, L. J., "MIMO radar with widely separated antennas," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 116–129, 2008.

[4] Godrich, H., Haimovich, A. M., and Blum, R. S., "Target localization accuracy gain in MIMO radar-based systems," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2783–2803, 2010.

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

[6] Baraniuk, R., Davenport, M., DeVore, R., and Wakin, M., "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[7] Eldar, Y. C., and Kutyniok, G., *Compressed sensing: theory and applications*, Cambridge University Press, 2012.

[8] Strohmer, T., and Friedlander, B., "Analysis of Sparse MIMO Radar," *Applied and Computational Harmonic Analysis*, vol. 37, no. 3, pp. 361–388, Nov. 2014.

[9] Hgel, M., Rauhut, H., and Strohmer, T., "Remote sensing via ℓ_1 minimization," *Foundations of Computational Mathematics*, vol. 14, no. 1, pp. 115–150, 2013.

[10] Rossi, M., Haimovich, A. M., and Eldar, Y. C., "Conditions for target recovery in spatial compressive sensing for MIMO radar," in *Proceedings of the IEEE International Conference on Acoustic, Speech, and Signal Processing*, 2013, pp. 4115–4119.

[11] Li, B., and Petropulu, A. P., "RIP analysis of the measurement matrix for compressive sensing-based MIMO radars," in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2014, pp. 497–500.

[12] Yu, Y., Petropulu, A. P., and Poor, H. V., "MIMO Radar Using Compressive Sampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp.146–163, Feb. 2010.

[13] Gogineni, S., and Nehorai, A., "Target estimation using sparse modeling for distributed MIMO radar," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5315–5325, 2011.

[14] Petropulu, A. P., Yu, Y., and Huang, J., "On exploring sparsity in widely separated MIMO radar," in *Proceedings of the 45th Asilomar Conference on Signals, Systems, and Computers*, 2011, pp. 1496–1500.

[15] Li, B., and Petropulu, A. P., "Performance guarantees for distributed MIMO radar based on sparse modeling," in *Proceedings of the IEEE Radar Conference*, 2014, pp. 1369–1372.

[16] Eldar, Y. C., and Mishali, M., "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, 2009.

[17] Haupt, J., Bajwa, W. U., Raz, G., and Nowak, R., "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.

[18] Yu, Y., Petropulu, A. P., and Poor, H. V., "CSSF MIMO RADAR: Compressive-Sensing and Step-Frequency Based MIMO Radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp.1490–1504, Apr. 2012.

[19] Duarte, M. F., and Eldar, Y. C., "Structured compressed sensing: from theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053-4085, Sep. 2011.

[20] Meier, L., Geer, S., and Buhlmann, P., "The group lasso for logistic regression," *Journal of the Royal Statistical Society: Series B*, vol. 70, no. 1, pp. 53–71, 2008.

[21] Eldar, Y. C., Kuppinger, P., and Bolcskei, H., "Block-sparse signals: uncertainty relations and efficient recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp.3042–3054, June 2010.

[22] Blumensath, T., and Davies, M. E., "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.

[23] Li, B., and Petropulu, A. P., "Structured sampling of structured signals," in *Proceedings of the IEEE Global Conference on Signal and Information Processing*, 2013, pp. 1009–1012.

[24] Li, B., and Petropulu, A. P., "Efficient target estimation in distributed MIMO radar via the ADMM," in *Proceedings of the 48th Conference on Information Sciences and Systems*, 2014, pp. 1–5.

[25] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., "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.

[26] Wang, Y., Yang, J., Yin, W., and Zhang, Y., "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

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

[28] Bertsekas, D. P., Nedic, A., and Ozdaglar, A. E., *Convex Optimization Theory*, Athena Scientific, Belmont, June 2009.

[29] Rauhut, H., "Compressive sensing and structured random matrices," in *Theoretical foundations and numerical methods for sparse recovery* Chapter 9, pp. 1–92, 2010.

[30] Richards, M. A., *Fundamentals of Radar Signal Processing*, McGraw-Hill, New York, 2005.

[31] Chen, C., and Vaidyanathan, P. P., "MIMO radar ambiguity properties and optimization using frequency-hopping waveforms," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp.5926–5936, Dec. 2008.

[32] Yu, Y., Sun, S., and Petropulu, A. P., "A capon beamforming method for clutter suppression in colocated compressive sensing based MIMO radars", in *Proceedings of SPIE 8717, Compressive Sensing II*, 87170J, 2013.

[33] Hunt, A. R., "Use of a frequency-hopping radar for imaging and motion detection through walls," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, pp. 1402–1408, May 2009.

[34] Ahmad, F., and Amin, M. G., "Through-the-wall human motion indication using sparsity-driven change detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51 , no. 2, pp. 881–890, 2013.