Phase Transitions in Frequency Agile Radar Using Compressed Sensing

Yuhan Li, Tianyao Huang, Xingyu Xu, *Graduate Student Member, IEEE*, Yimin Liu, *Member, IEEE*, Lei Wang, and Yonina C. Eldar, *Fellow, IEEE*

***Abstract*—Frequency agile radar (FAR) has improved anti-jamming performance over traditional pulse-Doppler radars under complex electromagnetic circumstances. To reconstruct the range-Doppler information in FAR, many compressed sensing (CS) methods including standard and block sparse recovery have been applied. In this paper, we study phase transitions of range-Doppler recovery in FAR using CS. In particular, we derive closed-form phase transition curves associated with block sparse recovery and complex Gaussian matrices, based on prior results of standard sparse recovery under real Gaussian matrices. We further approximate the obtained curves with elementary functions of radar and target parameters, facilitating practical applications of these curves. Our results indicate that block sparse recovery outperforms the standard counterpart when targets occupy more than one range cell, which are often referred to as extended targets. Simulations validate the availability of these curves and their approximations in FAR, which benefit the design of the radar parameters.**

***Index Terms*—Frequency agile radar, phase transition, block sparse recovery, $\ell_{2,1}$ norm minimization.**

I. Introduction

FREQUENCY agile radar (FAR) varies its carrier frequencies randomly in a pulse by pulse manner. It synthesizes a wide bandwidth by coherently processing echoes of different frequencies, achieving high range resolution (HRR) while requiring only a narrow-band hence low-cost receiver [2]. This facilitates applications including synthetic aperture radar (SAR) [3] and inverse synthetic aperture radar (ISAR) imaging [4], [5]. In addition, frequency agile radar (FAR) possesses excellent electronic counter-countermeasures performance [2], supports spectrum sharing [6], and enhances spectrum efficiency [7]. Owing to these advantages, FAR has drawn considerable attention in the radar community [8], [9].

Manuscript received December 30, 2020; revised May 18, 2021, June 14, 2021, and July 9, 2021; accepted July 15, 2021. Date of publication July 27, 2021; date of current version September 3, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bo Chen. This work was supported by the National Natural Science Foundation of China under Grant 61801258. This work was presented in part at the IEEE radar conference 2020 [DOI: 10.1109/RadarConf2043947.2020.9266406]. *(Corresponding author: Tianyao Huang.)*

Yuhan Li, Tianyao Huang, Xingyu Xu, Yimin Liu, and Lei Wang are with the Department of Electronic Engineering, Tsinghua University, Beijing, China (e-mail: liyh20@mails.tsinghua.edu.cn; huangtianyao2009@gmail.com; xy-xu19@mails.tsinghua.edu.cn; yiminliu@tsinghua.edu.cn; wangleixd@sina.com).

Yonina C. Eldar is with the Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

Digital Object Identifier 10.1109/TSP.2021.3099629

FAR relies on signal processing algorithms to recover the range-Doppler parameters of observed targets and clutter. Early works [2] employed the traditional matched filtering for range-Doppler reconstruction, which led to significant sidelobe pedestal. As a consequence, weak targets could be covered by the sidelobe of dominant ones or strong clutter [10]. To alleviate the sidelobe pedestal problem, compressed sensing (CS) methods (also known as sparse recovery [11]) have been suggested, which exploit the inherent sparsity of the targets [10], [12]. The authors in [13] further extended the standard sparse recovery approach to block sparse recovery to account for the situation of extended targets, where a target may be larger than the range resolution and therefore can occupy more than one range cell [14]. In this situation, the scatterers of a target share the same Doppler effect and gather along range, leading to block sparsity [15]. When confronted with such block sparse situations including extended targets, block sparse recovery is considered to perform better than the conventional sparse recovery [16], [17].

Precise conditions that guarantee reconstruction in FAR have been considered in several papers. In [12] and [13], the authors provided sufficient conditions (in terms of the numbers of targets K, radar pulses N and available frequencies M) that guarantee successful reconstruction of target scenes with high probability using standard and block sparse recovery, respectively. Nevertheless, those conditions, based on the well-known coherence property, are generally loose and pessimistic [18], and are therefore not accurate enough to predict the actual recovery performance given the radar and target parameters.

To obtain a tighter bound, we study here the phase transition, which emerges in many convex optimization problems [19]. In CS, phase transition means that there exist thresholds that divide the plane of parameters, i.e., the number of observations and the sparsity level, into regions, where recovery succeeds and fails with high probability [20]. These thresholds are called phase transition curves. Finding analytical expressions for these curves is an active research area. For standard sparse recovery, bounds on the phase transition curve of ℓ_1 norm minimization under standard Gaussian matrices were established in [21], [22]. Generalization to block sparse recovery and for complex numbers were given in [23] and [24], respectively. However, these approximate results assume the observation matrices to be large, and have complicated form. A more concise and tight bound, which has no requirement on the size of the observation matrix, was given in [19] using integral geometry. Nevertheless, it is

confined to standard sparse recovery under real-valued Gaussian matrices.

In FAR, block sparse recovery is preferred, and measurement matrices are complex-valued. Therefore, we first extend the results of [19] to block sparse situations and complex Gaussian matrices. While existing analyses are based on the Gaussian assumption, measurement matrices in FAR are not Gaussian but structured. Empirical experiments show that many random matrices exhibit identical phase transition curves as Gaussian matrices [25]. We demonstrate numerically in Section VI that the obtained bounds derived from Gaussian matrices are tight and accurate for FAR. Thus our results provide more precise conditions for exact range-Doppler reconstruction, compared to former works [12], [13].

Next, we approximate the obtained bounds, which involve minimization over an integral function, with some elementary functions. In particular, under relatively sparse scenes where there are only a few extended targets, we show that the required number of measurements when using block and standard sparse recovery are on the order of $2MK + O(K\sqrt{M \log \frac{N}{K\sqrt{M}}})$ and $2MK + O(KM \log \frac{N}{K})$, respectively. The former requires less radar measurements for exact reconstruction of extended targets, since $\sqrt{M \log \frac{N}{K\sqrt{M}}} < M \log \frac{N}{K}$ for reasonably large N and M. The accuracy of these approximations is validated by simulations. These approximations not only simplify the calculation of the bounds, facilitating their use in practical scenarios, but also explicitly and quantitatively reveal the dependency of the required number of measurements on the radar and target parameters. These explicit results enable theoretical performance comparison between block and standard sparse recovery methods, which demonstrates the superiority of block sparse recovery.

The main contributions of this paper are as follows:

- We derive the phase transition curves under Gaussian matrices for block sparse recovery, which are also numerically accurate for FAR, associated with structured and non-Gaussian matrices.
- We approximate the phase transition curves of both block and standard sparse recovery with elementary functions, facilitating their use in practical FAR, and demonstrating that block sparse recovery outperforms the standard one when reconstructing extended targets.

This paper extends [1] by providing proofs and discussions on approximating phase transition curves.

The rest of the paper is structured as follows. Section II introduces the signal model of FAR. In Section III, we briefly review basic concepts of standard and block sparse recovery as well as phase transitions for standard sparse recovery. Section IV extends the results of phase transitions in Section III to block sparse recovery and complex problems. Section V analyzes phase transitions for the FAR model, which are verified by simulations in Section VI. Section VII concludes the paper.

Throughout the paper, we use $\mathbb{R}$ and $\mathbb{C}$ to denote the real and complex number set, respectively. For $x \in \mathbb{R}$, $\lfloor x \rfloor$, represents the largest integer no greater than x. 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 $j := \sqrt{-1}$. Given a matrix $\boldsymbol{A}$, $[\boldsymbol{A}]_{m,n}$ is the (m,n)-th entry of $\boldsymbol{A}$. The transpose operator is $(\cdot)^T$ and $\mathbb{E}(\cdot)$ means the expectation of a random value. The real and imaginary parts of a complex-valued argument are written as $\Re\cdot$ and $\Im\cdot$, respectively. We use $\mathcal{N}(0,1)$ to denote standard Gaussian distribution.

II. FAR Signal Model

We begin by reviewing the signal model of FAR in Subsection II-A, followed by its matrix form in Subsection II-B.

A. Radar Model

We introduce the signal model of FAR following [12] and [13]. We start with the expressions of the transmissions and the received echoes from a single scattering point, representing target or clutter. We then extend the echo model to the case of multiple targets/clutter.

A FAR system changes the frequencies from pulse to pulse. Suppose that the radar transmits N monotone pulses during a coherent processing interval (CPI). The carrier frequency of the n-th pulse can be written as $f_n = f_c + C_n \Delta f$, where f_c represents the initial frequency, Δf represents the frequency step, and C_n is the n-th random modulation code. We assume that all the modulation codes are independently and identically distributed (i.i.d.) random variables with uniform density on $\mathcal{M} := \{0, 1, \ldots, M-1\}$, i.e., $C_n \sim U(\mathcal{M})$. Then the n-th transmitted pulse, $n \in \mathcal{N} := \{0, 1, \ldots, N-1\}$, is written as

$$S_T(n,t) = \mathrm{rect}\left((t - nT_r)/T_p\right) e^{j2\pi f_n (t - nT_r)}, \tag{1}$$

where T_r is the pulse repetition interval (PRI), T_p is the pulse duration, and $\mathrm{rect}(t)$ equals 1 when $0 \le t \le 1$ and 0 otherwise.

The received echoes can be seen as delays of transmissions. To clearly present the signal model, we assume that there exists only one ideal scatterer with complex scattering coefficient ζ, which has an initial range r from the radar and is moving along the radar's line of sight at a fixed speed of v. The time delay between the received and the transmitted signal at time instant t takes the form $\tau(t) := \frac{2(r+vt)}{c}$, where c is the velocity of light. Under the stop-and-hop assumption [26], it holds that $\tau(t) \approx \tau(nT_r)$ during the n-th pulse. Thus, the received echo of the n-th transmission can be expressed as

$$S_R(n,t) = \zeta S_T(n, t - \tau(t)) \approx \zeta S_T(n, t - \tau(nT_r)). \tag{2}$$

After down-conversion, the received echoes become

$$S_D(n,t) = S_R(n,t) e^{-j2\pi f_n (t - nT_r)}. \tag{3}$$

Substituting (2) and (1) into (3) and rearranging, we obtain

$$\begin{aligned} S_D(n,t) &= \zeta \mathrm{rect}\left(\frac{t - nT_r - \tau(nT_r)}{T_p}\right) e^{-j2\pi f_n \tau(nT_r)} \\ &= \ddot{\zeta} \mathrm{rect}\left(\frac{t - nT_r - \tau(nT_r)}{T_p}\right) e^{j2\pi f_r C_n + j2\pi f_v \epsilon_n n}, \end{aligned} \tag{4}$$

where $\epsilon_n := 1 + \frac{C_n \Delta f}{f_c}$. The parameters $\ddot{\zeta} := \zeta e^{-j4\pi f_c r/c}$, $f_r := -\frac{2\Delta f r}{c}$ and $f_v := -\frac{2 f_c v T_r}{c}$, representing the effective intensity, range frequency and velocity frequency, respectively, are unknown and need to be estimated.

The down-converted echoes are sampled at the Nyquist rate of a single pulse, $f_s = \frac{1}{T_p}$, at time instances, $t = nT_r + \frac{l_s}{f_s}$, $l_s = 0, 1, \ldots, \lfloor T_r f_s \rfloor$. Each sample corresponds to a coarse range resolution (CRR) bin. The coarse range will be refined by estimating f_r or r from the echoes. Here, f_r and r are referred to as the HRR information. Since data from those bins are processed identically and individually, without loss of generality, we assume that the target is located in the l_s-th CRR bin, and that the target does not move outside the bin during the CPI. Consequently, the $\mathrm{rect}(\cdot)$ term in (4) at the l_s sampling instance equals 1, so that the sampled $S_D(n,t)$ becomes

$$S_S(n) := S_D\left(n, nT_r + l_s/f_s\right) = \ddot{\zeta} e^{j2\pi f_r C_n + j2\pi f_v \epsilon_n n}. \quad (5)$$

The model (5), derived for the case of a single scatterer, can be extended to the setting in which K targets/clutter exist in a CRR bin. Particularly, we denote by v_k and f_{v_k} the velocity and velocity frequency of the k-th target (or clutter unit), respectively, with $k = 0, 1, \ldots, K-1$, and assume that the k-th target is composed of Q_k scatterers, moving at the same speed while the ranges and scattering intensities are different. For the i-th scatterer of the k-th target, $\ddot{\zeta}_{ki}$, r_{ki} and $f_{r_{ki}}$ denote the scattering coefficient, initial range and range frequency, respectively. We then extend (5) into a model for multiple targets as

$$S_S(n) = \sum_{k=0}^{K-1} \sum_{i=0}^{Q_k - 1} \ddot{\zeta}_{ki} e^{j2\pi f_{r_{ki}} C_n + j2\pi f_{v_k} \epsilon_n n}. \quad (6)$$

Here, $\{\ddot{\zeta}_{ki}, f_{r_{ki}}, f_{v_k}\}$ are unknown and need to be estimated from $S_S(n)$.

In the next subsection, we arrange the signal model in matrix form, which suggests a block sparse recovery approach for range-Doppler reconstruction.

B. Signal Model in Matrix Form

To write (6) in matrix form, we first divide the continuous f_r and f_v, representing range and Doppler parameters into grid points. In particular, since (f_r, f_v) is unambiguous in the domain $[0,1)^2$ and their resolutions are $1/M$ and $1/N$, respectively, we discretize f_r and f_v at the rates of $1/M$ and $1/N$, respectively, resulting in a series of grid points $(p/M, q/N)$, $p \in \mathcal{M}, q \in \mathcal{N}$.

We consider a discrete model, which assumes that all scatterers are situated exactly on the grid points, and use the matrix $\boldsymbol{X} \in \mathbb{C}^{M \times N}$ to encapsulate the scattering intensities, given by

$$[\boldsymbol{X}]_{p,q} := \begin{cases} \ddot{\zeta}_{ki}, & \exists (k,i), \text{ s.t. } (f_{r_{ki}}, f_{v_k}) = (\frac{p}{M}, \frac{q}{N}), \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

In practical scenes, scatterers of targets and clutter may be continuously distributed rather than located on the discrete grid. It is shown in [13] that radar returns of the discrete model well approximate the counterparts scattered from continuously located scatterers. Let $\boldsymbol{x}_q \in \mathbb{C}^M$ denote the q-th column of $\boldsymbol{X}$, representing the HRR profile of the target with Doppler frequency $q/N, q \in \mathcal{N}$. Vectorizing $\boldsymbol{X}$ yields $\boldsymbol{x} := [\boldsymbol{x}_0^T, \boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_{N-1}^T] \in \mathbb{C}^{MN}$. Thus, $\boldsymbol{x}$ contains N blocks each having M entries, which represent the HRR profile corresponding to a unique Doppler frequency [13], and is referred to as having a *block* structure. For the scenario that includes K targets, at most K blocks in $\boldsymbol{x}$ are nonzero, and $\boldsymbol{x}$ is a so-called K block sparse vector.

Following [13], we now write (6) in matrix form as

$$\boldsymbol{y} = \boldsymbol{\Theta} \boldsymbol{x}, \quad (8)$$

where the n-th entry of the measurement vector $\boldsymbol{y} \in \mathbb{C}^N$ is given by $[\boldsymbol{y}]_n := S_S(n)$. The observation matrix $\boldsymbol{\Theta} \in \mathbb{C}^{N \times MN}$, in accordance with $\boldsymbol{x}$, is separated into N blocks as

$$\boldsymbol{\Theta} := [\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_{N-1}], \quad (9)$$

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

$$[\boldsymbol{\Theta}_q]_{n,p} = e^{j2\pi \frac{p}{M} C_n + j2\pi \frac{q}{N} \epsilon_n n}, \quad q, n \in \mathcal{N}, p \in \mathcal{M}. \quad (10)$$

In some scenarios, only a random selection out of all N pulses are transmitted (in the aim of, e.g., reducing power consumption), or partial observations in $\boldsymbol{y}$ are abandoned and not processed because they have a strong interferer. This will lead to a compressive observation model, where only randomly selected entries of $\boldsymbol{y}$ and $\boldsymbol{\omega}$, as well as the corresponding rows of $\boldsymbol{\Psi}$, remain.

The matrix $\boldsymbol{\Theta}$ has more columns than rows, hence recovering $\boldsymbol{x}$ from $\boldsymbol{y}$ is an under-determined problem. Generally, there are only a small amount of targets occurring in a certain CRR, which means that K is small, causing $\boldsymbol{x}$ to be block sparse [16]. Therefore, block sparse recovery can be utilized to recover $\boldsymbol{x}$ and henceforth the target parameters. We will review some basic concepts on sparse recovery in Section III, and analyze the recovery performance in Section IV.

Since the purpose of this paper is to identify the fundamental limits on the recovery performance of FAR using sparse recovery methods, in most parts of the paper we focus on the noise free model (8), following typical approaches [19]–[24]. In the simulation section, we will numerically discuss the recovery performance under noisy circumstances, shown in Section VI-E.

III. Preliminaries on Compressed Sensing

In this section, we briefly introduce some preliminaries on CS and its phase transitions. In Section III-A, CS methods including non-block and block sparse recovery are reviewed. Then in Section III-B, we introduce the phase transition phenomenon in sparse recovery, which serves as a theoretical tool for performance analysis. To distinguish between the specific radar parameters like N, we use lower case letters such as n to denote the dimensions of matrices associated with a more general sparse recovery problem. Since some variables in this section can be either real or complex valued, and will be specified in later discussions, we use $\mathbb{S}$ to represent $\mathbb{R}$ or $\mathbb{C}$ for convenience.

A. Sparse Recovery

Consider an under-determined problem $\boldsymbol{y} = \boldsymbol{\Psi} \boldsymbol{x}$, where $\boldsymbol{y} \in \mathbb{S}^n$, $\boldsymbol{\Psi} \in \mathbb{S}^{n \times d}$ and $\boldsymbol{x} \in \mathbb{S}^d$. Here, $\boldsymbol{x}$ has no more than s nonzero

entries, and is called an s sparse vector. CS recovers $\boldsymbol{x}$ by harnessing the sparsity as in the following optimization program:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_0, \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x}. \tag{11}$$

Since the 'ℓ_0 norm' optimization (11) is NP-hard [11], under appropriate conditions, the problem can be solved more efficiently by ℓ_1 norm minimization, i.e.,

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1, \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x}. \tag{12}$$

The standard sparse recovery problem (12) can be extended to block sparse recovery. With some abuse of notation, we consider a block-structured vector $\boldsymbol{x} \in \mathbb{S}^{md}$ consisting of d blocks where each block has m entries, denoted by $\boldsymbol{x} := [\boldsymbol{x}_0^T, \boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_{d-1}^T]^T$. Here, $\boldsymbol{x}_q \in \mathbb{S}^m$ denotes the q-th block. We use s_B to represent the block sparsity, i.e., at most s_B blocks in $\boldsymbol{x}$ are nonzero. Similarly, we redefine $\boldsymbol{\Psi} \in \mathbb{S}^{n\times md}$ as the measurement matrix. To solve for $\boldsymbol{x}$ from observations $\boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x}$, we exploit the block sparsity in $\boldsymbol{x}$ by considering an $\ell_{2,1}$ minimization problem [16]

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{2,1}, \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x}. \tag{13}$$

Here, the $\ell_{2,1}$ norm, given by $\|\boldsymbol{x}\|_{2,1} := \sum_{q=0}^{d-1} \|\boldsymbol{x}_q\|_2$, is defined with respect to the block width m. When $m = 1$, (13) reduces to (12).

Both standard and block sparse recovery, i.e., (12) and (13), can be applied to FAR, and provide unique solutions with high probability under certain conditions [12], [13]. Particularly, the recoverable number of targets K is on the order of $O(\sqrt{\frac{N}{\log(MN)}})$ using ℓ_1 norm minimization [12] and $O(\frac{N}{M\log(MN)})$ using $\ell_{2,1}$ norm minimization [13], where we recall that N and M are the numbers of pulses and available frequencies, respectively. These obtained conditions are sufficient yet too pessimistic, as discussed in Section VI. Consequently, these conditions are difficult to use directly in practical FAR systems to guide the waveform design under given parameters. To seek appropriate conditions that guarantee exact recovery with high probability and are tight enough to provide design criterion for radar systems, we resort to phase transition curves, which we introduce next.

B. Phase Transition Phenomenon

In this section, we briefly introduce the phase transition phenomenon, following the ideas in [19].

Many CS works have focused on when exact recovery is possible using (12) and how these conditions change as a function of the problem parameters. Early works such as [21] observed that the probability of exact recovery possesses a phase transition with respect to the number of measurements, n, and the sparsity of the signal, s. Here, a phase transition means a dramatic change in the probability of exact recovery when these parameters, n and s, vary around certain values. To illustrate the phase transition phenomenon associated with the probability of exact recovery, we empirically present in Fig. 1 the probabilities of exact recovery, assuming an under-determined real-valued

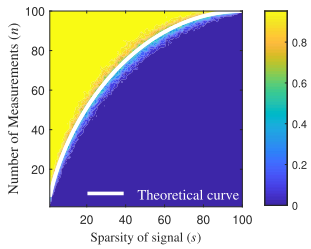


Fig. 1. Probabilities that (12) exactly solves $\boldsymbol{x}$ simulated from 50 trials. Here, $d = 100$, $\boldsymbol{\Psi} \in \mathbb{R}^{n\times d}$ has entries obeying i.i.d. $\mathcal{N}(0, 1)$, and the nonzero entries of $\boldsymbol{x}$ are randomly 1 or -1. Exact recovery is proclaimed when the estimate $\hat{\boldsymbol{x}}$ satisfies $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq 10^{-5}$. The theoretical curve is calculated analytically and is introduced in Proposition 2.

Gaussian measurement matrix $\boldsymbol{\Psi}$, applying (12) under different pairs (n, s). The corresponding phase transition points (n, s) where phase transitions happen, compose what we call the phase transition curve. This curve precisely characterizes the required conditions for exact recovery. While empirically calculating the curve is usually time consuming, expressions of the theoretical curve have been given in certain cases as we discuss next.

The paper [19] identified a theoretical phase transition curve for a more general optimization problem of the form

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} f(\boldsymbol{x}), \text{ s.t. } \boldsymbol{y} = \boldsymbol{\Psi}\boldsymbol{x}, \tag{14}$$

under a real-valued Gaussian measurement matrix $\boldsymbol{\Psi}$, based on integral geometry techniques. Here, $f(\cdot)$ is restricted to be convex and does not take the value $-\infty$. In general, $f(\cdot)$ indicates the 'structure' in a vector. For example, $f(\cdot) = \|\cdot\|_1$ characterizes the standard sparsity of a vector, and in this case (14) reduces to (12). To calculate the number of measurements n which causes a phase transition, two concepts are introduced, the descent cone and the statistical dimension. The descent cone of a proper convex function $f : \mathbb{R}^d \to \mathbb{R}$ at the point $\boldsymbol{x} \in \mathbb{R}^d$ is defined as

$$D(f, \boldsymbol{x}) := \bigcup_{\tau>0} \{\boldsymbol{y} \in \mathbb{R}^d : f(\boldsymbol{x} + \tau\boldsymbol{y}) \leq f(\boldsymbol{x})\}. \tag{15}$$

It depicts the conic hull of the perturbations which decrease or maintain f around $\boldsymbol{x}$. The number of measurements n which causes a phase transition, representing the phase transition curve, depends on the descent cone at the point $\boldsymbol{x}$, given by

$$n = \delta(D(f, \boldsymbol{x})) = \mathbb{E}\left[\text{dist}^2(\boldsymbol{g}, D(f, \boldsymbol{x}))\right]. \tag{16}$$

Here, $\delta(\cdot)$ is called the statistical dimension of a cone, the expectation is taken over the random vector $\boldsymbol{g} \in \mathbb{R}^d$, obeying the Gaussian distribution $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, and $\text{dist}(x, S)$ denotes the Euclidean distance from a vector $\boldsymbol{x} \in \mathbb{R}^d$ to a set $S \subset \mathbb{R}^d$

$$\text{dist}(\boldsymbol{x}, S) := \inf\{\|\boldsymbol{x} - \boldsymbol{y}\|_2 : \boldsymbol{y} \in S\}. \tag{17}$$

Now, identifying the phase transition curve becomes calculating the statistical dimension with respect to the ℓ_1 norm.

Directly calculating the statistical dimension is difficult, a tight upper bound on it is used instead in [19]. To obtain the bound, we first introduce the following definition. For an appropriate convex function $f : \mathbb{R}^d \to \mathbb{R}$, e.g., ℓ_1 norm of a d

dimensional vector, the subdifferential $\partial f(\boldsymbol{x})$ at a point $\boldsymbol{x} \in \mathbb{R}^d$ is defined as

$$\partial f(\boldsymbol{x}) := \{\boldsymbol{s} \in \mathbb{R}^d : f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{s}, \boldsymbol{y} - \boldsymbol{x} \rangle,\ \forall\, \boldsymbol{y} \in \mathbb{R}^d\}, \tag{18}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot product between two vectors. Next, the following upper bound is derived in Proposition 4.1 of [19]:

Proposition 1 ([19]): For $\boldsymbol{x} \in \mathbb{R}^d$ and a convex function f: $\mathbb{R}^d \to \mathbb{R}$, demanding the subdifferential $\partial f(\boldsymbol{x})$ to be compact, nonnull, and not containing the origin, the following function

$$J(\boldsymbol{x}, f) := \inf_{\tau \geq 0} \mathbb{E}[\mathrm{dist}^2(\boldsymbol{g}, \tau \cdot \partial f(\boldsymbol{x}))] \tag{19}$$

upper bounds $\delta(D(f, \boldsymbol{x}))$, where the expectation in (19) is taken over the random vector $\boldsymbol{g} \in \mathbb{R}^d$, obeying the Gaussian distribution $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

To apply Proposition 1, we substitute $f(\cdot)$ with $\|\cdot\|_1$ in (19) and calculate the infimum distance expectation $J(\boldsymbol{x}, \|\cdot\|_1)$, which implies the upper bound on $\delta(D(\|\cdot\|_1, \boldsymbol{x}))$. We then express the upper bound as a function of the sparsity and the dimension of $\boldsymbol{x}$, s and d, as claimed in the following proposition:

Proposition 2 ([19]): Given an s sparse signal $\boldsymbol{x} \in \mathbb{R}^d$, $\varphi(s, d)$, defined as

$$\varphi(s, d) := \inf_{\tau \geq 0} \left\{ s(1 + \tau^2) + (d - s) \int_\tau^\infty (u - \tau)^2 \phi(u) \mathrm{d}u \right\}, \tag{20}$$

upper bounds $\delta(D(\|\cdot\|_1, \boldsymbol{x}))$, where $\phi(u) := \sqrt{\frac{2}{\pi}} \exp(-\frac{u^2}{2}), u \geq 0$, is the probability density function of the folded normal distribution.

Proposition 2 offers a way to calculate the location of the phase transitions without performing complicated simulations running ℓ_1 norm minimization algorithms. The upper bound $\varphi(s, d)$ in (20) is tight, as shown in Fig. 1, denoted by 'theoretical curve'.

IV. Phase Transitions in Block Sparse Recovery

Here we extend the result of [19] to the block sparse setting. We present the theoretical results for real-valued Gaussian matrices in Subsection IV-B, and complex Gaussian matrices in Subsection IV-B.

A. Real-Valued Cases

Consider the model in (13) with $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{\Psi} \in \mathbb{R}^{n \times md}$ and $\boldsymbol{x} \in \mathbb{R}^{md}$. We assume that $\boldsymbol{\Psi}$ is a real-valued Gaussian matrix and $\boldsymbol{x}$ is block sparse with sparsity s_B. According to Proposition 1, we let $f(\boldsymbol{x})$ be the $\ell_{2,1}$ norm of $\boldsymbol{x}$ with a block size m. Thus, the phase transition curve becomes $\delta(D(\|\cdot\|_{2,1}, \boldsymbol{x}))$ and can be achieved by calculating the term in the right hand side of (19) with respect to $f(\boldsymbol{x})$. Following the steps used for the derivation of Proposition 2, we obtain the following proposition, which provides an upper bound on $\delta(D(\|\cdot\|_{2,1}, \boldsymbol{x}))$ in terms of s_B and d.

Proposition 3: Given an s_B block sparse signal $\boldsymbol{x} \in \mathbb{R}^{md}$ having a block size m, the function $\varphi_m(s_B, d)$, defined as

$$\varphi_m(s_B, d) := \inf_{\tau \geq 0} \left\{ s_B(m + \tau^2) + (d - s_B) \int_\tau^\infty (u - \tau)^2 \phi_m(u) \mathrm{d}u \right\}, \tag{21}$$

upper bounds $\delta(D(\|\cdot\|_{2,1}, \boldsymbol{x}))$. Here, $\phi_m(u)$ is the probability density function of the χ-distribution with m degrees of freedom, given by

$$\phi_m(u) := \begin{cases} \frac{u^{m-1} e^{-u^2/2}}{2^{m/2-1}\Gamma(\frac{m}{2})}, & u \geq 0, \\ 0, & \text{otherwise}, \end{cases} \tag{22}$$

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

Proof. See Appendix A. ∎

Proposition 3 offers a theoretical bound on the phase transition curve in block sparse recovery, which is empirically tight as will be shown in Section VI by experiments. It generalizes Proposition 2, because standard ℓ_1 norm minimization can be regarded as a special case of block sparse recovery with $m = 1$ and $s_B = s$. In this case, the curve $\varphi_m(s_B, d)$ becomes identical to $\varphi(s, d)$ in Proposition 2. Proposition 3 also paves the way to discussing sparse recovery with complex-valued measurement matrices. This is because both ℓ_1 and $\ell_{2,1}$ norm minimization under complex-valued measurement matrices can be expressed by $\ell_{2,1}$ norm minimization in real-valued formulations, as discussed in the next subsection.

B. Complex-Valued Cases

It is well known that complex-valued sparse recovery problems can be reformulated into real-valued ones [24]. We use the same model (13) with Subsection IV-A except that variables are complex valued: $\boldsymbol{y} \in \mathbb{C}^n$, $\boldsymbol{x} \in \mathbb{C}^{md}$, and $\boldsymbol{\Psi} \in \mathbb{C}^{n \times md}$ being a complex-valued Gaussian matrix. The model can be converted into a real-valued form $\boldsymbol{y}_r = \boldsymbol{\Psi}_r \boldsymbol{x}_r$ by introducing notations: $\boldsymbol{y}_r := [\Re\boldsymbol{y}^T, \Im\boldsymbol{y}^T]^T \in \mathbb{R}^{2n}$, $\boldsymbol{\Psi}_r := \begin{bmatrix} \Re\boldsymbol{\Psi} & -\Im\boldsymbol{\Psi} \\ \Im\boldsymbol{\Psi} & \Re\boldsymbol{\Psi} \end{bmatrix} \in \mathbb{R}^{2n \times 2md}$ and $\boldsymbol{x}_r := [\Re\boldsymbol{x}^T, \Im\boldsymbol{x}^T]^T \in \mathbb{R}^{2md}$. The q-th block in $\boldsymbol{x}$ is rewritten as $\bar{\boldsymbol{x}}_q = [\Re\boldsymbol{x}_q^T, \Im\boldsymbol{x}_q^T]^T \in \mathbb{R}^{2m}$. We then exchange the entries in $\boldsymbol{x}_r$ such that we obtain $\bar{\boldsymbol{x}} := [\bar{\boldsymbol{x}}_0^T, \ldots, \bar{\boldsymbol{x}}_{d-1}^T]^T \in \mathbb{R}^{2md}$. Applying the same arrangement to the columns of $\boldsymbol{\Psi}_r$ yields $\overline{\boldsymbol{\Psi}} \in \mathbb{R}^{2n \times 2md}$, and we have $\boldsymbol{y}_r = \overline{\boldsymbol{\Psi}} \bar{\boldsymbol{x}}$. Here, the real-valued vector $\bar{\boldsymbol{x}}$ has d blocks, which each contains $2m$ entries. The block sparsity of $\bar{\boldsymbol{x}}$ remains unchanged, s_B. Define the $\ell_{2,1}$ norm of $\bar{\boldsymbol{x}}$ with respect to the block size $2m$, i.e., $\|\bar{\boldsymbol{x}}\|_{2,1} := \sum_{q=0}^{d-1} \|\bar{\boldsymbol{x}}_q\|_2$. Then, it can be verified that $\|\bar{\boldsymbol{x}}_q\|_2 = \|\boldsymbol{x}_q\|_2$ and hence $\|\bar{\boldsymbol{x}}\|_{2,1} = \|\boldsymbol{x}\|_{2,1}$, where we recall that the latter $\ell_{2,1}$ norm is defined with respect to the block size of m. Consequently, the original complex-valued model (13) is equivalent to the optimization problem

$$\hat{\bar{\boldsymbol{x}}} = \arg\min_{\bar{\boldsymbol{x}}} \|\bar{\boldsymbol{x}}\|_{2,1},\ \text{s.t. } \boldsymbol{y}_r = \overline{\boldsymbol{\Psi}} \bar{\boldsymbol{x}}. \tag{23}$$

Since ℓ_1-norm based sparse recovery is a special case of block sparse recovery, following the same steps, we find that complex-valued ℓ_1 norm minimization is equivalent to a real-valued $\ell_{2,1}$ norm minimization with block width of 2.

We now calculate the bound on $\delta(D(\|\cdot\|_{2,1}, \boldsymbol{x}))$ for the complex-valued case, based on the real-valued representation (23). Proposition 1 assumes that entries in the measurement matrix are mutually independent Gaussian variables, while $\boldsymbol{\Psi}_r$ or $\overline{\boldsymbol{\Psi}}$ has duplicate entries, which are not independent. However, the dependence introduced by $\boldsymbol{\Psi}_r$ has little impact on its phase transition curve according to the empirical results in [24], which inspires us to apply Proposition 3, derived from Proposition 1, for the real-valued optimization problem (23). Therefore, the phase transitions of the optimization problem (23) emerge when $2n = \varphi_{2\,m}(s_B, d)$, where $\varphi_m(s, d)$ is defined in (21). We then denote by

$$\varphi_m^c(s_B, d) := \varphi_{2\,m}(s_B, d)/2 \tag{24}$$

an approximate bound on the phase transition curve. The accuracy of (24) will be verified in Section VI. It provides the location of phase transitions in complex-valued block sparse recovery, which can be applied to standard sparse recovery as well.

V. Phase Transition in FAR

In this section, we adapt the derived phase transition curves to FAR. We first present phase transition curves of FAR using block sparse recovery, followed by the counterpart using standard sparse recovery. We approximate and simplify the expressions of these curves under certain assumptions, which facilitates the calculation of these curves. In Subsection V-A and V-B, we show the results of block and standard sparse recovery, respectively. A discussion of these results is presented in Subsection V-C.

Phase transition curves of FAR are inspired by Propositions 2, 3 and (24), which however are given under Gaussian matrices. Generally, these curves are not theoretically applicable to FAR (10), because the measurement matrix in (10) is not Gaussian but highly structured. Currently, there is no theoretical evidence that Proposition 1 holds for such structured measurement matrices. However, we will show in the next section by simulations that (24) accurately indicates phase transitions in FAR.

In this section, we consider a FAR radar with the number of pulses and available frequencies being N and M, respectively. The radar illuminates K targets/clutter, of which each occupies an average of βM HRR bins within a CRR bin. Here, $\beta \in (0, 1]$ is a measure of the sizes of the targets compared with CRR bins and every CRR bin is regarded as a block of size M.

A. Block Sparse Recovery

Assuming only N_b observations out of the whole N radar echoes are available, we use (24) to identify the required N_b for exact target reconstruction with block sparse recovery. Substituting $n = N_b$, $m = M$, $d = N$ and $s_B = K$ into (24), we have

$$N_b = \frac{1}{2} \inf_{\tau \geq 0} \left\{ K(2M + \tau^2) + (N - K) \int_{\tau}^{\infty} (u - \tau)^2 \phi_{2\,M}(u) \mathrm{d}u \right\}. \tag{25}$$

The curve indicated in (25) fits the phase transitions in FAR, as numerically verified in Section VI.

The tightness of (25) makes it a powerful tool for guiding waveform design and evaluating recovery performance of FAR. For given system parameters M and N, as well as K, which means that we have some prior knowledge on the number of targets in a single CRR bin, (25) provides the minimum requirement on the number of observations to guarantee unique recovered targets. This is particularly useful when one aims to reduce the number of transmitted pulses N_b out of N, for the purposes of lowering power consumption [6], facilitating spectrum sharing between radar and communication [27], or interference rejection [26]. For a given tuple of parameters (M, N, N_b), (25) implies an equation with respect to K, the maximum number of recoverable targets, evaluating the performance of radars equipped with these parameters. This equation with respect to K can be efficiently solved by iterative methods, e.g., the bisection method [28], because N_b is a monotonic function with respect to K, as stated below.

Proposition 4: The right hand side of (25) increases monotonically with K.

Proof: See Appendix B. ■

Numerical calculation of (25) involves complicated integration operations, and when M is a slightly large number, the precise calculation of (25) is difficult. To avoid the computational burden and allow real-time calculation in practical scenarios, we approximate and simply (25) under different quantitative relations between N and K, as given in the following proposition.

Proposition 5: For large M and different orders of magnitude of $\frac{N}{K}$, N_b in (25) can be approximated by:

- i) when $\frac{N}{K} \gg \sqrt{M}$,

$$N_b \approx N_{b1} = 2MK - \frac{K}{4} + \frac{\sqrt{2}}{2} \cdot K \sqrt{(4M - 1) \log \frac{N - K}{K\sqrt{4M - 1}}}. \tag{26}$$

- ii) when $\frac{N}{K} \approx \sqrt{M}$,

$$N_b \approx N_{b2} = KM + \frac{K}{2} x_b(\tau_*), \tag{27}$$

where

$$x_b(\tau_*) = 2M - \frac{3}{4} + \frac{N}{4K} - \frac{(4M - 1)K}{N + K} + \frac{(N - K)^2}{2K\pi(N + K)} - \frac{(\sqrt{2} + 1)(N - K)}{2(N + K)} \sqrt{\frac{4M - 1}{\pi}}. \tag{28}$$

- iii) when $\frac{N}{K} \ll \sqrt{M}$,

$$N_b \approx N_{b3} = MN - \frac{(4M - 1)(N - K)^2}{4N}. \tag{29}$$

Proof: See Appendix C. ■

We will show in the next section that when $M \geq 4$, these approximations are quite accurate. Among these three relations between N and K, the case of $N/K \gg \sqrt{M}$ is of particular interest, representing that the observed target scene is relatively sparse. We will compare this curve with the counterpart of standard sparse recovery in the next subsection.

B. Standard Sparse Recovery

To compare FAR's recovery performance between non-block and block sparse recovery, i.e., (12) and (13), we also use (24) to indicate the required minimum number of radar echoes, denoted by N_s, when applying (12). In this case, the "block" size is $m = 1$, and the length of $\boldsymbol{x}$ is $md = d = MN$. The "block" sparsity becomes $s_B = \beta KM$, because each target occupies an average of βM HRR bins within a CRR bin, and thus leads to βM nonzero entries in $\boldsymbol{x}$. Substituting these variables into (24), we obtain

$$N_s = \frac{M}{2}\inf_{\tau\geq 0}\left\{\beta K(2+\tau^2)+(N-\beta K)\int_{\tau}^{\infty}(u-\tau)^2\phi_2(u)\mathrm{d}u\right\}. \tag{30}$$

The values in (25) and (30) are identical when $M = 1$ and $\beta = 1$. As a direct corollary of Proposition 4, N_s increases monotonically with β and K. This implies that the fewer HRR bins the targets occupy within a CRR bin, the better performance standard sparse recovery will have.

Similarly to Proposition 5, we have the following proposition that presents approximations to (30).

Proposition 6: When $\frac{N}{\beta K} \gg 1$ and $\frac{N}{\beta K} \approx 1$, (30) is approximated by

$$N_s \approx N_{s1} = 2\beta MK + \frac{\beta MK\tau_\star^2}{2}, \tag{31}$$

and

$$N_s \approx N_{s2} = MN - \frac{\pi M(N-\beta K)^2}{4N}, \tag{32}$$

respectively, where $\tau_\star$ is the solution of the following equation

$$\log(\tau_\star^2+1) = \log\frac{N-\beta K}{\beta K} - \frac{\tau_\star^2}{2}. \tag{33}$$

Proof: See Appendix D. ■

This proposition, like the counterpart for block sparse recovery, intuitively reveals the relationship between N_s and parameters (M, N, K, β), facilitating the comparison between block and standard sparse recovery.

C. Discussion

In the sequel, we discuss the obtained bounds when M is reasonably large and the observed target scene is relatively sparse, i.e., $N \gg K$, which occurs in many practical scenarios. Under such conditions, we adopt the approximations (26) and (31).

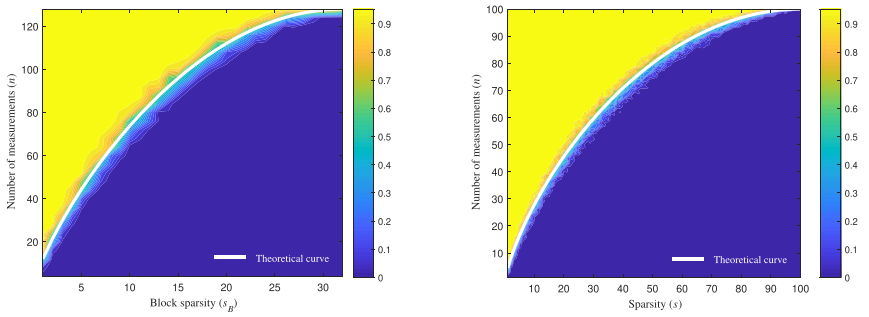
We first note that the obtained conditions that guarantee unique recovery are tighter than the previous counterparts presented in [12], [13]. The previous results, $K_b = O(\frac{N}{M\log(MN)})$ and $K_s = O(\frac{1}{\beta}\sqrt{\frac{N}{M^2\log(MN)}})$, (the subscripts denote block and standard sparse recovery, respectively), are based on coherence techniques, which lead to pessimistic bounds. To facilitate the comparison between our results and K_b, K_s, we set $N_{b1} = N$ and $N_{s1} = N$ in (26) and (31), respectively, because the intact observation models are considered in [12], [13], where all N pulses are transmitted, received and processed. Regarding (26), we have $K_{\mathrm{ib}} \approx \frac{N}{2M+\sqrt{2M}\log\frac{N}{2\sqrt{M}}}$ for intact block sparse recovery (hence the subscript 'ib'). Since in practice N is usually not extremely larger than M, we have $2M \geq \sqrt{2M}\log\frac{N}{2K\sqrt{M}}$. As a consequence, K_{ib} scales as $O(N/M)$, larger than $O(\frac{N}{M\log(MN)})$, indicating the tightness of K_{ib} over K_b. Similarly, from (31), we have $K_{\mathrm{is}} \approx \frac{1}{\beta}\frac{N}{2M+M\log N}$ for intact standard sparse recovery (hence the subscript 'is'), which is simplified as $K_{\mathrm{is}} = O(\frac{N}{\beta M\log N})$. In comparison with K_s, we find K_{is} scales larger than K_s. We will show by simulations in the next section that the approximations we derive in this section are tight for FAR, while the previous bounds [12], [13] are quite pessimistic.

The tightness enables these approximations to accurately characterize the recovery performance, and facilitates the performance comparison between block and standard sparse recovery for extended targets. In particular, from (26) and (31), we have $N_b = 2MK + O(K\sqrt{M\log\frac{N}{K\sqrt{M}}})$ and $N_s = 2\beta MK + O(\beta KM\log\frac{N}{K})$, respectively, suggesting $N_b < N_s$ for β approximate to 1 and reasonably large M. This means that for given (M, N, K), i.e., under the same system settings and sparse target scene, block sparse recovery requires fewer observations to guarantee unique recovery of extended targets, implying that block sparse recovery is generally more suitable for recovering extended targets with FAR. Meanwhile, in the case of $\beta \sim O(\frac{1}{\sqrt{M}})$, we have $N_s = O(K\sqrt{M}\log\frac{N}{K})$ less than N_b, which means that standard sparse recovery performs better when only a small portion of HRR bins are occupied in the considered CRR. In Section VI-C, we compare the results of standard and block sparse recovery with different β, which coincide with our conclusions above.

In summary, for scenes with a few extended targets block sparse recovery generally outperforms the standard one. The reason is that these scenes naturally possess block sparsity, captured well by the $\ell_{2,1}$ norm. However, in some special cases where only a small portion of HRR bins in the considered CRR bin are occupied by targets (this happens, e.g., when the extended targets have small size compared to the size of a CRR bin), indicating sparsity within a nonzero block, standard sparse recovery performs better than block sparse recovery.

In this paper, we mainly consider the cases where the carrier frequencies of transmitted pulses are assumed to be evenly distributed over the set $\mathcal{M}$. Empirical experiments show that when they are not evenly distributed, there is performance degradation. The phase transition curve in the cases of uneven distributions is left for future work.

Grid selection will also affect the performance results since the properties of the measurement matrices will vary with the grid. In this paper, we select grids according to the resolution of the radar system. However, when the continuous target



(a) Real-valued, block (b) Complex-valued, non-block

Fig. 2. Phase transitions under real and complex-valued Gaussian matrices using block and non-block sparse recovery, respectively.

parameters lie off the selected grids, the recovery performance may degrade. It is commonly believed that selecting denser grids is likely to relieve the off-the-grid problem [29]. Preliminary empirical results demonstrate that under denser grids there generally still exists a phase transition phenomenon and that the performance of CS approaches can be well described by the theoretical curves with some slight adjustments. Future investigations are required to provide a theoretical explanation of this phenomenon. Moreover, to address the off-the-grid problem, some off-grid CS-methods are proposed [29]; theoretical analyses of these off-grid methods are beyond the scope of this paper.

VI. Simulation Results

In this section, simulations are conducted to verify the theoretical curves derived for Gaussian matrices in Section IV and test their application in FAR. In Section VI-A and Section VI-B, we measure the success rates of recovering $\boldsymbol{x}$. Then in Section VI-C, we examine the relationship between the number of occupied HRR bins and the reconstruction performance. In the above three sections, we assert that $\boldsymbol{x}$ is recovered successfully when $\hat{\boldsymbol{x}}$, the estimation of $\boldsymbol{x}$, satisfies $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq 10^{-5}$. In Section VI-D, we examine the proximity of our approximations to the theoretical results. In Section VI-E, we discuss phase transitions in noisy settings, revealing the robustness of phase transitions to noise. To clarify the difference, we recall that $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ denote the Gaussian matrix and measurement matrix of FAR, respectively.

A. Phase Transitions Under Gaussian Matrices

This subsection carriers out simulation experiments to inspect the phase transitions on block sparse recovery and the phase transitions in complex-valued Gaussian matrices.

The first experiment considers real-valued block sparse recovery described in (13), where the entries of the observation matrix $\boldsymbol{\Psi} \in \mathbb{R}^{n \times md}$ obey i.i.d. $\mathcal{N}(0,1)$. The nonzero entries in $\boldsymbol{x} \in \mathbb{R}^{md}$ are 1 or -1 randomly with an identical probability $1/2$. We set $m = 4$ and $d = 32$, and vary (s_B, n) to calculate the probabilities of exact recovery. In the second simulation, we test the phase transition in complex-valued non-block sparse recovery (12), which can be solved by real-valued block sparse recovery as discussed in Subsection IV-B. Here, the entries of $\boldsymbol{\Psi} \in \mathbb{C}^{n \times d}$ have their real and imaginary parts obeying i.i.d. $\mathcal{N}(0,1)$. There are s nonzero entries in $\boldsymbol{x} \in \mathbb{C}^d$, whose phases are i.i.d. $U([0, 2\pi])$ and amplitudes equal 1. We set $d = 100$. In both experiments, 50 trials are performed on each pair (s_B, n) or (s, n) to calculate the success rates. The results for these two experiments are shown in Fig. 2 (a) and (b). The theoretical curves in (a) and (b) are computed by (21) and (24) with corresponding m and d, respectively.

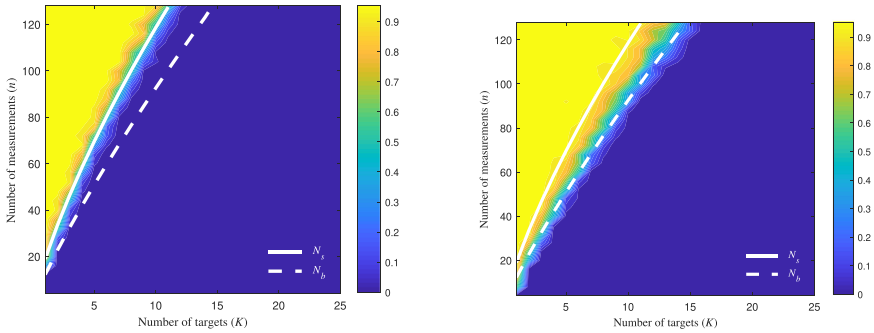
Fig. 2 shows that the theoretical curves conform to empirical phase transitions, which verifies Proposition 3 and (24).

B. Phase Transition in FAR Model

We next verify existence of phase transitions in FAR. Both standard and block sparse recovery methods are tested.

To inspect the phase transition in FAR (10), 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. Given the observations $\boldsymbol{y} = \hat{\boldsymbol{\Theta}}\boldsymbol{x}$, we use both standard (12) and block (13) sparse recovery to estimate $\boldsymbol{x}$. Recall that in standard sparse recovery, the sparsity is βKM, and we set $\beta = 1$ here. We set the parameters $M = 4$, $N = 128$, $\frac{\Delta f}{f_c} = 0.02$, and use 50 trials to calculate the success rates. The results of (12) and (13) are shown in Fig. 3 (a) and (b), respectively. The theoretical curves are calculated with corresponding M and N by (30) and (25), denoted by 'N_s' and 'N_b,' representing standard and block sparse recovery, respectively. For the sake of comparison between these two sparse recovery methods, we depict both theoretical curves in each figure of phase transition results.

From Fig. 3, we see that both theoretical curves well match their corresponding phase transition curves of FAR. Let $N_s = N_b = N$. The predicted numbers of recoverable targets under this setting are $K = 11.1$ and $K = 14.7$ for standard and block sparse recovery, respectively, whose corresponding success rates are 0.54 ($K = 11$, $n = 128$ in Fig. 3 (a)) and 0.50 ($K = 14$, $n = 128$ in Fig. 3 (b)). These rates are close to the threshold $1/2$ that divides the parameter plane into regions of success and failure, indicating that the obtained values of K are tight. However, the counterparts obtained from [12] and [13] are pessimistically $K = 0.4$ and $K = 0$, respectively. We also find that the curve of 'N_b' is generally lower than that of 'N_s,' revealing that block sparse recovery behaves better than standard sparse recovery in the tested cases.



(a) Standard sparse recovery (b) Block sparse recovery

Fig. 3. Phase transitions in FAR using standard and block sparse recovery.

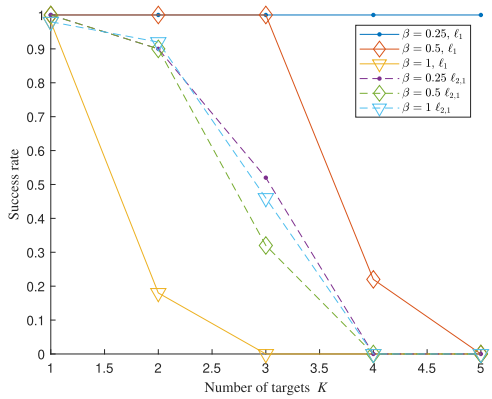


Fig. 4. Comparison of block and standard sparse recovery with different choices of β.

C. Phase Transitions and the Number of Occupied HRR Bins

In this experiment, we choose a fully observed FAR model $\boldsymbol{\Theta} \in \mathbb{C}^{N\times MN}$ as the measurement matrix, where $N = 128$, $M = 16$, $\frac{\Delta f}{f_c} = 0.02$. The phases of nonzero entries in the K block sparse vector $\boldsymbol{x}$ are i.i.d. $U([0, 2\pi])$ and the amplitudes equal 1. Given the observations $\boldsymbol{y} = \boldsymbol{\Theta x}$, we use both standard (12) and block (13) sparse recovery to estimate $\boldsymbol{x}$. Recall that in standard sparse recovery, the sparsity is βKM, and we set $\beta = 1/4, 1/2, 1$ here. For different β, we use 50 trials to calculate the success rates. The results are shown in Fig. 4.

From Fig. 4, the probabilities of exact reconstruction with block sparse recovery (labeled with $\ell_{2,1}$) are close, indicating that the phase transition of block sparse recovery is not sensitive to the value of β. In contrast, the probabilities of exact reconstruction via standard sparse recovery decrease with increasing β. When $\beta = 1$, block sparse recovery outperforms the standard one. For relatively small β (e.g., 1/2 and 1/4), standard sparse recovery performs better, because nonzero blocks of $\boldsymbol{x}$ also possess sparsity and ℓ_1 norm minimization well harnesses such sparsity.

D. Approximation of Phase Transition

In this subsection, we will show by simulations that our approximations of the phase transition curves are sufficiently close to the theoretical results, so that we can use them in practical applications to avoid time-consuming calculation.

We compute N_b and N_s versus K with (25) and (30) to represent the theoretical results under different N and M. Then we calculate the values of N_{b1}, N_{b2} and N_{s1} in (26), (27) and (31), respectively, to compare with the theoretical results. We set the parameters (N, M) to be (128,4), (128,6), (256,10) and (256,12), and the corresponding results are shown in Fig. 5(a)–(d), respectively. Since N_b and N_s cannot exceed N, we restrict their scales between 0 and N in these figures.

From Fig. 5, we see that N_{s1} well approximates N_s under all the tested scenarios. As expected in the discussions over Proposition 5, when both M and K are small, N_{b1} is closer to N_b than N_{b2}, while N_{b2} behaves better in fitting N_b when either M or K increases. Fig. 5 also shows that the larger M grows, the more block sparse recovery outperforms the standard counterpart, which is in accordance with our analysis in Section V-C.

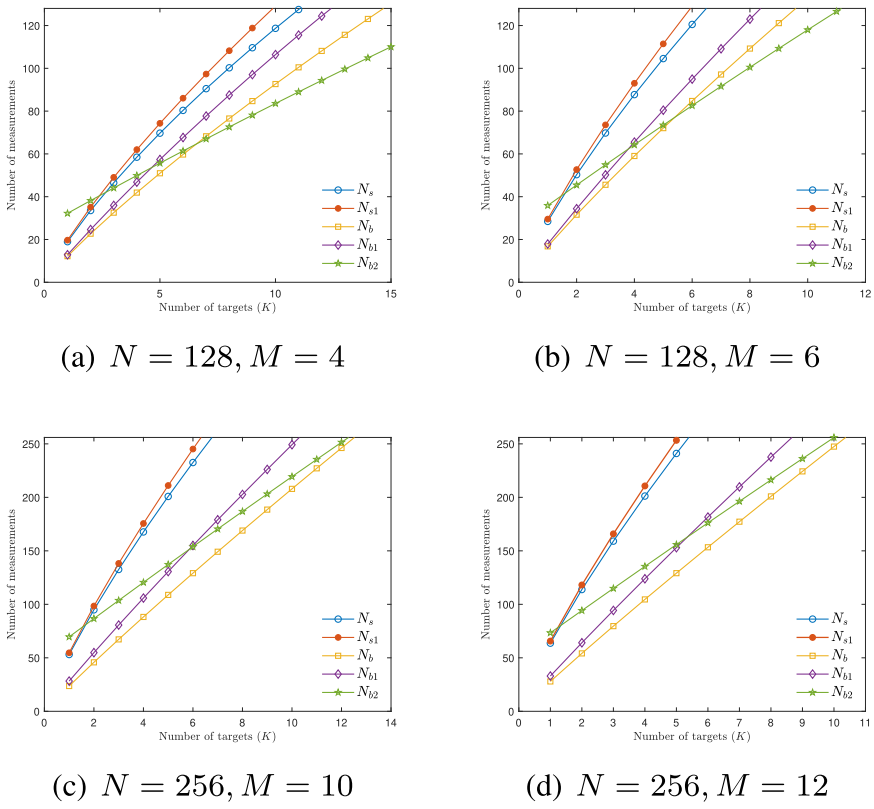
E. Phase Transition Under Noisy Conditions

In noisy cases, instead of "exact recovery," it would be more practical to introduce some "tolerance" in both the constraint condition and the evaluation of the performance of CS approaches. A noisy underdetermined system is denoted by $\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{v}$, where $\boldsymbol{v}$ denotes the noise and $\boldsymbol{x}$ is an unknown sparse/block sparse signal. This can be solved by the following optimization

$$\min_{\boldsymbol{x}} J(\boldsymbol{x}) \ \text{ s.t.} \|\boldsymbol{Ax} - \boldsymbol{y}\| \le \epsilon, \tag{34}$$

where $\epsilon \in \mathbb{R}^+$ is a constant chosen to tolerate the noise, satisfying $\|\boldsymbol{v}\| \le \epsilon$. And $J(\boldsymbol{x})$ is the cost function, which is the ℓ_1 norm or $\ell_{2,1}$ norm, representing the standard and block sparse recovery, respectively.

Then we evaluate the recovery results under two different criteria, respectively, i.e., "approximate recovery" and "robust recovery". Under noisy cases, the original signal $\boldsymbol{x}$ is regarded as "approximately recovered," if the result $\hat{\boldsymbol{x}}$ recovered from (34) satisfies $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \le C$; $\boldsymbol{x}$ is regarded as "robustly recovered," if $\hat{\boldsymbol{x}}$ satisfies $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \le C'\epsilon$ [30]. Here, C and C' are constants. The difference is whether the threshold depends on noise level



(a) $N = 128, M = 4$ (b) $N = 128, M = 6$

(c) $N = 256, M = 10$ (d) $N = 256, M = 12$

Fig. 5. Comparison between N_b, N_s, N_{b1}, N_{b2} and N_{s1}.

ϵ. We will show the phase transition phenomena and their relationships with the theoretical results in both criteria below.

For approximate recovery, consider a noisy intact FAR observation $\boldsymbol{y} = \boldsymbol{\Theta}\boldsymbol{x} + \boldsymbol{v}$, where $\boldsymbol{\Theta} \in \mathbb{C}^{N \times MN}$ is the measurement matrix of FAR with $M = 4, N = 128$ and $\frac{\delta f}{f_c} = 0.02$. The noise $\boldsymbol{v} \in \mathbb{C}^N$ obeys a complex circularly-symmetric Gaussian distribution with variance σ^2, i.e., $\boldsymbol{v} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. The unknown vector $\boldsymbol{x} \in \mathbb{C}^{MN}$ contains K nonzero blocks, each of which has block size M. The phases of entries in nonzero blocks of $\boldsymbol{x}$ are i.i.d $U([0, 2\pi])$ and the amplitudes equal 1. We use (34) to recover $\boldsymbol{x}$ by taking $J(\boldsymbol{x})$ as $\|\boldsymbol{x}\|_1$ and $\|\boldsymbol{x}\|_{2,1}$, respectively, and taking $\epsilon = \sqrt{N\sigma^2}$.

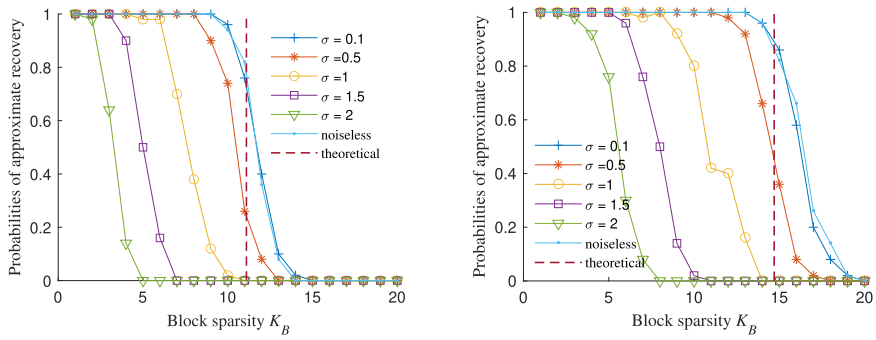
For standard and block sparse recovery, We use 50 trials to calcula te the probabilities of approximate recovery in the cases of $C = 1.5$ and $C = 2.5$, versus block sparsity K. Different noise variances σ are tested to examine the sensitivity of approximate recovery to noises.

The results are shown in Fig. 6. From the figure, we see that under different σ^2 and C, there are phase transition phenomena. Particularly, lower noise variances lead to better recovery performance, which right shift the probability curves. Increasing the tolerance C also improves the probabilities of approximate recovery in all cases. For reasonably low noise variances, the curves are close to the one under noiseless condition. We also note that the performance gap between the empirical and theoretical results is significant, which indicates that the criterion used in this figure, i.e., $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \leq C$, does not show the intrinsic connection between the performance of CS approaches in noisy cases and the theoretical phase transition curve. To better characterize the the behavior of CS methods, we turn to the second criterion, "robust recovery".

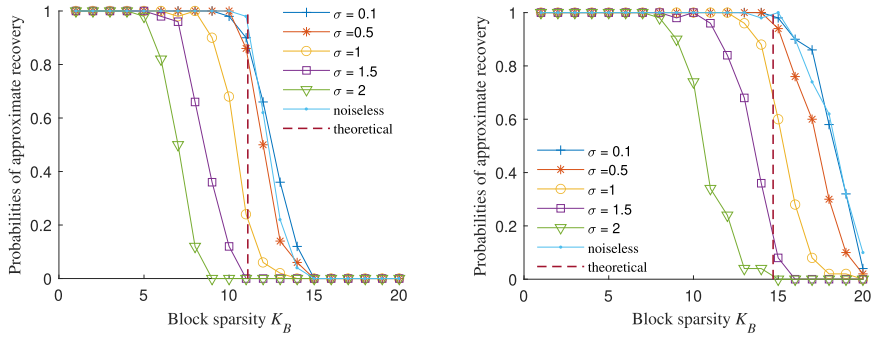
The tolerance of robust recovery, $C'\epsilon$, depends on noise level ϵ. Previous work [31] proves that a (large enough) Gaussian matrix $\boldsymbol{A}$ leading to exact recovery in noise-free cases also guarantees robust recovery in noisy cases using (34), indicating that their phase transitions show certain commonalities. To validate the phase transition of robust recovery in FAR model, we execute the same simulation as the above "approximate recovery," despite that we calculate the probabilities of robust recovery of x. For standard and block sparse recovery, the constants of robust recovery are set to $C' = 2.2/\sqrt{N}$ and $C' = 2/\sqrt{N}$, respectively.

The results are shown in Fig. 7. For ease of comparison, the probabilities of exact recovery under noise-free conditions are also depicted in Fig. 7, labeled with 'noiseless'.

From Fig. 7, in noisy cases, there is still a phase transition phenomenon of standard and block sparse recovery in the sense of probability of robust recovery. In addition, the phase transition points are close to their counterparts in noiseless cases, validating the importance of phase transition theory derived from noiseless conditions. In the results of both standard and block sparse recovery, we find that the increase of noise variance the sparsity K where the phase transition happens. The reason is that we set C' as a constant and ϵ changing with noise variance. As a result, the increase of variance makes the condition of robust recovery, i.e., $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \leq C'\epsilon$, become more tolerant, and hence increases the sparsity K where the phase transition happens.

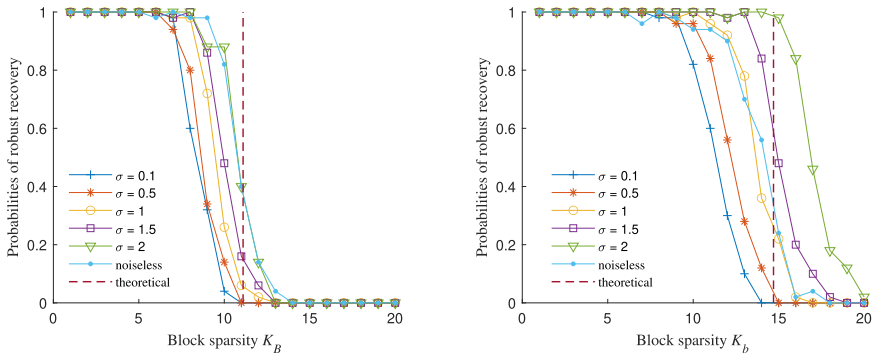


(a) Standard, $C = 1.5$ (b) Block, $C = 1.5$



(c) Standard, $C = 2.5$ (d) Block, $C = 2.5$

Fig. 6. Probabilities of approximate recovery, i.e., $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \leq C$, using standard and block sparse recovery under different variances of noise. The vertical dashed lines represent the theoretical phase transition points in the noiseless cases.



(a) Standard sparse recovery (b) Block sparse recovery

Fig. 7. Probabilities of robust recovery, i.e., $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \leq C'\epsilon$, using (a) standard and (b) block sparse recovery under different variances of noise. Particularly, the curves labeled with "noiseless" correspond to the probabilities of exact recovery, i.e., $\|\boldsymbol{x} - \hat{\boldsymbol{x}}\| \leq 10^{-5}$, in noiseless cases.

However, this does not indicate that the recovery algorithms perform better with the increase of noise variance.

The phase transition can be different with different criteria. The goal of this paper is not to analyse how the results varies according to the criterion, but to show that, with suitable criterion, the phase transition in noisy circumstance is relatively close to the phase transition of exact recovery in noiseless circumstance. Thereby, the theoretical phase transition curve is still instructive in practical scenarios.

VII. Conclusion

In this paper, standard and block sparse recovery for FAR are studied from the perspective of phase transitions. We generalize the phase transitions of standard sparse recovery under real Gaussian matrices to the cases associated with block sparse recovery and complex-valued Gaussian matrices, which numerically conform to phase transitions existing in FAR. Particularly, this paper extends [1] by providing complete theoretical proofs

and discussions on approximating phase transition curves, explicitly revealing the quantitative relationship between the required number of measurements and the numbers of radar pulses, frequencies and targets, as well as facilitating the calculation of these curves. These approximations with analytical expressions are tighter than previous results in [12], [13], and indicate that block sparse recovery requires less measurements to exactly reconstruct extended targets. Numerical results demonstrate the accuracy of the derived curves and their approximations.

APPENDIX A
PROOF OF PROPOSITION 3

In this section, we prove Proposition 3. According to Proposition 1, we first calculate the subdifferential of the $\ell_{2,1}$ norm defined in (18), and then the distance between the subdifferential and a Gaussian vector, as indicated in (19).

Regarding the subdifferential of $\ell_{2,1}$ norm of a s_B block sparse vector $\boldsymbol{x} \in \mathbb{R}^{md}$, we first reorganize the vector $\boldsymbol{x}$ for notation convenience. Without loss of generality, we assume that the support set of $\boldsymbol{x}$ is $\mathcal{B} = \{0, 1, \ldots, s_B - 1\}$, such that

$$\boldsymbol{x}^T = \left[\boldsymbol{x}_0^T, \boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_{s_B-1}^T, \boldsymbol{0}^T, \ldots, \boldsymbol{0}^T\right].$$

Here, the first s_B blocks $\boldsymbol{x}_i \in \mathbb{R}^m$, $i \in \mathcal{B}$, are nonzero. We use $x_{i,j}$ to represent the j-th element of the i-th block $\boldsymbol{x}_i$, i.e., the $(im + j)$-th entry of $\boldsymbol{x}$. Let $\mathcal{B}^c := \{s_B, s_B + 1, \ldots, d\}$ be the complementary set of $\mathcal{B}$.

We link the subdifferential of $\|\boldsymbol{x}\|_{2,1}$ and that of $\|\boldsymbol{x}_i\|_2$ by introducing the following lemma.

Lemma 1: Given two block vectors $\boldsymbol{x} = [\boldsymbol{x}_0^T, \boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_{d-1}^T]^T, \boldsymbol{s} = [\boldsymbol{s}_0^T, \boldsymbol{s}_1^T, \ldots, \boldsymbol{s}_{d-1}^T]^T \in \mathbb{R}^{md}$ with $\boldsymbol{x}_i, \boldsymbol{s}_i \in \mathbb{R}^m$, the following two statements are equivalent:

1) $\boldsymbol{s} \in \partial\|\boldsymbol{x}\|_{2,1}$,

2) $\boldsymbol{s}_i \in \partial\|\boldsymbol{x}_i\|_2,\ i = 0, 1, \ldots, d-1$.

Proof: According to (18), the definition of subdifferential, we rewrite the above statements into inequalities, respectively:

1) for $\forall \boldsymbol{y} \in \mathbb{R}^{md}$, $\|\boldsymbol{y}\|_{2,1} \geq \|\boldsymbol{x}\|_{2,1} + \langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{s}\rangle$,

2) for $\forall \boldsymbol{y}_i \in \mathbb{R}^d$, $\|\boldsymbol{y}_i\|_2 \geq \|\boldsymbol{x}_i\|_2 + \langle \boldsymbol{y}_i - \boldsymbol{x}_i, \boldsymbol{s}_i\rangle$, $i = 0, 1, \ldots, d-1$.

To prove 2) $\rightarrow$ 1), we set $\boldsymbol{y} = [\boldsymbol{y}_0^T, \boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_{d-1}^T]^T$. By summing both sides of the inequality in 2) with respect to i, we have

$$\sum_{i=0}^{d} \|\boldsymbol{y}_i\|_2 \geq \sum_{i=0}^{d} \|\boldsymbol{x}_i\|_2 + \sum_{i=0}^{d} \langle \boldsymbol{y}_i - \boldsymbol{x}_i, \boldsymbol{s}_i\rangle, \tag{35}$$

where the summation terms are equal to $\|\boldsymbol{y}\|_{2,1}$, $\|\boldsymbol{x}\|_{2,1}$ and $\langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{s}\rangle$, respectively, implying 1).

For the other direction 1) $\rightarrow$ 2), we construct $\boldsymbol{y} = [\boldsymbol{x}_0^T, \ldots, \boldsymbol{x}_{i-1}^T, \boldsymbol{y}_i^T, \boldsymbol{x}_{i+1}^T, \ldots, \boldsymbol{x}_{d-1}^T]^T$. Due to the arbitrariness of $\boldsymbol{y}$, the inequality in 1) still holds, directly yields the i-th inequality in 2) with some simple arrangement, $i = 0, 1, \ldots, d-1$.

Therefore, these two statements are equivalent. ■

Lemma 1 facilitates the calculation of $\partial\|\boldsymbol{x}\|_{2,1}$: we just need to calculate $\partial\|\boldsymbol{x}_i\|_2$. For a convex function $f(\cdot)$, if it is differentiable at a certain point $\boldsymbol{p}$, then the subdifferential of $f(\cdot)$ at $\boldsymbol{p}$ contains only one element, the differential of $f(\cdot)$ at $\boldsymbol{p}$ [32, §35]. Therefore, we discuss $\partial\|\boldsymbol{x}_i\|_2$ when $\boldsymbol{x}_i$ is a nonzero or zero block, respectively. In the former case, the function $\|\boldsymbol{x}_i\|_2$ is differentiable with respective to $\boldsymbol{x}_i$. The partial differential of $\|\boldsymbol{x}_i\|_2$ with respective to an entry $x_{i,j}$ is given by

$$\frac{x_{i,j}}{\sqrt{\sum_{k=0}^{m-1} x_{i,k}^2}}, i \in \mathcal{B}, j \in \mathcal{M}. \tag{36}$$

In the latter case, the partial differential of $\|\boldsymbol{x}_p\|_2$, $p \in \mathcal{B}^c$, does not exist, and we will calculate its subdifferential with the following lemma.

Lemma 2: Given the function $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2$, $\boldsymbol{x} \in \mathbb{R}^d$, the subdifferential of $f(\cdot)$ at $\boldsymbol{x} = \boldsymbol{0}$ is $\{\boldsymbol{s} \in \mathbb{R}^d : \|\boldsymbol{s}\|_2 \leq 1\}$.

Proof: Let $\mathcal{S} = \{\boldsymbol{s} \in \mathbb{R}^d : \|\boldsymbol{s}\|_2 \leq 1\}$. To prove $\mathcal{S} = \partial f(\boldsymbol{0})$, we need to show that $\mathcal{S} \subset \partial f(\boldsymbol{0})$ and $\partial f(\boldsymbol{0}) \subset \mathcal{S}$.

We first consider $\mathcal{S} \subset \partial f(\boldsymbol{0})$. Recall the definition of the subdifferential, $\partial f(\boldsymbol{x}) := \{\boldsymbol{s} \in \mathbb{R}^d : f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \boldsymbol{s}, \boldsymbol{y} - \boldsymbol{x}\rangle,\ \forall\ \boldsymbol{y} \in \mathbb{R}^d\}$. For $\forall \boldsymbol{s} \in \mathcal{S}$, using the Cauchy-Buniakowsky-Schwarz inequality, we have the following inequality

$$\|\boldsymbol{0}\|_2 + \langle \boldsymbol{s}, \boldsymbol{y} - \boldsymbol{0}\rangle \leq \|\boldsymbol{s}\|_2\|\boldsymbol{y}\|_2 \leq \|\boldsymbol{y}\|_2, \tag{37}$$

being true for $\forall \boldsymbol{y} \in \mathbb{R}^d$, which proves $\mathcal{S} \subset \partial f(\boldsymbol{0})$.

We then state $\partial f(\boldsymbol{0}) \subset \mathcal{S}$ by proving its contrapositive: for $\forall \boldsymbol{s} \in \mathbb{R}^d$, if $\boldsymbol{s} \notin \mathcal{S}$, then $\boldsymbol{s} \notin \partial f(\boldsymbol{0})$. For $\boldsymbol{s} \in \mathbb{R}^d$, $\|\boldsymbol{s}\|_2 > 1$, we choose $\boldsymbol{y} = \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|_2}$, and obtain the following inequality

$$f(\boldsymbol{y}) = \|\boldsymbol{y}\|_2 = 1 < \|\boldsymbol{s}\|_2 = \langle \boldsymbol{y}, \boldsymbol{s}\rangle = \langle \boldsymbol{y} - \boldsymbol{0}, \boldsymbol{s}\rangle + f(\boldsymbol{0}), \tag{38}$$

implying $\boldsymbol{s} \notin \partial f(\boldsymbol{0})$.

With the two parts above, we prove $\partial f(\boldsymbol{0}) = \mathcal{S}$. ■

Lemma 2 completes the subdifferential of $\|\boldsymbol{x}_p\|_2$ for $p \in \mathcal{B}^c$. Recall that when $i \in \mathcal{B}$ the subdifferential is given by (36).

We are now ready to derive $\partial\|\boldsymbol{x}\|_{2,1}$. Let $\boldsymbol{v} \in \mathbb{R}^{md}$ denote an element in the subdifferential of $\|\boldsymbol{x}\|_{2,1}$, i.e., $\boldsymbol{v} \in \partial\|\boldsymbol{x}\|_{2,1}$, with its i-th block denoted by $\boldsymbol{v}_i \in \mathbb{R}^m$ and $(im + j)$-th entry by $v_{i,j}$. The subdifferential $\partial\|\boldsymbol{x}\|_{2,1}$ forms a cone, given by

$$\partial\|\boldsymbol{x}\|_{2,1} = \left\{\boldsymbol{v} \in \mathbb{R}^{md} : v_{i,j} = \frac{x_{i,j}}{\sqrt{\sum_{k=0}^{m-1} x_{i,k}^2}}, \sum_{q=0}^{m-1} v_{p,q}^2 \leq 1, \right.$$
$$\left. i \in \mathcal{B}, j \in \mathcal{M}, p \in \mathcal{B}^c\right\}. \tag{39}$$

We then calculate the distance between a standard normal vector $\boldsymbol{g} \in \mathbb{R}^{md}$, $\boldsymbol{g} \sim \mathcal{N}(0, \boldsymbol{I}_{md})$, and the set $\partial\|\boldsymbol{x}\|_{2,1}$.

Let $\boldsymbol{g}_i \in \mathbb{R}^m$ and $g_{i,j}$ represent the i-th block and $(im + j)$th entry in $\boldsymbol{g}$, respectively. The distance is then calculated blockwise, given by

$$\operatorname{dist}^2\left(\boldsymbol{g}, \tau \cdot \{\boldsymbol{v}\}\right) = \sum_{i=0}^{d-1} \operatorname{dist}^2\left(\boldsymbol{g}_i, \tau \cdot \{\boldsymbol{v}_i\}\right). \tag{40}$$

For $i \in \mathcal{B}$, $\operatorname{dist}^2(\boldsymbol{g}_i, \tau \cdot \boldsymbol{v}_i) = \sum_{j=0}^{m-1}(g_{i,j} - \tau v_{i,j})^2$, because the subdifferential reduces to a single point. When $p \in \mathcal{B}^c$, we have $\{\boldsymbol{v}_p\} = \{\boldsymbol{s} \in \mathbb{R}^m : \|\boldsymbol{s}\|_2 \leq 1\}$. Therefore, the distance is

given by

$$\mathrm{dist}^2\left(\boldsymbol{g}_p u \cdot \{\boldsymbol{v}_p\}\right) = \inf_{\|\boldsymbol{v}_p\|_2 \le 1} \sum_{q=0}^{m-1} (g_{p,q} - \tau v_{p,q})^2, \quad (41)$$

which equals zero when $\|\boldsymbol{g}_p\|_2 \le \tau$ and $(\|\boldsymbol{g}_p\|_2 - \tau)^2$ otherwise, hence is expressed by $G_p(\tau) := \max((\|\boldsymbol{g}_p\|_2 - \tau)^2, 0)$. From the above discussion, we have

$$\mathrm{dist}^2\left(\boldsymbol{g}, \tau \cdot \{\boldsymbol{v}\}\right) = \sum_{i=0}^{s_B-1} \sum_{j=0}^{m-1} (g_{i,j} - \tau v_{i,j})^2 + \sum_{p=s_B}^{d-1} G_p(\tau). \quad (42)$$

We next calculate the expectation. We have $\mathbb{E}[g_{i,j}^2] = 1$, $\mathbb{E}[g_{i,j} v_{i,j}] = 0$ and $\sum_{j=0}^{m-1} v_{i,j}^2 = 1$, $i \in \mathcal{B}$, $j \in \mathcal{M}$. The expectation of the first term in (42) is rewritten as follows

$$\begin{aligned}\mathbb{E}\left[\sum_{i=0}^{s_B-1}\sum_{j=0}^{m-1}(g_{i,j} - \tau v_{i,j})^2\right] &= \mathbb{E}\left[\sum_{i=0}^{s_B-1}\sum_{j=0}^{m-1} g_{i,j}^2\right] \\ &\quad - 2\tau\mathbb{E}\left[\sum_{i=0}^{s_B-1}\sum_{j=0}^{m-1} g_{i,j} v_{i,j}\right] + \tau^2 \sum_{i=0}^{s_B-1}\sum_{j=0}^{m-1} v_{i,j}^2 \\ &= s_B\left(m + \tau^2\right). \end{aligned} \quad (43)$$

Let $u_p = \|\boldsymbol{g}_p\|_2$, which obeys the χ-distribution with m degrees of freedom. The expectation of the second term can be calculated as

$$\begin{aligned}\mathbb{E}\left[\sum_{i=s_B}^{d-1} G_p(\tau)\right] &= (d - s_B)\mathbb{E}\left[\max((u_p - \tau)^2, 0)\right] \\ &= (d - s_B)\int_\tau^\infty (u - \tau)^2 \phi_m(u)\mathrm{d}u \\ &\quad + (d - s_B)\int_0^\tau 0 \cdot \phi_m(u)\mathrm{d}u \\ &= (d - s_B)\int_\tau^\infty (u - \tau)^2 \phi_m(u)\mathrm{d}u, \end{aligned} \quad (44)$$

where $\phi_m(u)$ is the probability density function of the χ-distribution with m degrees of freedom. With the above results, the expectation is given by

$$\begin{aligned}&\mathbb{E}\left[\mathrm{dist}^2\left(\boldsymbol{g}, \tau \cdot \{\tilde{\boldsymbol{v}}\}\right)\right] \\ &\quad = s_B\left(m + \tau^2\right) + (d - s_B)\int_\tau^\infty (u - \tau)^2 \phi_m(u)\mathrm{d}u. \end{aligned} \quad (45)$$

According to Proposition 1, we obtain the upper bound on $\delta(D(\|\cdot\|_{2,1}, \boldsymbol{x}))$ as

$$\begin{aligned}\varphi_m(s_B, d) &= \inf_{\tau \ge 0} \mathbb{E}[\mathrm{dist}^2(\boldsymbol{g}, \tau \cdot \partial\|\boldsymbol{x}\|_{2,1})] \\ &= \inf_{\tau \ge 0}\{s_B(m + \tau^2) + (d - s_B)\int_\tau^\infty (u - \tau)^2 \phi_m(u)\mathrm{d}u\}, \end{aligned} \quad (46)$$

which completes the proof.

Appendix B
Proof of Proposition 4

Here, we prove the monotonicity of N_b in (25) with respective to K. For convenience, we define $H(K, \tau)$ as

$$\begin{aligned}H(K, \tau) &:= K(2M + \tau^2) \\ &\quad + (N - K)\int_\tau^\infty (u - \tau)^2 \phi_{2M}(u)\mathrm{d}u, \end{aligned} \quad (47)$$

where $\tau \ge 0$, such that

$$N_b = \frac{1}{2}\inf_{\tau \ge 0} H(K, \tau). \quad (48)$$

To reveal the monotonicity, we regard the integer K as a real number, and calculate the partial derivative of $H(K, \tau)$ with respect to K, given by

$$\frac{\partial H(K, \tau)}{\partial K} = 2M + \tau^2 - \int_\tau^\infty (u - \tau)^2 \phi_{2M}(u)\mathrm{d}u, \quad (49)$$

which is non-negative as shown below.

Since $(u - \tau)^2 \le u^2$ for $u \ge \tau \ge 0$ and $\phi_{2M}(u) \ge 0$ for $u \ge 0$, we have

$$\begin{aligned}\int_\tau^\infty (u - \tau)^2 \phi_{2M}(u)\mathrm{d}u &\le \int_\tau^\infty u^2 \phi_{2M}(u)\mathrm{d}u \\ &\le \int_0^\infty u^2 \phi_{2M}(u)\mathrm{d}u, \end{aligned} \quad (50)$$

which leads to

$$\begin{aligned}\frac{\partial H(K, \tau)}{\partial K} &\ge 2M + \tau^2 - \int_\tau^\infty u^2 \phi_{2M}(u)\mathrm{d}u \\ &\ge 2M + \tau^2 - \int_0^\infty u^2 \phi_{2M}(u)\mathrm{d}u. \end{aligned} \quad (51)$$

Recall that $\phi_{2M}(u)$ denotes the probability density function of the χ-distribution with $2M$ degrees of freedom. The integral term in (51) represents the second moment of χ-distribution, which equals the first moment of χ^2-distribution, i.e., the degrees of freedom $2M$ [33]. Therefore, (51) results in

$$\frac{\partial H(K, \tau)}{\partial K} \ge 2M + \tau^2 - 2M = \tau^2 \ge 0. \quad (52)$$

As a consequence, $H(K, \tau)$ is monotonically non-decreasing with the increase of K, which proves $H(K_1, \tau) \ge H(K_2, \tau)$, when $K_1 >= K_2$. Take the infimum of τ on both sides of the inequality, we have $N_b(K_1) >= N_b(K_2)$, completing the proof.

Appendix C
Proof of Proposition 5

To simplify N_b in (25), we first introduce

$$x_b(\tau) := \tau^2 + \frac{N - K}{K}\int_\tau^\infty (u - \tau)^2 \phi_{2M}(u)\mathrm{d}u, \quad (53)$$

such that

$$N_b = KM + \frac{K}{2}\inf_{\tau \ge 0} x_b(\tau). \quad (54)$$

In the following, we approximate $x_b(\tau)$ with a conciser form in X-A. We then seek the value of $\tau \ge 0$ that leads to the infimum

of $x_b(\tau)$ in X-B, followed by the calculation of $\inf_{\tau\geq 0} x_b(\tau)$ in X-C. Among the derivatives, some formulas corresponding to the integral over a normal distribution will be used [33], given by

$$\int_y^\infty u^3 e^{-\frac{u^2}{2}}\,\mathrm{d}u = y^2 e^{-\frac{y^2}{2}} + e^{-\frac{y^2}{2}}, \tag{55}$$

$$\int_y^\infty u^2 e^{-\frac{u^2}{2}}\,\mathrm{d}u = \sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\frac{y}{\sqrt{2}}\right) + y e^{-\frac{y^2}{2}}, \tag{56}$$

$$\int_y^\infty u e^{-\frac{u^2}{2}}\,\mathrm{d}u = e^{-\frac{y^2}{2}}, \tag{57}$$

$$\int_y^\infty e^{-\frac{u^2}{2}}\,\mathrm{d}u = \sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\frac{y}{\sqrt{2}}\right), \tag{58}$$

where $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-\eta^2 d\eta}$ is the complementary error function.

A. Approximation of $x_b(\tau)$

The approximation of $x_b(\tau)$ is based on the central limit theorem [33], indicating that the χ-distribution probability density function $\phi_{2\,M}(u)$ can be well approximated by a probability density function of a normal distribution when M is reasonably large. Particularly,

$$\phi_{2\,M}(u) \approx \frac{1}{\sigma_M\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u-\mu_M}{\sigma_M}\right)^2}, \tag{59}$$

where the mean and variance are denoted by μ_M and σ_M^2, respectively, given by

$$\mu_M = \frac{\sqrt{2}\Gamma(M+1/2)}{\Gamma(M)}, \tag{60}$$

$$\sigma_M^2 = 2\,M - \mu_M^2. \tag{61}$$

We note that for sufficiently large M, the mean and variance in (60) and (61) lead to

$$\mu_M^2 \approx 2M - 1/2, \tag{62}$$

$$\sigma_M^2 \approx 1/2, \tag{63}$$

respectively [34]. In this case, $\mu_M/\sigma_M = O(\sqrt{M})$.

By substituting (59) into (53), we approximate the integral term in the right hand side of (53) by

$$\int_\tau^\infty (u-\tau)^2 \phi_{2\,M}(u)\mathrm{d}u \approx \frac{1}{\sigma_M\sqrt{2\pi}}\int_\tau^\infty (u-\tau)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u. \tag{64}$$

Now,

$$\int_\tau^\infty (u-\tau)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u = \int_\tau^\infty (u-\mu_M+\mu_M-\tau)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u, \tag{65}$$

which can be expanded into a summation of three terms

$$\int_\tau^\infty (u-\mu_M)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u + 2(\mu_M-\tau)\int_\tau^\infty (u-\mu_M) e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u + \int_\tau^\infty (\mu_M-\tau)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u. \tag{66}$$

We next calculate these terms by using formulas (56)–(58) individually. The first term in (66) can be rewritten as

$$\begin{aligned}&\int_\tau^\infty (u-\mu_M)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u \\ &= \sigma_M^2\int_\tau^\infty \frac{(u-\mu_M)^2}{\sigma_M^2} e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u \\ &\overset{(a)}{=} \sigma_M^3\cdot\left(\sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\frac{\tau-\mu_M}{\sqrt{2}\sigma_M}\right) + \frac{\tau-\mu_M}{\sigma_M} e^{-\frac{(\tau-\mu_M)^2}{2\sigma_M^2}}\right),\end{aligned} \tag{67}$$

where (a) is a consequence of (56). The integral in the second term of (66) can be simplified by (57), implying

$$\begin{aligned}\int_\tau^\infty (u-\mu_M) e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u &= \sigma_M\int_\tau^\infty \frac{u-\mu_M}{\sigma_M} e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u \\ &= \sigma_M\cdot\sigma_M e^{-\frac{(\tau-\mu_M)^2}{2\sigma_M^2}}.\end{aligned} \tag{68}$$

With (58), we rewrite the integral of the third term in (66) as

$$\int_\tau^\infty e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u = \sigma_M\sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\frac{\tau-\mu_M}{\sqrt{2}\sigma_M}\right). \tag{69}$$

Substituting (67)–(69) into (66) yields

$$\begin{aligned}&\int_\tau^\infty (u-\tau)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u \\ &= \left(\sigma_M^3 + (\mu_M-\tau)^2\sigma_M\right)\sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\frac{\tau-\mu_M}{\sqrt{2}\sigma_M}\right) \\ &\quad + \left(\sigma_M^2(\tau-\mu_M) + 2(\mu_M-\tau)\sigma_M^2\right) e^{-\frac{(\tau-\mu_M)^2}{2\sigma_M^2}} \\ &= \left(\sigma_M^3 + (\mu_M-\tau)^2\sigma_M\right)\sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\frac{\tau-\mu_M}{\sqrt{2}\sigma_M}\right) \\ &\quad + (\mu_M-\tau)\sigma_M^2 e^{-\frac{(\tau-\mu_M)^2}{2\sigma_M^2}}.\end{aligned} \tag{70}$$

Plugging this integral into (53), yields

$$\begin{aligned}x_b(\tau) &= \tau^2 + \frac{N-K}{K\sigma_M\sqrt{2\pi}}\int_\tau^\infty (u-\tau)^2 e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\,\mathrm{d}u \\ &= \tau^2 + \frac{N-K}{2\,K}\left(\sigma_M^2 + (\mu_M-\tau)^2\right)\mathrm{erfc}\left(\frac{\tau-\mu_M}{\sqrt{2}\sigma_M}\right) \\ &\quad + \frac{N-K}{K\sqrt{2\pi}}(\mu_M-\tau)\sigma_M e^{-\frac{(\tau-\mu_M)^2}{2\sigma_M^2}}.\end{aligned} \tag{71}$$

B. Minimizer of $x_b(\tau)$

We denote by τ_* the minimizer of $x_b(\tau)$. Taking partial derivatives over both sides of (53) and letting $\frac{\partial x_b(\tau)}{\partial \tau}=0$, we have

$$\tau_* = \frac{N-K}{K}\int_{\tau_*}^{\infty}(u-\tau_*)\phi_{2\,M}(u)\mathrm{d}u. \tag{72}$$

In this subsection, we simplify the integral function (72), and then substitute the result into (71) to facilitate the calculation of $x_b(\tau_*)$.

Using (59), we approximate the integral in (72) by

$$\begin{aligned}&\int_{\tau_*}^{\infty}(u-\tau_*)e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\mathrm{d}u\\ &=\int_{\tau_*}^{\infty}(u-\mu_M+\mu_M-\tau_*)e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\mathrm{d}u\\ &=\sigma_M^2 e^{-\frac{(\tau_*-\mu_M)^2}{2\sigma_M^2}}+(\mu_M-\tau_*)\sigma_M\sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\frac{\tau_*-\mu_M}{\sqrt{2}\sigma_M}\right),\end{aligned} \tag{73}$$

from (68) and (69). Hence, we rewrite (72) as

$$\begin{aligned}\tau_* &= \frac{N-K}{K\sigma_M\sqrt{2\pi}}\int_{\tau_*}^{\infty}(u-\tau_*)e^{-\frac{(u-\mu_M)^2}{2\sigma_M^2}}\mathrm{d}u\\ &=\frac{N-K}{2\,K}(\mu_M-\tau_*)\mathrm{erfc}\left(\frac{\tau_*-\mu_M}{\sqrt{2}\sigma_M}\right)\\ &\quad+\frac{N-K}{K\sqrt{2\pi}}\sigma_M e^{-\frac{(\tau_*-\mu_M)^2}{2\sigma_M^2}}.\end{aligned} \tag{74}$$

Denote the solution to (74) by $\tau_*=\mu_M-\alpha\sigma_M$, where α is unknown indicating the normalized difference between τ_* and μ_M. Substituting $\tau_*=\mu_M-\alpha\sigma_M$ into (74), we obtain

$$\mu_M-\alpha\sigma_M=\frac{N-K}{2\,K}\alpha\sigma_M\mathrm{erfc}\left(-\frac{\alpha}{\sqrt{2}}\right)+\frac{N-K}{K\sqrt{2\pi}}\sigma_M e^{-\frac{\alpha^2}{2}}. \tag{75}$$

After some arrangement, (75) leads to

$$\begin{aligned}&\alpha-\left(\frac{\alpha}{2}\mathrm{erfc}\left(-\frac{\alpha}{\sqrt{2}}\right)+\frac{1}{\sqrt{2\pi}}e^{-\frac{\alpha^2}{2}}\right)\\ &\quad=\frac{\mu_M}{\sigma_M}-\frac{N}{K}\left(\frac{\alpha}{2}\mathrm{erfc}\left(-\frac{\alpha}{\sqrt{2}}\right)+\frac{1}{\sqrt{2\pi}}e^{-\frac{\alpha^2}{2}}\right),\end{aligned} \tag{76}$$

which can be rewritten as

$$\frac{\alpha}{2}\mathrm{erfc}\left(-\frac{\alpha}{\sqrt{2}}\right)+\frac{1}{\sqrt{2\pi}}e^{-\frac{\alpha^2}{2}}=\left(\frac{\mu_M}{\sigma_M}-\alpha\right)\cdot\frac{K}{N-K}. \tag{77}$$

Note that the function on the left hand side is monotonically increasing with respect to α while the one on the right hand side is monotonically decreasing. Therefore, there is only one solution to (77). We will substitute the values of α, i.e., $\alpha\gg 1$, $\alpha\approx 0$ and $\alpha\ll -1$, into (76), respectively, in order to check whether (76) holds and reveal the dependency between α and (K,N,M). We resort to series expansions of $\mathrm{erfc}(\cdot)$ at $\pm\infty$, respectively, given by

$$\frac{\alpha}{2}\mathrm{erfc}\left(\frac{-\alpha}{\sqrt{2}}\right)\Big|_{\alpha=+\infty}\approx\alpha-\frac{1}{\sqrt{2\pi}}e^{-\frac{\alpha^2}{2}}, \tag{78}$$

$$\frac{\alpha}{2}\mathrm{erfc}\left(\frac{-\alpha}{\sqrt{2}}\right)\Big|_{\alpha=-\infty}\approx-\frac{1}{\sqrt{2\pi}}e^{-\frac{\alpha^2}{2}}(1-1/\alpha^2). \tag{79}$$

i) We first consider the case $\alpha\ll -1$. Substitute (79) into (76), leads to

$$\alpha\approx\frac{\mu_M}{\sigma_M}-\frac{N-K}{K}\frac{e^{-\frac{\alpha^2}{2}}}{\sqrt{2\pi}\alpha^2}, \tag{80}$$

which indicates that satisfying $\alpha\ll -1$ requires $\frac{N}{K}\gg\frac{\mu_M}{\sigma_M}$. After arrangement of (80), we take logarithm on both sides of approximation, resulting in

$$\log\frac{N-K}{K}-\frac{\alpha^2}{2}-\log\sqrt{2\pi}\alpha^2\approx\log\left(\frac{\mu_M}{\sigma_M}-\alpha\right). \tag{81}$$

This implies

$$\begin{aligned}\alpha&\approx-\sqrt{2}\sqrt{\log\frac{N-K}{K}-\log\left(\frac{\mu_M}{\sigma_M}-\alpha\right)-\log\sqrt{2\pi}\alpha^2}\\ &\approx-\sqrt{2}\sqrt{\log\frac{(N-K)\sigma_M}{K\mu_M}}.\end{aligned} \tag{82}$$

Since in practice, neither N or M would be extremely large, μ_M/σ_M is comparable or larger than $\log N/K$, and therefore $\alpha=-O(\log\frac{N\sigma_M}{K\mu_M})$ from (82).

ii) In the second case $\alpha\approx 0$, which requires $\frac{N}{K}\approx\frac{\mu_M}{\sigma_M}$ such that (77) may hold. Applying Taylor expansion at $\alpha=0$, we approximate (77) with

$$\frac{1}{\sqrt{2\pi}}+\frac{\alpha}{2}\approx\left(\frac{\mu_M}{\sigma_M}-\alpha\right)\frac{K}{N-K}, \tag{83}$$

implying

$$\alpha\approx\frac{2(N-K)}{N+K}\left(\frac{\mu_M}{\sigma_M}\frac{K}{N-K}-\frac{1}{\sqrt{2\pi}}\right), \tag{84}$$

or simply $\alpha=O(\frac{\mu_M K}{\sigma_M N})$.

iii) We finally consider the third case $\alpha\gg 1$. Substituting (78) into (76) implies

$$\alpha\approx\frac{K\mu_M}{N\sigma_M}\gg 1, \tag{85}$$

which requires that $\frac{N}{K}\ll\frac{\mu_M}{\sigma_M}$.

Summarizing the three cases above, we find that: i) When $\frac{N}{K}\gg\frac{\mu_M}{\sigma_M}$, we have $\alpha\ll -1$ and (82); ii) When $\frac{N}{K}\approx\frac{\mu_M}{\sigma_M}$, we have $\alpha\approx 0$ and (84); iii) When $\frac{N}{K}\ll\frac{\mu_M}{\sigma_M}$, we have (85). We note that these three cases are generally complete, representing three kinds of relationship between the relative sparsity (N/K) and the number of available frequencies $(\mu_M/\sigma_M\approx 2\sqrt{M})$. With the obtained α, implying the miminizer τ_*, we next calculate the limit inferior $x_b(\tau_*)$.

C. The Infimum $\inf_{\tau\geq 0} x_b(\tau)$

Comparing $x_b(\tau)$ in (71) and τ_* in (74), we find the right hand side of (74) also appears in (71). We replace this term in (71) by τ_*, so that (71) becomes

$$x_b(\tau_*) = \tau_*^2 + \frac{N-K}{2\,K}\sigma_M^2 \mathrm{erfc}\left(\frac{\tau_*-\mu_M}{\sqrt{2}\sigma_M}\right) + (\mu_M - \tau_*)\tau_* = \mu_M\tau_* + \frac{N-K}{2\,K}\sigma_M^2 \mathrm{erfc}\left(\frac{\tau_*-\mu_M}{\sqrt{2}\sigma_M}\right). \quad (86)$$

Substituting $\tau_* = \mu_M - \alpha\sigma_M$, we rewrite (86) as

$$x_b(\tau_*) = \mu_M(\mu_M - \alpha\sigma_M) + \frac{N-K}{2\,K}\sigma_M^2 \mathrm{erfc}\left(-\frac{\alpha}{\sqrt{2}}\right). \quad (87)$$

For the three cases considered below (84), we calculate the limit inferior $x_b(\tau_*)$ with the obtained α, respectively.

1) In the first case when $\frac{N}{K} \gg \frac{\mu_M}{\sigma_M}$, we have $\alpha \ll -1$. Using Taylor expansion, we approximate (87) by

$$x_b(\tau_*) \approx \mu_M(\mu_M - \alpha\sigma_M) - \frac{N-K}{2\,K}\sigma_M^2 \frac{\sqrt{2}e^{-\frac{\alpha^2}{2}}}{\sqrt{\pi}\alpha}, \quad (88)$$

which can be further simplified by replacing the exponent term according to (80), given by

$$x_b(\tau_*) \approx \mu_M(\mu_M - \alpha\sigma_M) - \frac{N-K}{2\,K}\sigma_M^2 \frac{\sqrt{2}e^{-\frac{\alpha^2}{2}}}{\sqrt{\pi}\alpha} = \mu_M(\mu_M - \alpha\sigma_M) + \alpha\sigma_M^2\left(\alpha - \frac{\mu_M}{\sigma_M}\right) = (\mu_M - \alpha\sigma_M)^2. \quad (89)$$

Plugging (62), (63) and (82) into (89), we have

$$x_b(\tau_*) \approx \mu_M^2 - 2\alpha\mu_M\sigma_M \approx 2M - \frac{1}{2} + \sqrt{2}\sqrt{(4M-1)\log\frac{(N-K)}{K\sqrt{4M-1}}}. \quad (90)$$

Thus, the final N_b is

$$N_b \approx N_{b1} = 2MK - \frac{K}{4} + \frac{\sqrt{2}K}{2}\sqrt{(4M-1)\log\frac{(N-K)}{K\sqrt{4M-1}}}. \quad (91)$$

2) The second case corresponds to $\frac{N}{K} \approx \frac{\mu_M}{\sigma_M}$, leading to $\alpha \approx 0$. Expanding the $\mathrm{erfc}(\cdot)$ term at $\alpha = 0$, we approximate (87) by

$$x_b(\tau_*) \approx \mu_M(\mu_M - \alpha\sigma_M) + \frac{N-K}{2\,K}\sigma_M^2\left(1 - \sqrt{\frac{2}{\pi}}\alpha\right). \quad (92)$$

Substituting (84), (62) and (63) into (92) yields

$$x_b(\tau_*) = 2M - \frac{3}{4} + \frac{N}{4\,K} - \frac{(4M-1)K}{N+K} + \frac{(N-K)^2}{2K\pi(N+K)} - \frac{(\sqrt{2}+1)(N-K)}{2(N+K)}\sqrt{\frac{4M-1}{\pi}}. \quad (93)$$

Plugging the above result into (54), we have

$$N_b \approx N_{b2} = KM + \frac{K}{2}x_b(\tau_*). \quad (94)$$

3) In the third case when $\frac{N}{K} \ll \frac{\mu_M}{\sigma_M}$, we substitute (85) into (87), and obtain

$$x_b(\tau_*) = \frac{N-K}{N}\mu_M^2 + \frac{N-K}{2\,K}\sigma_M^2 \mathrm{erfc}\left(-\frac{K\mu_M}{\sqrt{2}N\sigma_M}\right). \quad (95)$$

Plugging (62) and (63) into (95), we have the approximation

$$x_b(\tau_*) \approx \frac{N-K}{N}\left(2M - \frac{1}{2}\right) + \frac{N-K}{2\,K} = 2M - 1 - \frac{K}{N}\left(2M - \frac{1}{2}\right) + \frac{N}{2\,K}. \quad (96)$$

With (96), we rewrite (54) approximately as

$$N_b \approx N_{b3} = KM + \frac{K}{2}\cdot\left(2M - 1 - \frac{K}{N}\left(2M - \frac{1}{2}\right) + \frac{N}{2\,K}\right) = \left(2M - \frac{1}{2}\right)K - \frac{(4M-1)K^2}{4\,N} + \frac{N}{4} = MN - \frac{(4M-1)(N-K)^2}{4\,N}. \quad (97)$$

Summarizing the three cases above, we have that i) $\frac{N}{K} \ll \frac{\mu_M}{\sigma_M}$, ii) $\frac{N}{K} \approx \frac{\mu_M}{\sigma_M}$, or iii) $\frac{N}{K} \gg \frac{\mu_M}{\sigma_M}$, N_b is approximated by (91), (94) or (97), respectively, completing the proof of Proposition 5.

Appendix D
Proof of Proposition 6

For notation purposes, we use $x_s(\tau)$ to represent the term inside the limit inferior operation of (30) as

$$x_s(\tau) := \tau^2 + \frac{N-\beta K}{\beta K}\int_\tau^\infty (u-\tau)^2\phi_2(u)\mathrm{d}u \quad (98)$$

$$= \tau^2 + \frac{N-\beta K}{\beta K}\int_\tau^\infty (u-\tau)^2\, u e^{-u^2/2}\mathrm{d}u \overset{(a)}{=} \tau^2 + \frac{N-\beta K}{\beta K}\left[2e^{-\tau^2/2} - \sqrt{2\pi}\tau\mathrm{erfc}(\tau/\sqrt{2})\right], \quad (99)$$

where (a) comes from (55), (56) and (57). Then, N_s in (30) is given by

$$N_s = \beta KM + \frac{\beta KM}{2}\inf_{\tau\geq 0} x_s(\tau). \quad (100)$$

Taking partial derivatives over both sides of (98) and letting $\frac{\partial x_s(\tau)}{\partial \tau} = 0$, we find the minimizer that leads to the limit inferior of $x_s(\tau)$

$$\tau_\star = \frac{N-\beta K}{\beta K}\int_{\tau_\star}^\infty (u-\tau_\star)\phi_2(u)\mathrm{d}u = \frac{N-\beta K}{\beta K}\int_{\tau_\star}^\infty (u-\tau_\star)u e^{-u^2/2}\mathrm{d}u \overset{(a)}{=} \frac{N-\beta K}{\beta K}\cdot\sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(\tau_\star/\sqrt{2}\right), \quad (101)$$

where (a) holds according to (56) and (57). Similarly to the technique used in X, we first approximately solve (101), and then calculate $x_s(\tau_\star)$.

The value $\tau_\star$ relies on N and βK. Since $N \geq \beta K$ (otherwise the unique recovery of the βK extended targets is not possible), we consider two cases i) $\frac{N}{\beta K} \approx 1$ and ii) $\frac{N}{\beta K} \gg 1$, representing the less sparse and relatively sparse cases, respectively.

i) In the first case, $\frac{N}{\beta K} \approx 1$, it is deduced from (101) that $\tau_\star$ takes values around 0. Approximating $\mathrm{erfc}(\tau/\sqrt{2})$ at $\tau = 0$ with first order Taylor expansion, we rewrite (101) as

$$\tau_\star \approx \frac{N-\beta K}{\beta K} \cdot \left(\sqrt{\pi/2} - \tau_\star\right), \tag{102}$$

which implies

$$\tau_\star \approx \frac{N-\beta K}{N} \cdot \sqrt{\pi/2}. \tag{103}$$

ii) In the second case, when $\frac{N}{\beta K} \gg 1$, we have that $\tau_\star$ is also sufficiently large $\tau_\star \gg 1$. Therefore, we expand the $\mathrm{erfc}(\cdot)$ function at $+\infty$, and rewrite (101) as

$$\begin{aligned}\tau_\star &\approx \frac{N-\beta K}{\beta K} \cdot \sqrt{\frac{\pi}{2}} \cdot \sqrt{\frac{2}{\pi}} e^{-\tau_\star^2/2} \left(\frac{1}{\tau_\star} - \frac{1}{\tau_\star^3}\right) \\ &\approx \frac{N-\beta K}{\beta K} e^{-\tau_\star^2/2} \frac{\tau_\star}{\tau_\star^2+1},\end{aligned} \tag{104}$$

implying

$$\log(\tau_\star^2+1) \approx \log\frac{N-\beta K}{\beta K} - \frac{\tau_\star^2}{2}. \tag{105}$$

This yields $\tau_\star = O(\sqrt{\log(N/\beta K)})$.

We then calculate $x_s(\tau_\star)$ with the substitution of $\tau_\star$. Note that the erfc term in (99) can be replaced by linear term according to (101), which simplifies (99) into

$$x_s(\tau_\star) = -{\tau_\star}^2 + \frac{N-\beta K}{\beta K} \cdot 2e^{-\tau_\star^2/2}. \tag{106}$$

We analyze the results in both cases i) $\frac{N}{\beta K} \gg 1$ and ii) $\frac{N}{\beta K} \approx 1$, respectively.

1) when $\frac{N}{\beta K} \gg 1$, we have that $\tau_\star$ is sufficiently large. Replacing the exponent term in (106) with a quadratic term according to (104), we simplify (106) into

$$x_s(\tau_\star) \approx \tau_\star^2 + 2. \tag{107}$$

Considering (105), we substitute (107) into (100), yielding

$$N_s \approx N_{s1} = 2M\beta K + \frac{M\beta K \tau_\star^2}{2}, \tag{108}$$

where $\tau_\star^2$ can be calculated from (105).

2) When $\frac{N}{\beta K} \approx 1$, we have $\tau_\star \approx 0$, thus we approximate $e^{-\tau_\star^2/2}$ at $\tau_\star = 0$, i.e., $e^{-\tau_\star^2/2}|_{\tau_\star=0} \approx 1 - \frac{\tau_\star^2}{2}$, and rewrite (106) as

$$x_s(\tau_\star) \approx -\frac{N\tau_\star^2}{\beta K} + \frac{2(N-\beta K)}{\beta K}. \tag{109}$$

Substituting (103) into (109) yields

$$x_s(\tau_\star) \approx -\frac{\pi(N-\beta K)^2}{2N\beta K} + \frac{2(N-\beta K)}{\beta K}. \tag{110}$$

Together with (100), this implies

$$N_s \approx N_{s2} = MN - \frac{\pi M(N-\beta K)^2}{4\,N}. \tag{111}$$

Combining the two cases above, we have approximations of N_s given by (108) or (111), respectively when i) $\frac{N}{\beta K} \gg 1$ or ii) $\frac{N}{\beta K} \approx 1$ holds, completing the proof of Proposition 6.

REFERENCES

[1] Y. Li, T. Huang, X. Xu, Y. Liu, and Y. C. Eldar, "Phase transition in frequency agile radar using compressed sensing," in *Proc. IEEE Radar Conf.*, 2020, pp. 1–6.

[2] S. R. Axelsson, "Analysis of random step frequency radar and comparison with experiments," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 890–904, Apr. 2007.

[3] J. Yang, J. Thompson, X. Huang, T. Jin, and Z. Zhou, "Random-frequency SAR imaging based on compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 983–994, Feb. 2013.

[4] L. Wang, T. Huang, and Y. Liu, "Phase compensation and image autofocusing for randomized stepped frequency ISAR," *IEEE Sensors J.*, vol. 19, no. 10, pp. 3784–3796, May 2019.

[5] Z. Liu, X. Wei, and X. Li, "Decoupled ISAR imaging using RSFW based on twice compressed sensing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 4, pp. 3195–3211, Oct. 2014.

[6] D. Ma, N. Shlezinger, T. Huang, Y. Liu, and Y. C. Eldar, "Joint radar-communication strategies for autonomous vehicles: Combining two key automotive technologies," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 85–97, Jul. 2020, 07 2020.

[7] D. Cohen, K. V. Mishra, and Y. C. Eldar, "Spectrum sharing radar: Coexistence via xampling," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 3, pp. 1279–1296, Jun. 2018.

[8] S. Liu, Y. Cao, T. S. Yeo, W. Wu, and Y. Liu, "Adaptive clutter suppression in randomized stepped-frequency radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 57, no. 2, pp. 1317–1333, Apr. 2021.

[9] T. Huang, N. Shlezinger, X. Xu, Y. Liu, and Y. C. Eldar, "Majorcom: A dual-function radar communication system using index modulation," *IEEE Trans. Signal Process.*, vol. 68, pp. 3423–3438, 2020.

[10] Y. Liu, H. Meng, G. Li, and X. Wang, "Range-velocity estimation of multiple targets in randomised stepped-frequency radar," *Electron. Lett.*, vol. 44, no. 17, pp. 1032–1034, 2008.

[11] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[12] 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.

[13] 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.

[14] K. Gerlach and M. J. Steiner, "Adaptive detection of range distributed targets," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1844–1851, Jul. 1999.

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

[16] 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.

[17] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.

[18] L. Carin, D. Liu, and B. Guo, "Coherence, compressive sensing, and random sensor arrays," *IEEE Antennas Propag. Mag.*, vol. 53, no. 4, pp. 28–39, Aug. 2011.

[19] 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: A J. IMA*, vol. 3, no. 3, pp. 224–294, 2014.

[20] S. Foucart and H. Rauhut, "Sparse recovery with random matrices," in *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer. 2013, pp. 271–310.

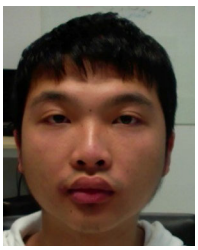
[21] D. L. Donoho, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," *Discrete Comput. Geometry*, vol. 35, no. 4, pp. 617–652, 2006.

[22] M. Stojnic, "Various thresholds for ℓ_1-optimization in compressed sensing," 2009, *arXiv:0907.3666*.

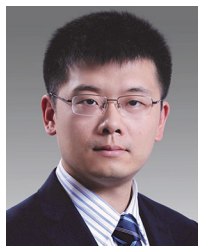
[23] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.
[24] Z. Yang, C. Zhang, and L. Xie, "On phase transition of compressed sensing in the complex domain," *IEEE Signal Process. Lett.*, vol. 19, no. 1, pp. 47–50, Jan. 2012.
[25] 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.
[26] T. Huang, N. Shlezinger, X. Xu, D. Ma, Y. Liu, and Y. C. Eldar, "Multi-carrier agile phased array radar," *IEEE Trans. Signal Process.*, vol. 68, pp. 5706–5721, 2020.
[27] D. Ma *et al.*, "Spatial modulation for joint radar-communications systems: Design, analysis, and hardware prototype," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2283–2298, Mar. 2021.
[28] A. Eiger, K. Sikorski, and F. Stenger, "A bisection method for systems of nonlinear equations," *ACM Trans. Math. Softw.*, vol. 10, no. 4, pp. 367–377, 1984.
[29] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
[30] J. Liu, J. Jin, and Y. Gu, "Robustness of sparse recovery via f-minimization: A topological viewpoint," *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 3996–4014, Jul. 2015.
[31] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6920–6941, Oct. 2011.
[32] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton University Press, 2015. https://doi.org/10.1515/9781400873173.
[33] K. L. Chung, *A. Course in Probability Theory*. New York, NY, USA: Academic Press, 2001.
[34] T. Burić and N. Elezović, "Bernoulli polynomials and asymptotic expansions of the quotient of gamma functions," *J. Comput. Appl. Math.*, vol. 235, no. 11, pp. 3315–3331, 2011.

Yuhan Li received the B.E. degree in electronic engineering from Tsinghua University, Beijing, China, in 2020. 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 Tsinghua University, Beijing, China, in 2014. From 2014 to 2017, he was a Radar Researcher with the Aviation Industry Corporation of China. Since July 2017, he is working with Intelligence Sensing Lab, Department of Electronic Engineering, Tsinghua University, as an Assistant Professor. His current research interests include signal processing, compressed sensing, and joint radar communications system design.



Xingyu Xu (Graduate Student Member, IEEE) received the B.E. degree in 2019 in electronic engineering from Tsinghua University, Beijing, China, where he is currently working toward the Ph.D. degree in electronic engineering. His research interests include learning theory, stochastic processes, and their application in signal processing.



Yimin Liu (Member, IEEE) received the B.S. and Ph.D. degrees (both with Hons.) in electronic engineering from Tsinghua University, Beijing, China, in 2004 and 2009, respectively. In 2004, he was with the Intelligence Sensing Lab, Department of Electronic Engineering, Tsinghua University. He is currently an Associate Professor with Tsinghua University, where his field of activity is 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.

Lei Wang received the B.S. degree in electronic information science and technology from Xidian University, Xi'an, China, in 2014 and the Ph.D degree in electronics engineering from Tsinghua University, Beijing, China, in 2020. He is currently a Postdoctoral Researcher with the Intelligence Sensing Lab, Department of Electronic Engineering, Tsinghua University. His current research interests include signal processing, compressive sensing, and radar 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, 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 Broad Institute, and an Adjunct Professor with Duke University, Durham, NC, USA, and was a Visiting Professor with Stanford. She is also the author of the book *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. She is a Member of the Israel Academy of Sciences and Humanities (elected 2017) and a EURASIP Fellow.

She was the recipient of many awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award (2013), the IEEE/AESS Fred Nathanson Memorial Radar Award (2014), and the IEEE Kiyo Tomiyasu Award (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 Murieland David Jacknow Award for Excellence in Teaching, and the Technions Award for Excellence in Teaching (two times). She was the recipient of 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 and the Israel Committee for Higher Education. She is the Editor-in-Chief of the *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 was an Associate Editor for the 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 Co-Chair and Technical Co-Chair of several international conferences and workshops.