Block Sparse Recovery With Redundant Measurement Matrices and Its Application in Frequency Agile Radar

YUHAN LI
Tsinghua University, Beijing, China

TIANYAO HUANG
University of Science and Technology Beijing, Beijing, China

YIMIN LIU
XIQIN WANG
Tsinghua University, Beijing, China

YONINA C. ELDAR, Fellow, IEEE
Weizmann Institute of Science, Rehovot, Israel

Block compressed sensing (or sparse recovery) and its performance bound, i.e., conditions that guarantee reconstruction of the original sparse vector, have been widely studied. Most scenarios assume that the blocks in the measurement matrix are full-rank. In this setting, phase transition theory provides a precise performance bound on the exact reconstruction of the original vector. However, in many practical applications, the blocks of the measurement matrices may not have full rank, and it becomes impossible to recover the original vector elementwise. In this article, we consider the compressed sensing problem with such redundant measurement matrices and derive a performance bound using phase transition theory. We focus on reconstructing the contribution of each block in the original vector to the observed signal instead of the vector itself. We show that this method is equivalent to a traditional $\ell_{2,1}$ norm minimization after a certain linear transformation. We theoretically prove the transformed $\ell_{2,1}$ norm minimization has a phase transition phenomenon. Based on this result, we derive the closed-form phase transition curve of the method as a tight performance bound. We also apply this result to frequency agile radar for performance evaluation and waveform design. Simulations validate our theoretical conclusions.

Manuscript received 26 February 2024; revised 6 May 2024 and 7 July 2024; accepted 25 July 2024. Date of publication 8 August 2024; date of current version 6 December 2024.

DOI. No. 10.1109/TAES.2024.3436708

Refereeing of this contribution was handled by D. Orlando.

This work was supported by the National Natural Science Foundation of China under Grant 62171259.

Authors' addresses: Yuhan Li, Yimin Liu, and Xiqin Wang are with Electronic Engineering Department, Tsinghua University, Beijing 100084, China, E-mail (liyh20@mails.tsinghua.edu.cn; yiminliu@tsinghua.edu.cn; wangxq_ee@tsinghua.edu.cn); Tianyao Huang is with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China, E-mail: (huangtianyao@ustb.edu.cn); Yonina C. Eldar is with the Faculty of Math and CS, Weizmann Institute of Science, Rehovot 7610001, Israel, E-mail: (yonina.eldar@weizmann.ac.il). *(Corresponding author: Tianyao Huang.)*

I. INTRODUCTION

Sparse recovery has attracted massive attention in many fields such as computer vision [1], radar signal processing [2], [3], [4], [5], [6], and machine learning [7]. The core idea of sparse recovery is to utilize the sparsity in the original $\boldsymbol{x}$ to recover it from the observed signal $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is a measurement matrix and known a priori. Block sparse recovery can be applied when $\boldsymbol{x}$ possesses additional block structure, i.e., nonzero entries of $\boldsymbol{x}$ cluster in several blocks. This arises in numerous situations such as extended target detection [8] in radar, machine learning [9], and data clustering [10].

Denote the ith block of $\boldsymbol{x}$ and $\boldsymbol{A}$ by $\boldsymbol{x}_i$ and $\boldsymbol{A}_i$, respectively. When all the blocks $\boldsymbol{A}_i$ are of full rank, referred to as nonredundant blocks, it is possible to use standard block sparse recovery methods to recover all the entries of $\boldsymbol{x}$. However, in many practical scenarios, one or more blocks $\boldsymbol{A}_i$ can be redundant or rank deficient, containing more columns than the dimension of the underlying subspace [1], [11]. Particularly, in this article, we focus on the target recovery problem with a frequency agile radar (FAR). As will be detailed later in Section IV, the blocks of its measurement matrix are redundant when only a portion of frequencies is transmitted by the radar, which is an important operating mode to counter electromagnetic interference. For a redundant block $\boldsymbol{A}_i$, exact recovery of the elements in $\boldsymbol{x}_i$ is impossible because there are multiple feasible solutions of $\boldsymbol{x}_i$ that generate the same observation $\boldsymbol{A}_i\boldsymbol{x}_i$ while not changing the block sparsity. Instead of exact elementwise reconstruction of the original signal, it is still possible and useful to identify the nonzero blocks in $\boldsymbol{x}$ and reconstruct the signal of each block $\boldsymbol{A}_i\boldsymbol{x}_i$, referred to as blockwise recovery.

Most works on block sparse recovery assume that the blocks in the measurement matrix are full-rank [12], [13], where the random Gaussian ensemble is commonly considered [14]. Block sparse recovery methods reconstruct $\boldsymbol{x}$ by finding the fewest blocks to represent $\boldsymbol{y}$. Under certain conditions [15], the unique block sparse solution can be found by solving

$$P_{\ell_{2,1}}: \hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Sigma_{i=0}^{d-1} \|\boldsymbol{x}_i\|_2, \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}. \tag{1}$$

In practical scenarios, some blocks may be redundant. While elementwise recovery is no longer possible, the authors in [16] and [17] instead focused on the reconstruction of the subspace of each block. Particularly, $\ell'_{2,1}$ norm minimization or $P'_{\ell_{2,1}}$ is proposed as

$$P'_{\ell_{2,1}}: \hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Sigma_{i=0}^{d-1} \|\boldsymbol{A}_i\boldsymbol{x}_i\|_2, \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{2}$$

which penalizes the norm of the reconstructed vectors from the blocks, i.e., $\Sigma_{i=0}^{d-1}\|\boldsymbol{A}_i\boldsymbol{x}_i\|_2$, referred to as $\ell'_{2,1}$ norm of $\boldsymbol{x}$ concerning the blocks $\boldsymbol{A}_i$, opposed to the $\ell_{2,1}$ norm in (1). Here, we concern the theoretical performance bound of $\ell'_{2,1}$ norm minimization method, i.e., the parameter conditions under which $\ell'_{2,1}$ methods exactly reconstruct the original signals. The theoretical conditions for block sparse recovery have been widely considered by former works. For the standard $\ell_{2,1}$ norm minimization, most early works [13], [15], [18] use incoherence-based techniques, including the restricted isometry property and the mutual incoherence property. For $\ell'_{2,1}$ norm minimization, Elhamifar and Vidal [17] analyzed the mutual and cumulative subspace coherence to provide a sufficient condition where it succeeds. This condition was further improved by [19]. However, these incoherence-based tools yield pessimistic and loose bounds. By using the phase transition theory, Li et al. [20] obtained a tight precise performance bound for $P_{\ell_{2,1}}$. In this article, we extend the tight performance bound to $P'_{\ell_{2,1}}$.

The phase transition phenomenon is commonly found in high-dimension geometry. Here, phase transition means an abrupt change in the probability of property occurring as the dimension parameter intersects a critical threshold [21]. For example, ℓ_1 norm minimization faces a phase transition when the number of measurements and nonzero entries in $\boldsymbol{x}$ vary [21], [22], [23], [24]. The authors in [24] and [25] provided a theoretical expression of the threshold, i.e., the theoretical phase transition curve, for ℓ_1 norm minimization when the measurement matrix is Gaussian or sub-Gaussian. Such curves serve as tight bounds guaranteeing recovery performance. Our former work [20] generalized the theoretical phase transition curves to $P_{\ell_{2,1}}$. Yet, these works aim at exactly recovering all the entries of $\boldsymbol{x}$ and assume that blocks $\boldsymbol{A}_i$ are of full rank. In this article, we consider exact blockwise recovery under redundant blocks and derive a corresponding theoretical phase transition curve. One of the difficulties lies in the change of measurement matrix structure. Since the latter measurement matrix has redundant blocks, it no longer satisfies the sub-Gaussian assumptions usually imposed by former works.

One major application scenario of the aforementioned block sparse recovery methods is the target reconstruction in FAR. Simultaneously, the performance bounds can be employed to guide the waveform design of FAR. FAR changes frequencies from pulse to pulse [20]. FAR typically operates in the so-called full frequency mode, where all the frequencies are utilized, covering the entire available bandwidth. In this radar mode, all the blocks of the measurement matrix are full-rank [8], and many sparse recovery methods are proposed due to the sparsity in the target scene. Zhou et al. [26] used a ℓ_1 norm penalty to enhance the sparse imaging with FAR. Wang et al. [27] capitalized on the sparsity through sparse Bayesian learning techniques. Meanwhile, Fei et al.[28] introduced an alternating direction method of multipliers (ADMM) framework to facilitate the computing complexity. Some machine learning methods, including LISTA, are also applied [29], [30]. Some theoretical bounds of sparse recovery methods, like ℓ_1 and $\ell_{2,1}$ norm minimization in for full frequency mode, are also established using mutual coherence of phase transition theory [3], [8], [20]. However, due to hostile or neutral interference, some in-band frequencies are not available, where the FAR should work in the incomplete frequency mode. In the partial frequency mode, some blocks are of deficient rank due to the lack of information, which is not considered by the above methods. This mode is important in the electronic counter-countermeasures (ECCM) scenarios but lacks discussion.

The goal of this article is to derive a tight performance bound for exact blockwise recovery under redundant blocks by phase transition theory. With the performance bound, we then examine which waveform design is optimal for target recovery with FAR.

The main contributions of this article are summarized as follows.

1) We show that the blockwise recovery with redundant blocks via $\ell'_{2,1}$ norm minimization, raised by former works [16], [17], is equivalent to the elementwise recovery via $\ell_{2,1}$ norm minimization with nonredundant blocks after certain linear reformulation.
2) We prove the reformulated problem has the property of phase transition, and thus obtain the performance bound for $\ell'_{2,1}$ norm minimization by deriving its phase transition curve. The entries in the measurement matrix of the reformulated problem are not independent, which is not analyzed in former works on phase transition.
3) We apply the obtained theoretical bounds to FAR and simplify the expressions with elementary function, which eases the calculations and helps guide the waveform design. Simulations validate the accuracy of the application in FAR.
4) The results indicate that using fewer frequencies can improve the reconstruction performance from the perspective of blockwise recovery. In other words, it sacrifices the range resolution to achieve a better Doppler estimation.

The rest of this article is organized as follows. In Section II, we introduce the $\ell'_{2,1}$ norm minimization method for block sparse recovery, as well as existing theoretical results based on phase transition theory. In Section III, we investigate the performance of $\ell'_{2,1}$ norm-based recovery and derive its theoretical phase transition curves. The obtained theoretical results are utilized for FAR in Section IV. In Section V, we present numerical results. Finally, Section VI concludes this article.

Throughout this article, we use $\mathbb{R}$ and $\mathbb{C}$ to denote the real and complex number sets, respectively, and $\mathbb{K}$ is used to represent either of them for convenience. Notations B^n and S^{n-1} are used to represent a unit ball and its sphere in n-d space, respectively. Vectors are written as lowercase boldface letters (e.g., $\boldsymbol{a}$), while matrices are written as uppercase boldface letters (e.g., $\boldsymbol{A}$). For a vector $\boldsymbol{a}$, $\|\boldsymbol{a}\|_i$ denotes the ℓ_i norm of $\boldsymbol{a}$ and the ℓ_2 norm is simplified as $\|\boldsymbol{a}\|$. For a set, both $|\cdot|$ and # denote cardinality. The transpose and

conjugate transpose operators are $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$, respectively. Denotations $\mathbb{E}(\cdot)$, $\mathbb{P}(\cdot)$ and $\mathrm{var}(\cdot)$ represent the expectation, probability, and variance of a random variable, respectively. We use $\mathcal{N}(0,\sigma^2)$ to represent the Gaussian distribution with zero mean and σ^2 variance, and use $\mathcal{N}_{\mathcal{C}}(0,\sigma^2)$ for circularly symmetric complex Gaussian distribution.

II. PRELIMINARY KNOWLEDGE

In this section, we briefly review some preliminary knowledge of sparse recovery. First in Section II-A, we detail the $P'_{\ell_{2,1}}$ method, which aims at blockwise recovery of sparse vectors and is applicable for both nonredundant and redundant measurement matrices. Then in Section II-B, we introduce the theoretical performance bound of sparse recovery based on the concept of phase transition. Current results focus on the standard or nonredundant block sparse recovery.

A. Blockwise Sparse Recovery and the $P'_{\ell_{2,1}}$ Method

Recall the block sparse recovery model $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$. We assume that the signal consist of d blocks, and use the notations $\boldsymbol{A}=[\boldsymbol{A}_0,\ldots,\boldsymbol{A}_{d-1}]\in\mathbb{K}^{n\times md}$ and $\boldsymbol{x}=[\boldsymbol{x}_0^{\mathrm{T}},\ldots,\boldsymbol{x}_{d-1}^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{K}^{md}$ to indicate the block structure. One says that $\boldsymbol{x}$ is s_B block sparse, if there are no more than s_B blocks $\boldsymbol{x}_i\in\mathbb{K}^m$ out of $\boldsymbol{x}$ being nonzero. Denote the estimation of $\boldsymbol{x}$ by $\hat{\boldsymbol{x}}$. We call the original signal $\boldsymbol{x}^*$ is blockwise recovered, if $\hat{\boldsymbol{x}}$ satisfies $\sum_i\|\boldsymbol{A}_i\boldsymbol{x}_i^*-\boldsymbol{A}_i\hat{\boldsymbol{x}}_i\|=0$. Compared to the "elementwise recovery," which requires $\sum_i\|\boldsymbol{x}_i^*-\hat{\boldsymbol{x}}_i\|=0$, the "blockwise recovery" concentrates on the correctness of the contribution of each block, $\boldsymbol{A}_i\boldsymbol{x}_i$. Due to the redundancy of $\boldsymbol{A}_i$ there exist different $\hat{\boldsymbol{x}}$ have the same $\boldsymbol{A}_i\boldsymbol{x}_i$ for all i. They are all considered exact recovery under the blockwise criterion. These two criteria are equivalent when $\boldsymbol{A}_i$ is full-rank.

Corresponding to the blockwise recovery, Elhamifar et al. [17] referred to $P'_{\ell_{2,1}}$, penalizing the norm of the reconstructed vectors from the blocks, repeated as

$$P'_{\ell_{2,1}}:\hat{\boldsymbol{x}}=\arg\min_{\boldsymbol{x}}\Sigma_{i=0}^{d-1}\|\boldsymbol{A}_i\boldsymbol{x}_i\|,\ \text{s.t.}\ \boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}.$$

Here, $P'_{\ell_{2,1}}$ is applicable for both nonredundant and redundant measurement matrices. In nonredundant cases, $\|\boldsymbol{A}_i\boldsymbol{x}_i\|$ equals 0 only when $\boldsymbol{x}_i=\boldsymbol{0}$. In redundant cases, the noncontributing blocks $\boldsymbol{x}_i$, which satisfy $\boldsymbol{A}_i\boldsymbol{x}_i=\boldsymbol{0}$ can be set to $\boldsymbol{x}_i=\boldsymbol{0}$ [17]. Thus, $P'_{\ell_{2,1}}$ results in block sparse solutions.

B. Performance Bound Based on Phase Transition

The phase transition phenomenon was observed by [22] in compressed sensing problems, which shows that the solvabilities of the problems change vastly around certain parameter settings [31]. For example, as shown in Fig. 1, there is a dramatic change in the probability of exact recovery via $P_{\ell_{2,1}}$ from 0 to 1, when d and m are fixed (the numbers of blocks and elements in a block, respectively), and the number of measurements n and the sparsity of the signal s_B vary. One says that the sparse recovery problem has phase transitions on these point pairs (n,s_B). These phase

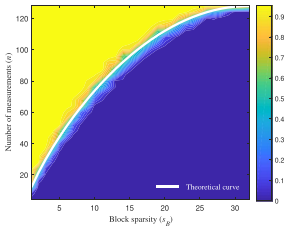


Fig. 1. Probability that $P_{\ell_{2,1}}$ exactly solves $\boldsymbol{x}$. Here, $d=32$, $m=4$, $\boldsymbol{A}\in\mathbb{R}^{n\times md}$ has entries obeying i.i.d. $\mathcal{N}(0,1)$. Exact recovery is proclaimed when the estimate $\hat{\boldsymbol{x}}$ satisfies $\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|_2\le 10^{-5}$. For each (n,s_B), the probability is calculated with 50 trials using different $\boldsymbol{A}$ and $\boldsymbol{x}$. The recovery succeeds with a probability near 1 in the yellow region while hardly succeeds in the deep blue region.

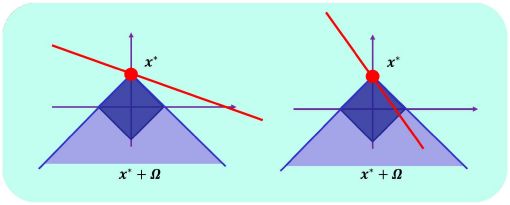


Fig. 2. Geometrical illustration that shows how the result of the recovery depends on the null space and the descent cone.

transition points on the plane of parameters compose the so-called phase transition curve.

Whether the recovery succeeds has a geometrical illustration as shown in Fig. 2. The red points $\boldsymbol{x}^*$ are the true value, and the red lines denote the constraints $\boldsymbol{y}=\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}^*)$ or equivalently, $\boldsymbol{x}^*+null(\boldsymbol{A})$ for random measurement matrices $\boldsymbol{A}$. We use Ω to denote the descent cone of $\|\cdot\|_{2,1}$ at $\boldsymbol{x}$, $\Omega=\{\boldsymbol{u}:\exists\tau>0,\|\boldsymbol{x}+\tau\boldsymbol{u}\|_{2,1}\le\|\boldsymbol{x}\|_{2,1},\ \boldsymbol{u}\ne\boldsymbol{0}\}$, which is the light blue region in Fig. 2. On the left subfigure, $\boldsymbol{x}^*+null(\boldsymbol{A})$ does not intersect with $\boldsymbol{x}^*+\Omega$, i.e., $\boldsymbol{0}\notin\boldsymbol{A}(\Omega)$. Thus, $\boldsymbol{x}^*$ has the smallest $\ell_{2,1}$ norm in the set $\boldsymbol{x}^*+null(\boldsymbol{A})$, which indicates that the solution to $P_{\ell_{2,1}}$ is exactly $\boldsymbol{x}^*$ and the recovery succeeds. On the right, one finds $\boldsymbol{0}\in\boldsymbol{A}(\Omega)$ and the recovery fails. The phase transition phenomenon in Fig. 1 is now transformed into the phase transition phenomenon of the possibility of $null(\boldsymbol{A})$ and Ω intersecting, as the parameter (n,s_B) varies.

With the above illustrations, we can depict the phase transition in a more concise and general way. Consider a random linear map $\boldsymbol{H}\in\mathbb{R}^{n\times d}$, whose null space is drawn uniformly from the Haar measure on the Grassmann manifold of subspace in $\mathbb{R}^d$ [24], [25] (The measurements matrix $\boldsymbol{A}$, often assumed as Gaussian in compressed sensing, meets this requirement almost surely). There is a phase transition in the behavior of the map $\boldsymbol{H}$ as [24]

1) $n\le\delta(\Omega)-C\sqrt{\delta(\Omega)}$ implies $\boldsymbol{0}\in\boldsymbol{H}(\Omega)$ with high probability;
2) $n\ge\delta(\Omega)+C\sqrt{\delta(\Omega)}$ implies $\boldsymbol{0}\notin\boldsymbol{H}(\Omega)$ with high probability.

Here, C is a small positive constant, and Ω is a convex set, which does not contain the origin. The key step to obtain the phase transition curve is to calculate the statistical

dimension, $\delta(\Omega)$, defined as

$$\delta(\Omega) = \delta(\boldsymbol{\theta}(\Omega)) = \mathbb{E}\left[\left(\sup_{\boldsymbol{t}\in\boldsymbol{\theta}(\Omega)} \boldsymbol{t}\cdot\boldsymbol{g}\right)_+^2\right] \tag{3}$$

where $\boldsymbol{g} \in \mathcal{N}(0, I)$ is a standard normal vector and the map $\boldsymbol{\theta}$ retracts the region to a unit sphere, defined as

$$\boldsymbol{\theta}(\boldsymbol{t}) = \begin{cases} \frac{\boldsymbol{t}}{\|\boldsymbol{t}\|}, & \boldsymbol{t} \neq \boldsymbol{0} \\ \boldsymbol{0}, & \boldsymbol{t} = \boldsymbol{0}. \end{cases} \tag{4}$$

In our former work, inspired by the theoretical phase transition curve for standard sparse recovery in [24], we calculated the statistical dimension of the descent cone of $\ell_{2,1}$ norm and obtained the theoretical phase transition curve for block sparse recovery [20]. We repeat the results from [20] as follows.

PROPOSITION 1 ([20]) Given a measurement matrix $\boldsymbol{A} \in \mathbb{K}^{n\times md}$ whose entries are i.i.d. normally distributed and the observed signal $\boldsymbol{y} \in \mathbb{K}^n$, the recovery of the original s_B block sparse signal $\boldsymbol{x} \in \mathbb{K}^{md}$ with block size m via $P_{\ell_{2,1}}$ faces phase transitions when the number of measurements n equals the function

$$\varphi_m(s_B, d) := \inf_{\tau\geq 0}\left\{s_B(m+\tau^2) + (d-s_B)\int_\tau^\infty (u-\tau)^2\phi_m(u)\mathrm{d}u\right\} \tag{5}$$

if $\mathbb{K} = \mathbb{R}$. Here, $\phi_m(u)$ is the probability density function of the χ-distribution with m degrees of freedom, given by

$$\phi_m(u) := \begin{cases} \frac{u^{m-1}e^{-u^2/2}}{2^{m/2-1}\Gamma(\frac{m}{2})}, & u \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

and $\Gamma(\cdot)$ is the gamma function.

If $\mathbb{K} = \mathbb{C}$, the recovery faces phase transitions when n equals

$$\varphi_m^c(s_B, d) = \frac{1}{2}\varphi_{2m}(s_B, d). \tag{7}$$

Phase transition theory is primarily based on the Gaussian measurement matrix. However, other random measurement matrices have also been observed to possess universality [21], [25], which corroborates the validity of the theory in various scenarios, such as Bernoulli matrices, randomly selected Fourier matrices, and randomly selected Hadamard matrices.

Note that this proposition considers $\ell_{2,1}$ norm minimization with a measurement matrix having full-rank blocks. Although the full-rank assumption is met with overwhelming probability by Gaussian measurement matrices, practical measurement matrices usually have structured blocks, which might be rank-deficient or redundant. Therefore, the current results cannot be directly used to obtain the theoretical phase transition for blockwise recovery via $P'_{\ell_{2,1}}$ in redundant cases.

III. PHASE TRANSITION CURVE WITH $P'_{\ell_{2,1}}$

Our goal is to derive the theoretical phase transition curve of blockwise recovery via $P'_{\ell_{2,1}}$ in redundant cases. We first present in Section III-A Theorem 2 and Proposition 3, paving the way to the theoretical curve, of which the mathematical form is shown at the end of Section III-A. Sections III-B and III-C give proof of the theorem and the proposition, respectively.

A. Main Results

1) *Problem Reformation:* Proposition 1 identifies the phase transition curve of elementwise recovery via $P_{\ell_{2,1}}$ in nonredundant cases, inspiring us to reform (2) into an equivalent $\ell_{2,1}$ problem.

THEOREM 2 Consider a measurement matrix $\boldsymbol{A} = [\boldsymbol{A}_0, \boldsymbol{A}_1, \ldots, \boldsymbol{A}_{d-1}] \in \mathbb{K}^{n\times md}$, with the rank of $\boldsymbol{A}_i \in \mathbb{K}^{n\times m}$ being r and a block sparse signal $\boldsymbol{x} = [\boldsymbol{x}_0^{\mathrm{T}}, \boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_{d-1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{K}^{md}$. Then, $\boldsymbol{x}$ can be blockwise recovered via $P'_{\ell_{2,1}}$ from the observation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, if and only if $\boldsymbol{z} \in \mathbb{K}^{rd}$ can be elementwise recovered via $\ell_{2,1}$ norm minimization from the observation $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{z}$. Here, the vectors of the block $\boldsymbol{S}_i \in \mathbb{K}^{n\times r}$ in the matrix $\boldsymbol{S} = [\boldsymbol{S}_0, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_{d-1}] \in \mathbb{K}^{n\times rd}$ forms an orthonormal basis of the space spanned by $\boldsymbol{A}_i$. The vector $\boldsymbol{z} = [\boldsymbol{z}_0^{\mathrm{T}}, \boldsymbol{z}_1^{\mathrm{T}}, \ldots, \boldsymbol{z}_{d-1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{K}^{rd}$ satisfies $\boldsymbol{S}_i\boldsymbol{z}_i = \boldsymbol{A}_i\boldsymbol{x}_i$.

The detailed proof is given in Section III-B. This theorem indicates that the phase transition curve of the blockwise recovery via $P'_{\ell_{2,1}}$ with a redundant measurement matrix $\boldsymbol{A}$ is equivalent to the counterpart of the elementwise recovery via $\ell_{2,1}$ norm minimization with a nonredundant measurement matrix $\boldsymbol{S}$ generated from $\boldsymbol{A}$. Note that we do not require an explicit form of the matrix $\boldsymbol{S}$, though a simple way to obtain $\boldsymbol{S}_i$ is by performing a Gram–Schmidt process on $\boldsymbol{A}_i$. In this sense, $P'_{\ell_{2,1}}$ is same to $P_{\ell_{2,1}}$ after a blockwise orthonormalization as the preprocessing.

2) *Phase Transition With Structured S:* Based on Theorem 2, we then use Proposition 1 to analyze the phase transition of $l_{2,1}$ norm minimization problem, with $\boldsymbol{S}$ being the measurement matrix. While Proposition 1 holds for a Gaussian measurement matrix, $\boldsymbol{S}$ has a certain structure: each column has a unit norm and is orthogonal to other columns in the same block. By imposing some assumptions on $\boldsymbol{S}$ that are easy to satisfy, we show in the following Proposition 3 that there still exists phase transition phenomenon for such a structured measurement matrix, and the phase transition curve is the same as the one indicated in Proposition 1.

PROPOSITION 3 Consider a random matrix $\boldsymbol{S} = [\boldsymbol{S}_0, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_{d-1}] \in \mathbb{K}^{n\times rd}$, where $\boldsymbol{S}_i \in \mathbb{K}^{n\times r}$ satisfies $\boldsymbol{S}_i^{\mathrm{H}}\boldsymbol{S}_i = \boldsymbol{I}$, each column in $\boldsymbol{S}_i$ is uniformly distributed on the unit sphere $\mathcal{S}^{n-1}$, and $\boldsymbol{S}_i$ is independent from $\boldsymbol{S}_j$ for any $i \neq j$. Then for a nonempty, compact, and convex subset of $\mathbb{K}^{rd}$, E that does not contain the origin, and a small constant ϵ, there exists a number n, such that:

1) $\mathbb{P}\{\boldsymbol{0} \notin \{\boldsymbol{y} | \boldsymbol{y} = \boldsymbol{S}\boldsymbol{x}, \boldsymbol{x} \in E\}\}$ is close to 1 if $n \geq (1+\epsilon)\delta(E)$;

2) $\mathbb{P}\{\mathbf{0} \in \{\boldsymbol{y}|\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x}, \boldsymbol{x} \in E\}\}$ is close to 1 if $n \leq (1 - \epsilon)\delta(E)$.

The proof is given in Section III-C, assuming $\boldsymbol{S}$ being real numbered. It is easy to extend to the complex number case following the technique in [20].

In this proposition, we assume that any block $\boldsymbol{S}_i$ is orthonormalized and independent from other blocks. This is satisfied if $\boldsymbol{S}$ is generated by performing Gram–Schmidt operation on $\boldsymbol{A}$ block wisely, with the distribution of the subspace of $\boldsymbol{A}_i$ being a Haar measure on the Grassmann manifold.

This proposition reveals that such random matrix $\boldsymbol{S}$ has almost the same phase transition property as the Gaussian matrices. Take E in Proposition 3 as the descent conic of the $\ell_{2,1}$ norm $\Sigma_{i=0}^{d-1}\|\boldsymbol{z}_i\|$. Therefore, the phase transition curve of the problem is demonstrated by $n = \delta(E)$, which is the same as the phase transition curve demonstrated in (5) and (7).

3) *Deriving the Phase Transition Curve:* Combing Theorem 2 and Proposition 3, the phase transition curve of the blockwise recovery of the s_B block sparse signal $\boldsymbol{x} \in \mathbb{K}^{md}$ via $P'_{\ell_{2,1}}$ problem (2) is given by

$$\varphi^{RC}(r, s_B, d) := \begin{cases} \varphi_r(s_B, d), & \mathbb{K} = \mathbb{R} \\ \varphi_r^c(s_B, d), & \mathbb{K} = \mathbb{C} \end{cases} \tag{8}$$

where both $\boldsymbol{A}$ and $\boldsymbol{x}$ have block size m, and the rank of each block in $\boldsymbol{A}$ is r.

B. Proof of Theorem 2

The proof is divided into two main steps, converting (2) to a nonredundant counterpart, and further converting it to a $\ell_{2,1}$ norm minimization problem.

We first show that some linear transformation over $\boldsymbol{A}_i$ in $P'_{\ell_{2,1}}$ does not change its solutions and its performance bound. This will pave the road to rewriting the redundant problem (2) into its nonredundant counterpart.

For $\boldsymbol{A}_i$ and $\boldsymbol{A}$ defined in Theorem 2, we consider a linear operator $\boldsymbol{L}_i \in \mathbb{K}^{m\times t}$ over $\boldsymbol{A}_i$, $t \geq r$, the rank of $\boldsymbol{A}_i$, yielding $\boldsymbol{B}_i = \boldsymbol{A}_i\boldsymbol{L}_i \in \mathbb{K}^{n\times t}$. Assume that $\boldsymbol{L}_i$ does not change the rank, i.e., $\operatorname{rank}(\boldsymbol{B}_i) = \operatorname{rank}(\boldsymbol{A}_i)$. Then, the original $P'_{\ell_{2,1}}$ given in (2) is equivalent to the following problem:

$$\arg\min_{z} \Sigma_{i=0}^{d-1}\|\boldsymbol{B}_i\boldsymbol{z}_i\|, \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{B}\boldsymbol{z} \tag{9}$$

where $\boldsymbol{B} = [\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_{d-1}]$. Note that $\boldsymbol{B}_i$ can be either redundant or nonredundant. Since $\boldsymbol{B}_i$ shares the same column space of $\boldsymbol{A}_i$, there also exists a linear operator $\boldsymbol{R}_i$ such that $\boldsymbol{A}_i = \boldsymbol{B}_i\boldsymbol{R}_i$.

The mapping between the solutions of (2) and (9) is stated more precisely in the following lemma.

LEMMA 4 For a given vector $\boldsymbol{y}$, if $\boldsymbol{x}^*$ is one solution to (2), then $\boldsymbol{R}\boldsymbol{x}^*$ is a solution to (9). Symmetrically, $\boldsymbol{L}\boldsymbol{z}^*$ is a solution to (2) if $\boldsymbol{z}^*$ is a solution to (9). Here, $\boldsymbol{L} = \operatorname{diag}(\boldsymbol{L}_0, \boldsymbol{L}_1, \ldots, \boldsymbol{L}_{d-1})$ and $\boldsymbol{R} = \operatorname{diag}(\boldsymbol{R}_0, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_{d-1})$.

PROOF See Appendix A. ■

This indicates that (2) and (9) have the same optimal value. The optimal solution space of either problem can be embedded in the space of the other one with a linear operator $\boldsymbol{L}$ or $\boldsymbol{R}$.

The above relationship between solutions to these two problems leads to the consistency of their performance in signal reconstruction, as stated in the following corollary.

COROLLARY 5 The problem (2) can be blockwise recovered, if and only if (9) can be blockwise recovered.

PROOF See Appendix B. ■

Particularly, for the special case that $\boldsymbol{B}_i$ is nonredundant (with $t = r$), blockwise recovery of (9) is equivalent to elementwise recovery. Thus, with Corollary 5, we complete the first step of the proof, i.e., converting the blockwise recovery of the underdetermined problem $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ to the blockwise recovery (or equivalently elementwise recovery) of $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{z}$, where each block in $\boldsymbol{B}$ is of full rank and has the same column space as its corresponding block in $\boldsymbol{A}$. Then, we further convert the $P'_{\ell_{2,1}}$ problem to $P_{\ell_{2,1}}$ by selecting specific full-column-rank measurement matrices $\{\boldsymbol{B}_i\}$. Consider a measurement matrix $\boldsymbol{S} = [\boldsymbol{S}_0, \boldsymbol{S}_1, \ldots, \boldsymbol{S}_{d-1}] \in \mathbb{K}^{n\times rd}$, where $\boldsymbol{S}_i \in \mathbb{K}^{n\times r}$ satisfies $\boldsymbol{S}_i^{\mathrm{H}}\boldsymbol{S}_i = \boldsymbol{I}$. As a special case of $\boldsymbol{B}_i$, $\boldsymbol{S}_i$ is also nonredundant, and we obtain $\|\boldsymbol{S}_i\boldsymbol{z}_i\| = \|\boldsymbol{z}_i\|$. Therefore, in this case, the optimization problem (9), using $P'_{\ell_{2,1}}$ to solve $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{z}$, is equivalent to the following $P_{\ell_{2,1}}$ optimization form:

$$\hat{\boldsymbol{z}} = \arg\min_{z} \Sigma_{i=0}^{d-1}\|\boldsymbol{z}_i\|, \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{S}\boldsymbol{z}. \tag{10}$$

Given $\boldsymbol{A}$, we can always find a linear operator $\boldsymbol{L}$, so that $\boldsymbol{S}_i = \boldsymbol{A}_i\boldsymbol{L}_i$ is an orthonormal basis of the space spanned by the vectors in $\boldsymbol{A}_i$. In such cases, the nonredundant $P_{\ell_{2,1}}$ problem (10) is elementwise recovered, if and only if its $P'_{\ell_{2,1}}$ counterpart (9) is elementwise or blockwise recovered, if and only if the redundant problem (2) is blockwise recovered, as we assert in Corollary 5. Hence, we complete the proof of Theorem 2.

C. Proof of Proposition 3

The basic idea to prove Proposition 3 inherits from [25], which mathematically shows that the phase transition phenomenon exists in sub-Gaussian measurement matrices. Since in Proposition 3, we assume $\boldsymbol{S}_i^{\mathrm{H}}\boldsymbol{S}_i = \boldsymbol{I}$ introducing some dependency between the elements in $\boldsymbol{S}$, the original proof in [25], requiring independent measurement matrix, cannot be directly applied. Our goal of the proof is to tackle the dependency problem.

We start from a real numbered $\boldsymbol{S}$. According to [25, Prop. 3.8], given a nonempty, compact, and convex subset of $\mathbb{R}^{rd}$, E that does not contain the origin, as well as its projection on the sphere $\mathcal{S}^{rd-1}$ $\Omega = \boldsymbol{\theta}(E)$, and the linear map $\boldsymbol{S} \in \mathbb{R}^{n\times rd}$ defined in Proposition 3, the following two lemmas lead to Proposition 3.

LEMMA 6 For a small number ϵ, if $n \geq (1 + \epsilon)\delta(\Omega)$, one has $\min_{\boldsymbol{t}\in\Omega}\|\boldsymbol{S}\boldsymbol{t}\| > 0$ with high probability.

PROOF See Appendix C. ■

LEMMA 7 For a small number ϵ, if $n \leq (1-\epsilon)\delta(\Omega)$, one has $\min_{\|\boldsymbol{t}\|=1} \min_{\boldsymbol{s}\in cone(\Omega^\circ)} \|\boldsymbol{s}-\boldsymbol{S}^{\mathrm{H}}\boldsymbol{t}\| > 0$ with high probability, where $\Omega^\circ := \{\boldsymbol{x} \in \mathcal{S}^{rd-1} : \boldsymbol{x}\cdot\boldsymbol{t} \leq 0 \quad \forall \boldsymbol{t} \in \Omega\}$ is the polar of Ω and $cone(\Omega^\circ) := \{\boldsymbol{x} : \boldsymbol{x} = a\boldsymbol{t} \quad \forall \boldsymbol{t} \in \Omega^\circ \quad \forall a > 0\}$.

PROOF See Appendix D. ■

The relationship between the above lemmas and Proposition 3 is shown in [25, Prop. 3.8]. For completeness, we make some illustrations below. Lemma 6 is somehow trivial as $\|\boldsymbol{St}\| \geq 0$ indicates that $\boldsymbol{St} \neq 0$ for any $\boldsymbol{t}$. Therefore, $\boldsymbol{0} \notin \{\boldsymbol{y}|\boldsymbol{y} = \boldsymbol{Sx}, \boldsymbol{x} \in E\}$. Lemma 7 guarantees that cone(Ω°) does not intersect range$(\boldsymbol{S}^{\mathrm{H}})$. By duality, cone$(\Omega)$ intersects null$(\boldsymbol{S})$, i.e., there exists $\boldsymbol{x} \in E$ so that $\boldsymbol{Sx} = \boldsymbol{0}$. In conclusion, combining Lemmas 6 and 7, one can reach Proposition 3.

The proofs of Lemmas 6 and 7 are greatly inspired by the proofs of [25, Ths. 9.1 and 16.2], which are founded for the bounded random matrix model. We make efforts to show the same conclusions are true for matrices like $\boldsymbol{S}$ in Proposition 3.

IV. APPLICATION OF THEORETICAL RESULTS IN FAR

In this part, we apply the above theoretical results to the FAR model to obtain precise conditions that guarantee reconstruction. We first make a brief introduction about the FAR model, including both full-frequency and partial-frequency cases, in Section IV-A. Since the full-frequency cases have been sufficiently discussed in [20], we focus on the partial-frequency cases and present phase transition curves of FAR using block sparse recovery in Section IV-B. Then we compare these two cases in Section IV-C and our results indicate that FAR has better detection performance in the partial frequency case.

A. Signal Model of FAR

An FAR system changes the frequencies from pulse to pulse. We consider a radar transmitting N monotone pulses during a coherent processing interval (CPI). The frequency of the nth pulse, $n \in \mathcal{N} := \{0, 1, \ldots, N-1\}$, is represented by $f_n = f_c + C_n\Delta f$. Here, f_c denotes the initial frequency, Δf denotes the frequency step and C_n is the nth random modulation code. All modulation codes are selected from $\mathcal{M} = \{0, 1, \ldots, M-1\}$, where M represents the number of available frequencies and is less than N.

According to [20], the observing model of FAR can be rewritten with a matrix form as follows:

$$\boldsymbol{y} = \boldsymbol{\Theta x}. \tag{11}$$

Here, the nth entry of the measurement vector $\boldsymbol{y} \in \mathbb{C}^N$ is the received signal after down-conversion and sampling. The unknown vector $\boldsymbol{x} \in \mathbb{C}^{MN}$ contains the range-Doppler information of the targets. It has a block structure with the form $\boldsymbol{x} := [\boldsymbol{x}_0^{\mathrm{T}}, \boldsymbol{x}_1^{\mathrm{T}}, \ldots, \boldsymbol{x}_{N-1}^{\mathrm{T}}]$. Each block $\boldsymbol{x}_q \in \mathbb{C}^M$ represents the high range resolution (HRR) profile of the target with Doppler frequency q/N. The pth entry in $\boldsymbol{x}_q$ is nonzero when there exists a scatterer with Doppler frequency q/N and range frequency p/M. For the scenario that includes K targets, at most K blocks in $\boldsymbol{x}$ are nonzero, and thus $\boldsymbol{x}$ is a so-called K block sparse vector. The observation matrix $\boldsymbol{\Theta} \in \mathbb{C}^{N\times MN}$, in accordance with $\boldsymbol{x}$, is separated into N blocks as

$$\boldsymbol{\Theta} := \left[\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_{N-1}\right]. \tag{12}$$

Each block $\boldsymbol{\Theta}_q \in \mathbb{C}^{N\times M}$ has the (n, p)th entry given by

$$[\boldsymbol{\Theta}_q]_{n,p} = e^{j2\pi\frac{p}{M}C_n + j2\pi\frac{q}{N}\epsilon_n n}, \quad q, n \in \mathcal{N}, p \in \mathcal{M} \tag{13}$$

where $\epsilon_n := 1 + \frac{C_n\Delta f}{f_c}$.

Our goal is to recover the unknown vector $\boldsymbol{x}$ from (11). Recall that $\boldsymbol{x}$ is block-sparse, and the matrix $\boldsymbol{\Theta}$ has more columns than rows, hence recovering $\boldsymbol{x}$ from $\boldsymbol{y}$ is an underdetermined problem. Therefore, block sparse recovery methods can be utilized to recover $\boldsymbol{x}$, such that the high-resolution range and Doppler information of targets are obtained.

While previous literature mainly considered full-frequency FAR, this article focuses on the partial-frequency case, introduced below.

1) Full-frequency case: All the M frequencies are transmitted within a CPI.
2) Partial-frequency case: Only part of the M frequencies are transmitted within a CPI, while the rest frequencies are discarded.

The motivations to use the partial-frequency rather than full-frequency FAR are twofold.

1) Enhance the ECCM performance: In practice, some of the available frequencies are occupied by other radars/devices or even by hostile interference. To increase the signal-to-interference ratio, it is natural to avoid transmitting at those frequencies, such that the echoes would not be submerged by interference.
2) Improve the target detection performance: This will be later discussed in detail in Section IV-C. An intuitive explanation is that in FAR, increasing the number of used frequencies enlarges the number of unknown range parameters to estimate, while not increasing the number of observations. As a consequence, the partial-frequency mode achieves better target reconstruction performance and thus the detection performance than the full-frequency case.

To differentiate the partial-frequency case from the full counterpart, we use different denotations here. Let the new observation matrix be $\boldsymbol{\Theta}'$, with the qth block $\boldsymbol{\Theta}'_q$, whose (n, p)th entry is given by

$$[\boldsymbol{\Theta}'_q]_{n,p} = e^{j2\pi\frac{p}{M}C'_n + j2\pi\frac{q}{N}\epsilon_n n}, \quad q, n \in \mathcal{N}, p \in \mathcal{M} \tag{14}$$

where C'_n is the nth modulation code distributed on the new code set $\mathcal{R}$, $\mathcal{R} \subset \mathcal{M}$. Denote by $R = |\mathcal{R}| \leq M$ the number of used frequencies. The full-frequency FAR and traditional fixed frequency radar can be regarded as a special case of partial-frequency FAR, with $R = M$ and $R = 1$, respectively.

Inheriting the denotations, we rewrite [8, Lemma 2] as follows.

LEMMA 8 The mth eigenvalue of $\mathbf{\Theta}_q'^{\mathrm{T}}\mathbf{\Theta}_q'$ can be represented as

$$\lambda_m = \frac{M}{N}\sum_{n=0}^{N-1}\delta(C_n' - m) \tag{15}$$

where $\delta(\cdot)$ denotes the discrete delta function, being 1 at 0 and zeros otherwise.

This lemma leads to the following result regarding the distribution of $\mathrm{rank}(\mathbf{\Theta}_q')$.

LEMMA 9 When C_n' is evenly distributed on $\mathcal{R}$, $\mathrm{Rank}(\mathbf{\Theta}_q')$ equals R with probability more than $1 - \frac{\epsilon R^2}{(N - R\log R - R)^2}$, where ϵ is a small constant.

PROOF This is a direct conclusion from basic probability theory. ■

For a relatively large N, $\mathrm{rank}(\mathbf{\Theta}_q')$ equals R with a probability close to 1. Hence, for convenience for analysis, it is reasonable to assume that the rank is R in the rest of the article. Consequently, $\mathbf{\Theta}_q$ is nonredundant in full-frequency cases, while $\mathbf{\Theta}_q'$ is redundant in partial-frequency cases. Therefore, elementwise block sparse recovery $P_{\ell_{2,1}}$ is applied to the full-frequency FAR, and blockwise block sparse recovery $P'_{\ell_{2,1}}$ is preferred in the partial-frequency counterpart.

B. Phase Transition Curve in Partial-Frequency FAR

In this section, we apply our theoretical results about the phase transition to the FAR model. The measurement matrix is structured and can be viewed as a randomly selected Fourier matrix, such a structure greatly weakens its independence, and thus the mathematical tools we currently use are difficult to strictly prove the existence of phase transition properties in it (in fact, even the concentration property in the distribution of its singular value is difficult to prove). Strict proof is an open problem that needs much more effort. Yet, there are some works giving support to applying the theoretical phase transition curve to analyze the performance of $P_{\ell_{2,1}}$ and $P'_{\ell_{2,1}}$ in the FAR model. Donoho and Tanner [21] showed that the phase transition phenomenon exists in the randomly selected Fourier matrix by simulations. Wang et al. [8] showed that the measurement matrix of FAR behaved in certain isotropy, as each block was unified and sufficiently orthogonal to other blocks. This property is likely to $\boldsymbol{S}$ in Proposition 3. Therefore, it is reasonable to use the theoretical phase transition curve in the FAR model. We state our results in the rest part of this section, and these results are validated by simulations in Section V.

Only N_* out of all the N echoes will guarantee exact blockwise reconstruction, where N_* is given by

$$N_* := \inf_{\tau\geq 0}\left\{KR + \frac{K\tau^2}{2} + \frac{N-K}{2}\int_{\tau}^{\infty}(u-\tau)^2\phi_{2R}(u)\mathrm{d}u\right\}. \tag{16}$$

This is obtained by substituting $d = N$, $s_B = K$ and $r = R$ into (8).

The tightness of (16) is validated by simulations in Section V, which makes (16) useful in radar waveform design. For a given K, which implies that some prior knowledge of potential targets is acquired, one can determine the number of frequencies used for detection, R, when other parameters are given. In comparison with the traditional full-frequency FAR, the partial-frequency mode saves the spectrum resources, while guaranteeing performance in detection. The conserved frequencies are valuable for other usages, e.g., communications and interference rejection.

To alleviate the computational burden, we further approximate and simplify (16) according to [20] as follows.

PROPOSITION 10 For large N and K, if $\frac{N}{K} \gg \sqrt{R}$, namely, the target scene is sufficiently sparse, (16) can be approximated by

$$N_* \approx 2RK - \frac{K}{4} + \frac{\sqrt{2}K}{2}\sqrt{(4R-1)\log\frac{N-K}{K\sqrt{4R-1}}}. \tag{17}$$

The approximation is represented by elementary functions and has no inferior or integral operations required in (16). This intuitively indicates the contribution of each parameter to the required number of radar echoes and also facilitates the performance comparison between the full and partial-frequency FARs.

C. Discussions

Based on Proposition 10, we analyze the relationship between the recovery performance and the used frequencies, which shows that the use of fewer frequencies improves the performance in terms of blockwise recovery, at some cost of range resolution.

First, under a fixed K and with some simple calculation, we conclude from Proposition 10 that a smaller R leads to a smaller N_*, which means fewer echoes (out of N echoes) are required to guarantee the exact blockwise recovery of K targets. Conversely, when we fixed the number of available echoes, more targets can be blockwise recovered by reducing R. Therefore, fewer frequencies lead to better performance. The calculation to reveal the relationship between R and N_* is given below. Use α and β to denote $4R-1$ and $\frac{N-K}{K}$, respectively, where $\alpha > 0$ and $\beta \gg \sqrt{\alpha}$ according to the premise of Proposition 10. This simplifies the term in the right-hand side of (17) as

$$(4R-1)\log\frac{N-K}{K\sqrt{4R-1}} = \alpha\log\beta - \frac{1}{2}\alpha\log\alpha. \tag{18}$$

To reveal the monotonicity of N_* with respect to R, we calculate the following deviation:

$$\frac{\partial\left(\alpha\log\beta - \frac{1}{2}\alpha\log\alpha\right)}{\partial\alpha} = \log\frac{\beta}{\sqrt{\alpha}} - 1 > 0 \tag{19}$$

implying that this term monotonically increases with α, and thus increases with R. Consequently, we see that N_* increases with R.

As a direct consequence of the above conclusion, the partial-frequency mode outperforms the full-frequency counterpart in terms of blockwise recovery, as will be validated by simulation results in the next section. This phenomenon can be interpreted from the perspective of the number of unknowns to estimate. From Theorem 2, the target reconstruction is equivalent to solving $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{C}^{RN}$ contains RN unknowns. A larger R means that there are more parameters to solve in $\boldsymbol{x}$ (and that S becomes fatter), increasing the difficulty of target recovery.

In contrast to the full-frequency FAR, the price of partial-frequency mode and the corresponding blockwise recovery method is the degradation of range resolution. This can be intuitively explained by the fact that more frequencies indicate more detailed range information. From the perspective of mathematics, recall that the qth block $\boldsymbol{x}_q$ represents the HRR profile of the target with the normalized Doppler frequency q/N. The blockwise recovery disregards the details within the block $\boldsymbol{x}_q$, but instead treats the block as a whole and focuses on the reconstruction of the Doppler information. In the case when the intact HRR profiles are demanded, elementwise recovery of $\boldsymbol{x}$ is required, and thus the full-frequency mode is preferred.

V. SIMULATIONS

Our simulations contain three parts. In the first part, we mainly validate the correctness of Theorem 2, where the measurement matrices are generated from Gaussian. In the second part, we consider the FAR model and verify the accuracy of the phase transition curve (8) for the partial-frequency cases. In the last part, we show that our theoretical results are robust to small noise.

A. Validating Theorems 2 and 3 With Gaussian Matrices

We inherit the notation in Section III and consider three models

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \ell(\boldsymbol{A}, \boldsymbol{x}), \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{20}$$

$$\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \ell(\boldsymbol{B}, \boldsymbol{z}), \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{B}\boldsymbol{z} \tag{21}$$

$$\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \ell(\boldsymbol{S}, \boldsymbol{z}), \ \text{s.t.}\ \boldsymbol{y} = \boldsymbol{S}\boldsymbol{z} \tag{22}$$

where $\ell(\cdot)$ represents the objective function of $\ell_{2,1}$ or $\ell'_{2,1}$ norm, assigned later. Here, the entries in $\boldsymbol{S} \in \mathbb{R}^{n\times rd}$ are i.i.d. as standard Gaussian distribution, and thus, $\boldsymbol{S}$ is nonredundant. The redundant observation matrix $\boldsymbol{A} \in \mathbb{R}^{n\times md}$ with $m > r$ is generated from $\boldsymbol{S}$, as $\boldsymbol{A}_i = \boldsymbol{S}_i\boldsymbol{T}_i$. Here, we generate Gaussian matrices and perform normalization on their columns to obtain $\boldsymbol{T}_i \in \mathbb{R}^{r\times m}$. We set the linear operator defined in Section III $\boldsymbol{L}_i = [\boldsymbol{I}_r, \boldsymbol{0}]^{\mathrm{T}}$ for $i = 0, 1, \ldots, d-1$. The nonredundant measurement matrix $\boldsymbol{B} \in \mathbb{R}^{n\times rd}$ is given by $\boldsymbol{B}_i = \boldsymbol{A}_i\boldsymbol{L}_i$.

We focus on exact blockwise recovery: the recovery is asserted as a success if the estimation $\hat{\boldsymbol{k}} \in \mathbb{R}^{md}$ satisfies $\Sigma_{i=0}^{d-1}\|\boldsymbol{A}_i\boldsymbol{x}_i - \boldsymbol{K}_i\hat{\boldsymbol{k}}_i\|_2 \le 10^{-4}$, where $(\boldsymbol{K}, \hat{\boldsymbol{k}})$ represents $(\boldsymbol{A}, \hat{\boldsymbol{x}})$, $(\boldsymbol{B}, \hat{\boldsymbol{z}})$ or $(\boldsymbol{S}, \hat{\boldsymbol{z}})$. In the simulations, $\boldsymbol{y}$ is given by $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, and the nonzero entries in $\boldsymbol{x} \in \mathbb{R}^{md}$ also obey i.i.d. standard Gaussian distribution. We set the parameters $m = 4$, $r = 2$, and $d = 32$.

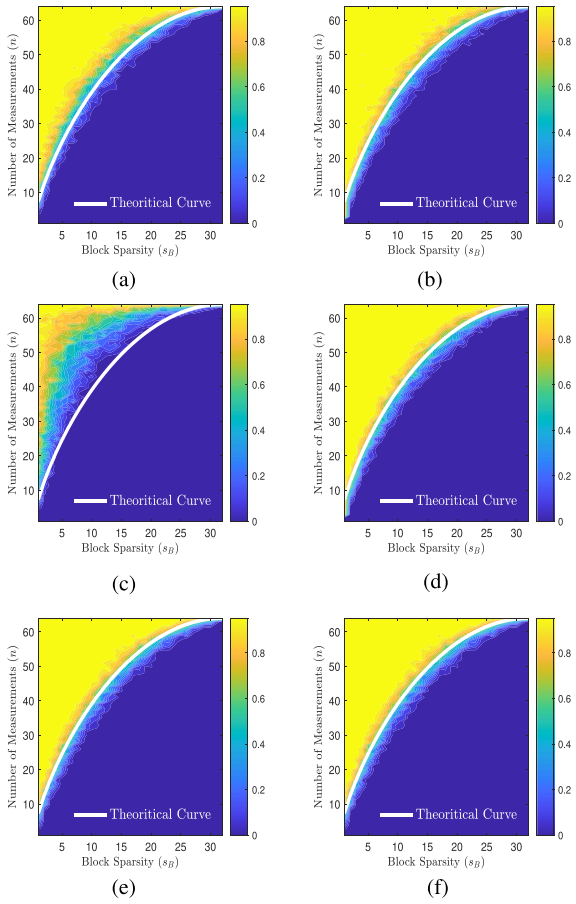


Fig. 3. Probabilities of exact recovery (represented by different colors) with Gaussian-based measurement matrices. (a) $\ell_{2,1}, \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. (b) $\ell'_{2,1}$, $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. (c) $\ell_{2,1}, \boldsymbol{y} = \boldsymbol{B}\boldsymbol{z}$. (d) $\ell'_{2,1}, \boldsymbol{y} = \boldsymbol{B}\boldsymbol{z}$. (e) $\ell_{2,1}, \boldsymbol{y} = \boldsymbol{S}\boldsymbol{z}$. (f) $\ell'_{2,1}, \boldsymbol{y} = \boldsymbol{S}\boldsymbol{z}$.

With the above parameter setting, we then vary s_B and n, i.e., the block sparsity of $\boldsymbol{x}$ and the number of measurements, respectively, to calculate the probabilities of exact recovery with different methods and different measurement matrices. The range of variables spans s_B from 1 to 32 in increments of 1, while the range of variable n extends from 4 to 128 in increments of 4.

The probabilities are calculated from 50 Monte-Carlo trials on each pair (s_B, n). The results are shown in Fig 3, where the theoretical curves are also presented for comparison, obtained by (5) and (8) with corresponding m, d, and r, respectively. Particularly, these curves are identical for they have the same parameters.

First, we observe Fig 3(b)–(f), which numerically validates the proof of Theorem 2 step by step. Different methods and models yield the same phase transitions. which also match with the theoretical curves from Proposition 1 and (8). The consistency of (b) and (d), as well as (b) and (e), validates the declarations in Lemma 4 and Corollary 5 that linear transformation in measurement matrices does not affect the solvability of $P'_{\ell_{2,1}}$ as long as the subspace of each block is not changed. The same phase transitions in (f) and

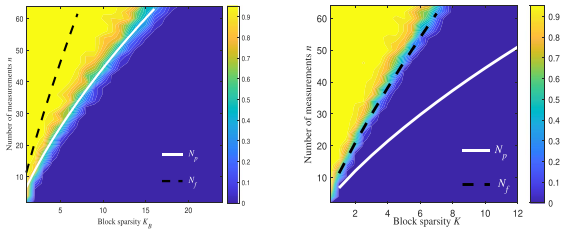


Fig. 4. Phase transitions with FAR models. (a) $R = 2$. (b) $R = 4$.

(e) verify that results of (22) with $\ell(\cdot)$ being $\ell_{2,1}$ and $\ell'_{2,1}$ are the same, as explained in Section III-B.

Then we observe Fig. 3(a) and (c), which comes from (20) and (21) with $\ell(\cdot)$ being $\ell_{2,1}$, respectively. It can be seen that they perform worse than their counterparts with $\ell'_{2,1}$, as they have less area of exact recovery because they demand more strict parameter conditions. Further, (c) behaves worse than (a) since $\boldsymbol{B}_i$ has fewer columns than $\boldsymbol{A}_i$. To summarize, the simulation results validate our inference on the phase transitions of different block sparse recovery methods and encourage us to use (8) to estimate the performance bound of the redundant block sparse recovery.

B. Simulations With FAR Models

We next verify phase transitions in FAR. Both full-frequency and par-frequency cases are tested.

To inspect the phase transition of the FAR models (13) and (14), we randomly select n rows from $\boldsymbol{\Theta}' \in \mathbb{C}^{N\times MN}$ to form a partial measurement matrix $\hat{\boldsymbol{\Theta}}' \in \mathbb{C}^{n\times MN}$. The phases of nonzero entries in $\boldsymbol{x}$ are i.i.d. $U([0, 2\pi])$ and the amplitudes equal 1. We consider two cases: $R = M$ and $R = M/2$, representing the full- and partial-frequency cases, respectively. We set the parameters $M = 4$, $N = 64$, $\frac{\Delta f}{f_c} = 0.02$, and use 50 Monte-Carlo trials to calculate the success rates of exact blockwise recovery. The results of $R = M/2$ and $R = M$ are shown in Fig 4(a) and (b), respectively. The theoretical phase transition curves of full- and partial-frequency cases are calculated with corresponding M, N, and R by (16), denoted by "N_f" and "N_p," respectively. For the sake of comparison between these two cases, we depict both theoretical curves in each figure of phase transition results.

From Fig. 4, it is evident that when $n > N_p$ in (a) or $n > N_f$ in (b), the recovery succeeds with high probabilities. Otherwise, the recovery fails with high probabilities. This observation indicates that both theoretical curves N_f and N_p well match their respective phase transitions, thereby affirming the accuracy of (14) in both full-frequency and partial-frequency FAR modes. Additionally, the transition band from 0 to 1 for success rate, which is depicted in green on the graphs, is relatively narrow, validating the tightness of (16). We also observe that the black curve N_f is always beyond the white curve N_p (when K is fixed), implying that the partial-frequency mode outperforms the full-frequency counterpart in the aspect of blockwise recovery.

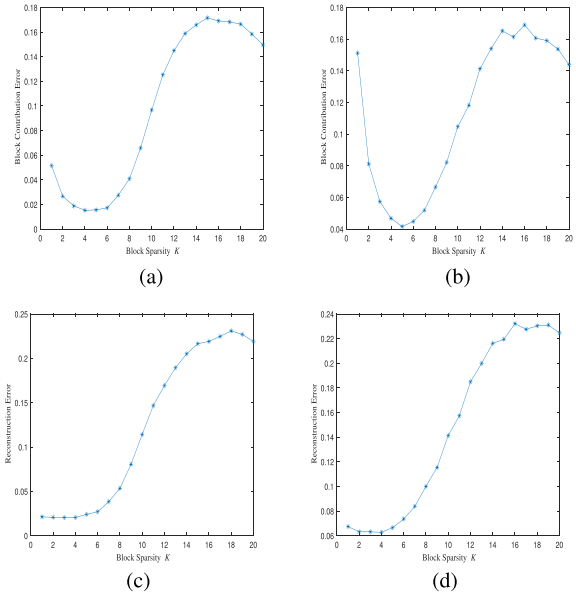


Fig. 5. Reconstruction error and block contribution error of P'_{ℓ_δ} with different noises. (a) Block contribution error, $\sigma^2 = 0.01$. (b) Block contribution error, $\sigma^2 = 0.1$. (c) Reconstruction error, $\sigma^2 = 0.01$. (d) Reconstruction error, $\sigma^2 = 0.1$.

C. Simulations With Noise

Generally, our conclusions about phase transitions are derived from noiseless assumptions. Yet, it can be employed to characterize the variation in the error between the estimated and true values under noisy conditions, which is a significant aspect when considering the practical utility. In this section, we aim at showing by simulations that the errors also have a phase transition phenomenon and the phase transition points can be approximately predicted by our theoretical conclusions.

Given the observation, $\boldsymbol{y} = \boldsymbol{\Theta}'\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{v}$ is the additive white Gaussian noise whose real part and complex part obey $\mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, we consider the following optimization from [16]:

$$P'_{\ell_\delta} : \hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \sum_{i=0}^{d-1} \|\boldsymbol{\Theta}'_i \boldsymbol{x}_i\|_2, \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{\Theta}'\boldsymbol{x}\|_2 \le \delta. \quad (23)$$

In the simulation, we set the parameters $N = 32$, $M = 4$, $R = 2$, and $\frac{\delta f}{f_c} = 0.02$. The tolerance in (23) is set to be $\delta = \sqrt{N}\sigma$. We vary the block sparsity K of $\boldsymbol{x}$ and apply both P'_{ℓ_δ} and P_{ℓ_δ} to obtain the estimations $\hat{\boldsymbol{x}}$ under different noise conditions. For each estimation $\hat{\boldsymbol{x}}$, we examine its "reconstruction error" and "block contribution error" [17], which are given by $\frac{\|\boldsymbol{y} - \sum_{i\in\Lambda_{\boldsymbol{x}}} \boldsymbol{\Theta}'_i \hat{\boldsymbol{x}}_i\|_2}{\|\boldsymbol{y}\|_2}$, $1 - (\sum_{i\in\Lambda_{\boldsymbol{x}}} \|\boldsymbol{\Theta}'_i \hat{\boldsymbol{x}}_i\|_2)/(\sum_{i=0}^{d-1} \|\boldsymbol{\Theta}'_i \hat{\boldsymbol{x}}_i\|_2)$, respectively. Here, $\Lambda_{\boldsymbol{x}}$ represents the set composed by the index of the nonzero blocks in $\boldsymbol{x}$.

Substitute $N* = N = 32$, $M = 4$, and $R = 2$ into (16). We then obtain a solution $K = 8.4$. From the results shown in Fig. 5, we observe that under both noise conditions, the errors start to increase rapidly when $K >= 9$, indicating the reconstructions are no longer closed to the true values. In

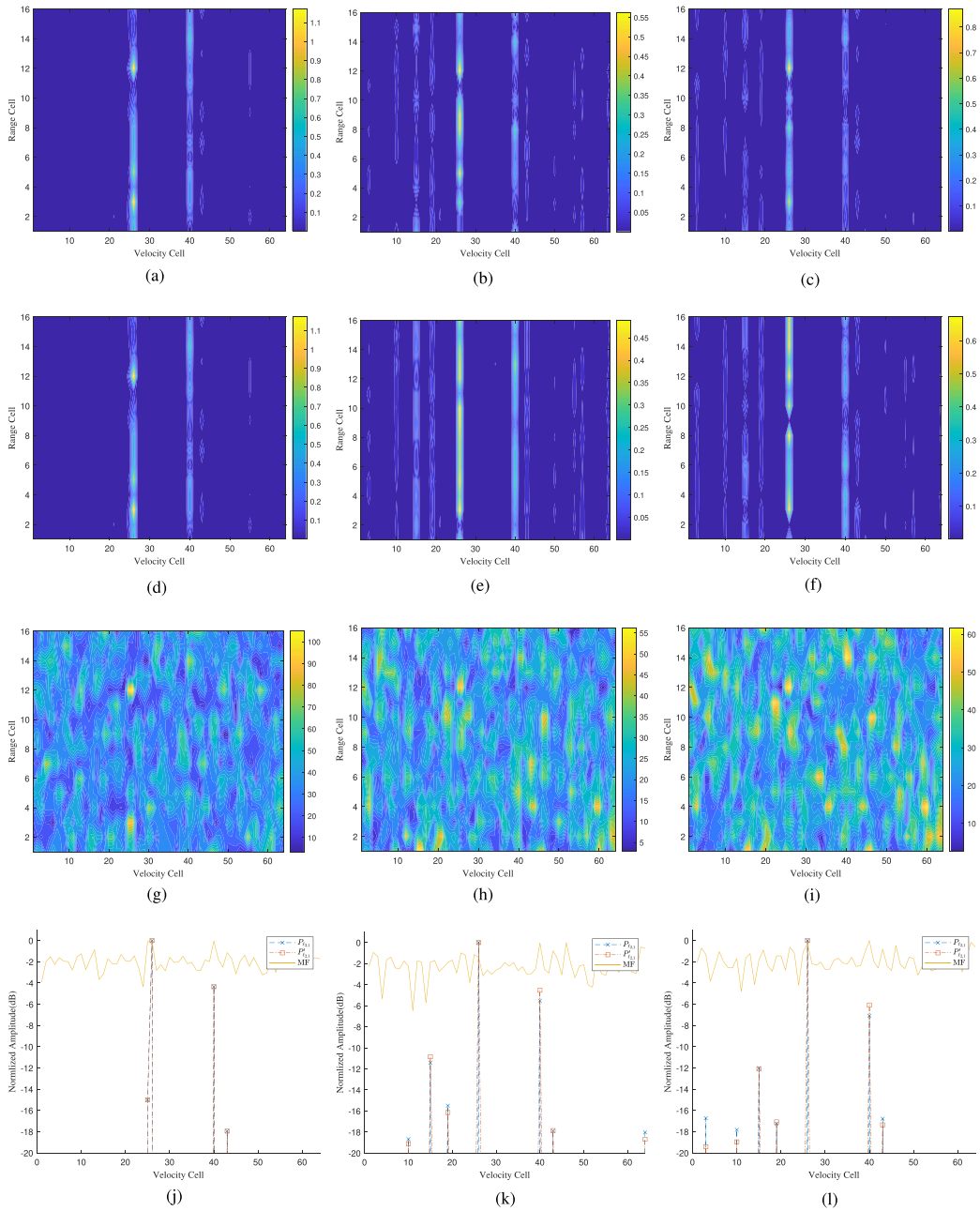


Fig. 6. Reconstruction of $P_{\ell_{2,1}}$, $P'_{\ell_{2,1}}$ and the matched filter in three setups. (a) Results of $P_{\ell_{2,1}}$ in case (1). (b) Results of $P_{\ell_{2,1}}$ in case (2). (c) Results of $P_{\ell_{2,1}}$ in case (3). (d) Results of $P'_{\ell_{2,1}}$ in case (1). (e) Results of $P'_{\ell_{2,1}}$ in case (2). (f) Results of $P'_{\ell_{2,1}}$ in case (3). (g) Results of the matched filter in case (1). (h) Results of the matched filter in case (2). (i) Results of the matched filter in case (3). (j) Normalized Amplitude of Velocity Bins in case (1). (k) Normalized Amplitude of Velocity Bins in case (2). (l) Normalized Amplitude of Velocity Bins in case (3).

Fig. 5(a), (c), and (d), we see that when $K \leq 8$, both the block contribution errors and reconstruction errors remain at a relatively low level. These observations are in accordance with our theoretical analysis that $K = 8.4$ is a phase transition point. In Fig. 5(b), the block contribution error is relatively high when $K = 1, 2$. This observation is due to the definition of block contribution error. The denominator $(\sum_{i=0}^{d-1} \|\boldsymbol{\Theta}'_i \hat{\boldsymbol{x}}_i\|_2)$ including the energy of noise and targets. Since we fix the noise variance in our simulation, $(\sum_{i\in\Lambda_x} \|\boldsymbol{\Theta}'_i \hat{\boldsymbol{x}}_i\|_2)/(\sum_{i=0}^{d-1} \|\boldsymbol{\Theta}'_i \hat{\boldsymbol{x}}_i\|_2)$ can be small when the targets are extremely sparse, consequently amplifying the block contribution error. As a consequence, Fig. 5 validates that the theoretical phase transition curve is applicable and reliable in the presence of noise, which highlights its practical values.

D. Field Experiment Results

In this section, real measured data are utilized to depict the performance difference between $P_{\ell_{2,1}}$ and $P'_{\ell_{2,1}}$. The radar transmits $N = 64$ pulses across $M = 16$ frequency points,

with each frequency point corresponding to four pulses. Two aircraft are flying away from the radar with relative velocities of about 35 and 45 m/s. We take a certain coarse range resolution (CRR) bin as an example and recover the range-Doppler map of the target. Particularly, to simulate full-frequency and partial-frequency scenarios, we considered the following three experimental setups.

1) All the $N = 64$ pulses are used.
2) We randomly select eight frequency points and eliminate all the corresponding pulses. Only 32 pulses across the other eight frequency points are used.
3) We randomly select four frequency points and eliminate all corresponding pulses. We randomly select four other frequency points and eliminate two pulses with each frequency point. There are 40 pulses across 12 frequency points are used.

In Fig. 6(a)–(g), we show the reconstruction results with $P_{\ell_{2,1}}$, $P'_{\ell_{2,1}}$ and the matched filter under the above three settings. The targets are located in the 26th and the 40th velocity cell, aligning with the speeds of 35 and 45 m/s. Denote the recovery result in each subfigure by $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_0, \hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_{N-1}] \in \mathbb{C}^{M\times N}$, where $\hat{\boldsymbol{x}}_i$ is the recovered high-range resolution profile in the ith velocity cell.

We also calculate its normalized amplitude in the ith velocity cell by $\|\hat{\boldsymbol{x}}_i\|_2 / \max_i \|\hat{\boldsymbol{x}}_i\|_2$. The results are shown in Fig. 6(j)–(l). We observe that both $P_{\ell_{2,1}}$ and $P'_{\ell_{2,1}}$ have better reconstructions than the traditional matched filter. For case (1), $P_{\ell_{2,1}}$ and $P'_{\ell_{2,1}}$ have the same reconstruction results, which is consistent with our analysis in full-frequency cases. For cases (2) and (3), where only a part of the frequency points are transmitted, compared to that of $P'_{\ell_{2,1}}$, the recovery results of $P_{\ell_{2,1}}$ have higher energy (sidelobes) on the nontarget velocity grids. The observation validates that $P'_{\ell_{2,1}}$ outperforms $P_{\ell_{2,1}}$ in partial-frequency cases."

VI. CONCLUSION

In this article, block sparse recovery with deficient measurement matrices is studied. We show that the $\ell'_{2,1}$ norm minimization problem is equivalent to the $\ell_{2,1}$ counterpart after certain linear reformulation. Based on this result, we strictly prove the existence of phase transition and derive the phase transition curve of block sparse recovery in the case that the measurement matrices have redundant blocks, which extends [20]. The results also heuristically indicate the superiority of $P'_{\ell_{2,1}}$ over $P_{\ell_{2,1}}$. Particularly, the theoretical phase transition curve is applied to the target detection of FAR, which is powerful for guiding waveform design and suggests that partial-frequency cases outperform full-frequency ones in target detection. Numerical results validate the correctness of the derivation and the accuracy of the derived curve.

APPENDIX A
PROOF OF LEMMA 4

In this part, we prove Lemma 4. The proof is mainly organized as follows. First, we find that the optimal solutions of (2) can be converted to a vector satisfying the constraints of (9). Next, oppositely, the solutions of (9) can be converted to a vector meeting the constraints of (2). Then, we show that these original solutions and converted vectors are the optimizers of their corresponding objective functions.

We introduce the following notations. Let $\boldsymbol{x}^*$ and $\boldsymbol{z}^*$ represent the optimal solutions of (2) and (9), respectively, where both vectors have a block structure given by $\boldsymbol{x}^* = [\boldsymbol{x}_0^{*\mathrm{T}}, \boldsymbol{x}_1^{*\mathrm{T}}, \ldots, \boldsymbol{x}_{d-1}^{*\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{z}^* = [\boldsymbol{z}_0^{*\mathrm{T}}, \boldsymbol{z}_1^{*\mathrm{T}}, \ldots, \boldsymbol{z}_{d-1}^{*\mathrm{T}}]^{\mathrm{T}}$. Let $\bar{\boldsymbol{z}}_i = \boldsymbol{R}_i \boldsymbol{x}_i^*$, and concatenating $\bar{\boldsymbol{z}}_i$ yields $\bar{\boldsymbol{z}} = [\bar{\boldsymbol{z}}_0^{\mathrm{T}}, \bar{\boldsymbol{z}}_1^{\mathrm{T}}, \ldots, \bar{\boldsymbol{z}}_{d-1}^{\mathrm{T}}]^{\mathrm{T}}$. Similarly, let $\bar{\boldsymbol{x}}_i = \boldsymbol{L}_i \boldsymbol{z}_i^*$.

Recall that $\boldsymbol{A}_i = \boldsymbol{B}_i \boldsymbol{R}_i$. Thus, we have $\boldsymbol{y}_i = \boldsymbol{A}_i \boldsymbol{x}_i^* = \boldsymbol{B}_i \bar{\boldsymbol{z}}_i$, and $\bar{\boldsymbol{z}} = \boldsymbol{R}\boldsymbol{x}^*$, satisfying $\boldsymbol{y} = \boldsymbol{B}\bar{\boldsymbol{z}}$, the constraint of (9). Since $\boldsymbol{z}^*$ minimizes the objective function, we have

$$\sum \|\boldsymbol{A}_i \boldsymbol{x}_i^*\| = \sum \|\boldsymbol{B}_i \bar{\boldsymbol{z}}_i\| \geq \sum \|\boldsymbol{B}_i \boldsymbol{z}_i^*\|. \tag{24}$$

Through the same procedure above, in consideration of $\boldsymbol{B}_i = \boldsymbol{A}_i \boldsymbol{L}_i$, we obtain

$$\sum \|\boldsymbol{B}_i \boldsymbol{z}_i^*\| = \sum \|\boldsymbol{A}_i \bar{\boldsymbol{x}}_i\| \geq \sum \|\boldsymbol{A}_i \boldsymbol{x}_i^*\|. \tag{25}$$

Combining (24) and (25), we find

$$\sum \|\boldsymbol{A}_i \bar{\boldsymbol{x}}_i^*\| = \sum \|\boldsymbol{B}_i \bar{\boldsymbol{z}}_i\| = \sum \|\boldsymbol{B}_i \bar{\boldsymbol{z}}_i^*\| = \sum \|\boldsymbol{A}_i \bar{\boldsymbol{x}}_i\|. \tag{26}$$

Therefore, $\bar{\boldsymbol{z}} = \boldsymbol{R}\boldsymbol{x}^*$ is also a solution to (9).

APPENDIX B
PROOF OF COROLLARY 5

In this section, we prove Corollary 5. Since (2) and (9) is symmetrical, we just prove the "only if" part and the "if" part can be derived with the same procedure.

To prove the "only if" part, we inherit the denotations in Section III-B, and use $\boldsymbol{x}^{\#}$ and $\boldsymbol{z}^{\#} = \boldsymbol{R}\boldsymbol{x}^{\#}$ to denote the true values. If $\boldsymbol{x}^{\#}$ can be blockwise recovered via (2), i.e., all the solutions, $\boldsymbol{x}^*$, satisfy $\sum_i \|\boldsymbol{A}_i \boldsymbol{x}_i^* - \boldsymbol{A}_i \boldsymbol{x}_i^{\#}\| = 0$, then $\boldsymbol{z}^*$, the solution to (9) satisfies the following:

$$\begin{aligned}\sum_i \|\boldsymbol{B}_i \boldsymbol{z}_i^* - \boldsymbol{B}_i \boldsymbol{z}_i^{\#}\| &= \sum_i \|\boldsymbol{A}_i \boldsymbol{R}_i \boldsymbol{z}_i^* - \boldsymbol{A}_i \boldsymbol{R}_i \boldsymbol{z}_i^{\#}\| \\ &= \sum_i \|\boldsymbol{A}_i \boldsymbol{x}_i^* - \boldsymbol{A}_i \boldsymbol{x}_i^{\#}\| = 0.\end{aligned} \tag{27}$$

The second equation is from Lemma 4. It implies the correctness of the "only if" part and completes the proof.

APPENDIX C
PROOF OF LEMMA 6

We prove Lemma 6 by proving the following inequality:

$$\begin{aligned}&\mathbb{P}\left\{\left(\min_{t\in T} \|\boldsymbol{S}\boldsymbol{t}\|\right)^2 \leq \frac{1}{n}\left((\mathcal{E}(T))_+^2 - C_0(n+rd)^{1-\kappa}\right) - C_1\xi\right\} \\ &\leq \max(e^{-\xi^2 n}, e^{-\xi n})\end{aligned} \tag{28}$$

where T is a closed subset of the unit ball $\mathcal{S}^{rd}$.

Here, the functions are defined as $(x)_+ := \max(x, 0)$ and $\mathcal{E}(\Omega) := \mathbb{E}\inf_{\boldsymbol{x}\in\Omega}(\sqrt{n}\|\boldsymbol{x}\| + \boldsymbol{g}\cdot\boldsymbol{x})$ is called excess width. The small constant κ satisfies $\kappa > 0$. By taking $T = \Omega$

in (28), $(\mathcal{E}(T))_+^2$ are proved to be $O(\epsilon(n+rd))$ [25]. So $(\min_{t\in T}\|\boldsymbol{St}\|)^2$ is greater than 0 for a high probability by taking appropriate n, ξ, and ϵ.

Defining $\sigma_{\min}^2(\boldsymbol{S};T) := (\min_{t\in T}\|\boldsymbol{St}\|)^2$, we prove (28) with two steps. First, we show that $\sigma_{\min}^2(\boldsymbol{S};T)$ is concentrated on its mean, and second, we bound its mean. Naturally, we separate the original inequality into two parts.

1) For any $\xi \geq 0$

$$\mathbb{P}\left\{\sigma_{\min}^2(\boldsymbol{S};T) \leq \mathbb{E}\sigma_{\min}^2(\boldsymbol{S};T) - C\xi\right\} \leq C_0 e^{-\xi^2 n} + C_1 e^{-\xi n}. \tag{29}$$

2) There exists some $\kappa \geq 0$ such that

$$(\mathcal{E}(T))_+^2 - (n+rd)^{1-\kappa} \leq n\mathbb{E}\sigma_{\min}^2(\boldsymbol{S};T). \tag{30}$$

In [25], the proof of (29) with a sub-Gaussian matrix is derived by Massart's modified logarithmic Sobolev inequality [32], which is too loose in our case. Some other mathematical tools need to be adopted. The proof of (30) is almost done by [25]. Our derivations are highly based on them and differ only in certain details. So we only repeat the main procedures and remark the difference in the proof of (30). Please refer to [25] for more details.

The rest of this appendix is arranged as follows. In Appendix A, we underline some important facts about $\boldsymbol{S}$. In Appendix B, we use the Levy–Gromov theorem from [33] to show the concentration property (29). In Appendix C, we follow the procedures in [25] to prove (30) and show how the proof for sub-Gaussian models can be used in the proof of structured and normalized $\boldsymbol{S}$.

A. Preliminary Facts

Before we start dealing with the above two inequalities, we introduce two facts about the matrix $\boldsymbol{S}$, which are useful for the subsequent proof.

FACT 1 With probability greater than $1-2\exp(-Ct^2 n)$, for a constant c, the singular values λ of $\boldsymbol{S}$ are bounded by

$$\sqrt{\frac{rd}{n}} - c - t \leq \lambda_{\min} \leq \lambda_{\max} \leq \sqrt{\frac{rd}{n}} + c + t. \tag{31}$$

Fixed any $\boldsymbol{x} \in \mathbb{S}^{n-1}$, we obtain $\|\boldsymbol{S}^{\mathrm{T}}\boldsymbol{x}\|_2^2 = \sum_{i=0}^{N-1}\|\boldsymbol{S}_i^{\mathrm{T}}\boldsymbol{x}\|_2^2 = \sum_{i=0}^{N-1} Z_i^2$. Here, Z_i is an i.i.d sub-Gaussian random variable satisfying $\mathbb{E}Z_i^2 = \frac{r}{n}$. Then, following [34, Th. 5.39], one yields Fact 1. Such distribution of the singular values exists in many other random matrices.

FACT 2 For each element s from the matrix $\boldsymbol{S}$, s satisfies $\mathbb{E}(s^4) = O(\frac{1}{n^2})$ and $\mathbb{E}(s^6) = O(\frac{1}{n^3})$.

PROOF

$$\begin{aligned}\mathbb{E}(s^4) &= \int_{\pi/2}^{-\pi/2} \cos^{n-2}\theta \sin^4\theta d\theta \Big/ \int_{\pi/2}^{-\pi/2} \cos^{n-2}\theta d\theta \\ &= \frac{1}{I_{n-2}} \int_{\pi/2}^{-\pi/2} \cos^{n-2}\theta(1-\cos^2\theta)^2 d\theta \\ &= \frac{I_{n-2} - 2I_n + I_{n+2}}{I_{n-2}} = \frac{3}{n(n+2)}.\end{aligned} \tag{32}$$

Algorithm 1: Map $\boldsymbol{g}_i$ to $\boldsymbol{S}_i$ via Coordinate Rotating.

Require: $\boldsymbol{g}_{i,j} \in \mathbb{R}^{n-j}$, $j = 0, 1, \ldots, r-1$; $\boldsymbol{e}_j$, $j = 0, 1, \ldots, n-1$.
Output: $\boldsymbol{S}_i \in \mathbb{R}^{n\times r}$.

```
set k = 0
while k < r − 1 do
  s_k = [e_k, e_{k+1}, ..., e_{n−1}] · g_{i,k};
  if s_k == e_k then
    continue;
  else if s_k == −e_k then
    e_{k+1} = -e_{k+1};
  else
    U = [e_k, (s_k − <e_k, s_k> e_k) / ||s_k − <e_k, s_k> e_k||_2];
    θ = ∠(U^T s_k);
    R = I − U(U^T U)^{-1} U^T + U[cos θ, − sin θ; − sin θ, cos θ]U^T;
    [e_{k+1}, ..., e_{n−1}] = R · [e_{k+1}, ..., e_{n−1}]
  end if
  k = k + 1;
end while
S_i = [s_0, s_1, ..., s_{r−1}]
```

$$\begin{aligned}\mathbb{E}(s^6) &= \int_{\pi/2}^{-\pi/2} \cos^{n-2}\theta \sin^6\theta d\theta \Big/ \int_{\pi/2}^{-\pi/2} \cos^{n-2}\theta d\theta \\ &= \frac{1}{I_{n-2}} \int_{\pi/2}^{-\pi/2} \cos^{n-2}\theta(1-\cos^2\theta)^3 d\theta \\ &= \frac{I_{n-2} - 3I_n + 3I_{n+2} - I_{n+4}}{I_{n-2}} = \frac{15}{n(n+2)(n+4)}.\end{aligned} \tag{33}$$

■

B Proof of (29)

To prove the (29), we first show $\sigma_{\min}(\boldsymbol{S};T)$ is concentrated on its mean, using the Lipschitz property and the concentration on the high-dimension sphere [33]. Then we combine Fact 1 to show $\sigma_{\min}^2(\boldsymbol{S};T)$ is concentrated. For ease of presentation, we abuse the letter C to denote certain constants in our following derivations.

For each block $\boldsymbol{S}_i$ in $\boldsymbol{S}$, the first column in $\boldsymbol{S}_i$ is a random vector that is uniformly distributed on the sphere $\mathcal{S}^{n-1}$. Due to the second column being orthogonal to the first one, it is uniformly distributed on a sphere $\mathcal{S}^{n-2}$. Thus, $\boldsymbol{S}_i$ can be demonstrated by a $(nr - \frac{r^2-r}{2})$-dimension vector $\boldsymbol{g}_i = [\boldsymbol{g}_{i,0}^{\mathrm{T}}, \ldots, \boldsymbol{g}_{i,r-1}^{\mathrm{T}}]^{\mathrm{T}}, \boldsymbol{g}_{i,k} \in \mathbb{R}^{n-k}$, which uniformly distributed on $\mathcal{S}^{n-1} \times \mathcal{S}^{n-2} \times \cdots \times \mathcal{S}^{n-r}$. Below, we show how $\boldsymbol{S}_i$ is determined by $\boldsymbol{g}_i$, i.e., we construct a map h from $\boldsymbol{g}_i$ to $\boldsymbol{S}_i$, with a given standard basis $\boldsymbol{e}_j$, $j \in 0, 1, \ldots, n-1$, whose jth entry is 1 and the others are 0. The central idea underlying the mapping is to rotate the entire coordinate axis, ensuring that the corresponding base $\boldsymbol{e}_k$ and target vector $\boldsymbol{s}_k$ coincide. This guarantees that the vector formed by the remaining bases $\boldsymbol{e}_j$, $j > k$ is orthogonal to the $\boldsymbol{s}_k$ while maintaining a "uniform distribution."

LEMMA 11 The map h from $\boldsymbol{g}_i$ to $\boldsymbol{S}_i$ is r-Lipschitz, i.e., for $\boldsymbol{S}_1 = h(\boldsymbol{g}_1)$ and $\boldsymbol{S}_2 = h(\boldsymbol{g}_2)$, it holds that $\|\boldsymbol{S}_1 - \boldsymbol{S}_2\|_F \leq r\|\boldsymbol{g}_1 - \boldsymbol{g}_2\|$.

PROOF First, a small change in $\boldsymbol{g}_{i,j}$ will not change $\boldsymbol{S}_{i,k}$ for $k < j$. Then we assume $\boldsymbol{g}_{i,j}$ rotates $\Delta\theta$, and thus the columns $\boldsymbol{S}_{i,k}$ rotates $\Delta\theta_k \leq \Delta\theta$. The equality holds only when the rotation of $\boldsymbol{g}_{i,j}$ is exactly on the plane spanned by $\boldsymbol{S}_{i,j}$ and $\boldsymbol{S}_{i,k}$. In this case, an infinitesimal disturbance Δd on $\boldsymbol{g}_{i,j}$ leads to a deviation less than $\sqrt{r-j+1}\Delta d$ on $\boldsymbol{S}$. Sum it over j, one get that the map from $\boldsymbol{g}_i$ to $\boldsymbol{S}_i$ is r-Lipschitz.■

Using Lemma 11, we find that the map from $\boldsymbol{g} = [\boldsymbol{g}_0^{\mathrm{T}}, \ldots, \boldsymbol{g}_{d-1}^{\mathrm{T}}]^{\mathrm{T}}$ to $\boldsymbol{S}_i$ is r-Lipschitz. The map $f(\boldsymbol{g}) = |\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)|$ is $r-$Lipschitz as T is unit sphere. Using Levy–Gromov inequality [33], one bounds $f(\boldsymbol{g})$

$$\mathbb{P}\left(|\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)| > \xi\right) \leq Ce^{-n\xi^2}. \tag{34}$$

Then we show that $\mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)^2$ is close to $\mathbb{E}^2\sigma_{\min}(\boldsymbol{S}; T)$, and thus show the concentration of $\sigma_{\min}(\boldsymbol{S}; T)^2$.

Given any $\delta = \frac{1}{n} > 0$, we have

$$\begin{aligned}
&\mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)^2 - \mathbb{E}^2\sigma_{\min}(\boldsymbol{S}; T) \\
&= \mathbb{E}\left(|\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)|^2\right) \\
&\leq \sum_{i=1}^{\infty} i\delta^2 \mathbb{P}(\sqrt{i-1}\delta \leq |\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)| \leq \sqrt{i}\delta) \\
&\leq \sum_{i=1}^{\infty} \delta^2 \mathbb{P}\left(|\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)| > \sqrt{i-1}\delta\right) \\
&\leq \sum_{i=1}^{\infty} \delta^2 Ce^{-n(i-1)\delta^2} = \frac{1}{n}\sum_{i=0}^{\infty} Ce^{-i} = \frac{C}{n(1+e^{-1})}.
\end{aligned} \tag{35}$$

For a fixed vector $\boldsymbol{t} \in T$, we have

$$\mathbb{E}(\sigma_{\min}(\boldsymbol{S}; T)) \leq \mathbb{E}\|\boldsymbol{S}\boldsymbol{t}\| \leq \mathbb{E}(\lambda_{\max}). \tag{36}$$

Using the concentration of the singular value of $\boldsymbol{S}$ in Fact 1, we conclude that there exists a constant B such that $\mathbb{E}(\sigma_{\min}(\boldsymbol{S}; T)) \leq B$.

Combining (34) and (35), for any constant ξ and $n \geq 2/\xi$, we have

$$\begin{aligned}
&\mathbb{P}\left(|\sigma_{\min}^2(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}^2(\boldsymbol{S}; T)| > \xi\right) \\
&\leq \mathbb{P}\left(|\sigma_{\min}^2(\boldsymbol{S}; T) - \mathbb{E}^2\sigma_{\min}(\boldsymbol{S}; T)| \right. \\
&\quad \left. + |\mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)^2 - \mathbb{E}^2\sigma_{\min}(\boldsymbol{S}; T)| > \xi\right) \\
&\leq \mathbb{P}\left(|\sigma_{\min}^2(\boldsymbol{S}; T) - \mathbb{E}^2\sigma_{\min}(\boldsymbol{S}; T)| + \frac{1}{n} > \xi\right) \\
&\leq \mathbb{P}\left(|\sigma_{\min}^2(\boldsymbol{S}; T) - \mathbb{E}^2\sigma_{\min}(\boldsymbol{S}; T)| > \frac{\xi}{2}\right) \\
&\leq \mathbb{P}\left(|(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))^2 \right. \\
&\quad \left. + 2\mathbb{E}\sigma_{\min}(\boldsymbol{S}; T)(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))| > \frac{\xi}{2}\right).
\end{aligned} \tag{37}$$

Substitute the bound B into the inequality, we obtain

$$\begin{aligned}
&\mathbb{P}\left(|\sigma_{\min}^2(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}^2(\boldsymbol{S}; T)| > \xi\right) \\
&\leq \mathbb{P}\left(|(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))|^2 \right. \\
&\quad \left. + 2B|(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))| > \frac{\xi}{2}\right) \\
&\leq \mathbb{P}\left(\frac{\xi}{6B} < |(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))| < B\right) \\
&\quad + \mathbb{P}\left(\max(B, \sqrt{\frac{\xi}{6}}) < |(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))|\right) \\
&\leq \mathbb{P}\left(\frac{\xi}{6B} < |(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))|\right) \\
&\quad + \mathbb{P}\left(\sqrt{\frac{\xi}{6}} < |(\sigma_{\min}(\boldsymbol{S}; T) - \mathbb{E}\sigma_{\min}(\boldsymbol{S}; T))|\right) \\
&\leq C_0 e^{-\frac{n\xi^2}{6B}} + C_1 e^{-\frac{n\xi}{6}}.
\end{aligned} \tag{38}$$

C. Proof of (30)

Note that (30) is a consequence of (29). We mimic the proof technique from [25, Sec 9.3], which is less relevant to the structure of $\boldsymbol{S}$. The preliminary step is to separate the original set T into several subsets. For each set $J \subset \{0, 1, \ldots, rd-1\}$, a closed subset T_J is defined as

$$T_J := \left\{\boldsymbol{t} \in T : |t_j| \leq (\#J)^{-1/2} \quad \forall j \in J^c\right\} \tag{39}$$

where $J^c := \{0, 1, \ldots, rd-1\}\backslash J$. Fixing $k \leq rd$, we have

$$T = \bigcup_{\#J=k} T_J, \quad \#\{T_J : \#J = k\} \leq \left(\frac{erd}{k}\right)^k. \tag{40}$$

Step 1. Bound the expectation on T with that on T_j: First, using [25, Prop. 12.2] and taking $\log k < \log n$, we have the following inequality for all $\zeta \geq 0$:

$$\mathbb{E}\sigma_{\min}^2(\boldsymbol{S}; T) \geq \min_{\#J=k} \mathbb{E}\sigma_{\min}^2(\boldsymbol{S}; T_J) - C\sqrt{\frac{k}{n}\log(rd/k)}. \tag{41}$$

Note that [25, Prop. 12.2] does not strictly require the matrix $\boldsymbol{S}$ to be a random Gaussian matrix but rely on the inequality (29). Besides, when the item $e^{-n\xi^2}$ dominates in (29), the second item on the right hand of (41) is $\sqrt{\frac{k}{n}\log(rd/k)}$. When the item $e^{-n\xi}$ dominates, it is $\frac{k}{n}\log(rd/k)$, instead. By choosing k satisfying $\log k < \log n$, we obtain a uniform bound in both cases, as (41) shows.

Step 2. Bound the expectation with $\boldsymbol{S}$ with that on a new matrix with Gaussian entries: We next bound the expectation item on the right hand side of (41). Following [25, Prop. 13.1], let $\boldsymbol{\Phi}$ be an $n \times rd$ Gaussian matrix with entries obeying i.i.d $\mathcal{N}(0, \frac{1}{n})$. Define a new matrix $\boldsymbol{\Psi} := \boldsymbol{\Psi}(J)$, where $\boldsymbol{\Psi}_J = \boldsymbol{S}_J$ and $\boldsymbol{\Psi}_{J^c} = \boldsymbol{\Phi}_{J^c}$, i.e., replacing some columns of $\boldsymbol{S}$ by Gaussian entries. From Fact 2 that every entry $S_{i,j}$ in $\boldsymbol{S}$ is bounded by $\mathbb{E}(S_{i,j}^4) \leq B\frac{1}{n^2}$ and $\mathbb{E}(S_{i,j}^6) \leq B\frac{1}{n^3}$, we obtain

the following inequality by applying [25, Prop. 13.1]:

$$\begin{aligned}&\left|\mathbb{E}\sigma_{\min}^2(\boldsymbol{S};T_J)-\mathbb{E}\sigma_{\min}^2(\boldsymbol{\Psi};T_J)\right|\\&\le C/n+C\beta^{-1}rd\log(nrd)+\frac{C(\beta/n+\beta^2/n^2)r^2d}{k^{3/2}}.\end{aligned}\tag{42}$$

Here, β is a smoothing parameter. By selecting $\beta = k^{1/2}n^{2/3}$, the right hand side of the above inequation becomes

$$C/n+C\frac{rd\log(nrd)}{k^{1/2}n^{2/3}}+\frac{C(k^{1/2}n^{-1/3}+kn^{-2/3})r^2d}{k^{3/2}}.\tag{43}$$

Since $n/d = O(1)$ and $n^{2/3} \le k \le rd$, the second term dominates, yielding

$$\mathbb{E}\sigma_{\min}^2(\boldsymbol{S};T_J)\ge\mathbb{E}\sigma_{\min}^2(\boldsymbol{\Psi};T_J)-C\frac{rd\log(nrd)}{k^{1/2}n^{2/3}}.\tag{44}$$

Step 3. Bound the expectation with the new matrix on T_J: Next, we bound $\mathbb{E}\sigma_{\min}^2(\boldsymbol{\Psi};T_J)$. Using [25, Prop. 14.1], we have the following bound:

$$\begin{aligned}\mathbb{E}\sigma_{\min}^2(\boldsymbol{\Psi};T_J)&\ge\frac{1}{n}\left(\mathcal{E}(T_J)-C\sqrt{k}\right)_+^2\\&\ge\frac{1}{n}\left(\mathcal{E}(T)-C\sqrt{k}\right)_+^2\ge\frac{1}{n}(\mathcal{E}(T))_+^2-C\sqrt{k/n}.\end{aligned}\tag{45}$$

Step 4. Obtain the target bound: Finally, combining (41), (44), and (45) leads to

$$\mathbb{E}\sigma_{\min}^2(\boldsymbol{S})\ge(\mathcal{E}(T))_+^2-C\frac{rd\log(nrd)}{k^{1/2}n^{2/3}}-C\sqrt{\frac{k}{n}\log(rd/k)}.\tag{46}$$

Taking $k = n^{-1/6}rd$, we conclude

$$\begin{aligned}\mathbb{E}\sigma_{\min}^2(\boldsymbol{S})&\ge(\mathcal{E}(T))_+^2-Cn^{-7/12}(rd)^{1/2}\log(nrd)\\&\ge(\mathcal{E}(T))_+^2-\frac{1}{n}C(n+rd)^{11/12}\log(nrd).\end{aligned}\tag{47}$$

Appendix D
Proof of Lemma 7

Since Lemma 7 shares the similar structure with Lemma 6, we inherit the proof technique in Appendix C. We define

$$\tau_{\min}(\boldsymbol{S},K):=\min_{\|\boldsymbol{t}\|=1}\min_{\boldsymbol{s}\in K^\circ}\|\boldsymbol{s}-\boldsymbol{S}^{\mathrm{H}}\boldsymbol{t}\|_2.\tag{48}$$

It is sufficient to prove

$$\begin{aligned}&\mathbb{P}\left\{\tau_{\min}^2(\boldsymbol{S},K)\le\frac{1}{n}\left(\frac{(\mathcal{E}(\Omega))_-^2}{C_2\log(n+rd)}-C_3(n+rd)^{1-\kappa}\right)-C_4\xi\right\}\\&\le\max(e^{-\xi^2n},e^{-\xi n})\end{aligned}\tag{49}$$

where K is a closed convex cone in $\mathbb{R}^n$ and $\Omega = K\cap\mathcal{S}^{n-1}$.

The proof consists of two steps.

1) For any $\xi \ge 0$

$$\begin{aligned}&\mathbb{P}\left\{\tau_{\min}^2(\boldsymbol{S};K)\le\mathbb{E}\tau_{\min}^2(\boldsymbol{S};K)-C\xi\right\}\\&\le C_2e^{-\xi^2n}+C_3e^{-\xi n}.\end{aligned}\tag{50}$$

2) There exists some $\kappa \ge 0$ such that

$$\frac{(\mathcal{E}(\Omega))_-^2}{C_4\log(n+rd)}-(n+rd)^{1-\kappa}\le n\mathbb{E}\sigma_{\min}^2(\boldsymbol{S};K).\tag{51}$$

Since the proof of (50) is almost the same as that of (29), relying on the Lipschiz property and the Levy–Gromov theorem [33], we omit the details.

We focus on the proof of (51), which is completed similarly by using Facts 1 and 2 and the derivation in [25, Sec. 16]. We reserve some important derivations in our case, and the readers can refer to [25, Sec. 16] for detailed proof.

Define the set $P := K^\circ\cap R\mathcal{B}^n$ and $Q := \mathcal{S}^{rd-1}$, where $R := C_5\sqrt{1+\frac{rd}{n}}$ for some constant C_5. Then for $I\subset\{1,\dots,n\}$ and $J\subset\{1,\dots,rd\}$, we define

$$P_I:=\{\boldsymbol{p}\in P:|p_i|\le R(\#I)^{-1/2}\ \ \forall i\in I^c\}\tag{52}$$

$$Q_J:=\{\boldsymbol{q}\in Q:|q_i|\le(\#J)^{-1/2}\ \ \forall j\in J^c\}.\tag{53}$$

Given a fixed integer $1 \le k \le \min(n, rd)$, one has

$$P=\bigcup_{\#I=k}P_I\ \text{ and }\ Q=\bigcup_{\#J=k}Q_J\tag{54}$$

$$\#\{P_I:\#I=k\}\cdot\#\{Q_J:\#J=k\}\le\left(\frac{e(n+rd)}{k}\right)^k.\tag{55}$$

After the partition, we derive the bound by separating it into several steps. First, using [25, Prop. 18.1] and Fact 1, we have

$$\mathbb{E}\tau_{\min}^2(\boldsymbol{S},K)\ge\mathbb{E}\min_{\boldsymbol{q}\in Q}\min_{\boldsymbol{p}\in P}\|\boldsymbol{p}-\boldsymbol{S}\boldsymbol{q}\|-\frac{C}{n}.\tag{56}$$

Then, using [25, Prop. 12.2], one has

$$\begin{aligned}&\mathbb{E}\min_{\boldsymbol{q}\in Q}\min_{\boldsymbol{p}\in P}\|\boldsymbol{p}-\boldsymbol{S}\boldsymbol{q}\|^2\ge\\&\min_{\#I=\#J=k}\mathbb{E}\min_{\boldsymbol{q}\in Q_J}\min_{\boldsymbol{p}\in P_I}\|\boldsymbol{p}-\boldsymbol{S}\boldsymbol{q}\|^2-C\sqrt{\frac{k}{n}\log(n+rd)/k}.\end{aligned}\tag{57}$$

Given the index sets I and J, we introduce a new matrix

$$\boldsymbol{\Psi}(I,J):=\begin{bmatrix}\boldsymbol{S}_{I,J}&\boldsymbol{S}_{I,J^c}\\\boldsymbol{S}_{I^c,J}&\boldsymbol{\Gamma}_{I^c,J^c}\end{bmatrix}\tag{58}$$

where the elements in $\boldsymbol{\Gamma}\in\mathbb{R}^{n\times rd}$ satisfy i.i.d. $\mathcal{N}(0,\frac{1}{n})$. By [25, Prop. 19.2], one obtains the following bound:

$$\begin{aligned}&\mathbb{E}\min_{\boldsymbol{q}\in Q_J}\min_{\boldsymbol{p}\in P_I}\|\boldsymbol{p}-\boldsymbol{S}\boldsymbol{q}\|^2\ge\\&\mathbb{E}\min_{\|\boldsymbol{q}\|=1}\min_{\boldsymbol{p}\in K^\circ}\|\boldsymbol{p}-\boldsymbol{\Psi}(I,J)\boldsymbol{q}\|^2-\frac{C(n+rd)^{11/6}\log(nrd)}{kn}.\end{aligned}\tag{59}$$

The expectation item in the right-hand side of (59) is bounded by

$$\begin{aligned}&\mathbb{E}\min_{\|\boldsymbol{q}\|=1}\min_{\boldsymbol{p}\in K^\circ}\|\boldsymbol{p}-\boldsymbol{\Psi}(I,J)\boldsymbol{q}\|^2\\&\ge\frac{1}{n}\left(\frac{(\mathcal{E}(\Omega))_-}{C\sqrt{\log n}}-C\sqrt{k\log n}\right)_+^2\ge\frac{1}{n}\frac{(\mathcal{E}(\Omega))_-^2}{C\sqrt{\log n}}-C\sqrt{k/n}\end{aligned}\tag{60}$$

by applying [25, Prop. 20.1].

Combining the above inequalities, we obtain the final bound

$$n\mathbb{E}\tau_{\min}^2(\boldsymbol{S},K) \geq \frac{(\mathcal{E}(\Omega))_-^2}{C\sqrt{\log n}} - C\Big(\frac{(n+rd)^{11/6}\log(nrd)}{k} + \sqrt{kn\log(n+rd)/k}\Big). \tag{61}$$

By taking $k=(n+rd)^{8/9}$, we have

$$n\mathbb{E}\tau_{\min}^2(\boldsymbol{S},K) \geq \frac{(\mathcal{E}(\Omega))_-^2}{C\sqrt{\log(n+rd)}} - C(n+rd)^{17/18}\log(n+rd) \tag{62}$$

completing the proof of (51).

ACKNOWLEDGMENT

The authors would like to thank X. Xu from Carnegie Mellon University for the discussion.

REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[2] L. C. Potter, E. Ertin, J. T. Parker, and M. Cetin, "Sparsity and compressed sensing in radar imaging," *Proc. IEEE*, vol. 98, no. 6, pp. 1006–1020, Jun. 2010.

[3] T. Huang, Y. Liu, X. Xu, Y. C. Eldar, and X. Wang, "Analysis of frequency agile radar via compressed sensing," *IEEE Trans. Signal Process.*, vol. 66, no. 23, pp. 6228–6240, Dec. 2018.

[4] A. Aubry, V. Carotenuto, A. De Maio, and M. A. Govoni, "Multi-snapshot spectrum sensing for cognitive radar via block-sparsity exploitation," *IEEE Trans. Signal Process.*, vol. 67, no. 6, pp. 1396–1406, Mar. 2019.

[5] A. Aubry, V. Carotenuto, A. De Maio, and L. Pallotta, "High range resolution profile estimation via a cognitive stepped frequency technique," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 1, pp. 444–458, Feb. 2019.

[6] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[7] D. J. Hsu, S. M. Kakade, J. Langford, and T. Zhang, "Multi-label prediction via compressed sensing," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 772–780.

[8] L. Wang, T. Huang, and Y. Liu, "Randomized stepped frequency radars exploiting block sparsity of extended targets: A theoretical analysis," *IEEE Trans. Signal Process.*, vol. 69, pp. 1378–1393, 2021.

[9] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse Bayesian learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 300–309, Feb. 2013.

[10] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1600–1607.

[11] S. R. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[12] Y. C. Eldar and H. Bolcskei, "Block-sparsity: Coherence and efficient recovery," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2009, pp. 2885–2888.

[13] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[14] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

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

[16] A. Ganesh, Z. Zhou, and Y. Ma, "Separation of a subspace-sparse signal: Algorithms and conditions," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2009, pp. 3141–3144.

[17] E. Elhamifar and R. Vidal, "Block-sparse recovery via convex optimization," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4094–4107, Aug. 2012.

[18] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[19] P. Saidi, G. Atia, and A. Vosoughi, "Improved block-sparse recovery bound using cumulative block coherence," in *Proc. 54th Asilomar Conf. Signals Syst. Comput.*, 2020, pp. 1385–1389.

[20] Y. Li, T. Huang, X. Xu, Y. Liu, L. Wang, and Y. C. Eldar, "Phase transitions in frequency agile radar using compressed sensing," *IEEE Trans. Signal Process.*, vol. 69, pp. 4801–4818, 2021.

[21] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philos. Trans. Roy. Soc. A: Math. Phys. Eng. Sci.*, vol. 367, no. 1906, pp. 4273–4293, 2009.

[22] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete Comput. Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[23] M. Stojnic, "Block-length dependent thresholds in block-sparse compressed sensing," 2009, *arXiv:0907.3679*.

[24] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Inf. Inference: J. IMA*, vol. 3, no. 3, pp. 224–294, 2014.

[25] S. Oymak and J. A. Tropp, "Universality laws for randomized dimension reduction, with applications," *Inf. Inference: J. IMA*, vol. 7, no. 3, pp. 337–446, 2018.

[26] K. Zhou, D. Li, F. He, S. Quan, and Y. Su, "A sparse imaging method for frequency agile SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5223616.

[27] J. Wang, B. Shan, S. Duan, Y. Zhao, and Y. Zhong, "An off-grid compressive sensing algorithm based on sparse Bayesian learning for RFPA radar," *Remote Sens.*, vol. 16, no. 2, 2024, Art. no. 403.

[28] Z. Fei, J. Zhang, X. Zhu, and X. Ding, "Joint phase error estimation and sparse scene restoration algorithms for frequency agile radar," *IET Radar, Sonar Navigation*, vol. 15, no. 11, pp. 1525–1534, 2021.

[29] R. Fu, Y. Liu, T. Huang, and Y. C. Eldar, "Structured LISTA for multidimensional harmonic retrieval," *IEEE Trans. signal Process.*, vol. 69, pp. 3459–3472, 2021.

[30] R. Fu, V. Monardo, T. Huang, and Y. Liu, "Deep unfolding network for block-sparse signal recovery," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2021, pp. 2880–2884.

[31] S. Foucart and H. Rauhut, "Sparse recovery with random matrices," in *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013, pp. 271–310.

[32] P. Massart, "About the constants in talagrand's concentration inequalities for empirical processes," *Ann. Probability*, vol. 28, no. 2, pp. 863–884, 2000.

[33] V. D. Milman and G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces: Isoperimetric Inequalities in Riemannian Manifolds*. Berlin, Germany: Springer, 2009.

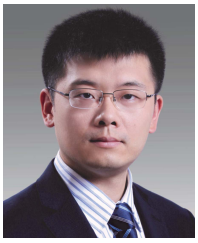
[34] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," 2010, *arXiv:1011.3027*.

Yuhan Li received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2020, where he is currently working toward the Ph.D. degree in electronic engineering.

His research interests include signal processing, compressed sensing, and optimization methods.

Tianyao Huang received the B.S. degree in telecommunication engineering from the Harbin Institute of Technology, Heilongjiang, China, in 2009, and the Ph.D. degree in electronics engineering from the Tsinghua University, Beijing, China, in 2014, respectively.

He is currently a Professor with the School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, P. R. China. From 2017 to 2023, he was with the Intelligence Sensing Lab, Department of Electronic Engineering, Tsinghua University, as an Assistant Professor. From 2014 to 2017, he was a radar researcher in Aviation Industry Corporation of China (AVIC). His current research interests include signal processing, compressed sensing, and joint radar communications system design.



Yimin Liu received the B.S. and Ph.D. degrees (both with Hons.) in electronics engineering from Tsinghua University, Beijing, China, in 2004 and 2009, respectively.

From 2004, he was with the Intelligence Sensing Lab (ISL), Department of Electronic Engineering, Tsinghua University. He is currently an Professor with Tsinghua University, where his field of activity is study on new concept radar and other microwave sensing technologies. His current research interests include radar theory, statistic signal processing, compressive sensing and their applications in radar, spectrum sensing, and intelligent transportation systems.

Xiqin Wang received the B.S. and Ph.D. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1991 and 1996, respectively.

He was a Visiting Scholar and an Assistant Research Engineer with PATH/ITS, UC Berkeley, from 2000 to 2003. He is also interested in the structure of knowledge and curricula reforming in electronic engineering. His current research interests include radar and communications signal processing, image processing, compressed sensing and cognitive signal processing, and electronic system design.

Yonina C. Eldar (Fellow, IEEE) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 2002.

She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. She was previously a Professor with the Department of Electrical Engineering, Technion, where she held the Edwards Chair in Engineering. She is also a Visiting Professor with MIT, a Visiting Scientist with the Broad Institute, and an Adjunct Professor with Duke University, and was a Visiting Professor with Stanford. She is also the author of the book titled *Sampling Theory: Beyond Bandlimited Systems* and coauthor of four other books published by Cambridge University Press. Her research interests include statistical signal processing, sampling theory and compressed sensing, learning and optimization methods, and their applications to biology, medical imaging, and optics.

Dr. Eldar has received many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award in 2013, the IEEE/AESS Fred Nathanson Memorial Radar Award in 2014, and the IEEE Kiyo Tomiyasu Award in 2016. She was a Horev Fellow of the Leaders in Science and Technology program with the Technion and an Alon Fellow. She was the recipient of the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair with the Technion, the Muriel and David Jacknow Award for Excellence in Teaching, and the Technions Award for Excellence in Teaching (two times). She received several best paper awards and best demo awards together with her research students and colleagues including the SIAM Outstanding Paper Prize, the UFFC Outstanding Paper Award, the Signal Processing Society Best Paper Award and the IET Circuits, Devices and Systems Premium Award, and was selected as one of the 50 most influential women in Israel and in Asia. She is also a highly cited Researcher. She was a Member of the Young Israel Academy of Science and Humanities (elected 2017) and the Israel Committee for Higher Education. She is the Editor-in-Chief for *Foundations and Trends in Signal Processing*, a Member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees. In the past, she was a Signal Processing Society Distinguished Lecturer, a Member of the IEEE Signal Processing Theory and Methods and Bio-Imaging Signal Processing technical committees, and served as an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, *EURASIP Journal of Signal Processing*, *SIAM Journal on Matrix Analysis and Applications*, and *SIAM Journal on Imaging Sciences*. She was the Cochair and Technical Cochair of several international conferences and workshops. She is a EURASIP Fellow.