Analysis of Sparse MIMO Radar

Thomas Strohmer*
Dept. Mathematics
UC Davis
Davis CA 95616

Benjamin Friedlander †
Dept. Elec. Eng.
UC Santa Cruz
Santa Cruz, CA 95064

Abstract

We consider a multiple-input-multiple-output radar system and derive a theoretical framework for the recoverability of targets in the azimuth-range domain and the azimuth-range-Doppler domain via sparse approximation algorithms. Using tools developed in the area of compressive sensing, we prove bounds on the number of detectable targets and the achievable resolution in the presence of additive noise. Our theoretical findings are validated by numerical simulations.

Keywords: *Sparsity, Radar, Compressive Sensing, Random Matrix, MIMO*

1 Introduction

While radar systems have been in use for many decades, radar is far from being a 'solved problem'. Indeed, exciting new developments in radar pose great challenges both to engineers and mathematicians [6]. Two such developments are the advent of MIMO (multi-input multi-output) radar [10], and the application of compressed sensing to radar signal processing [15].

MIMO radar is characterized by using multiple antennas to simultaneously transmit diverse, usually orthogonal, waveforms in addition to using multiple antennas to receive the reflected signals. MIMO radar has the potential for enhancing spatial resolution and improving interference and jamming suppression. The ability of MIMO radar to shape the transmit beam post facto allows for adapting the transmission based on the received data in a way which is not possible in non-MIMO radar.

A radar system illuminates a given area and attempts to detect and determine the location of objects of interest in its field of view, and to estimate their strength (radar reflectivity). The space of interest may be divided into range-azimuth (distance and direction) cells, or range-Doppler-azimuth (distance, direction and speed) cells in the case there is relative motion between the radar

*T.S. was supported by the National Science Foundation under grant DMS-0811169 and by DARPA under grant N66001-11-1-4090.

†B.F. was supported by the National Science Foundation under grant CCF-0725366.

and the object. In many cases the radar scene is sparse in the sense that only a small fraction (often a *very* small fraction) of the cells is occupied by the objects of interest.

Conventional radar processing does not take into account the a-priori knowledge that the radar scene is sparse. Recent works, such as [15, 21] developed techniques which attempt to exploit this sparsity using tools from the area of compressed sensing [4, 8]. The exploitation of sparsity has the potential to improve the performance of radar systems under certain conditions and is therefore of considerable practical interest.

In this paper we study the issue of sparsity in the specific context of a MIMO radar system employing multiple antennas at the transmitter the receiver, where the two arrays are co-located. We note that related work on the application of compressive sensing techniques to MIMO radar can be found in [30, 31]. Our emphasis here is on developing the basic theory needed to apply sparse recovery techniques for the detection of the locations and reflectivities of targets for MIMO radar.

The basic model for the problem we are considering involves a linear measurement equation $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ where $\mathbf{y}$ is a vector of measurements collected by the receiver antennas over an observation interval, $\mathbf{A}$ is a measurement matrix whose columns correspond to the signal received from a single unit-strength scatterer at a particular range-azimuth (or range-azimuth-Doppler) cell, $\mathbf{x}$ is a vector whose elements represent the complex amplitudes of the scatterers, and $\mathbf{w}$ is a noise vector. The measurement equation is assumed to be under-determined, possibly highly under-determined. The sparsity of the radar scene is introduced by assuming that only K elements of the vector $\mathbf{x}$ are non-zero, where K is much smaller than the dimension of the vector. The measurement matrix $\mathbf{A}$ embodies in it the details of the radar system such as the transmitted waveforms and the structure of antenna array.

In this paper we study the conditions under which this problem has a satisfactory solution. This is a fundamental issue of both theoretical and practical importance. More specifically, the analysis presented in the following sections addresses the following issues:

- It is known from the theory of compressed sensing [4, 8] that the matrix $\mathbf{A}$ must satisfy certain conditions in order that the solution computed via an appropriate convex program will indeed coincide with the desired sparsest solution (whose computation is in general an NP-hard problem). In our problem the characteristics of this matrix depend on the choice of the radar waveforms and the number and positions of the transmit and receive antennas. We develop the results necessary for understanding how the selection of the parameters of the radar system affects the conditions mentioned above.

- The ability of the algorithm to correctly detect targets depends on the number of these targets, K, and the signal to noise ratio. We show that as long as the number of the targets is less than a maximal value $K_{\max}$, and the signal to noise is larger than some minimal value $\mathrm{SNR}_{\min}$, the targets can be correctly detected with high probability by solving an ℓ_1-regularized least squares problem known under the name *lasso*. Explicit formulas are presented for $K_{\max}$ and $\mathrm{SNR}_{\min}$ as a function of the number of transmit and receive antennas and the number of azimuth and range cells.

The structure of the paper is as follows. Subsection 1.1 introduces notation used throughout the paper. In Section 2 we describe the problem formulation and the setup. We derive conditions for the recovery of targets in the Doppler-free case in Section 3, and the case of detecting targets in presence of Doppler is analyzed in Section 4. Our theoretical results are supported by numerical simulations, see Section 5. We conclude in Section 6. Finally, some auxiliary results are collected in the appendices.

1.1 Notation

Let $\mathbf{v} \in \mathbb{C}^n$. As usual, we define $\|\mathbf{v}\|_1 := \sum_{k=1}^n |\mathbf{v}_k|$ and $\|\mathbf{v}\|_2 := \sqrt{\sum_{k=1}^n |\mathbf{v}_k|^2}$. For a given matrix $\mathbf{A}$ we denote its k-th column by $\mathbf{A}_k$ and the element in the i-th row and k-th column by $\mathbf{A}_{[i,k]}$. The operator norm of $\mathbf{A}$ is the largest singular value of $\mathbf{A}$ and is denoted by $\|\mathbf{A}\|_{\text{op}}$, the Frobenius norm of $\mathbf{A}$ is $\|\mathbf{A}\|_F = \sqrt{\sum_{i,k} |\mathbf{A}_{[i,k]}|^2}$. The coherence of $\mathbf{A}$ is defined as

$$\mu(\mathbf{A}) := \max_{k \neq l} \frac{|\langle \mathbf{A}_k, \mathbf{A}_l \rangle|}{\|\mathbf{A}_k\|_2 \|\mathbf{A}_l\|_2}. \tag{1}$$

For $\mathbf{x} \in \mathbb{C}^n$, let $\mathbf{T}_\tau$ denote the circulant translation operator, defined by

$$\mathbf{T}_\tau \mathbf{x}(l) = \mathbf{x}(l - \tau), \tag{2}$$

where $l - \tau$ is understood modulo n, and let $\mathbf{M}_f$ be the modulation operator defined by

$$\mathbf{M}_f \mathbf{x}(l) = \mathbf{x}(l) e^{2\pi i l f}. \tag{3}$$

2 Problem formulation and signal model

We refer to [24, 6] for the mathematical foundations of radar and to [18] for an introduction to MIMO radar. However, the reader needs only a very basic knowledge of the mathematical concepts underlying radar to be able to follow our approach.

We consider a MIMO radar employing N_T antennas at the transmitter and N_R antennas at the receiver. We assume that the element spacing is sufficiently small so that the radar return from a given scatterer is fully correlated across the array. In other words, this is a coherent propagation scenario.

To simplify the presentation we assume that the two arrays are co-located, i.e. this is a mono-static radar. The extension to the bi-static case is straightforward as long as the coherency assumption holds for each array. The arrays are characterized by the array manifolds: $\mathbf{a}_R(\beta)$ for the receive array and $\mathbf{a}_T(\beta)$ for the transmit array, where $\beta = \sin(\theta)$ is the direction relative to the array. We assume that the arrays and all the scatterers are in the same 2-D plane. The extension to the 3-D case is straightforward and all of the following results hold for that case as well.

For convenience we formulate our theorems and analysis in terms of delay τ instead of range r. This is no loss of generality, as delay and range are related by $\tau = 2r/c$, with c denoting the speed of light.

2.1 The model for the azimuth-delay domain

The i-th transmit antenna repeatedly transmits the signal $s_i(t)$. Let $\mathbf{Z}(t; \beta, \tau)$ be the $N_R \times N_t$ noise-free received signal matrix from a unit strength target at direction β and delay τ, where N_t

is the number of samples in time. Then

$$\mathbf{Z}(t;\beta,\tau) = \mathbf{a}_R(\beta)\mathbf{a}_T^T(\beta)\mathbf{S}_\tau^T,$$

where $\mathbf{S}_\tau$ is an $N_t \times N_T$ matrix whose columns are the circularly delayed signals $s_i(t-\tau)$, sampled at the discrete time points $t = n\Delta_t, n = 1,\ldots,N_t$. If $\tau = 0$, we often write simply $\mathbf{S}$ instead of $\mathbf{S}_0$.

Assuming uniformly spaced linear arrays, the array manifolds are given by

$$\mathbf{a}_T(\beta) = \begin{bmatrix} 1 \\ e^{j2\pi d_T\beta} \\ \vdots \\ e^{j2\pi d_T\beta(N_T-1)} \end{bmatrix} \tag{4}$$

and

$$\mathbf{a}_R(\beta) = \begin{bmatrix} 1 \\ e^{j2\pi d_R\beta} \\ \vdots \\ e^{j2\pi d_R\beta(N_R-1)} \end{bmatrix} \tag{5}$$

where d_T and d_R are the normalized spacings (distance divided by wavelength) between the elements of the transmit and receive arrays, respectively.

The spatial characteristics of a MIMO radar are closely related to that of a virtual array with $N_T N_R$ antennas, whose array manifold is $\mathbf{a}(\beta) = \mathbf{a}_T(\beta)\otimes\mathbf{a}_R(\beta)$. It is known [11] that the following choices for the spacing of the transmit and receive array spacing will yield a uniformly spaced virtual array with half wavelength spacing:

$$\begin{aligned} d_R = 0.5, d_T = 0.5N_R; \\ d_T = 0.5, d_R = 0.5N_T. \end{aligned} \tag{6}$$

Both of these choices lead to a virtual array whose aperture is $0.5(N_T N_R - 1)$ wavelengths. This is the largest virtual aperture free of grating lobes. The choices (6) and (7) will also show up in our theoretical analysis, e.g. see Theorem 1.

Next let $\mathbf{z}(t;\beta,\tau) = \text{vec}\{\mathbf{Z}\}(t;\beta,\tau)$ be the noise-free vectorized received signal. We set up a discrete delay-azimuth grid $\{(\beta_i,\tau_j)\}, 1 \le i \le N_\beta, 1 \le j \le N_\tau$, where Δ_β and Δ_τ denote the corresponding discretization stepsizes. Using vectors $\mathbf{z}(t;\beta_i,\tau_j)$ for all grid points (β_i,τ_j) we construct a complete response matrix $\mathbf{A}$ whose columns are $\mathbf{z}(t;\beta_i,\tau_j)$ for $1 \le i \le N_\beta$ and $1 \le j \le N_\tau$. In other words, we have N_τ delay values and N_β azimuth values, so that $\mathbf{A}$ is a $N_R N_t \times N_\tau N_\beta$ matrix.

Assume that the radar illuminates a scene consisting of K scatterers located on K points of the (β,τ_j) grid. Let $\mathbf{x}$ be a sparse vector whose non-zero elements are the complex amplitudes of the scatterers in the scene. The zero elements corresponds to grid points which are not occupied by scatterers. We can then define the radar signal $\mathbf{y}$ received from this scene by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v} \tag{7}$$

where $\mathbf{y}$ is a $N_R N_t \times 1$ vector, $\mathbf{x}$ is a $N_\tau N_\beta \times 1$ sparse vector, $\mathbf{v}$ is a $N_R N_t \times 1$ complex Gaussian noise vector, and $\mathbf{A}$ is a $N_R N_t \times N_\tau N_\beta$ matrix.

2.2 The model for the azimuth-delay-Doppler domain

The discussion so far was for the case of a stationary radar scene and a fixed radar, in which case there is no Doppler shift. The extension of this signal model to include the Doppler effect is conceptually straightforward, but leads to a significant increase in the problem dimension.

The signal model for the return from a unit strength scatterer at direction β, delay τ, and Doppler f (corresponding to its radial velocity with respect to the radar) is given by

$$\mathbf{Z}(t;\beta,\tau,f) = \mathbf{a}_R(\beta)\mathbf{a}_T^T(\beta)\mathbf{S}_{\tau,f}^T,$$

where $\mathbf{S}_{\tau,f}$ is a $N_t \times N_T$ matrix whose columns are the circularly delayed and Doppler shifted signals $s_i(t-\tau)e^{j2\pi ft}$.

As before we let $\mathbf{z}(t;\beta,\tau,f) = \text{vec}\{\mathbf{Z}\}(t;\beta,\tau,f)$ be the noise-free vectorized received signal. We extend the discrete delay-azimuth grid by adding a discretized Doppler component (with stepsize Δ_f and corresponding Doppler values $f = k\Delta_f, k = 1,\dots,N_f$) and obtain a uniform delay-azimuth-Doppler grid $\{(\beta_i,\tau_j,f_k)\}$. Using vectors $\mathbf{z}(t;\beta_i,\tau_j,f_k)$ for all discrete (β_i,τ_j,f_k) we construct a complete response matrix $\mathbf{A}$ whose columns are $\mathbf{z}(t;\beta_i,\tau_j,f_k)$ for $1 \le i \le N_\beta$, $1 \le j \le N_\tau$, $1 \le k \le N_f$.

Assume that the radar illuminates a scene consisting of K scatterers located on K points of the (β,τ_j,f_k) grid. Let $\mathbf{x}$ be a sparse vector whose non-zero elements are the complex amplitudes of the scatterers in the scene. The zero elements corresponds to grid points which are not occupied by scatterers. We can then define the radar signal received from this scene $\mathbf{y}$ by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v} \tag{8}$$

where $\mathbf{y}$ is a $N_RN_t \times 1$ vector, $\mathbf{x}$ is a $N_\tau N_\beta N_f \times 1$ sparse vector, $\mathbf{v}$ is a $N_RN_t \times 1$ complex Gaussian noise vector, and $\mathbf{A}$ is a $N_RN_t \times N_\tau N_\beta N_f$ matrix.

2.3 The target model

We define the *sign function* for a vector $z \in \mathbb{C}^n$ as

$$\text{sgn}(z_k) = \begin{cases} z_k/|z_k| & \text{if } z_k \neq 0, \\ 0 & \text{else.} \end{cases} \tag{9}$$

We introduce the following *generic K-sparse target model*:

- The support $I_K \subset \{1,\dots,N_\tau N_\beta\}$ of the K nonzero coefficients of $\mathbf{x}$ is selected uniformly at random.

- The non-zero coefficients of $\text{sgn}(\mathbf{x})$ form a Steinhaus sequence, i.e., the phases of the non-zero entries of $\mathbf{x}$ are random and uniformly distributed in $[0,2\pi)$.

We do not impose any condition on the amplitudes of the non-zero entries of $\mathbf{x}$. We do assume however that the targets are exactly located at the discretized grid points. This is certainly an idealized assumption, that is not satisfied in this strict sense in practice, resulting in a "gridding error". We refer the reader to [16, 7] for an initial analysis of the associated perturbation error, and to [9] for an interesting numerical approach to deal with this issue.

2.4 The recovery algorithm – Debiased Lasso

A standard approach to find a sparse (and under appropriate conditions *the sparsest*) solution to a noisy system $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ is via

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{10}$$

which is also known as lasso [26]. Here $\lambda > 0$ is a regularization parameter.

In this paper we adopt the following two-step version of lasso. In the first step we compute an estimate $\tilde{I}$ for the support of $\mathbf{x}$ by solving (10). In the second step we estimate the amplitudes of $\mathbf{x}$ by solving the reduced-size least squares problem $\min \|\mathbf{A}_{\tilde{I}}\mathbf{x}_{\tilde{I}} - \mathbf{y}\|_2$, where $\mathbf{A}_{\tilde{I}}$ is the submatrix of $\mathbf{A}$ consisting of the columns corresponding to the index set $\tilde{I}$, and similarly for $\mathbf{x}_{\tilde{I}}$. This is a standard way to "debias" the solution, we thus will call this approach in the sequel *debiased lasso*.

3 Recovery of targets in the Doppler-free case

We assume that $s_i(t)$ is a periodic, continuous-time white Gaussian noise signal of period-duration T seconds and bandwidth B. The transmit waveforms are normalized so that the total transmit power is fixed, independent of the number of transmit antennas. Thus, we assume that the entries of $s_i(t)$ have variance $\frac{1}{N_T}$. It is convenient to introduce the finite-length vector $\mathbf{s}_i$ associated with s_i, via $\mathbf{s}_i(l) := s_i(l\Delta_t), l = 1, \dots, N_t$, where $\Delta_t = \frac{1}{2B}$ and $N_t = T/\Delta_t$.

Theorem 1 *Consider* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$*, where* $\mathbf{A}$ *is as defined in Subsection 2.1 and* $\mathbf{w}_i \in \mathcal{CN}(0, \sigma^2)$*. Choose the discretization stepsizes to be* $\Delta_\beta = \frac{2}{N_R N_T}$ *and* $\Delta_\tau = \frac{1}{2B}$*. Let* $d_T = 1/2, d_R = N_T/2$ *or* $d_T = N_R/2, d_R = 1/2$*, and suppose that*

$$N_t \geq 128, \qquad N_\tau \geq \sqrt{N_\beta}, \qquad \textit{and} \qquad \left(\log(N_\tau N_\beta)\right)^3 \leq N_t. \tag{11}$$

If $\mathbf{x}$ *is drawn from the generic* K*-sparse target model with*

$$K \leq K_{\max} := \frac{c_0 N_\tau N_R}{3N_T \log(N_\tau N_\beta)} \tag{12}$$

for some constant $c_0 > 0$*, and if*

$$\min_{k \in I} |\mathbf{x}_k| > \frac{10\sigma}{\sqrt{N_R N_t}}\sqrt{2\log N_\tau N_\beta}, \tag{13}$$

then the solution $\tilde{\mathbf{x}}$ *of the debiased lasso computed with* $\lambda = 2\sigma\sqrt{2\log(N_\tau N_\beta)}$ *obeys*

$$\operatorname{supp}(\tilde{\mathbf{x}}) = \operatorname{supp}(\mathbf{x}), \tag{14}$$

with probability at least

$$(1 - p_1)(1 - p_2)(1 - p_3)(1 - p_4),$$

and

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\sigma\sqrt{12 N_t N_R}}{\|\mathbf{y}\|_2} \tag{15}$$

with probability at least

$$(1-p_1)(1-p_2)(1-p_3)(1-p_4)(1-p_5),$$

where

$$p_1 = e^{-\frac{(1-\sqrt{1/3})^2 N_t}{2}} + N_t^{1-CN_T},$$

$$p_2 = 2e^{-\frac{N_t(\sqrt{2}-1)^2}{4}} + 2(N_R N_T)^{-1} - 6(N_t N_\beta)^{-1},$$

$$p_3 = e^{-\frac{(1-\sqrt{1/3})^2 N_t}{2}}, \quad p_4 = N_R N_T e^{-\frac{N_R N_t}{25}},$$

and

$$p_5 = 2(N_\tau N_\beta)^{-1}(2\pi \log(N_\tau N_\beta) + K(N_\tau N_\beta)^{-1}) + \mathcal{O}((N_\tau N_\beta)^{-2\log 2}).$$

Remark:

(i) While the expressions for the probability of success in the above theorem are admittedly somewhat unpleasant, we point out that the individual terms are fairly small. Moreover, the probabilities can easily be made smaller by slightly increasing the constants in the assumptions on N_t, N_R, N_T.

(ii) The assumptions in (11) are fairly mild and easy to satisfy in practice.

(iii) We emphasize that there is no constraint on the dynamic range of the target amplitudes. The lasso estimate will recover all target locations correctly as long as they exceed the noise level (13), regardless of the dynamical range between the targets.

(iv) We note that $|x_k|^2/\sigma^2$ is the signal-to-noise ratio for the k-th scatterer at the receiver array input. The measurement vector $\mathbf{y}$ provides $N_R N_t$ measurements of x_k. Therefore it is useful to define the signal-to-noise ratio associated with the k-th scatterer as $\mathrm{SNR}_k = N_R N_t |x_k|^2/\sigma^2$. This is often referred to as the output SNR because it is the effective SNR at the output of a matched-filter receiver. Equation (13) can thus be written as $\mathrm{SNR}_k > 200 \log N_\tau N_\beta$, However, the factor 200 is definitely way too conservative. As is evident from the comments following Theorem 1.3 in [3], one can replace the factor 10 in (13) by a factor $(1+\varepsilon)$ for some $\varepsilon > 0$, at the cost of a somewhat reduced probability of success and some slightly stronger conditions on the coherence and sparsity. This indicates that the SNR condition for which perfect target detection can be achieved is

$$\mathrm{SNR} \geq \mathrm{SNR}_{\min} := C \log N_\tau N_\beta, \tag{16}$$

where C is a constant of size $\mathcal{O}(1)$.

(v) The condition that the target locations are assumed to be random can likely be removed by using a different proof technique that relies on a dual certificate approach (e.g. see [5]) and tools developed in [22]. We do not pursue this direction in this paper.

The proof of Theorem 1 is carried out in several steps. We need two key estimates, one concerns a bound for the operator norm of $\mathbf{A}$, the other one concerns a bound for the coherence of $\mathbf{A}$. We start with deriving a bound for $\|\mathbf{A}\|_{\mathrm{op}}$.

Lemma 2 *Let* $\mathbf{A}$ *be as defined in Theorem 1. Then*

$$\mathbb{P}\Big(\|\mathbf{A}\|_{op}^2 \geq N_t N_R N_T (1+\log N_t)\Big) \leq N_t^{1-CN_T}, \tag{17}$$

where $C > 0$ *is some numerical constant.*

Proof: There holds $\|\mathbf{A}\|_{\text{op}}^2 = \|\mathbf{A}\mathbf{A}^*\|_{\text{op}}$. It is convenient to consider $\mathbf{A}\mathbf{A}^*$ as block matrix

$$\begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \dots & \mathbf{B}_{1,N_R} \\ \vdots & \ddots & & \vdots \\ \mathbf{B}_{N_R,1}^* & & & \mathbf{B}_{N_R,N_R} \end{bmatrix},$$

where the blocks $\{\mathbf{B}_{i,i'}\}_{i,i'=1}^{N_R}$ are matrices of size $N_t \times N_t$. We claim that $\mathbf{A}\mathbf{A}^*$ is a block-Toeplitz matrix (i.e., $\mathbf{B}_{i,i'} = \mathbf{B}_{i+1,i'+1}, i = 1,\dots,N_R-1$) and the individual blocks $\mathbf{B}_{i,i'}$ are circulant matrices. To see this, recall the structure of $\mathbf{A}$ and consider the entry $\mathbf{B}_{[i,l;i',l']}$, $i,i' = 1,\dots,N_R; l,l' = 1,\dots,N_t$:

$$\begin{aligned}
\mathbf{B}_{[i,l;i',l']} = (\mathbf{A}\mathbf{A}^*)_{[i,l;i',l']} &= \sum_\beta \sum_\tau \mathbf{A}_{[i,l;\beta,\tau]} \overline{\mathbf{A}_{[i',l';\beta,\tau]}} \\
= \sum_\beta \sum_{n=1}^{N_\tau} \mathbf{a}_R(\beta)_i \sum_{k=1}^{N_T} \mathbf{a}_T(\beta)_k s_k(l\Delta_t - n\Delta_\tau) &\overline{\mathbf{a}_R(\beta)}_{i'} \overline{\sum_{k'=1}^{N_T} \mathbf{a}_T(\beta)_{k'} s_{k'}(l'\Delta_t - n\Delta_\tau)} \\
= \sum_\beta \mathbf{a}_R(\beta)_i \overline{\mathbf{a}_R(\beta)}_{i'} \sum_{k=1}^{N_T}\sum_{k'=1}^{N_T} &\mathbf{a}_T(\beta)_k \overline{\mathbf{a}_T(\beta)}_{k'} \sum_{n=1}^{N_\tau} s_k(l\Delta_t - n\Delta_\tau)\overline{s_{k'}(l'\Delta_t - n\Delta_\tau)} \\
= \sum_\beta e^{j2\pi d_R(i-i')\beta} \sum_{k=1}^{N_T}\sum_{k'=1}^{N_T} &e^{j2\pi d_T(k-k')\beta} \sum_{n=1}^{N_\tau} s_k(l\Delta_t - n\Delta_\tau)\overline{s_{k'}(l'\Delta_t - n\Delta_\tau)},
\end{aligned} \tag{18}$$

where we used the delay discretization $\tau = n\Delta_\tau, n = 1,\dots,N_\tau$. The block-Toeplitz structure, $\mathbf{B}_{i,i'} = \mathbf{B}_{i+1,i'+1}$, follows from observing that the expression (18) depends on the difference $i - i'$, but not on the individual values of i, i'. The circulant structure of an individual block $\mathbf{B}_{i,i'}$ (i, i' are now fixed) follows readily from noting that

$$\sum_{n=1}^{N_\tau} s_k(l\Delta_t - n\Delta_\tau)\overline{s_{k'}(l'\Delta_t - n\Delta_\tau)} = \sum_{n=1}^{N_\tau} s_k((l+1)\Delta_t - n\Delta_\tau)\overline{s_{k'}((l'+1)\Delta_t - n\Delta_\tau)},$$

since we have chosen $\Delta_t = \Delta_\tau$ and since the shifts are circulant in this case.

We will now show that the blocks $B_{i,i'}$ are actually zero-matrices for $i \neq i'$. For convenience we introduce the notation

$$G_{k,k'}(l,l') := \sum_{n=1}^{N_\tau} s_k(l\Delta_t - n\Delta_\tau)\overline{s_{k'}(l'\Delta_t - n\Delta_\tau)}, \qquad l,l' = 1,\dots,N_t; k,k' = 1,\dots,N_T,$$

Substituting $d_T = 1/2, d_R = N_T/2$ (the very similar calculation for $d_R = 1/2, d_T = N_R/2$ is left to the reader) and the discretization $\beta = n\Delta_\beta, n = 1, \ldots, N_\beta$, with $\Delta_\beta = \frac{2}{N_R N_T}$ in (18) we can write

$$\begin{aligned} \mathbf{B}_{[i,l;i',l']} &= \sum_{n=-\frac{N_R N_T}{2}}^{\frac{N_R N_T}{2}-1} e^{j2\pi \frac{N_T}{2}(i-i')\frac{2n}{N_R N_T}} \sum_{k=1}^{N_T}\sum_{k'=1}^{N_T} e^{j2\pi \frac{1}{2}(k-k')\frac{2n}{N_R N_T}} G_{k,k'}(l,l') \\ &= \sum_{k=1}^{N_T}\sum_{k'=1}^{N_T} G_{k,k'}(l,l') \sum_{n=0}^{N_R N_T-1} e^{j2\pi N_T(i-i')\frac{n}{N_R N_T}} e^{j2\pi(k-k')\frac{n}{N_R N_T}}. \end{aligned} \tag{19}$$

We analyze the inner summation in (19) separately.

$$\begin{aligned} \sum_{n=0}^{N_R N_T-1} e^{j2\pi N_T(i-i')\frac{n}{N_R N_T}} e^{j2\pi(k-k')\frac{n}{N_R N_T}} &= \sum_{n_1=0}^{N_T-1}\sum_{n_2=0}^{N_R-1} e^{j2\pi(k-k')\frac{n_1 N_R+n_2}{N_R N_T}} e^{j2\pi N_T(i-i')\frac{n_1 N_R+n_2}{N_R N_T}} \\ &= \sum_{n_2=0}^{N_R-1} e^{j2\pi(k-k')\frac{n_2}{N_R N_T}} e^{j2\pi(i-i')\frac{n_2 N_T}{N_R N_T}} \sum_{n_1=0}^{N_T-1} e^{j2\pi(k-k')\frac{n_1 N_R}{N_R N_T}} e^{j2\pi(i-i')\frac{n_1 N_R N_T}{N_R N_T}} \\ &= \sum_{n_2=0}^{N_R-1} e^{j2\pi(k-k')\frac{n_2}{N_R N_T}} e^{j2\pi(i-i')\frac{n_2}{N_R}} \sum_{n_1=0}^{N_T-1} e^{j2\pi(k-k')\frac{n_1}{N_T}} \underbrace{e^{j2\pi(i-i')n_1}}_{=1 \text{ for all } i,i'} \\ &= \sum_{n_2=0}^{N_R-1} e^{j2\pi(k-k')\frac{n_2}{N_R N_T}} e^{j2\pi(i-i')\frac{n_2}{N_R}} \sum_{n_1=0}^{N_T-1} e^{j2\pi(k-k')\frac{n_1}{N_T}} \\ &= \sum_{n_2=0}^{N_R-1} e^{j2\pi(k-k')\frac{n_2}{N_R N_T}} e^{j2\pi(i-i')\frac{n_2}{N_R}} N_T \delta_{k-k'}. \end{aligned}$$

Hence

$$\begin{aligned} \mathbf{B}_{[i,l;i',l']} &= N_T \sum_{k=1}^{N_T}\sum_{k'=1}^{N_T} \delta_{k-k'} G_{k,k'}(l,l') \sum_{n_2=0}^{N_R-1} \underbrace{e^{j2\pi(k-k')\frac{n_2}{N_R N_T}}}_{=1 \text{ for } k=k'} e^{j2\pi(i-i')\frac{n_2}{N_R}} \\ &= N_T \sum_{k=1}^{N_T} G_{k,k}(l,l') \sum_{n_2=0}^{N_R-1} e^{j2\pi(i-i')\frac{n_2}{N_R}} = N_T N_R \sum_{k=1}^{N_T} G_{k,k}(l,l')\delta_{i-i'}. \end{aligned}$$

Thus, $\mathbf{B}_{i,i'} = \mathbf{0}$ for $i \neq i'$, and $\mathbf{A}^*\mathbf{A}$ is indeed a block-diagonal matrix, which in turn implies $\|\mathbf{A}\|_{\text{op}}^2 = \max_i \|\mathbf{B}_{i,i}\|_{\text{op}}$. But due to the block-Toeplitz structure of $\mathbf{A}^*\mathbf{A}$ we have $\mathbf{B}_{1,1} = \mathbf{B}_{2,2} = \cdots = \mathbf{B}_{N_R,N_R}$. Therefore

$$\|\mathbf{A}\|_{\text{op}}^2 = \|\mathbf{B}_{1,1}\|_{\text{op}}. \tag{20}$$

To bound $\|\mathbf{B}_{1,1}\|_{\text{op}}$ we utilize its circulant structure as well as tail bounds of quadratic forms. Let $\mathbf{b}$ be the first column of $\mathbf{B}_{1,1}$, then $\|\mathbf{B}_{1,1}\|_{\text{op}} = \sqrt{N_t}\|\hat{\mathbf{b}}\|_\infty$ where $\hat{\mathbf{b}}$ is the Fourier transform of $\mathbf{b}$. From our previous computations we have (after a change of variables)

$$\mathbf{b}(l) = N_T N_R \sum_{k=1}^{N_T} G_{k,k}(l,0) = N_T N_R \sum_{k=1}^{N_T}\sum_{n=1}^{N_\tau} s_k(n\Delta_\tau - l\Delta_t)\overline{s_k(n\Delta_\tau)}, \qquad l = 0, \ldots, N_t - 1.$$

We will rewrite this expression so that we can apply Lemma 12 to bound $\|\hat{\mathbf{b}}\|_\infty$. Let $\mathbf{T}_{N_t}$ denote the translation operator on $\mathbb{C}^{N_t}$ as introduced in (2) and define the $N_tN_T \times N_tN_T$ block-diagonal matrix $\mathbf{U}^{(l)} = \{u_{ii'}^{(l)}\}$ by

$$\mathbf{U}^{(l)} := N_R N_T \sqrt{N_t} \mathbf{I}_{N_T} \otimes \mathbf{T}_{N_t}^l, \qquad \text{for } l = 0, \dots, N_t - 1. \tag{21}$$

Furthermore, let $\mathbf{z} = [\mathbf{s}_1^T, \mathbf{s}_2^T, \dots, \mathbf{s}_{N_T}^T]^T$, then

$$\sqrt{N_t}\mathbf{b}(l) = \sqrt{N_t} N_T N_R \sum_{k=1}^{N_T} \langle \mathbf{s}_k, \mathbf{T}_{N_t}^l \mathbf{s}_k \rangle = \langle \mathbf{z}, \mathbf{U}^{(l)} \mathbf{z} \rangle, = \sum_{i,i'=1}^{N_tN_T} u_{ii'}^{(l)} \bar{\mathbf{z}}_i \mathbf{z}_{i'}.$$

and therefore

$$\sqrt{N_t}\hat{\mathbf{b}}(k) = \frac{1}{\sqrt{N_t}} \sum_{l=0}^{N_t-1} \sum_{i,i'=1}^{N_tN_T} u_{ii'}^{(l)} \bar{\mathbf{z}}_i \mathbf{z}_{i'} e^{j2\pi kl/N_t} = \sum_{i,i'=1}^{N_tN_T} \bar{\mathbf{z}}_i \mathbf{z}_{i'} \frac{1}{\sqrt{N_t}} \sum_{l=0}^{N_t-1} u_{ii'}^{(l)} e^{j2\pi kl/N_t} = \sum_{i,i'=1}^{N_tN_T} \bar{\mathbf{z}}_i \mathbf{z}_{i'} v_{ii'}^{(k)},$$

where we have denoted $v_{ii'}^{(k)} := \frac{1}{\sqrt{N_t}} \sum_{l=0}^{N_t-1} u_{ii'}^{(l)} e^{j2\pi kl/N_t}$ for $i, i' = 0, \dots, N_tN_T - 1$ and $k = 0, \dots, N_t - 1$. It follows from (21) and standard properties of the Fourier transform that the matrix $\mathbf{V}^{(k)} := \{v_{ii'}^{(k)}\}$ is a block-diagonal matrix with N_T blocks of size $N_t \times N_t$, where each non-zero entry of such a block has absolute value N_RN_T. Furthermore, a little algebra shows that $\|\mathbf{V}^{(k)}\|_F = \sqrt{N_t^2 N_R^2 N_T^3}$, $\|\mathbf{V}^{(k)}\|_{\text{op}} = N_t N_R N_T$, $\text{trace}(\mathbf{V}^{(k)}) = N_t N_R N_T^2$, and

$$\mathbb{E}\Big(\sum_{i,i'=1}^{N_tN_T} \bar{\mathbf{z}}_i \mathbf{z}_{i'} v_{ii'}^{(k)} \Big) = \frac{1}{N_T} \text{trace}(\mathbf{V}^{(k)}) = N_t N_R N_T.$$

We can now apply Lemma 12 (keeping in mind that $x_i \sim \mathcal{CN}(0, \frac{1}{N_T})$) and obtain

$$\mathbb{P}\big(|\sqrt{N_t}\hat{\mathbf{b}}(l)| \geq N_t N_R N_T + t \big) \leq \exp\Big(-C \min\Big\{ \frac{tN_T}{N_tN_RN_T}, \frac{t^2 N_T^2}{N_t^2 N_R^2 N_T^3} \Big\} \Big),$$

where $C > 0$ is some numerical constant.

Choosing $t = N_t N_R N_T \log N_t$ gives

$$\mathbb{P}\big(|\sqrt{N_t}\hat{\mathbf{b}}(l)| \geq N_t N_R N_T (1 + \log N_t) \big) \leq \exp(-C N_T \log N_t),$$

for $l = 0, \dots, N_t - 1$. Forming the union bound over the N_t possibilities for l gives

$$\mathbb{P}\big(\max_l \{ |\sqrt{N_t}\hat{\mathbf{b}}(l)| \} \geq N_t N_R N_T (1 + \log N_t) \big) \leq \sum_{l=0}^{N_t-1} \exp(-C \sqrt{N_T} \log N_t) = N_t^{1 - CN_T}. \tag{22}$$

We recall that $\|\mathbf{B}_{1,1}\|_{\text{op}} = \max_l |\sqrt{N_t}\hat{\mathbf{b}}(l)|$, and substitute (22) into (20) to complete the proof. ❑

Next we estimate the coherence of $\mathbf{A}$. Since the columns of $\mathbf{A}$ do not all have the same norm, we will proceed in two steps. First we bound the modulus of the inner product of any two columns of $\mathbf{A}$ and then use this result to bound the coherence of a properly normalized version of $\mathbf{A}$. Since the columns of $\mathbf{A}$ depend on azimuth and delay, we index them via the double-index (τ, β). Thus the (τ, β)-th column of $\mathbf{A}$ is $\mathbf{A}_{\tau,\beta}$.

Lemma 3 *Let* $\mathbf{A}$ *be as defined in Theorem 1. Assume that*

$$N_\tau \geq \sqrt{N_\beta} \qquad \text{and} \qquad \log(N_\tau N_\beta) \leq \frac{N_t}{30}, \tag{23}$$

then

$$\max_{(\tau,\beta)\neq(\tau',\beta')} \left|\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau',\beta'}\rangle\right| \leq 3N_R\sqrt{N_t \log(N_\tau N_\beta)} \tag{24}$$

with probability at least $1 - 2(N_R N_T)^{-1} - 6(N_\tau N_R N_T)^{-1}$.

Proof: We assume $d_T = \frac{1}{2}, d_R = \frac{N_T}{2}$ and leave the case $d_T = \frac{N_R}{2}, d_R = \frac{1}{2}$ to the reader. We need to find an upper bound for

$$\max |\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau',\beta'}\rangle| \qquad \text{for } (\tau, \beta) \neq (\tau', \beta').$$

It follows from the definition of $z(t; \beta, r)$ via a simple calculation that

$$\mathbf{A}_{\tau,\beta} = \mathbf{a}_R(\beta) \otimes (\mathbf{S}_\tau \mathbf{a}_T(\beta)),$$

from which we readily compute

$$\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau',\beta'}\rangle = \langle \mathbf{a}_R(\beta), \mathbf{a}_R(\beta')\rangle\langle \mathbf{S}_\tau \mathbf{a}_T(\beta), \mathbf{S}_{\tau'} \mathbf{a}_T(\beta')\rangle. \tag{25}$$

We use the discretization $\beta = n\Delta_\beta$, $\beta' = n'\Delta_\beta$, where $\Delta_\beta = \frac{2}{N_R N_T}$, $n, n' = 1, \ldots, N_\beta$, with $N_\beta = N_R N_T$, and obtain after a standard calculation

$$\langle \mathbf{a}_R(\beta), \mathbf{a}_R(\beta')\rangle = \begin{cases} N_R & \text{if } n - n' = kN_R \text{ for } k = 0, \ldots, N_T - 1, \\ 0 & \text{if } n - n' \neq kN_R, \end{cases} \tag{26}$$

and

$$\langle \mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle = \begin{cases} 0 & \text{if } n - n' = kN_R \text{ for } k = 1, \ldots, N_T - 1, \\ \langle \mathbf{a}_T(\beta), \mathbf{a}_T(\beta)\rangle & \text{if } n - n' = 0. \end{cases} \tag{27}$$

As a consequence of (26), concerning β, β' we only need to focus on the case $n - n' = kN_R$ for $k = 1, \ldots, N_T - 1$. Moreover, since

$$\langle \mathbf{S}_\tau \mathbf{a}_T(\beta), \mathbf{S}_{\tau'} \mathbf{a}_T(\beta')\rangle = \langle \mathbf{S}_{\tau-\tau'} \mathbf{a}_T(\beta), \mathbf{S} \mathbf{a}_T(\beta')\rangle, \quad \text{for } \tau, \tau' = 0, \ldots, N_\tau - 1,$$

and $|\langle \mathbf{S}_\tau \mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle| = |\langle \mathbf{S}_{N_t-\tau} \mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle|$, we can confine the range of values for τ, τ' to $\tau' = 0, \tau = 0, \ldots, N_t/2$.

We split our analysis into three cases, (i) $\beta \neq \beta', \tau = 0$, (ii) $\beta \neq \beta', \tau \neq 0$, and (iii) $\beta = \beta', \tau \neq 0$.

Case (i) $\beta \neq \beta', \tau = 0$**:** We will first find a bound for $|\langle \mathbf{a}_R(\beta), \mathbf{a}_R(\beta')\rangle\langle \mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle|$ and then invoke Lemma 11 to obtain a bound for $|\langle \mathbf{a}_R(\beta), \mathbf{a}_R(\beta')\rangle\langle \mathbf{S}\mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta')\rangle|$.

Based on (26) and (27), to bound $|\langle \mathbf{a}_R(\beta), \mathbf{a}_R(\beta')\rangle\langle \mathbf{S}\mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta')\rangle|$ we only need to consider those n, n' for which $n-n'$ is not a multiple of N_R, in which case $\mathbf{a}_T(\beta)$ and $\mathbf{a}_T(\beta')$ are orthogonal. We have

$$|\langle \mathbf{a}_R(\beta), \mathbf{a}_R(\beta')\rangle\langle \mathbf{S}\mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta')\rangle| \leq N_R\, |\langle \mathbf{S}^*\mathbf{S}\mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle|. \tag{28}$$

By Lemma 11 there holds

$$\mathbb{P}\Big(|\langle \mathbf{S}^*\mathbf{S}\mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle| \geq tN_t\Big) \leq 2\exp\Big(-N_t\frac{t^2}{C_1+C_2t})\Big) \tag{29}$$

for all $0 < t < 1$, where $C_1 = \frac{4e}{\sqrt{6\pi}}$ and $C_2 = \sqrt{8}e$. We choose $t = 3\sqrt{\frac{1}{N_t}\log(N_\tau N_R N_T)}$ in (29) and get

$$\mathbb{P}\Big(|\langle \mathbf{S}^*\mathbf{S}\mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle| \geq 3\sqrt{N_t\log(N_\tau N_R N_T)}\Big) \leq 2\exp\Big(-\frac{9\log(N_\tau N_R N_T)}{C_1+\frac{3C_2}{\sqrt{N_t}}\sqrt{\log(N_\tau N_R N_T)}}\Big). \tag{30}$$

We claim that

$$\frac{9\log(N_\tau N_R N_T)}{C_1+\frac{3C_2}{\sqrt{N_t}}\sqrt{\log(N_\tau N_R N_T)}} \geq 2\log(N_R N_T). \tag{31}$$

To verify this claim we first note that (31) is equivalent to

$$9\log N_\tau \geq \log(N_R N_T)(2C_1 + \frac{6C_2}{\sqrt{N_t}}\sqrt{\log(N_\tau N_\beta)} - 9).$$

Using both assumptions in (23) and the fact that $2C_1 + \frac{6C_2}{\sqrt{30}} - 9 \leq \frac{9}{2}$ we obtain

$$9\log N_\tau \geq \log N_\beta(2C_1 + \frac{6C_2}{\sqrt{30}} - 9) \geq \log N_\beta(2C_1 + \frac{6C_2}{\sqrt{N_t}}\sqrt{\log(N_t N_\beta)} - 9),$$

which establishes (31). Substituting now (31) into (30) gives

$$\mathbb{P}\Big(|\langle \mathbf{S}^*\mathbf{S}\mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle| \geq 3\sqrt{N_t\log(N_\tau N_R N_T)}\Big) \leq 2\exp\big(-2\log(N_R N_T)\big). \tag{32}$$

To bound $\max|\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau,\beta'}\rangle|$ we only have to take the union bound over $N_R N_T$ different possibilities associated with β, β', as $\tau = \tau' = 0$. Forming now the union bound, and using (28), yields

$$\mathbb{P}\Big(|\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau,\beta'}\rangle| \leq 3N_R\sqrt{N_t\log(N_\tau N_R N_T)}\Big) \geq 1 - 2(N_R N_T)^{-1}. \tag{33}$$

Case (ii) $\beta \neq \beta', \tau \neq 0$**:** We need to consider the case $|\langle \mathbf{S}_\tau\mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta')\rangle|$ where $\beta = n\Delta_\beta$, $\beta' = n'\Delta_\beta$, with $n - n' = kN_R$ for $k = 1, \ldots, N_T - 1$. Since the entries of $\mathbf{S}$ are i.i.d. Gaussian random variables, it follows that the entries of $\mathbf{S}_\tau\mathbf{a}_T(\beta)$ are i.i.d. $\mathcal{CN}(0,1)$-distributed, and similar for $\mathbf{S}\mathbf{a}_T(\beta')$. Moreover, the fact that $\langle \mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle = 0$ implies that $\mathbf{S}_\tau\mathbf{a}_T(\beta)$ and $\mathbf{S}\mathbf{a}_T(\beta')$ are independent. Consequently, the entries of $\sum_{l=0}^{N_t-1}\overline{(\mathbf{S}_\tau\mathbf{a}_T(\beta))_l}(\mathbf{S}\mathbf{a}_T(\beta'))_l$ are jointly independent. Therefore, we can apply Lemma 14 with $t = 3\sqrt{N_t\log(N_\tau N_R N_T)}$, form the union bound over the $N_\tau N_R N_T$ possibilities associated with τ (we do not take advantage of the fact we actually have only $N_\tau - 1$ and not N_τ possibilities for τ) and β, β' (here, we take again into account property (26)), and eventually obtain

$$\mathbb{P}\Big(|\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau',\beta'}\rangle| \le 3N_R\sqrt{N_t \log(N_\tau N_R N_T)}\Big) \ge 1 - 2(N_\tau N_R N_T)^{-1}. \tag{34}$$

Case (iii) $\beta = \beta', \tau \neq 0$: We need to find an upper bound for $|\langle \mathbf{S}_\tau \mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta)\rangle|$ where $\tau = 1, \dots, N_t - 1$. Since Since each of the entries of $\mathbf{S}_\tau \mathbf{a}_T(\beta)$ and of $\mathbf{S}\mathbf{a}_T(\beta)$ is a sum of N_T i.i.d. Gaussian random variables of variance $1/N_T$, we can write

$$|\langle \mathbf{S}_\tau \mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta)\rangle| = |\sum_{l=0}^{N_t-1} \bar{g}_{l-\tau} g_l|, \tag{35}$$

where $g_l \sim \mathcal{N}(0,1)$. Note that the terms $\bar{g}_{l-\tau} g_l$ in this sum are no longer all jointly independent. But similar to the proof of Theorem 5.1 in [20] we observe that for any $\tau \neq 0$ we can split the index set $0, \dots, N_t - 1$ into two subsets $\Lambda_\tau^1, \Lambda_\tau^2 \subset \{0, \dots, N_t - 1\}$, each of size $N_t/2$, such that the $N_t/2$ variables $\bar{g}(l-\tau)g(l)$ are jointly independent for $l \in \Lambda_\tau^1$, and analogous for Λ_τ^2. (For convenience we assume here that N_t is even, but with a negligible modification the argument also applies for odd N_t.) In other words, each of the sums $\sum_{l\in\Lambda_\tau^r} \bar{g}(l-\tau)g(l), r = 1, 2$, contains only jointly independent terms. Hence we can apply Lemma 14 and obtain

$$\mathbb{P}\Big(\big|\sum_{l\in\Lambda_\tau^r} \bar{g}(l-\tau)g(l)\big| > t\Big) \le 2\exp\Big(-\frac{t^2}{N_t/2 + 2t)}\Big)$$

for all $t > 0$. Choosing $t = \frac{3}{2}\sqrt{N_t \log(N_t N_R N_T)}$ gives

$$\begin{aligned}\mathbb{P}\Big(\big|\sum_{l\in\Lambda_\tau^r} \bar{g}(l-\tau)g(l)\big| > \frac{3}{2}\sqrt{N_t \log(N_t N_R N_T)}\Big) &\le 2\exp\Big(-\frac{\frac{9}{4}N_t \log(N_t N_R N_T)}{\frac{N_t}{2} + 3\sqrt{N_t \log(N_t N_R N_T)}}\Big)\\ &\le 2\exp\Big(-\frac{9\log(N_t N_R N_T)}{2 + 12\sqrt{\frac{\log(N_t N_R N_T)}{N_t}}}\Big).\end{aligned} \tag{36}$$

Condition (23) implies that $12\sqrt{\frac{\log(N_t N_R N_T)}{N_t}} \le \frac{5}{2}$, hence the estimate in (36) becomes

$$\begin{aligned}\mathbb{P}\Big(\big|\sum_{l\in\Lambda_\tau^r} \bar{g}(l-\tau)g(l)\big| > \frac{3}{2}\sqrt{\log(N_t N_R N_T)}\sqrt{N_t}\Big) &\le 2\exp\Big(-\frac{9\log(N_t N_R N_T)}{2+\frac{5}{2}}\Big)\\ &= 2\exp\big(-2\log(N_t N_R N_T)\big)\\ &= 2(N_t N_R N_T)^{-2}.\end{aligned} \tag{37}$$

Using equation (35), inequality (37), and the pigeonhole principle, we obtain

$$\mathbb{P}\Big(|\langle \mathbf{S}_\tau \mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta)\rangle| > 3\sqrt{N_t \log(N_t N_R N_T)}\Big) \le 4(N_t N_R N_T)^{-2},$$

Combining this estimate with (25) yields

$$\mathbb{P}\Big(|\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau',\beta}\rangle| \geq 3N_R\sqrt{N_t \log(N_\tau N_R N_T)}\Big) \leq 4(N_t N_R N_T)^{-2},$$

We apply the union bound over the $\frac{N_t}{2}N_T N_R$ different possibilities and arrive at

$$\mathbb{P}\Big(\max|\langle \mathbf{A}_{\tau,\beta}, \mathbf{A}_{\tau',\beta}\rangle| \leq 3N_R\sqrt{N_t \log(N_\tau N_R N_T)}\Big) \geq 1 - 4(N_t N_R N_T)^{-1}, \tag{38}$$

where the maximum is taken over all $\tau, \tau', \beta, \beta'$ with $\tau \neq \tau'$.

An inspection of the bounds (33), (34), and (38) establishes (24), which is what we wanted to prove. ❏

The key to proving Theorem 1 is to combine Lemma 2 and Lemma 3 with Theorem 15. The latter theorem requires the matrix to have columns of unit-norm, whereas the columns of our matrix $\mathbf{A}$ have all different norms (although the norms concentrate nicely around $\sqrt{N_t N_R N_T}$). Thus instead of $\mathbf{A}\mathbf{x} = \mathbf{y}$ we now consider

$$\tilde{\mathbf{A}}\mathbf{z} = \mathbf{y}, \qquad \text{where } \tilde{\mathbf{A}} := \mathbf{A}\mathbf{D}^{-1} \text{ and } \mathbf{z} := \mathbf{D}\mathbf{x}. \tag{39}$$

Here $\mathbf{D}$ is the $N_\tau N_\beta \times N_\tau N_\beta$ diagonal matrix defined by

$$\mathbf{D}_{(\tau,\beta),(\tau,\beta)} = \|\mathbf{A}_{\tau,\beta}\|_2. \tag{40}$$

In the noise-free case we can easily recover $\mathbf{x}$ from $\mathbf{z}$ via $\mathbf{x} = \mathbf{D}^{-1}\mathbf{z}$. In the noisy case we will utilize the fact that for proper choices of λ the associated lasso solutions of (10) and (50), respectively, have the same support, see also the proof of Theorem 1.

The following lemma gives a bound for $\mu(\tilde{\mathbf{A}})$ and $\|\tilde{\mathbf{A}}\|_{\text{op}}$ in terms of the corresponding bounds for $\mathbf{A}$.

Lemma 4 *Let $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}^{-1}$, where the $\mathbf{D}$ the diagonal matrix is defined by (40). Under the conditions of Theorem 1, there holds*

$$\mathbb{P}\Big(\|\tilde{\mathbf{A}}\|_{op}^2 < 3(1 + \log N_t)\Big) \geq 1 - p_1, \tag{41}$$

where $p_1 = e^{-N_t\frac{(\sqrt{1/3}-1)^2}{2}} - N_t^{1-C\sqrt{N_T}}$, and

$$\mathbb{P}\Big(\mu(\tilde{\mathbf{A}}) \leq 6\sqrt{\frac{1}{N_t}\log(N_\tau N_R N_T)}\Big) \geq 1 - p_2, \tag{42}$$

where $p_2 = 2e^{-\frac{N_t(\sqrt{2}-1)^2}{4}} - 2(N_R N_T)^{-1} - 6(N_t N_R N_T)^{-1}$.

Proof: We have

$$\|\tilde{\mathbf{A}}\|_{\text{op}}^2 \leq \frac{\|\mathbf{A}\|_{\text{op}}^2}{\max_{\tau,\beta}\|\mathbf{A}_{\tau,\beta}\|_2^2}. \tag{43}$$

Recall that

$$\mathbf{A}_{\tau,\beta} = \mathbf{a}_R(\beta) \otimes (\mathbf{S}_\tau \mathbf{a}_T(\beta)), \tag{44}$$

hence $\|\mathbf{A}_{\tau,\beta}\|_2^2 = \|\mathbf{a}_R(\beta)\|_2^2 \|\mathbf{S}_\tau \mathbf{a}_T(\beta)\|_2^2$. Since the entries $(\mathbf{S}_\tau \mathbf{a}_T(\beta))_k \sim \mathcal{CN}(0, N_T)$, we have $\mathbb{E}\|\mathbf{S}_\tau \mathbf{a}_T(\beta)\| = \sqrt{N_t}$, and thus by Lemma 9

$$\mathbb{P}\Big(\sqrt{N_t} - \|\mathbf{S}_\tau \mathbf{a}_T(\beta)\|_2 > t\Big) \le e^{-\frac{t^2}{2}}, \tag{45}$$

for all $t > 0$, hence

$$\mathbb{P}\Big(\frac{1}{\|\mathbf{S}_\tau \mathbf{a}_T(\beta)\|_2^2} < \frac{1}{(\sqrt{N_t} - t)^2}\Big) \ge 1 - e^{-\frac{t^2}{2}}, \tag{46}$$

Choosing $t = (1 - \sqrt{1/3})\sqrt{N_t}$ in (46) and forming the union bound only over the $N_R N_T$ different possibilities associated with β (note that $\|\mathbf{S}_\tau \mathbf{a}_T(\beta)\|_2 = \|\mathbf{S}\mathbf{a}_T(\beta)\|_2$ for all τ), gives

$$\mathbb{P}\Big(\frac{1}{\max\limits_{\tau,\beta} \|\mathbf{A}_{\tau,\beta}\|_2^2} < \frac{3}{N_t N_R}\Big) \ge 1 - N_R N_T e^{-\frac{N_t(1-\sqrt{1/3})^2}{2}}. \tag{47}$$

The diligent reader may convince herself that the probability in (47) is indeed close to one under the condition (11). We insert (17) and (47) into (43) and obtain

$$\mathbb{P}\Big(\|\tilde{\mathbf{A}}\|_{\text{op}}^2 < 3N_T(1 + \log N_t)\Big) \ge 1 - e^{-\frac{N_t(1-\sqrt{1/3})^2}{2}} - N_t^{1-C\sqrt{N_T}}. \tag{48}$$

which proves (41).

To establish (42) we first note that

$$\mu(\tilde{\mathbf{A}}) \le \max_{(\tau,\beta)\neq(\tau',\beta')} \Big\{\mathbf{D}^{-1}_{(\tau,\beta),(\tau,\beta)} |(\mathbf{A}^*\mathbf{A})_{(\tau,\beta),(\tau',\beta')}| \mathbf{D}^{-1}_{(\tau',\beta'),(\tau',\beta')}\Big\}, \tag{49}$$

where $\mathbf{D}^{-1}_{(\tau,\beta),(\tau,\beta)} = \|\mathbf{A}_{\tau,\beta}\|_2^{-1}$. Using Lemma 9 and (44) we compute

$$\mathbb{P}\Big(\|\mathbf{A}_{\tau,\beta}\|_2 > \sqrt{N_t N_R} - \sqrt{N_R} t\Big) \ge 1 - e^{-\frac{t^2}{2}}.$$

Therefore

$$\mathbb{P}\Big(\frac{1}{\|\mathbf{A}_{\tau,\beta}\|_2} < \frac{1}{\sqrt{N_t N_R} - \sqrt{N_R} t}\Big) \ge 1 - e^{-\frac{t^2}{2}},$$

and thus

$$\mathbb{P}\Big(|\tilde{\mathbf{A}}^*\tilde{\mathbf{A}})_{(\tau,\beta),(\tau',\beta')}| \le \frac{1}{(\sqrt{N_t N_R} - \sqrt{N_R} t)^2} |(\mathbf{A}^*\mathbf{A})_{(\tau,\beta),(\tau',\beta')}|\Big) \ge 1 - 2e^{-\frac{t^2}{2}},$$

By choosing $t = (1 - 1/\sqrt{2})\sqrt{N_t}$, we can write (50) as

$$\mathbb{P}\Big(|\tilde{\mathbf{A}}^*\tilde{\mathbf{A}})_{(\tau,\beta),(\tau',\beta')}| \le \frac{2}{N_t N_R} |(\mathbf{A}^*\mathbf{A})_{(\tau,\beta),(\tau',\beta')}|\Big) \ge 1 - 2e^{-\frac{N_t(\sqrt{2}-1)^2}{4}}.$$

Finally, plugging (50) into (49) and using (24) we arrive at

$$\mathbb{P}\Big(\mu(\tilde{\mathbf{A}}) \le 6\sqrt{\frac{1}{N_t}\log(N_\tau N_R N_T)}\Big) \ge 1 - 2e^{-\frac{N_t(\sqrt{2}-1)^2}{4}} - 2(N_R N_T)^{-1} - 6(N_t N_R N_T)^{-1}.$$

❑

We are now ready to prove Theorem 1. Among others it hinges on a (complex version of a) theorem by Candès and Plan [3], which is stated in Appendix B.

Proof of Theorem 1: We first point out that the assumptions of Theorem 1 imply that the conditions of Lemma 2 and Lemma 3 are fulfilled. For Lemma 2 this is obvious. Concerning Lemma 3, an easy calculation shows that the conditions $(\log(N_\tau N_R N_T))^3 \leq N_t$ and $N_t \geq 128$ indeed yield that $\log(N_t N_R N_T) \leq \frac{N_t}{23}$.

Note that the solution $\tilde{\mathbf{x}}$ of (10) and the solution $\tilde{\mathbf{z}}$ of the following lasso problem

$$\min_{\mathbf{z}} \frac{1}{2}\|\mathbf{A}\mathbf{D}^{-1}\mathbf{z}-\mathbf{y}\|_2^2+\lambda\|\mathbf{z}\|_1, \qquad \text{with } \lambda = 2\sigma\sqrt{2\log(N_\tau N_R N_T)}, \tag{50}$$

satisfy $\operatorname{supp}(\tilde{\mathbf{x}}) = \operatorname{supp}(\mathbf{D}^{-1}\tilde{\mathbf{z}})$.

We will first establish the claims in Theorem 1 for the system $\tilde{\mathbf{A}}\mathbf{z} = \mathbf{y}$ in (39) where $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}^{-1}$, $\mathbf{z} = \mathbf{D}\mathbf{x}$ and then switch back to $\mathbf{A}\mathbf{x} = \mathbf{y}$.

We verify first condition (77). Property (13) and the fact that $\mathbf{z} = \mathbf{D}\mathbf{x}$ imply that

$$|z_k| \geq \frac{10\|\mathbf{A}_{\tau,\beta}\|_2}{\sqrt{N_R N_t}}\sigma\sqrt{2\log(N_\tau N_\beta)}, \qquad \text{for } (\tau,\beta)\in S. \tag{51}$$

Using Lemma 9 we get that

$$\mathbb{P}\Big(\|\mathbf{A}_{\tau,\beta}\| \geq \sqrt{N_R N_t} - t\Big) \geq 1 - e^{-\frac{t^2}{2}}. \tag{52}$$

Choosing $t = \frac{2}{10}\sqrt{N_R N_t}$ and combining (52) with (51) gives

$$|z_k| \geq 8\sigma\sqrt{2\log(N_\tau N_\beta)}, \qquad \text{for } k \in S,$$

with probability at least $1 - e^{-\frac{N_R N_t}{25}}$, thus establishing condition (77).

Note that $\tilde{\mathbf{A}}$ has unit-norm columns as required by Theorem 15. It remains to verify condition (75). Using the assumption (11), and the coherence bound (42) we compute

$$\mu^2(\tilde{\mathbf{A}}) \leq 36\frac{1}{N_t}\log(N_\tau N_R N_T) \leq 36\frac{\log(N_\tau N_R N_T)}{\log^3(N_\tau N_R N_T)} = \frac{36}{\log^2(N_\tau N_R N_T)},$$

which holds with probability as in (42), and thus the coherence property (75) is fulfilled.

Furthermore, using (41) we see that condition (12) implies

$$K \leq \frac{c_0 N_\tau N_R}{3(1+\log N_t)\log(N_\tau N_R N_T)} \leq \frac{c_0 N_\tau N_R}{\|\tilde{\mathbf{A}}\|_{\text{op}}^2 \log(N_\tau N_R N_T)}$$

with probability as stated in (41). Thus assumption (76) of Theorem 15 is also fulfilled (with high probability) and we obtain that

$$\operatorname{supp}(\tilde{\mathbf{z}}) = \operatorname{supp}(\mathbf{z}). \tag{53}$$

We note that the relation $\operatorname{supp}(\tilde{\mathbf{x}}) = \operatorname{supp}(\mathbf{x})$ holds with the same probability as the relation $\operatorname{supp}(\tilde{\mathbf{z}}) = \operatorname{supp}(\mathbf{z})$ (see equation (53)), since $\operatorname{supp}(\mathbf{z}) = \operatorname{supp}(\mathbf{x})$ and multiplication by an invertible diagonal matrix does not change the support of a vector. This establishes (14) with the corresponding probability.

As a consequence of (79) we have the following error bound

$$\frac{\|\tilde{\mathbf{z}} - \mathbf{z}\|_2}{\|\mathbf{z}\|_2} \le \frac{3\sigma\sqrt{N_\tau N_\beta}}{\|\mathbf{y}\|_2} \tag{54}$$

which holds with probability at least

$$\big(1-p_1)(1-p_2\big)(1-e^{-\frac{N_R N_t}{25}}\big)\big(1-2(N_\tau N_\beta)^{-1}(2\pi\log(N_\tau N_\beta)+K(N_\tau N_\beta)^{-1}) - \mathcal{O}((N_\tau N_\beta)^{-2\log 2})\big),$$

where the probabilities p_1, p_2 are as in Lemma 4. Using the fact that $\tilde{\mathbf{z}} = \mathbf{D}\tilde{\mathbf{x}}$, we compute

$$\frac{1}{\kappa(\mathbf{D})}\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \le \frac{\|\mathbf{D}(\tilde{\mathbf{x}} - \mathbf{x})\|_2}{\|\mathbf{D}\mathbf{x}\|_2} = \frac{\|\tilde{\mathbf{z}} - \mathbf{z}\|_2}{\|\mathbf{z}\|_2},$$

or, equivalently,

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \le \kappa(\mathbf{D})\frac{\|\tilde{\mathbf{z}} - \mathbf{z}\|_2}{\|\mathbf{z}\|_2}. \tag{55}$$

Proceeding along the lines of (45)-(47), we estimate

$$\mathbb{P}\big(\kappa(\mathbf{D}) \le 2\big) \ge 1 - N_R N_T e^{-\frac{N_t(1-\sqrt{1/3})^2}{2}}. \tag{56}$$

The bound (15) follows now from combining (54) with (55) and (56). ❑

4 Recovery of targets in the Doppler case

In this section we analyze the case of moving targets/antennas, as described in 2.2. As in the stationary setting, we assume that $s_i(t)$ is a periodic, continuous-time white Gaussian noise signal of period-duration T seconds and bandwidth B. The transmit waveforms are normalized so that the total transmit power is fixed, independent of the number of transmit antennas. Thus, we assume that the entries of $s_i(t)$ have variance $\frac{1}{N_T}$.

Theorem 5 *Consider* $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$*, where* $\mathbf{A}$ *is as defined in Subsection 2.2 and* $\mathbf{w}_i \in \mathcal{CN}(0, \sigma^2)$*. Choose the discretization stepsizes to be* $\Delta_\beta = \frac{2}{N_R N_T}$, $\Delta_\tau = \frac{1}{2B}$ *and* $\Delta_f = \frac{1}{T}$*. Let* $d_T = 1/2, d_R = N_T/2$ *or* $d_T = N_R/2, d_R = 1/2$*, and suppose that*

$$N_t \ge 128, \qquad \max\{N_\tau, N_f, \sqrt{N_\tau, N_f}\} \ge \sqrt{N_\beta}, \qquad \text{and} \qquad \big(\log(N_\tau N_\beta)\big)^3 \le N_t.$$

If $\mathbf{x}$ *is drawn from the generic* K*-sparse target model with*

$$K \le K_{\max} := \frac{c_0 N_\tau N_f N_R}{6\log(N_\tau N_f N_\beta)}$$

for some constant $c_0 > 0$*, and if*

$$\min_{k\in I} |\mathbf{x}_k| > \frac{10\sigma}{\sqrt{N_R N_t}}\sqrt{2\log N_\tau N_f N_\beta},$$

then the solution $\tilde{\mathbf{x}}$ *of the debiased lasso computed with* $\lambda = 2\sigma\sqrt{2\log(N_\tau N_f N_\beta)}$ *obeys*

$$\operatorname{supp}(\tilde{\mathbf{x}}) = \operatorname{supp}(\mathbf{x}),$$

with probability at least

$$(1-p_1)(1-p_2)(1-p_3)(1-p_4),$$

and

$$\frac{\|\tilde{\mathbf{x}}-\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \le \frac{\sigma\sqrt{12N_tN_R}}{\|\mathbf{y}\|_2}$$

with probability at least

$$(1-p_1)(1-p_2)(1-p_3)(1-p_4)(1-p_5),$$

where

$$p_1 = e^{-\frac{(1-\sqrt{1/3})^2N_t}{2}} + N_T e^{-(\sqrt{3/2}-\sqrt{2})N_t},$$

$$p_2 = 2(N_RN_T)^{-1} + 2(N_\tau N_R N_T)^{-1} + 2(N_f N_R N_T)^{-1} + 6(N_\tau N_f N_R N_T)^{-1} + 2e^{-\frac{N_t(\sqrt{2}-1)^2}{4}},$$

$$p_3 = N_R N_T e^{-\frac{(1-\sqrt{1/3})^2N_t}{2}}, \qquad p_4 = e^{-\frac{N_RN_t}{25}},$$

and

$$p_5 = 2(N_\tau N_\beta)^{-1}(2\pi\log(N_\tau N_\beta) + S(N_\tau N_\beta)^{-1}) + \mathcal{O}((N_\tau N_\beta)^{-2\log 2}).$$

Proof: The proof is very similar to that of Theorem 1. Below we will establish the analogs of the key steps, Lemma 2, Lemma 3, and Lemma 4, and leave the rest to the reader. ❏

Lemma 6 *Let* $\mathbf{A}$ *be as defined in Theorem 5. Then*

$$\mathbb{P}\Big(\|\mathbf{A}\|_{op}^2 \le 2N_tN_fN_RN_T\Big) \ge 1 - N_T e^{-N_t(\frac{3}{2}-\sqrt{2})}. \tag{57}$$

Proof: We proceed as in the proof of Lemma 2. There holds $\|\mathbf{A}\|_{\mathrm{op}}^2 = \|\mathbf{A}\mathbf{A}^*\|_{\mathrm{op}}$. It is convenient to consider $\mathbf{A}\mathbf{A}^*$ as block matrix

$$\begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \dots & \mathbf{B}_{1,N_R} \\ \vdots & \ddots & & \vdots \\ \mathbf{B}^*_{N_R,1} & & & \mathbf{B}_{N_R,N_R} \end{bmatrix},$$

where the blocks $\{\mathbf{B}_{i,i'}\}_{i,i'=1}^{N_R}$ are matrices of size $N_t \times N_t$. We claim that $\mathbf{A}\mathbf{A}^*$ is a block-Toeplitz matrix (i.e., $\mathbf{B}_{i,i'} = \mathbf{B}_{i+1,i'+1}, i = 1,\dots,N_R-1$) and the individual blocks $\mathbf{B}_{i,i'}$ are circulant matrices. To see this, recall the structure of $\mathbf{A}$ and consider the entry $\mathbf{B}_{[i,l;i',l']}$, $i,i' = 1,\dots,N_R; l,l' =$

$1, \ldots, N_t$:

$$\mathbf{B}_{[i,l;i',l']} = (\mathbf{A}\mathbf{A}^*)_{[i,l;i',l']} = \sum_\beta \sum_\tau \sum_f \mathbf{A}_{[i,l;\tau,f,\beta]} \mathbf{A}_{[i',l';\tau,f,\beta]}$$

$$= \sum_\beta e^{j2\pi d_R (i-i')\beta} \sum_{k=1}^{N_T} \sum_{k'=1}^{N_T} e^{j2\pi d_T (k-k')\beta} G_{k,k'}(l,l') \sum_{m=1}^{N_f} e^{j2\pi(l-l')\Delta_t m \Delta_f}$$

$$= \sum_{n=0}^{N_R N_T - 1} e^{j2\pi(i-i')\frac{nN_T}{N_R N_T}} \sum_{k=1}^{N_T} \sum_{k'=1}^{N_T} e^{j2\pi(k-k')\frac{n}{N_R N_T}} G_{k,k'}(l,l') N_f \delta_{l-l'} \tag{58}$$

$$= N_T N_R N_f \sum_{k=1}^{N_T} \|\mathbf{s}_k\|^2 \delta_{i-i'} \delta_{l-l'} \tag{59}$$

where we have used in (58) that $N_f = \frac{2B}{\Delta_f} = 2BT$, whence $\sum_{m=1}^{N_f} e^{j2\pi(l-l')m\Delta_t \Delta_f} = N_f \delta_{l-l'}$. Thus

$$\mathbf{A}\mathbf{A}^* = (N_T N_R N_f \sum_{k=1}^{N_T} \|\mathbf{s}_k\|^2)\, \mathbf{I}, \tag{60}$$

i.e., $\mathbf{A}\mathbf{A}^*$ is just a scaled identity matrix. Since $\mathbf{s}_k$ is a Gaussian random vector with $\mathbf{s}_k(j) \sim \mathcal{CN}(0,1)$, Lemma 9 yields

$$\mathbb{P}\Big(\|\mathbf{s}_k\|_2^2 - (\mathbb{E}\|\mathbf{s}_k\|_2)^2 \geq t(t + 2\mathbb{E}\|\mathbf{s}_k\|_2)\Big) \leq e^{-t^2/2}, \tag{61}$$

where we note that $\mathbb{E}\|\mathbf{s}_k\|_2 = \sqrt{\frac{N_t}{N_T}}$. We choose $t = (\sqrt{2}-1)\sqrt{N_t}$, and obtain, after forming the union bound over $k = 1, \ldots, N_t - 1$,

$$\mathbb{P}\Big(\sum_{k=1}^{N_T} \|\mathbf{s}_k\|_2^2)^2 \geq 2N_t\Big) \leq N_T e^{-N_t(\frac{3}{2}-\sqrt{2})}. \tag{62}$$

The bound (57) now follows from (60). ❏

Next we establish a coherence bound for $\mathbf{A}$.

Lemma 7 *Let* $\mathbf{A}$ *be as defined in the Doppler case. Assume that*

$$N \geq \sqrt{N_\beta} \log(NN_\beta) < \frac{N_t}{30}, \tag{63}$$

where $N := \max\{N_\tau, N_f, \sqrt{N_\tau N_f}\}$. *Then*

$$\max_{(\tau,f,\beta)\neq(\tau',f',\beta')} \big|\langle \mathbf{A}_{\tau,f,\beta}, \mathbf{A}_{\tau',f',\beta'} \rangle\big| \leq 3N_R \sqrt{N_t \log(N_\tau N_f N_\beta)}$$

with probability at least $1 - 2(N_R N_T)^{-1} - 2(N_\tau N_R N_T)^{-1} - 2(N_f N_R N_T)^{-1} - 6(N_\tau N_f N_R N_T)^{-1}$.

Proof:

We have that $\mathbf{A}_{\tau,f,\beta} = \mathbf{a}_R(\beta) \otimes (\mathbf{S}_{\tau,f}\mathbf{a}_T(\beta))$. A standard calculation shows that

$$|\langle \mathbf{S}_{\tau,f}\mathbf{a}_T(\beta), \mathbf{S}_{\tau',f'}\mathbf{a}_T(\beta')\rangle| = |\langle \mathbf{S}_{\tau-\tau',f-f'}\mathbf{a}_T(\beta), \mathbf{a}_T(\beta')\rangle| \tag{64}$$

for $\tau, \tau' = 0, \ldots, N_\tau - 1, f, f' = 0, \ldots, N_f - 1$, thus we only need to consider $|\langle \mathbf{S}_{\tau,f}\mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta')\rangle|$. As in the proof of Lemma 3 we distinguish several cases.

Case (a) $\beta \neq \beta', \tau = 0, f = 0$**:** In this case we are concerned with $|\langle \mathbf{S}\mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta')\rangle|$, which is the same as Case (i) of Lemma 3, except that in the present case we have a bit more flexibility in choosing t in the analogous version of (29). Here we can choose $t = 3\sqrt{\frac{1}{N_t}\log(NN_RN_T)}$, where $N = \max\{N_\tau, N_f, \sqrt{N_\tau N_f}\}$. Proceeding then as in the proof of Case (i) of Lemma 3 we obtain

$$\mathbb{P}\Big(|\langle \mathbf{A}_{\tau,f,\beta}, \mathbf{A}_{\tau,f,\beta'}\rangle| \leq 3N_R\sqrt{N_t \log(N_\tau N_R N_T)}\Big) \geq 1 - 2(N_R N_T)^{-1}. \tag{65}$$

Case (b) $\beta \neq \beta', \tau \neq 0, f = 0$**:** This is exactly the same as Case (ii) of Lemma 3. We obtain

$$\mathbb{P}\Big(|\langle \mathbf{A}_{\tau,f,\beta}, \mathbf{A}_{\tau',f,\beta'}\rangle| \leq 3N_R\sqrt{N_t \log(N_\tau N_R N_T)}\Big) \geq 1 - 2(N_\tau N_R N_T)^{-1}. \tag{66}$$

Case (c) $\beta \neq \beta', \tau = 0, f \neq 0$**:** It is well known that $(\mathbf{T}_\tau \mathbf{x})^\wedge = \mathbf{M}_{-\tau}\hat{\mathbf{x}}$. Hence, by Parseval's theorem, $\langle \mathbf{T}_\tau \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{M}_{-\tau}\hat{\mathbf{x}}, \hat{\mathbf{y}}\rangle$. Since the normal distribution is invariant under Fourier transform, this case is therefore already covered by Case (b), and we leave the details to the reader. We get

$$\mathbb{P}\Big(|\langle \mathbf{A}_{\tau,f,\beta}, \mathbf{A}_{\tau,f',\beta'}\rangle| \leq 3N_R\sqrt{N_t \log(N_f N_R N_T)}\Big) \geq 1 - 2(N_f N_R N_T)^{-1}. \tag{67}$$

Case (d) $\beta \neq \beta', \tau \neq 0, f \neq 0$**:** This is similar to Case (ii) of Lemma 3. The only difference is that we have $N_t N_f N_R N_T$ different possibilities to consider when forming the union bound (the additional factor N_f is of course due to frequency shifts associated with the Doppler effect). Thus in this case the bound reads

$$\mathbb{P}\Big(|\langle \mathbf{A}_{\tau,f,\beta}, \mathbf{A}_{\tau',f',\beta'}\rangle| \leq 3N_R\sqrt{N_t \log(N_\tau N_f N_R N_T)}\Big) \geq 1 - 2(N_\tau N_f N_R N_T)^{-1}. \tag{68}$$

Case (e) $\beta = \beta'$**:** We need to bound $|\langle \mathbf{T}_\tau \mathbf{M}_f \mathbf{S}\mathbf{a}_T(\beta), \mathbf{S}\mathbf{a}_T(\beta)\rangle|$, where we recall that $\mathbf{S}\mathbf{a}_T(\beta)$ is a Gaussian random vector with variance N_T. (We note that a related case is covered by Theorem 5.1 in [20], which considers $\langle T_\tau M_f h, h\rangle$, where h is a Steinhaus sequence.) This case is essentially taken care off by Case (iii) of Lemma 3, by noting that a Gaussian random vector of variance σ remains Gaussian (with the same σ) when pointwise multiplied by a fixed vector with entries from the torus. The only difference is that, as in Case (d) above, we have $N_t N_f N_R N_T$ different possibilities to consider when forming the union bound. Hence, the bound in this case becomes

$$\mathbb{P}\Big(\max|\langle \mathbf{A}_{\tau,f,\beta}, \mathbf{A}_{\tau',f',\beta}\rangle| \leq 3N_R\sqrt{N_t \log(N_\tau N_f N_R N_T)}\Big) \geq 1 - 4(N_t N_f N_R N_T)^{-1}. \tag{69}$$

❏

Lemma 8 *Let* $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}^{-1}$, *where the entries of the* $N_\tau N_f N_\beta \times N_\tau N_f N_\beta$ *diagonal matrix are given by* $\mathbf{D}_{(\tau,f,\beta),(\tau,f,\beta)} = \|\mathbf{A}_{\tau,\beta}\|_2$. *Under the conditions of Theorem 1 there holds*

$$\mathbb{P}\Big(\|\tilde{\mathbf{A}}\|_{op}^2 < 6N_T\Big) \geq 1 - p_1, \tag{70}$$

where

$$p_1 = e^{-\frac{(1-\sqrt{1/3})^2 N_t}{2}} + N_T e^{-(\sqrt{3/2}-\sqrt{2})N_t},$$

and

$$\mathbb{P}\Big(\mu(\tilde{\mathbf{A}}) \leq 6\sqrt{\frac{1}{N_t}\log(N_\tau N_f N_R N_T)}\Big) \geq 1 - p_2, \tag{71}$$

where

$$p_2 = 2(N_R N_T)^{-1} + 2(N_\tau N_R N_T)^{-1} + 2(N_f N_R N_T)^{-1} + 6(N_\tau N_f N_R N_T)^{-1} + 2e^{-\frac{N_t(\sqrt{2}-1)^2}{4}},$$

Proof: Since the proof of this lemma follows closely that of Lemma 4, we omit it. ❏

5 Numerical Experiments

Next we illustrate the performance of the compressive MIMO radar developed in previous sections. We consider a Doppler-free scenario. The following parameters are used in this example: $N_T = 8$ transmit antennas, $N_R = 8$ receive antennas, $N_t = 64$ samples, $N_\tau = N_t$ range values.

At each experiment K scatterers of unit amplitude are placed randomly on the range/azimuth grid, i.e the vector $\mathbf{x}$ has K unit entries at random locations along the vector. White Gaussian noise is added to the composite data vector $\mathbf{Ax}$ with variance σ^2 determined to as to produce the specified output signal-to-noise ratio (see also item (iv) of the Remark after Theorem 1). The lasso solution $\hat{\mathbf{x}}$ is calculated with λ as specified in Theorem 1. The numerical algorithm to solve (10) was implemented in Matlab using TFOCS [1]. The experiment is repeated 100 times using independent noise realizations.

The probabilities of detection P_d and false alarm P_{fa} are computed as follows. The values of the estimated vector $\hat{\mathbf{x}}$ corresponding to the true scatterer locations are compared to a threshold. Detection is declared whenever a value exceeds the threshold. The probability of detection is defined as the number of detections divided by the total number of scatterers K. Next the values of the estimated vector $\hat{\mathbf{x}}$ corresponding to locations not containing scatterers are compared to a threshold. A false alarm is declared whenever one of these values exceeds the threshold. The probability of false alarm is defined as the number of false alarms divided by the total number of scatterers K. The probabilities of detection and false alarm are averaged over the 100 repetitions of the experiment.

The probabilities are re-computed for a range of values of the threshold to produce the so-called Receiver Operating Characteristics (ROC) [14, 28, 25] - the graph of P_d vs. P_{fa}. As the threshold decreases, the probability of detection increases and so does the probability of false alarm. In practice the threshold is usually adjusted to as to achieve a specified probability of false alarm.

Figures 1, 2, 3 and 4 depict the ROC for different values of the output signal to noise ratio. We note that the probability of detection increases as the SNR increases and decreases as K, the number of scatterers increases.

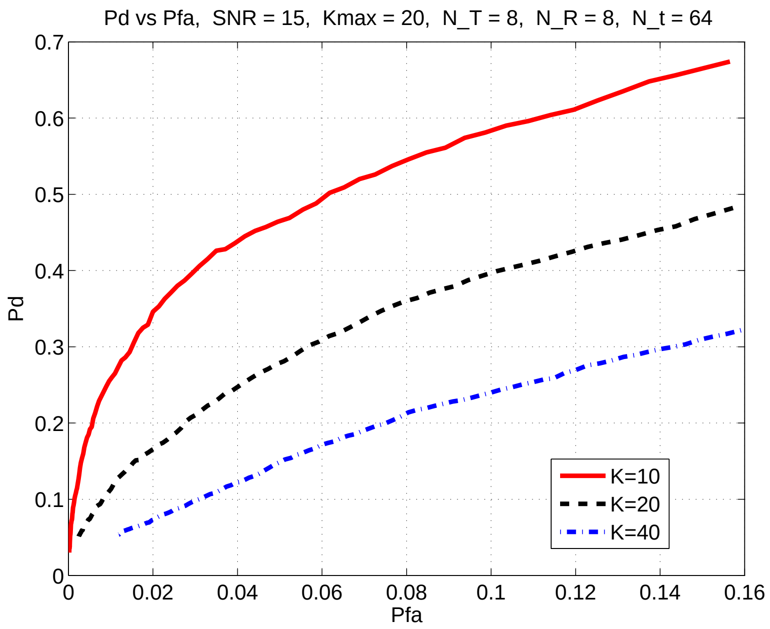

Figure 1. Probability of detection vs. probability of false alarm for SNR = 15 dB, and three values of K: $K_{\max}/2, K_{\max}, 2K_{\max}$.

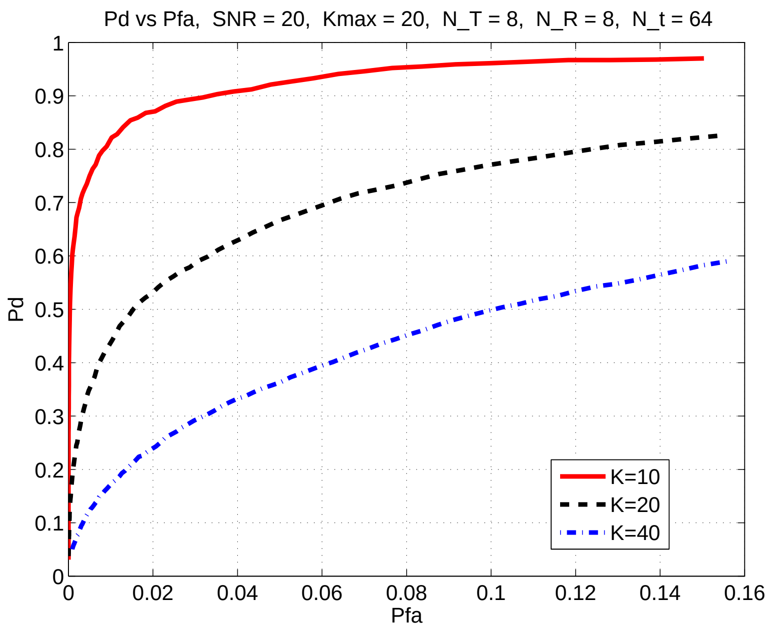

Figure 2. Probability of detection vs. probability of false alarm for SNR = 20 dB, and three values of K: $K_{\max}/2, K_{\max}, 2K_{\max}$.

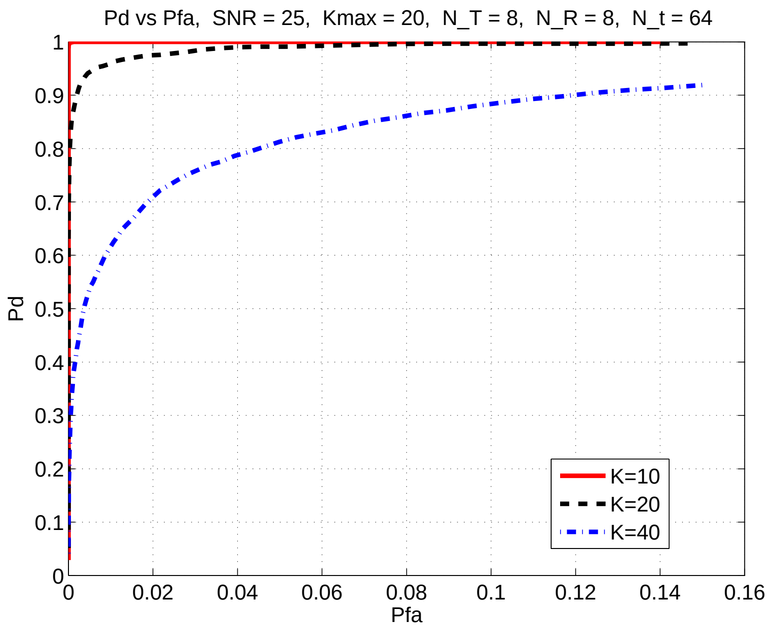

Figure 3. Probability of detection vs. probability of false alarm for SNR = 25 dB, and three values of K: $K_{\max}/2, K_{\max}, 2K_{\max}$.

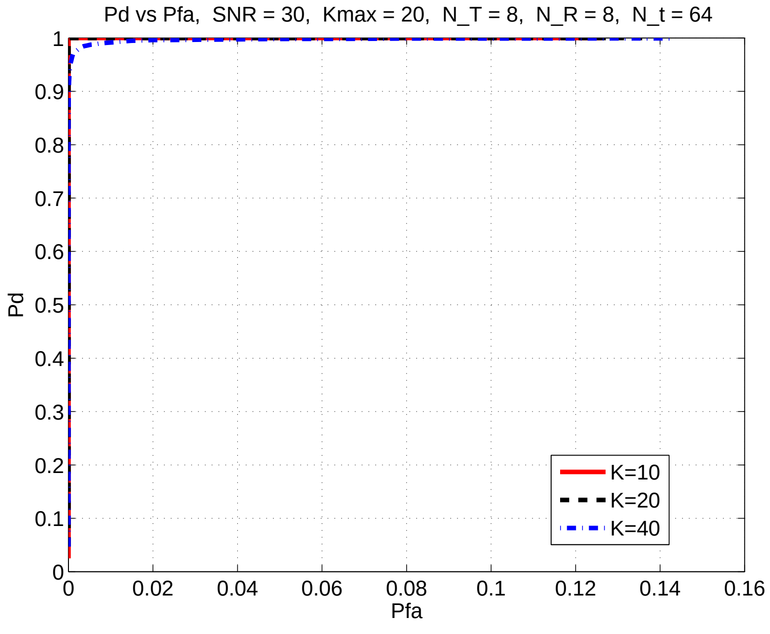

Figure 4. Probability of detection vs. probability of false alarm for SNR = 30 dB, and three values of K: $K_{\max}/2, K_{\max}, 2K_{\max}$.

6 Conclusion

Techniques from compressive sensing and sparse approximation make it possible to exploit the sparseness of radar scenes to potentially improve system performance of MIMO radar. In this paper we have derived a mathematical framework that yields explicit conditions for the radar waveforms and the transmit and receive arrays so that the radar sensing matrix has small coherence and robust sparse recovery in the presence of noise becomes possible. Our approach relies on a deterministic (and very specific) positioning of transmit and receive antennas and random waveforms. It seems plausible that results similar to the ones derived in this paper can be established for the case where the antenna locations are chosen at random and the transmission signals are deterministic. This would be of interest, since one could then potentially take advantage of specific properties of recently designed deterministic radar waveforms such as in [2, 19].

Appendix A

In this appendix we collect some auxiliary results.

Lemma 9 *[29, Proposition 34] Let* $\mathbf{x} \in \mathbb{C}^n$ *be a vector with* $x_k \sim \mathcal{CN}(0, \sigma^2)$*, then for every* $t > 0$ *one has*

$$\mathbb{P}\Big(\|\mathbf{x}\|_2 - \mathbb{E}\|\mathbf{x}\|_2 > t\Big) \le e^{-\frac{t^2}{2\sigma^2}}. \tag{72}$$

The following lemma, which relates moments and tails, can be found e.g. in [22, Proposition 6.5].

Lemma 10 *Suppose* Z *is a random variable satisfying*

$$(\mathbb{E}|Z|^p)^{1/p} \le \alpha \beta^{1/p} p^{1/\gamma} \qquad \text{for all } p \ge p_0$$

for some constants $\alpha, \beta, \gamma, p_0 > 0$*. Then*

$$\mathbb{P}(|Z| \ge e^{1/\gamma} \alpha u) \le \beta e^{-u^\gamma/\gamma}$$

for all $u \ge p_0^{1/\gamma}$*.*

The following lemma is a rescaled version of Lemma 3.1 in [23].

Lemma 11 *Let* $\mathbf{A} \in \mathbb{C}^{n\times m}$ *be a Gaussian random matrix with* $A_{i,j} \sim \mathcal{CN}(0, \sigma^2)$*. Then for all* $\mathbf{x}, \mathbf{y} \in \mathbb{C}^m$ *with* $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = \sqrt{m}$ *and all* $t > 0$

$$\mathbb{P}\Big\{|\frac{1}{n\sigma^2}\langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y}\rangle - \langle \mathbf{x}, \mathbf{y}\rangle| > tm\Big\} \le 2\exp\Big(-n\frac{t^2}{C_1 + C_2 t}\Big),$$

with $C_1 = \frac{4e}{\sqrt{6\pi}}$ *and* $C_2 = \sqrt{8}e$*.*

The next lemma is a slight generalization of a result by Hanson and Wright on tail bounds for quadratic forms [12].

Lemma 12 *Let* $\mathbf{M} = \{m_{ij}\}_{i,j=1}^{n}$ *be a normal matrix and let* $X_i, i = 0, \dots, n-1$ *be independent,* $\mathcal{CN}(0,1)$*-distributed random variables. Denote*

$$S_n = \sum_{i,j=0}^{n-1} m_{ij} X_i \bar{X}_j.$$

Then for all $t > 0$

$$\mathbb{P}\Big(S_n \geq t + \mathbb{E}S_n\Big) \leq \exp\big(-C\min\{\frac{t}{\sigma\|\mathbf{M}\|_{op}}, \frac{t^2}{\sigma^2\|\mathbf{M}\|_F^2}\}\big),$$

where C *is a numerical constant independent of* $\mathbf{M}$ *and* n.

Proof: The proof follows essentially the same steps as the proof of the main theorem in [12], which considers the case where $\mathbf{M}$ is hermitian and the x_i are real-valued. Extending the x_i to the complex case is trivial, thus the only modification that needs to be addressed is the extension of $\mathbf{M}$ from the hermitian to the normal case. But Lemma 5 in [12] holds for normal matrices as well, therefore the lemma follows. ❑

For convenience we state the following version of Bernstein's inequality, which will be used in the proof of Lemma 14.

Theorem 13 (See e.g. [27]) *Let* $X_1, \dots, X_n$ *be independent random variables with zero mean such that*

$$\mathbb{E}|X_i|^p \leq \frac{1}{2}p! K^{p-2} v_i, \qquad \text{for all } i = 1, \dots, n; p \in \mathbb{N}, p \geq 2,$$

for some constants $K > 0$ *and* $v_i > 0, i = 1, \dots, n$. *Then, for all* $t > 0$

$$\mathbb{P}\Big(|\sum_{i=1}^{n} X_i| \geq t\Big) \leq 2\exp\Big(-\frac{t^2}{2v + Kt}\Big), \tag{73}$$

where $v := \sum_{i=1}^{n} v_i$.

We also need the following deviation inequality for unbounded random variables. It is a complex-valued and slightly sharpened version of Lemma 6 in [13], the better constant will be useful when we apply Lemma 14 in the proof of Lemma 3.

Lemma 14 *Let* X_i *and* Y_i, $i = 1, \dots, n$, *be sequences of i.i.d. complex Gaussian random variables with variance* σ. *Then,*

$$\mathbb{P}\Big(|\sum_{i=1}^{n} \bar{X}_i Y_i| > t\Big) \leq 2\exp\big(-\frac{t^2}{\sigma^2(n\sigma^2 + 2t)}\big). \tag{74}$$

Proof: In order to apply Bernstein's inequality, we need to compute the moments $\mathbb{E}|X_i Y_i|^p$. Since X_i and Y_i are independent, there holds

$$\mathbb{E}(|X_i Y_i|^p) = \mathbb{E}(|X_i|^p)\mathbb{E}(|Y_i|^p) = (\mathbb{E}(|X_i|^p))^2.$$

The moments of X_i are well-known:

$$\mathbb{E}|X_i|^{2p} = p!\,\sigma^{2p},$$

hence

$$(\mathbb{E}|X_i|^{2p})^2 = (2p!)^2(\sigma^{2p})^2 \leq \frac{1}{4}(2p)!(\sigma^2)^{2p} \leq \frac{1}{2}(2p)!(\sigma^2)^{2p-2}\frac{(\sigma^2)^2}{2}.$$

We apply Bernstein's inequality (73) with $K = \sigma^2$ and $v_i = \frac{(\sigma^2)^2}{2}, i = 1, \ldots, n$ and obtain (74).
❏

Appendix B

We consider a general linear system of equations $\mathbf{\Psi x} = \mathbf{y}$, where $\mathbf{\Psi} \in \mathbb{C}^{n\times m}$, $\mathbf{x} \in \mathbb{C}^m$ and $n \leq m$. We introduce the following generic K-sparse model:

- The support $I \subset \{1, \ldots, m\}$ of the K nonzero coefficients of $\mathbf{x}$ is selected uniformly at random.
- The non-zero entries of $\text{sgn}(\mathbf{x})$ form a Steinhaus sequence, i.e., $\text{sgn}(\mathbf{x}_k) := \mathbf{x}_k/|\mathbf{x}_k|, k \in I$, is a complex random variable that is uniformly distributed on the unit circle.

The following theorem is a slightly extended version of Theorem 1.3 in [3].

Theorem 15 *Given $\mathbf{y} = \mathbf{\Psi x} + \mathbf{w}$, where $\mathbf{\Psi}$ has all unit-ℓ_2-norm columns, $\mathbf{x}$ is drawn from the generic K-sparse model and $\mathbf{w}_i \sim \mathcal{CN}(0, \sigma^2)$. Assume that*

$$\mu(\mathbf{\Psi}) \leq \frac{C_0}{\log m}, \tag{75}$$

where $C_0 > 0$ is a constant independent of n, m. Furthermore, suppose

$$K \leq \frac{c_0 m}{\|\mathbf{\Psi}\|_{op}^2 \log m} \tag{76}$$

for some constant $c_0 > 0$ and that

$$\min_{k\in I} |\mathbf{x}_k| > 8\sigma\sqrt{2\log m}. \tag{77}$$

Then the solution $\hat{\mathbf{x}}$ to the debiased lasso computed with $\lambda = 2\sigma\sqrt{2\log m}$ obeys

$$\text{supp}(\hat{\mathbf{x}}) = \text{supp}(\mathbf{x}), \tag{78}$$

and

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\sigma\sqrt{3n}}{\|\mathbf{y}\|_2} \tag{79}$$

with probability at least

$$1 - 2m^{-1}(2\pi\log m + Km^{-1}) - \mathcal{O}(m^{-2\log 2}). \tag{80}$$

Proof: The paper [3] treats only the real-values case. However it is not difficult to see that the results by Candès and Plan can be extended to the complex setting if their definition of the sign-function is replaced by (9) and consequently their generic sparse model is replaced by the generic sparsity model introduced in the beginning of this appendix. The proofs of the theorems in [3] can then be easily adapted to the complex case via some straightforward modifications, such as replacing in many steps $\langle\cdot,\cdot\rangle$ by its real part, $\mathrm{Re}\langle\cdot,\cdot\rangle$ and replacing certain scalar quantities by its conjugate analogs. To give a concrete example of such a modification, consider (in the notation of [3]) the inequality right before eq.(3.10) in [3],

$$|\hat{\beta}_i| = |\beta_i + h_i| \geq |\beta_i| + \mathrm{sgn}(\beta_i)h_i.$$

This inequality needs to be replaced by its complex counterpart

$$|\hat{\beta}_i| = |\beta_i + h_i| \geq |\beta_i| + \mathrm{Re}(\mathrm{sgn}(\beta_i)\overline{h_i}).$$

By carrying out these easy modifications (the details of which are left to the reader) we can readily establish (78) analogous to (1.11) of Theorem 1.3 in [3].

Once we have recovered the support of $\mathbf{x}$, call it I, we can solve for the coefficients of $\mathbf{x}$ by solving the standard least squares problem $\min \|\mathbf{A}_I\mathbf{x}_I - \mathbf{y}\|_2$, where $\mathbf{A}_I$ is tbe submatrix of $\mathbf{A}$ whose columns correspond to the support set I, and similarly for $\mathbf{x}_I$. Statement (79) follows by noting that the proof of Theorem 3.2 in [3] yields as side result that with high probability the eigenvalues of any submatrix $\mathbf{A}_I^*\mathbf{A}_I$ with $|I| \leq K$ are contained in the interval $[1/2, 3/2]$, which of course implies that $\kappa(\mathbf{A}_I) \leq \sqrt{3}$. The statement follows now by substituting this bound into the standard error bound, eq. (5.8.11) in [17]. ❑

Acknowledgements

T.S. wants to thank Sasha Soshnikov for helpful discussions on random matrix theory and Haichao Wang for a careful reading of the manuscript.

References

[1] S. Becker, E. Candès, and M. Grant. Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation* 3(3), 165–218, 2011.

[2] J.J. Benedetto and S. Datta. Construction of infinite unimodular sequences with zero auto correlation. *Advances in Computational Mathematics*, 32:191–207, 2010.

[3] E.J. Candès and Y. Plan. Near-ideal model selection by ℓ_1 minimization. *Annals of Statistics*, 37(5A):2145–2177, 2009.

[4] E.J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies. *IEEE Trans. on Information Theory*, 52:5406–5425, 2006.

[5] E.J. Candès and Y. Plan. A Probabilistic and RIPless Theory of Compressed Sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.

[6] M. Cheney and B. Borden. *Fundamentals of Radar Imaging*. Society for Industrial and Applied Mathematics, 2009.

[7] Y. Chi, A. Pezeshki, L. Scharf, and R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Processing*, 59(5): 2182-2195, 2011.

[8] D. L. Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52(4):1289–1306, 2006.

[9] A. Fannjiang and W. Liao. Coherence pattern-guided compressive sensing with unresolved grids. *SIAM J. Imaging Sci.*, 5:179–202, 2012.

[10] B. Friedlander. Adaptive Signal Design for MIMO Radar. In J. Li and P. Stoica, editors, *MIMO Radar Signal Processing*, chapter 5. John Wiley & Sons, 2009.

[11] B. Friedlander. On the relationship between MIMO and SIMO radars. *IEEE Trans. Signal Processing*, 57(1):394–398, January 2009.

[12] D.L. Hanson and F.T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

[13] J. Haupt, W. Bajwa, G. Raz, and R. Nowak. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE Trans. Inform. Theory*, 56(11):5862–5875, 2010.

[14] C. Helstrom. *Elements of Signal Detection and Estimation*. Prentice Hall, 1995, 2005.

[15] M. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. on Signal Processing*, 57(6):2275–2284, 2009.

[16] M. Herman and T. Strohmer. General deviants: an analysis of perturbations in compressed sensing. *IEEE Journal of Selected Topics in Signal Processing: Special Issue on Compressive Sensing*, 4(2):342–349, 2010.

[17] R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[18] J. Li and P. Stoica, editors. *MIMO Radar Signal Processing*. John Wiley & Sons, 2009.

[19] A. Pezeshki, A.R. Calderbank, W. Moran, and S.D. Howard. Doppler resilient Golay complementary waveforms. *IEEE Transactions on Information Theory*, 54(9):4254–4266, 2008.

[20] G.E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Processing*, 56(11):5376–5388, 2008.

[21] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin. Sparsity and compressed sensing in radar imaging. *Proceedings of the IEEE*, 98(6):1006 –1020, 2010.

[22] H. Rauhut. Compressive sensing and structured random matrices. In *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.

[23] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory*, 54(5):2210–2219, 2008.

[24] A. W. Rihaczek. *High-Resolution Radar*. Artech House, Boston, 1996. (originally published: McGraw-Hill, NY, 1969).

[25] L.L. Scharf. *Statistical Signal Processing*. Prentice Hall, 1990.

[26] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.

[27] A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.

[28] H.L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley-Interscience, 2001.

[29] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Yonina C. Eldar and Gitta Kutyniok, editors, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2010. To Appear. Preprint available at `http://www-personal.umich.edu/~romanv/papers/papers.html`.

[30] Y. Yu, A. Petropulu, and V. Poor. Measurement matrix design for compressive sensing-based MIMO radar. *IEEE Trans. on Signal Processing*, 59(11):5338–5352, 2011.

[31] Y. Yu, A. Petropulu, and V. Poor. CSSF MIMO RADAR: Low-complexity compressive sensing based MIMO radar that uses step frequency. *IEEE Trans. on Aerospace and Electronic Systems*, to appear.