

Compressed sensing radar detectors under the row-orthogonal design model: a statistical mechanics perspective

Siqi Na, Tianyao Huang, Yimin Liu, Takashi Takahashi,
Yoshiyuki Kabashima, and Xiqin Wang

June 30, 2022

Abstract

Compressed sensing (CS) model of complex-valued data can represent the signal recovery process of a large amount types of radar systems, especially when the measurement matrix is row-orthogonal. Based on debiased least absolute shrinkage and selection operator (LASSO), detection problem under Gaussian random design model, i.e. the elements of measurement matrix are drawn from Gaussian distribution, is studied by literature. However, we find that these approaches are not suitable for row-orthogonal measurement matrices. In view of statistical mechanics approaches, we provide derivations of more accurate test statistics and thresholds (or p-values) under the row-orthogonal design model, and theoretically analyze the detection performance of the present detector. Such detector can analytically provide the threshold according to given false alarm rate, which is not possible with the conventional CS detector, and the detection performance is proved to be better than that of the traditional LASSO detector. Comparing with other debiased LASSO based detectors, simulation results indicate that the proposed approach can achieve more accurate probability of false alarm when the measurement matrix is row-orthogonal, leading to better detection performance under Neyman-Pearson principle.

Key words- Compressed sensing, radar detection, LASSO, row-orthogonal matrix, replica method, statistical mechanics

Contents

1	Introduction	2
2	Design and analysis of the debiased LASSO detector	6
2.1	Debiased LASSO detector	6
2.1.1	Analytical expression of false alarm probability	7
2.1.2	Better detection performance	10
2.2	Comparison with existing debiased LASSO detectors	10

3	Derivation of the debiased LASSO estimator and the estimation of its variance	12
3.1	Main results	12
3.2	Proof of Claim 3.1	14
3.3	Proof of Claim 3.2	16
3.4	Proof of Claim 3.5	18
4	Numerical Experiments	20
4.1	Settings	20
4.2	Gaussianity of \mathbf{w} in the case of row-orthogonal matrix design	20
4.3	Accuracy of estimating σ_w^2	22
4.4	Detection performance of debiased LASSO detector	24
5	Conclusion	25
A	Proof of Theorem 2.3	25
B	Proof of Theorem 2.7	27
C	Proof of Claim 3.6	29
C.1	Derivation of the limit $N \rightarrow \infty$	30
C.1.1	The derivation of $\mathcal{T}(\mathbf{k})$	32
C.1.2	The derivation of $\mathcal{S}(\mathbf{k}, \mathbf{x}_0)$	34
C.2	Final result	35
D	Derivation of averages with respect to row-orthogonal matrices	38
E	Derivation of χ through linear response argument	40

1 Introduction

CS model of complex-valued data assumes a scenario of recovering an N -dimensional vector $\mathbf{x}_0 \in \mathbb{C}^N$ from a M -dimensional vector $\mathbf{y} \in \mathbb{C}^M$, given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \boldsymbol{\xi}, \quad (1)$$

where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is the measurement matrix (or the sensing matrix) and $\boldsymbol{\xi} \in \mathbb{C}^M$ is the complex additive white Gaussian noise (AWGN) with i.i.d. components $\xi_i \sim \mathcal{CN}(0, \sigma^2)$. Basically, for the CS model, M is less than N , and we regard $\gamma = M/N$ as *compression rate*. The original signal \mathbf{x}_0 is a sparse vector containing only $k = 2\rho N$ non-zero entries, where 2ρ ($0 \leq 2\rho \leq 1$) is referred to as the *signal density*.

In many practical applications, the signal processing can be modeled as above, while the sensing matrix has a specific structure. Particularly, in this paper, we focus on radar applications, in which \mathbf{y} , \mathbf{x}_0 and \mathbf{A} refers to the sampled received signal, radar observation scene and observation (or steering) matrix, respectively. Generally, in the

radar application scenarios, the number of non-zero entries in \mathbf{x}_0 , which represents the intensity of scattering points, is small enough to satisfy $k \ll N$. The indices of the non-zero entries indicate the “position” of the scattering points (or the targets), such as range, azimuth, radial velocity, etc.

The observation matrix \mathbf{A} incorporates the geometry of the observation scene and the design of the transmitting waveform. In the present paper, we are concerned with the case where \mathbf{A} is row-orthogonal matrix. That is, $\mathbf{A}_i \mathbf{A}_j^H = 0$ for $i \neq j$ and $\mathbf{A}_i \mathbf{A}_i^H = 1$, in which \mathbf{A}_i represents the i -th column of matrix \mathbf{A} . A large amount types of radar transmitting waveform utilize a row-orthogonal steering matrix, and we list several below for illustration:

- 1) The partial observation problem of a pulse Doppler radar system, such as working in complex electromagnetic environments [1], leads to a result of partial Fourier steering matrix, which is apparently row-orthogonal.
- 2) The steering matrix of frequency agile radar system is similarly row-orthogonal [2, 3], which possesses a variety of merits such as good electronic counter-countermeasures (ECCM) performance, low hardware system cost and convenience of spectrum sharing.
- 3) Sub-Nyquist radar systems [4, 5] realize the observation and compression on several dimensions, such as temporal domain, spatial domain and spectral domain. The steering matrix of sub-Nyquist radar, which is proved to be the Kronecker product of several partial Fourier matrices (depends on the number of dimensions of the observation scene) in [5], satisfies the row-orthogonal property.

For modern radar systems, it is crucial to detect targets element-wisely. As previously mentioned, the indices of the non-zero entries in \mathbf{x}_0 represent the information of the targets, and judging whether each entry in \mathbf{x}_0 is non-zero can inform us about the existence of the targets and their location. We summarize such requirement as solving the following hypothesis testing problems:

$$\begin{cases} \mathcal{H}_{0,i} : x_{0,i} = 0, \\ \mathcal{H}_{1,i} : x_{0,i} \neq 0, \end{cases} \quad (2)$$

for $i = 1, 2, \dots, N$ and designing thresholds for these tests. Note that it is different from the task to detect whether there exists any target in the whole observed scene. The latter is generally more simple, which can be expressed as

$$\begin{cases} \mathcal{H}_0 : \mathbf{y} = \boldsymbol{\xi}, \\ \mathcal{H}_1 : \mathbf{y} = \mathbf{A}\mathbf{x}_0 + \boldsymbol{\xi}, \end{cases} \quad (3)$$

and can be easily solved by conventional approaches such as generalized likelihood ratio test (GLRT) or Wald test. While since the linear model (1) is underdetermined, the maximum likelihood estimation (MLE) of $x_{0,i}$ is unavailable. Consequently, one cannot directly apply conventional detectors to solve detection problem (2).

Therefore, it is natural that we estimate \mathbf{x}_0 by CS approaches and complete the tests on the basis of estimation, in which LASSO [6] is a frequently used technique. The complex-valued LASSO estimator $\hat{\mathbf{x}}^{\text{LASSO}}$ is given by

$$\hat{\mathbf{x}}^{\text{LASSO}} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}, \quad (4)$$

where

$$\|\mathbf{x}\|_1 \doteq \sum_{i=1}^N \sqrt{(\text{Re}(x_i))^2 + (\text{Im}(x_i))^2}. \quad (5)$$

As a convex optimization problem, LASSO can be solved by standard techniques such as interior point and homotopy methods [7, 8]. There are also plenty of threshold class fast algorithms such as approximate message passing (AMP), iterative shrinkage-thresholding algorithm (ISTA), etc. Besides the computationally feasible feature, the nature of $\hat{\mathbf{x}}^{\text{LASSO}}$ has been studied in a large literature, which mainly focus on: the prediction error $\|\mathbf{A}(\hat{\mathbf{x}}^{\text{LASSO}} - \mathbf{x}_0)\|_2^2/M$ [9], the estimation error $\|\hat{\mathbf{x}}^{\text{LASSO}} - \mathbf{x}_0\|_q$ with $q \in \{1, 2\}$ [10–12] and variable selection (or support estimation) of \mathbf{x}_0 [13–15], denoted by the support set $S_0 = \{i \leq N : x_{0,i} \neq 0\}$ such that $\mathbb{P}(\hat{S} \neq S_0)$ is bounded. However, the above study is not enough to solve detection problems (2): we still cannot get their p-values (or the threshold of given false alarm rate, which is extremely important in radar applications). Naturally, one would wonder about the distribution of \hat{x}_i^{LASSO} in purpose of designing the detector.

A certain linear transformation of the LASSO estimator, called *debiased LASSO*, is given by

$$\hat{\mathbf{x}}^{\text{d}} = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda} \mathbf{A}^H (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}}), \quad (6)$$

where $\Lambda > 0$ is the *debiased coefficient* computed from known variables and contains information about the structure of sensing matrix \mathbf{A} . Asymptotic analysis of the LASSO solution based on AMP algorithm was developed in [16] and [17], which proves that the empirical distribution of the difference

$$\mathbf{w} \doteq \hat{\mathbf{x}}^{\text{d}} - \mathbf{x}_0, \quad (7)$$

converges weakly to Gaussian distribution under Gaussian random design model (the entries of observation matrix \mathbf{A} are i.i.d. drawn from Gaussian distribution). This conclusion derives two studies for the design of detectors for solving the detection problem (2) based on debiased LASSO: [18] for the real-valued CS model and [19] for the complex-valued one. While all the studies above restrict the measurement matrix to be Gaussian, work in [20] based on statistical mechanics methods as well provides different ways for obtaining $\hat{\mathbf{x}}^{\text{d}}$. The methodologies used allows the asymptotic analysis of LASSO solutions and the construction of debiased LASSO estimator for multiple real-valued observation matrix ensembles such as Gaussian, row-orthogonal, random discrete cosine transformation (DCT). We aim to derive debiased LASSO in complex-valued form based on [20], especially under row-orthogonal design model, and analyze the detection performance of the resulting constructed detector.

The main contributions of this paper are twofold. First, we present a general debiased LASSO detector framework and analyze its detection performance. We summarize the existing research on detection problems, including [18] and complex approximate message passing (CAMP) based [19], and find that the test statistics used are the same: all debiased LASSO. Therefore, we construct a general detector framework based on debiased LASSO, and analyze its detection performance. We prove that the detection performance of debiased LASSO detector is better than that of traditional LASSO detector under the Neyman-Pearson principle, and its detection threshold can be analytically calculated by the given false alarm rate, thus the detection rate can be further quantified. In contrast, most CS methods do not have closed-form solutions, and the distribution is not available. Consequently, the threshold cannot be given by the false alarm rate either. Therefore, we believe that the debiased lasso detector can completely replace the traditional CS detector when solving the detection problem (2).

Second, we extend the results of [20] to the complex domain to enhance their applicability to engineering applications. Some methods in statistical mechanics are used in reference [20], which sacrifice part of the mathematical rigor to obtain more attractive results: the debiased LASSO estimator for some non-Gaussian sensing matrices can be derived, and a more accurate estimation medium for the variance σ_w^2 of \mathbf{w} (as defined in (7)) can be provided. These results make it possible to implement debiased LASSO detector under row-orthogonal design model. Therefore, for engineering applications, we extend these derivation processes to the complex domain. We find that the debiased coefficient under the row-orthogonal assumption is different from the results in [18] and [19], and the correctness of the results obtained by our method is verified by simulation experiments. Numerical results also verify that our method can estimate the variance σ_w^2 of \mathbf{w} more accurately than [18] and [19]. This leads to a more accurate threshold for the designed debiased LASSO detector and therefore to a better reaching of a given false alarm rate. Such conclusion is also verified by simulation results.

The organization of the present paper is as follows. In Section 2, we provide our results on the design of the detector for compressed sensing radar system and the analysis on its detection performance. The derivation of the test statistic and the threshold of the presented detector is elaborated in Section 3. Section 4 furnishes some numerical validation of the previous asymptotic analysis. We conclude the paper in Section 5.

Throughout the paper, we use a , \mathbf{a} and \mathbf{A} as a number, a vector and a matrix, respectively. For a set S , $\#S$ denotes its cardinality. Denote by $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ for the real and imaginary parts of a complex-valued component, respectively. Function $\delta(\cdot)$ is the Dirac's delta function and $\Theta(x)$ is Heaviside's step function. The operators $(\cdot)^T$, $(\cdot)^*$ and $(\cdot)^H$ represent the transpose, conjugate and conjugate transpose of a component, respectively. Denote by $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ the Gaussian density and $\Phi(x) = \int_{-\infty}^x \varphi(u)du$ the Gaussian distribution.

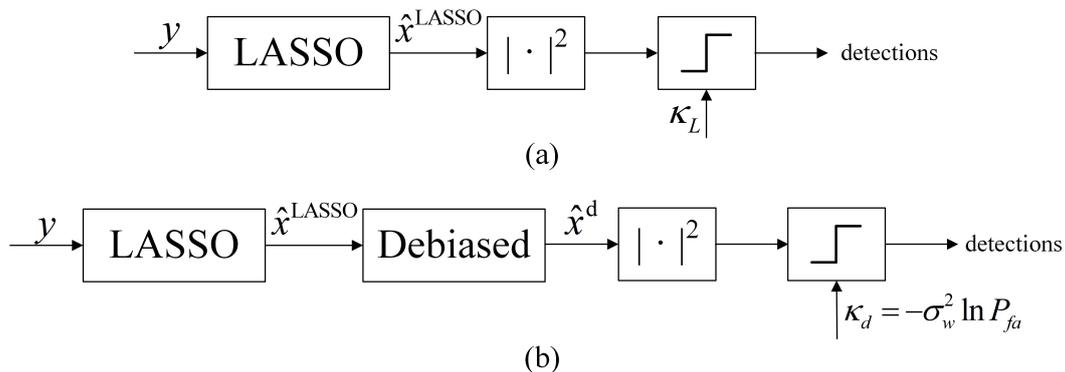


Figure 1: Frameworks of two detectors for detection problem (2): (a) LASSO detector; (b) Debiased LASSO detector.

2 Design and analysis of the debiased LASSO detector

In this section, we first introduce the debiased LASSO detector. The advantage of the debiased LASSO detector is elaborated mainly by comparing with the LASSO detector in Section 2.1. While both the proposed detector and the ones in [18–20] are debiased LASSO detectors, the difference is explained in Section 2.2. Particularly, this paper inherits some derivation from [20], and we will briefly describe the contribution of this work over [20] in Section 2.2.

2.1 Debiased LASSO detector

With regard to detection problem (2), we here define the following performance metrics. Denote by

$$\varphi_i = \begin{cases} 1, & \text{if detector reject } H_{0,i}; \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

Define the false alarm probability P_{fa} as

$$P_{fa} = \lim_{N \rightarrow \infty} \frac{1}{N-k} \sum_{i \in S^c} \varphi_i, \quad (9)$$

and the detection probability P_d as

$$P_d = \lim_{N \rightarrow \infty} \frac{1}{k} \sum_{i \in S} \varphi_i, \quad (10)$$

where $S = \{i : x_{0,i} \neq 0\}$ is the support set of \mathbf{x}_0 with $\#S = k$ and $S^c = \{1, \dots, N\} \setminus S$.

The LASSO detector, which is a natural idea, possesses the structure shown in Fig. 1 (a), in which the test statistic is provided by LASSO estimator. Indeed, other traditional CS radar detectors have the similar form with the LASSO detector, which contain the

compressed sensing module and a test of judging whether the amplitude is greater than a threshold.

On the contrary, the debiased LASSO detector for compressed sensing radar has the structure shown in Fig. 1 (b). The test statistic is given by debiased estimator $\hat{\mathbf{x}}^d$ and the threshold κ_d can fix the probability of false alarm. In the present paper, we propose the debiased LASSO detector for complex-valued row-orthogonal observation matrix called Complex Row-Orthogonal Debiased detector (CROD), whose procedure is shown in Algorithm 1. The detailed derivation of the involved parameters will be presented later in Section 3.

We declare here that the benefits of the debiased LASSO detector are mainly twofold. First, the relationship between the threshold κ_d and the false alarm rate P_{fa} can be computed analytically, as will be detailed later in Section 2.1.1. This is usually not possible for traditional CS detectors, because the distribution of the solution obtained by CS methods is unavailable. Second, its detection performance is better than that of the LASSO detector, which will be proved in Section 2.1.2.

2.1.1 Analytical expression of false alarm probability

The most significant advantage of the debiased LASSO detector over the conventional CS detector is that the analytical relationship between the threshold κ_d and the false alarm rate P_{fa} can be obtained. Controlling false alarm rate is particularly important in many radar applications due to resource allocation and other reasons. Asymptotic analysis shows that as $N \rightarrow \infty$, if Λ is suitably chosen, \hat{x}_i^d , the i -th entry of $\hat{\mathbf{x}}^d$, approximately follows Gaussian distribution with mean $x_{0,i}$, where $x_{0,i}$ is the i -th entry of \mathbf{x}_0 . For the convenience of readers, we here restate the definition and conclusion in [16] (, which is similar to that in [17],) to describe this result more precisely.

Definition 2.1 ([16]). *For a given $(\gamma, 2\rho) \in [0, 1]^2$, a sequence of instances $\{\mathbf{x}_0(N), \boldsymbol{\xi}(N), \mathbf{A}(N)\}_{N \in \mathbb{N}}$ indexed by N is said to be a converging sequence of Gaussian design model if the empirical distribution of the entries $\mathbf{x}_0(N) \in \mathbb{R}^N$ converges weakly to a probability measure p_X with bounded second moment, the empirical distribution of the entries $\boldsymbol{\xi}(N) \in \mathbb{R}^M$ ($M = \gamma N$) converges weakly to a probability measure p_ξ with bounded second moment and the elements of $\mathbf{A}(N) \in \mathbb{R}^{M \times N}$ are i.i.d. drawn from a Gaussian distribution.*

Lemma 2.2 ([16]). *Let $\{\mathbf{x}_0(N), \boldsymbol{\xi}(N), \mathbf{A}(N)\}_{N \in \mathbb{N}}$ be a converging sequence of Gaussian design model. The empirical law of $\mathbf{w}(N) = \hat{\mathbf{x}}^d(N) - \mathbf{x}_0(N)$ converges to a zero-mean Gaussian distribution almost surely as $N \rightarrow \infty$ for a specific Λ .*

In the present paper, we further prove that such Λ is unique.

Theorem 2.3. *Let $\{\mathbf{x}_0(N), \boldsymbol{\xi}(N), \mathbf{A}(N)\}_{N \in \mathbb{N}}$ be a converging sequence of Gaussian design model. Let $\Lambda_1 \in \mathbb{R}$ and $\hat{\mathbf{x}}_1^d = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda_1} \mathbf{A}^H(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}})$, such that the empirical law of $\mathbf{w}_1(N) = \hat{\mathbf{x}}_1^d(N) - \mathbf{x}_0(N)$ converges to a zero-mean Gaussian distribution almost surely as $N \rightarrow \infty$. Then for all $\Lambda_2 \neq \Lambda_1$ and $\hat{\mathbf{x}}_2^d = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda_2} \mathbf{A}^H(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}})$, the empirical law of $\mathbf{w}_2(N) = \hat{\mathbf{x}}_2^d(N) - \mathbf{x}_0(N)$ does not converge to a Gaussian distribution as $N \rightarrow \infty$.*

Algorithm 1 CROD

Input: \mathbf{y} , \mathbf{A} , regularization parameter λ , probability of false alarm P_{fa} , variance of noise σ^2

Output: debiased LASSO estimator $\hat{\mathbf{x}}^d$, threshold κ_d

1: Let

$$\hat{\mathbf{x}}^{\text{LASSO}} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right\}.$$

2: The debiased LASSO estimator is obtained from

$$\hat{\mathbf{x}}^d = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda_{\text{CROM}}} \mathbf{A}^H (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}}), \quad (11)$$

with Λ_{ROM} and ρ_{CA} given by

$$\begin{aligned} \Lambda_{\text{CROM}} &= \frac{\gamma - \rho_{\text{CA}}}{1 - \rho_{\text{CA}}}, \\ \rho_{\text{CA}} &= \frac{1}{2N} \sum_{i=1}^N \left[\left(2 - \frac{\lambda}{\Lambda_{\text{CROM}} |\hat{x}_i^{\text{LASSO}}| + \lambda} \right) \cdot \Theta(|\hat{x}_i^{\text{LASSO}}|) \right]. \end{aligned} \quad (12)$$

3: The threshold κ_d is given by

$$\kappa_d = -\sigma_w^2 \ln P_{fa},$$

where

$$\chi = \frac{\rho_{\text{CA}}(1 - \rho_{\text{CA}})}{\gamma - \rho_{\text{CA}}}, \quad (13)$$

$$G'(-\chi; \mathbf{J}) = \frac{1 + \chi - \sqrt{(\chi + 1)^2 - 4\gamma\chi}}{2\chi}, \quad (14)$$

$$G''(-\chi; \mathbf{J}) = \frac{2\gamma\chi - \chi - 1 + \sqrt{(\chi + 1)^2 - 4\gamma\chi}}{2\chi^2 \sqrt{(\chi + 1)^2 - 4\gamma\chi}}, \quad (15)$$

$$\overline{\text{RSS}} = \frac{1}{M} \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}}\|_2^2, \quad (16)$$

$$\hat{\chi} = \frac{\gamma G'''(-\chi; \mathbf{J})}{2G'(-\chi; \mathbf{J}) - 2G''(-\chi; \mathbf{J})\chi} \overline{\text{RSS}} + \frac{-G'''(-\chi; \mathbf{J})\gamma + (G'(-\chi; \mathbf{J}))^2}{2G'(-\chi; \mathbf{J}) - 2G''(-\chi; \mathbf{J})\chi} \sigma^2, \quad (17)$$

$$\sigma_w^2 = \frac{2\hat{\chi}}{\Lambda_{\text{CROM}}^2}. \quad (18)$$

Proof. The proof is given in Appendix A. \square

In addition, we claim that such uniqueness is not limited to Gaussian sensing matrix, but other matrix manifolds also have the same conclusion such as row-orthogonal matrices. The following definition of a converging sequence of row-orthogonal design model is required.

Definition 2.4. For a given $(\gamma, 2\rho) \in [0, 1]^2$, a sequence of instances $\{\mathbf{x}_0(N), \boldsymbol{\xi}(N), \mathbf{A}(N)\}_{N \in \mathbb{N}}$ indexed by N is said to be a converging sequence of row-orthogonal design model if the empirical distribution of the entries $\mathbf{x}_0(N) \in \mathbb{R}^N$ converges weakly to a probability measure p_X with bounded second moment, the empirical distribution of the entries $\boldsymbol{\xi}(N) \in \mathbb{R}^M$ ($M = \gamma N$) converges weakly to a probability measure p_ξ with bounded second moment and $\mathbf{A}(N) \in \mathbb{R}^{M \times N}$ is randomly drawn from all the row-orthogonal matrices with a size of $M \times N$.

Corollary 2.5. Let $\{\mathbf{x}_0(N), \boldsymbol{\xi}(N), \mathbf{A}(N)\}_{N \in \mathbb{N}}$ be a converging sequence of row-orthogonal design model. Let $\Lambda_1 > 0$ and $\hat{\mathbf{x}}_1^d = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda_1} \mathbf{A}^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}})$, such that the empirical law of $\mathbf{w}_1(N) = \hat{\mathbf{x}}_1^d(N) - \mathbf{x}_0(N)$ converges to a zero-mean Gaussian distribution almost surely as $N \rightarrow \infty$. Then for all $\Lambda_2 \neq \Lambda_1$ and $\hat{\mathbf{x}}_2^d = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda_2} \mathbf{A}^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}})$, the empirical law of $\mathbf{w}_2(N) = \hat{\mathbf{x}}_2^d(N) - \mathbf{x}_0(N)$ does not converge to a Gaussian distribution as $N \rightarrow \infty$.

The proof of the corollary is the same as Theorem 2.3.

Denote the sample variance of $\mathbf{w}(N)$ by σ_w^2 , it is natural to get the following analytical relationship between the probability of false alarm and the threshold κ_d of the detector.

Theorem 2.6. The false alarm probability P_{fa} of the debiased LASSO detector satisfies:

$$\kappa_d = -\sigma_w^2 \ln P_{fa}, \quad (19)$$

where the test is

$$\varphi_i = \begin{cases} 1, & |\hat{x}_i^d|^2 > \kappa_d; \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

Proof. For $i \in S^c$, which means that $x_{0,i} = 0$, the empirical law of $\{\hat{x}_i^d\}$ converges to $\mathcal{CN}(0, \sigma_w^2)$ as $N \rightarrow \infty$. Therefore, the empirical distribution of $\{|\hat{x}_i^d|^2\}$ converges to exponential distribution of rate parameter $1/\sigma_w^2$, leading to:

$$P_{fa} = \lim_{N \rightarrow \infty} \frac{1}{N - k} \sum_{i \in S^c} \varphi_i = \exp\left(-\frac{\kappa_d}{\sigma_w^2}\right), \quad (21)$$

which proves (19). \square

From Theorem 2.6, we conclude that the threshold κ_d of the debiased LASSO detector can be analytically calculated by the false alarm probability. Such mission cannot be achieved by traditional CS detector for the reason that almost all of the solutions

Table 1: Comparison of several detectors.

approach	Λ	σ_w^2	Gaussian	row-orthogonal	complex
SDL-test [18]	(22)	(23)	✓	×	×
CAMP [19]	(24)	(25)	✓	×	✓
ROD [20]	(26)	(13) - (18)	✓	✓	×
CROD	(27)	(13) - (18)	✓	✓	✓

obtained by CS methods does not have a closed form, neither the distribution of the solutions. In the present paper, we will provide an asymptotic analysis from a statistical mechanics perspective and the process of calculating the coefficients Λ with variance σ_w^2 in Section 3, which can be adapted to the row-orthogonal matrix design of \mathbf{A} mentioned for radar applications.

2.1.2 Better detection performance

At first glance, the debiased LASSO estimator destroys the sparsity brought by the LASSO results, but we will next prove theoretically that treating $|\hat{x}_i^d|$ as a test statistic compared to $|\hat{x}_i^{\text{LASSO}}|$ will improve the detection performance of the detector.

Theorem 2.7. *For the same false alarm probability P_{fa} , the detection probability of debiased LASSO detector $P_{d,1}$ is not less than that of LASSO detector $P_{d,2}$.*

Proof. See Appendix B for the proof. □

Theorem 2.7 suggests that applying such non-sparse solution in the detector instead lead to better detection performance.

2.2 Comparison with existing debiased LASSO detectors

In this subsection, we compare the proposed detector CROD with SDL-test [18], CAMP [19], and the Row-Orthogonal Debiased detector (ROD) constructed by the conclusions obtained from [20]. Due to the frameworks of all the debiased LASSO detectors are the same, we mainly list the differences of the debiased coefficient Λ and the approaches to estimate the variance σ_w^2 .

1. SDL-test [18] suggests that

$$\Lambda_G = \gamma - \rho_a, \tag{22}$$

where $\rho_a = \#\{i|\hat{x}_i^{\text{LASSO}} \neq 0\}/N$ denotes the *active component density* and

$$\hat{\sigma}_w = \frac{\sqrt{\gamma}}{\Phi^{-1}(0.75)(\gamma - \rho_a)} \text{median}(|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}}|), \tag{23}$$

for real-valued Gaussian random sensing matrix.

2. CAMP [19] provides

$$\Lambda_{\text{CG}} = \gamma - \rho_{\text{CA}}, \quad (24)$$

where ρ_{CA} given by (12) denotes the *active component density* and

$$\hat{\sigma}_w = \frac{1}{\sqrt{\ln 2}} \text{median} \left(\left| \hat{\mathbf{x}}^{\text{d}} \right| \right). \quad (25)$$

for complex-valued Gaussian random sensing matrix.

3. ROD from [20] claims that

$$\Lambda_{\text{ROM}} = \frac{\gamma - \rho_a}{1 - \rho_a}, \quad (26)$$

and σ_w^2 is given by exchanging all the ρ_{CA} in (13) - (18) to ρ_a for real-valued row-orthogonal sensing matrix.

4. The proposed CROD claims that

$$\Lambda_{\text{CROM}} = \frac{\gamma - \rho_{\text{CA}}}{1 - \rho_{\text{CA}}}, \quad (27)$$

and σ_w^2 is given by (13) - (18) for complex-valued row-orthogonal sensing matrix.

Moreover, the methodologies in [20] can obtain the same result as (22) under Gaussian random matrix design. Our work also gives the same debiased coefficient as (24) for complex-valued Gaussian sensing matrix in Corollary 3.4. Based on the derivation in Section 3, we believe that Λ is related to the asymptotic eigenvalue distribution $\rho_{\mathbf{J}}(s)$ of $\mathbf{J} = \mathbf{A}^T \mathbf{A}$ (or $\mathbf{J} = \mathbf{A}^H \mathbf{A}$), which is referred to Claim 3.2. For a Gaussian matrix, $\rho_{\mathbf{J}}(s)$ can be calculated by (36), which yields (24). For other sensing matrices, the proposed method is also applicable as $\rho_{\mathbf{J}}(s)$ is obtainable. We summarize the results of the comparison in Table 1.

Recall that we have proved the uniqueness of the debiased coefficient Λ . When the steering matrix \mathbf{A} is row-orthogonal, the correctness of the debiased coefficient Λ_{CROM} given in the present paper will be verified by numerical result in Section 4. Besides the construction of the debiased LASSO estimator, the approach to estimate σ_w^2 in the present paper is different from CAMP and SDL-test. There is also numerical result of the accuracy of the three estimation mediums presented in Section 4.

The derivation, including the construction of the debiased LASSO estimator and the estimation of the variance, in Section 3 is primarily based on [20]. However, since the signal model in radar systems are composed of complex vectors or matrices, our next derivations are in complex form, which differs from both the process and the results of [20]. We will verify in our simulation experiments that the direct use of the real-valued version results is incorrect. This is because the complex LASSO differs from the real one with regard to the computation of the ℓ_1 -norm of \mathbf{x} in the regularization term.

3 Derivation of the debiased LASSO estimator and the estimation of its variance

In Algorithm 1, we simply present the results of the debiased LASSO estimator and estimation of its variance σ_w^2 . In this section, we will provide the detailed derivation of these results, based on some statistical mechanics approaches. Led by [21], more and more work apply statistical mechanics approaches for theoretical analysis in information theory and communications theory. Although a mathematically rigorous justification of replica method is still in-progress, it has been proved extensively successful in very difficult problems and applied to derive a number of captivating results, see [22–25]. In addition, we refer to the results in this Section as claims.

Particularly, the asymptotic analysis of the distribution of LASSO solutions will first be introduced, which tells us that a central limit obeying Gaussian distribution with a mean of \mathbf{x}_0 exists. Then, we aim to get the close-form expression of the central limit, which will be regarded as the debiased LASSO estimator. At the end of this section, the procedure for estimating the variance of the debiased estimator will be presented.

Throughout the analysis, we will generally examine the typical reconstruction performance in the limit $N, M \rightarrow \infty$, but keeping the compression rate γ constant. The assumption of large system limit $N \rightarrow \infty$ will be omitted in the following part to avoid wording. We list our main results in Section 3.1. The proof details are referred to Section 3.2, 3.3 and 3.4.

3.1 Main results

Our first result provides the asymptotic distribution of the LASSO solution $\hat{\mathbf{x}}^{\text{LASSO}}$.

Claim 3.1. *When the sensing matrix \mathbf{A} is randomly drawn from row-orthogonal matrices, the asymptotic distribution of the LASSO solution $\hat{\mathbf{x}}^{\text{LASSO}}$ can be inferred as follow:*

$$\hat{x}_i^{\text{LASSO}} = \text{ST}_{\lambda, \hat{Q}}(h_i) \doteq \frac{h_i}{|h_i|} \cdot \frac{|h_i| - \lambda}{\hat{Q}} \Theta(|h_i| - \lambda), \quad (28)$$

where

$$h_i = \hat{m}x_{0,i} + \sqrt{2\hat{\chi}}z_i, \quad (29)$$

in which $z_i \sim \mathcal{CN}(0, 1)$ are *i.i.d.* standard complex variables and \hat{m} , \hat{Q} , $\hat{\chi}$ are positive real numbers.

Such conclusion reminds us that the debiased LASSO estimator can be easily obtained if h_i is available. We next introduce the construction of debiased LASSO $\hat{\mathbf{x}}^{\text{d}}$ for certain matrix ensembles through deriving h_i .

Claim 3.2. *Suppose the Gram matrix $\mathbf{J} = \mathbf{A}^H \mathbf{A}$ has deterministic asymptotic eigenvalue distribution $\rho_{\mathbf{J}}(s)$, the debiased LASSO estimator $\hat{\mathbf{x}}^{\text{d}}$ is given by*

$$\hat{\mathbf{x}}^{\text{d}} = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda} \mathbf{A}^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}}). \quad (30)$$

The debiased coefficient is

$$\Lambda = \frac{t \cdot \rho_{\mathbf{CA}}}{\rho_{\mathbf{CA}} - 1}, \quad (31)$$

in which t is the solution of

$$\int \frac{\rho_{\mathbf{J}}(s)}{t-s} ds = \frac{1 - \rho_{\mathbf{CA}}}{t}, \quad (32)$$

and $\rho_{\mathbf{CA}}$ is complex active component density of the LASSO solution defined in (12).

Setting the sensing matrix ensemble to row-orthogonal, then the asymptotic eigenvalue distribution of \mathbf{J} is given by

$$\rho_{\mathbf{J}}(s) = (1 - \gamma)\delta(s) + \gamma\delta(s - 1). \quad (33)$$

Therefore, one can obtain the following corollary.

Corollary 3.3. *When the complex-valued sensing matrix \mathbf{A} is row-orthogonal, the debiased LASSO estimator $\hat{\mathbf{x}}^d$ is given by*

$$\hat{\mathbf{x}}^d = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda_{\text{CROM}}} \mathbf{A}^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}}), \quad (34)$$

where

$$\Lambda_{\text{CROM}} = \frac{\gamma - \rho_{\mathbf{CA}}}{1 - \rho_{\mathbf{CA}}}, \quad (35)$$

and $\rho_{\mathbf{CA}}$ is complex active component density of the LASSO solution defined in (12).

According to [26], when the entries of \mathbf{A} are all i.i.d. Gaussian ensembles with mean 0 and variance $1/N$, the asymptotic eigenvalue distribution is given by

$$\rho_{\mathbf{J}}(s) = (1 - \gamma)\delta(s) + \frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - s)(s - \lambda_-)}}{s} \mathbb{I}_{[\lambda_-, \lambda_+]}(s), \quad (36)$$

$$\lambda_{\pm} = (1 \pm \sqrt{\gamma})^2, \quad (37)$$

$$\mathbb{I}_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}. \quad (38)$$

Then comes the following corollary.

Corollary 3.4. *When the sensing matrix \mathbf{A} is of random i.i.d. complex Gaussian ensemble in which all entries of \mathbf{A} are i.i.d. complex Gaussian variables with mean 0 and variance $1/N$, the debiased LASSO estimator $\hat{\mathbf{x}}^d$ is given by*

$$\hat{\mathbf{x}}^d = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda_{\text{CG}}} \mathbf{A}^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}}), \quad (39)$$

where $\Lambda_{\text{CG}} = \gamma - \rho_{\mathbf{CA}}$.

In addition, we find that the debiased coefficient Λ_{CG} is the same as the result in [19] in the case of complex Gaussian matrix design.

The third result presents a method for estimating the variance σ_w^2 .

Claim 3.5. *When the sensing matrix \mathbf{A} is randomly drawn from row-orthogonal matrices, the sample variance σ_w^2 of $\mathbf{w} = \hat{\mathbf{x}}^{\text{d}} - \mathbf{x}_0$ converges to $2\hat{\chi}/\hat{Q}^2$, where*

$$\hat{Q} = G'(-\chi; \mathbf{J}), \quad (40)$$

and

$$\hat{\chi} = \frac{\gamma G''(-\chi; \mathbf{J})}{2G'(-\chi; \mathbf{J}) - 2G''(-\chi; \mathbf{J})\chi} \overline{\text{RSS}} + \frac{-G''(-\chi; \mathbf{J})\gamma + (G'(-\chi; \mathbf{J}))^2}{2G'(-\chi; \mathbf{J}) - 2G''(-\chi; \mathbf{J})\chi} \sigma^2. \quad (41)$$

Here,

$$G'(-\chi; \mathbf{J}) = t(-\chi) + \frac{1}{\chi}, \quad (42)$$

$$G''(-\chi; \mathbf{J}) = t'(-\chi) + \frac{1}{\chi^2}, \quad (43)$$

$$\overline{\text{RSS}} = \lim_{N \rightarrow \infty} \frac{1}{N\gamma} \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}}\|_2^2. \quad (44)$$

in which

$$-\chi = \int \frac{\rho_{\mathbf{J}}(s)}{t(-\chi) - s} ds, \quad (45)$$

$$t(-\chi) = \frac{\rho_{\text{CA}} - 1}{\chi}, \quad (46)$$

$$t'(-\chi) = - \left[\int \frac{\rho_{\mathbf{J}}(s)}{(t(-\chi) - s)^2} ds \right]^{-1}. \quad (47)$$

The definition of the residual sum of squares $\overline{\text{RSS}}$ is given in (84), while it is unrealistic to calculate in practice. In a single hypothesis testing, we estimate the value of $\overline{\text{RSS}}$ in this way:

$$\overline{\text{RSS}} = \frac{1}{M} \|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}}\|_2^2. \quad (48)$$

3.2 Proof of Claim 3.1

We evaluate the free energy density corresponding to the LASSO Hamiltonian $H(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2/2 + \lambda\|\mathbf{x}\|_1$ at a zero-temperature limit:

$$f(\lambda) \doteq - \lim_{\beta \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{1}{\beta N} \mathbb{E}_{\mathbf{A}, \xi} [\ln Z(\mathbf{y}, \mathbf{A}; \lambda, \beta)], \quad (49)$$

where β is the inverse temperature and Z is the partition function:

$$Z(\mathbf{y}, \mathbf{A}; \lambda, \beta) = \int \exp \left(-\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 - \beta\lambda\|\mathbf{x}\|_1 \right) d\mathbf{x}. \quad (50)$$

In the zero-temperature limit $\beta \rightarrow \infty$, the Boltzmann distribution $e^{-\beta H(\mathbf{x})}/Z$ is dominated by the configurations of the LASSO solution. Hence, one can evaluate how the LASSO estimator depends on \mathbf{x}_0 , \mathbf{A} , $\boldsymbol{\xi}$ by analyzing the macroscopic behavior of the typical free energy density (49) using statistical mechanics.

Based on the replica method and the replica symmetric (RS) ansatz, the following is claimed.

Claim 3.6. *When the steering matrix \mathbf{A} is row-orthogonal, the free energy density (49) can be evaluated:*

$$f = \underset{Q, \hat{Q}, \chi, \hat{\chi}, m, \hat{m}}{\text{extr}} \left\{ G'(-\chi; \mathbf{J}) \left(Q - 2m + \rho - \frac{\chi}{2} \sigma^2 \right) + \frac{\gamma}{2} \sigma^2 - \hat{Q}Q + \hat{\chi}\chi + 2\hat{m}m \right. \\ \left. + \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \min_{x_i} \left[\frac{\hat{Q}}{2} |x_i|^2 - \text{Re} \left(\left(\hat{m}x_{0,i} + \sqrt{2\hat{\chi}}z_i \right)^* x_i \right) + \lambda |x_i| \right] Dz_i \right\} \quad (51)$$

where $\text{extr}_X \{F(X)\}$ denotes extremization of a function $F(X)$ with respect to X and $G'(x; \mathbf{J})$ is the derivative of $G(x; \mathbf{J})$ with respect to x . We define $\int (\dots) Dz$, \mathbf{J} and $G(x)$ as follows:

$$\int (\dots) Dz \doteq \int (\dots) \frac{\exp(-|z|^2)}{\pi} dz, \quad (52)$$

$$\mathbf{J} \doteq \mathbf{A}^H \mathbf{A}, \quad (53)$$

$$G(x; \mathbf{J}) \doteq \text{extr}_t \left[- \int \rho_{\mathbf{J}}(s) \ln |t - s| ds + tx \right] - \ln |x| - 1, \quad (54)$$

where z is a complex number and $\rho_{\mathbf{J}}(s)$ is the asymptotic eigenvalue distribution of \mathbf{J} . The derivative of the function $G(x; \mathbf{J})$ has the following form:

$$G'(x; \mathbf{J}) = t(x) - \frac{1}{x}, \quad (55)$$

where $t(x)$ is implicitly determined by the extreme value condition of (54):

$$x = \int \frac{\rho_{\mathbf{J}}(s)}{t(x) - s} ds. \quad (56)$$

Proof. The proof is given in Appendix C. \square

We can then obtain the possible distribution of LASSO solution by the free energy density.

On one hand, from the extreme value condition of the free energy density (51) of variable \hat{m} , that is $\frac{\partial f}{\partial \hat{m}} = 0$, we have

$$m = \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{i=1}^N \int \text{Re} (x_{0,i}^* \hat{x}_i) Dz_i, \quad (57)$$

where \hat{x}_i satisfies

$$\hat{x}_i = \arg \min_{x_i} \left[\frac{\hat{Q}}{2} |x_i|^2 - \operatorname{Re} \left(\left(\hat{m} x_{0,i} + \sqrt{2\hat{\chi} z_i} \right)^* x_i \right) + \lambda |x_i| \right]. \quad (58)$$

The above optimization problem has its analytical solution, given by

$$\hat{x}_i = \operatorname{ST}_{\lambda, \hat{Q}}(h_i), \quad (59)$$

where h_i is given in (29).

On the other hand, from definition of the macroscopic physical observable m , that is $m = \frac{1}{2N} \operatorname{Re}(\mathbf{x}_0^H \mathbf{x}_a)$, we have

$$m = \lim_{N \rightarrow \infty} \frac{1}{2N} \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [\operatorname{Re}(\mathbf{x}_0^H \hat{\mathbf{x}}^{\text{LASSO}})]. \quad (60)$$

Comparing (57) with (60), the distribution of the LASSO solution $\hat{\mathbf{x}}^{\text{LASSO}}$ can be inferred as follow:

$$\hat{x}_i^{\text{LASSO}} = \hat{x}_i = \operatorname{ST}_{\lambda, \hat{Q}}(h_i). \quad (61)$$

3.3 Proof of Claim 3.2

Denote by $\mathbf{h} = [h_1, h_2, \dots, h_N]^T$, where h_i is defined in (29). In the Thouless-Anderson-Palmer (TAP) analysis, \mathbf{h} is called local field. Denote by $\langle \mathbf{x} \rangle$ the average of \mathbf{x} taken by the Boltzmann distribution $e^{-\beta H(\mathbf{x})}/Z$, which is given by

$$\langle \mathbf{x} \rangle = \frac{\int \mathbf{x} \exp \left(-\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 - \beta \lambda \|\mathbf{x}\|_1 \right) d\mathbf{x}}{\int \exp \left(-\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 - \beta \lambda \|\mathbf{x}\|_1 \right) d\mathbf{x}}. \quad (62)$$

In the zero-temperature limit $\beta \rightarrow \infty$, the average $\langle \mathbf{x} \rangle$ is equal to the LASSO solution $\hat{\mathbf{x}}^{\text{LASSO}}$. TAP approach is to obtain the mean field equation, which describes the relation between the local field \mathbf{h} and the average $\langle \mathbf{x} \rangle$.

Consider the following alternative Gibbs free energy [20]:

$$Gi(\mathbf{m}, Q) = \operatorname{extr}_{\mathbf{h}, \Lambda} \left\{ \operatorname{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \frac{1}{\beta} \ln \int e^{-\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \beta \operatorname{Re}(\mathbf{h}^H \mathbf{x}) - \frac{\beta}{2} \Lambda \|\mathbf{x}\|_2^2 - \beta \lambda \|\mathbf{x}\|_1} d\mathbf{x} \right\}. \quad (63)$$

From the extreme value condition on \mathbf{h} , \mathbf{m} , Λ and Q of $G(\mathbf{m}, Q)$, one can conclude that

$$\langle \mathbf{x} \rangle, \left\langle \|\mathbf{x}\|_2^2 \right\rangle / (2N) = \arg \min_{\mathbf{m}, Q} G(\mathbf{m}, Q). \quad (64)$$

We will evaluate it by expectation consistent approximate inference [27].

Claim 3.7. *The alternative Gibbs free energy (63) possesses the following expression under the limit $\beta \rightarrow \infty$:*

$$Gi(\mathbf{m}, Q) = \text{extr}_{\mathbf{h}, \Lambda} \left\{ \text{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \sum_{i=1}^N \frac{(|h_i| - \lambda)^2}{2\Lambda} \cdot \Theta(|h_i| - \lambda) \right\} - \frac{N}{\beta} G(-\chi; \mathbf{J}) + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{m}\|_2^2, \quad (65)$$

where $\chi = \beta(Q - q)$ and $q = \sum_i m_i^2/N$.

Proof. The expectation consistent inference provides the following approximation:

$$Gi(\mathbf{m}, Q) \simeq \phi_{ada}(\mathbf{m}, Q) = \tilde{\phi}(\mathbf{m}, Q; l=0) + \phi^G(\mathbf{m}, Q; l=1) - \phi^G(\mathbf{m}, Q; l=0), \quad (66)$$

where

$$\tilde{\phi}(\mathbf{m}, Q; l) = \text{extr}_{\mathbf{h}, \Lambda} \left\{ \text{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \frac{1}{\beta} \ln \int e^{-\frac{\beta l}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \beta \text{Re}(\mathbf{h}^H \mathbf{x}) - \frac{\beta}{2} \Lambda \|\mathbf{x}\|_2^2 - \beta \lambda \|\mathbf{x}\|_1} d\mathbf{x} \right\}, \quad (67)$$

and

$$\phi^G(\mathbf{m}, Q; l) = \text{extr}_{\mathbf{h}, \Lambda} \left\{ \text{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \frac{1}{\beta} \ln \int e^{-\frac{\beta l}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \beta \text{Re}(\mathbf{h}^H \mathbf{x}) - \frac{\beta}{2} \Lambda \|\mathbf{x}\|_2^2} d\mathbf{x} \right\}. \quad (68)$$

With such approximation, $\tilde{\phi}(\mathbf{m}, Q; l=1)$, $\phi^G(\mathbf{m}, Q; l=1)$ and $\phi^G(\mathbf{m}, Q; l=0)$ can be easily calculated as follow:

$$\tilde{\phi}(\mathbf{m}, Q; l=0) = \text{extr}_{\mathbf{h}, \Lambda} \left\{ \text{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \sum_{i=1}^N \frac{(|h_i| - \lambda)^2}{2\Lambda} \cdot \Theta(|h_i| - \lambda) \right\}, \quad (69)$$

$$\phi^G(\mathbf{m}, Q; l=1) = -\frac{N}{\beta} G(-\chi; \mathbf{J}) + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{m}\|_2^2 + \frac{N}{\beta} \ln \frac{\beta}{2\pi} - \frac{N}{\beta} \ln |-\chi| - \frac{N}{\beta}, \quad (70)$$

$$\phi^G(\mathbf{m}, Q; l=0) = \frac{N}{\beta} \ln \frac{\beta}{2\pi} - \frac{N}{\beta} \ln |-\chi| - \frac{N}{\beta}, \quad (71)$$

where (69) is calculated under the limit $\beta \rightarrow \infty$. Therefore,

$$\phi_{ada}(\mathbf{m}, Q) = \text{extr}_{\mathbf{h}, \Lambda} \left\{ \text{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \sum_{i=1}^N \frac{(|h_i| - \lambda)^2}{2\Lambda} \cdot \Theta(|h_i| - \lambda) \right\} - \frac{N}{\beta} G(-\chi; \mathbf{J}) + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{m}\|_2^2. \quad (72)$$

□

Taking the limit $\beta \rightarrow \infty$ and $N \rightarrow \infty$, we can get the mean field equation by the extreme value condition on \mathbf{h} , Λ , \mathbf{m} , Q of (72) and linear response argument [21, 28],

given by

$$\mathbf{h} = \Lambda \mathbf{m} + \mathbf{A}^H (\mathbf{y} - \mathbf{A} \mathbf{m}), \quad (73)$$

$$m_i = \frac{h_i}{|h_i|} \cdot \frac{|h_i| - \lambda}{\Lambda} \cdot \Theta(|h_i| - \lambda), \quad (74)$$

$$\Lambda = G'(-\chi; \mathbf{J}), \quad (75)$$

$$\chi = \frac{1}{2N\Lambda} \sum_{i=1}^N \left[\left(2 - \frac{\lambda}{|h_i|} \right) \cdot \Theta(|h_i| - \lambda) \right], \quad (76)$$

where (76), derived in Appendix E, is obtained by linear response argument. Denote by ρ_{CA} the complex active component density of the LASSO solution as follow:

$$\rho_{\text{CA}} = \frac{1}{2N} \sum_{i=1}^N \left[\left(2 - \frac{\lambda}{|h_i|} \right) \cdot \Theta(|h_i| - \lambda) \right], \quad (77)$$

which is the same as (12). According to (73), one can obtain the debiased estimator $\hat{\mathbf{x}}^{\text{d}}$ as follow:

$$\hat{\mathbf{x}}^{\text{d}} = \hat{\mathbf{x}}^{\text{LASSO}} + \frac{1}{\Lambda} \mathbf{A}^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}}), \quad (78)$$

where $\Lambda = G'(-\chi; \mathbf{J})$ and χ can be obtained by solving

$$\frac{\rho_{\text{CA}}}{\chi} = G'(-\chi; \mathbf{J}). \quad (79)$$

Then we get the following equations according to the extreme condition in function $G(-\chi; \mathbf{J})$ (56) and (79):

$$\Lambda = G'(-\chi; \mathbf{J}) = \frac{\rho_{\text{CA}}}{\chi} = t(-\chi) + \frac{1}{\chi}, \quad (80)$$

$$-\chi = \frac{1 - \gamma}{t(-\chi)} + \frac{\gamma}{t(-\chi) - 1}, \quad (81)$$

which lead to (31) and (32).

3.4 Proof of Claim 3.5

We first define some macroscopic observables.

Claim 3.8. *The following relationships between the free energy density, regularization term, and residual sum of squares hold:*

$$f = \frac{\gamma}{2} \overline{\text{RSS}} + \bar{r}, \quad (82)$$

$$\bar{r} = \mathbb{E}_{\mathbf{A}, \xi} \left[\left\langle \frac{\lambda}{N} \sum_{i=1}^N |x_i| \right\rangle \right] = 2\hat{\chi}\chi + 2\hat{m}m - 2\hat{Q}Q, \quad (83)$$

$$\overline{\text{RSS}} = \mathbb{E}_{\mathbf{A}, \xi} [\text{RSS}] = \mathbb{E}_{\mathbf{A}, \xi} \left[\left\langle \frac{1}{M} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \right\rangle \right] \quad (84)$$

$$= \frac{2}{\gamma} \left[G'(-\chi; \mathbf{J}) (Q - 2m + \rho - \frac{\chi}{2} \sigma^2) + \frac{\gamma}{2} \sigma^2 - \hat{\chi}\chi \right], \quad (85)$$

where \bar{r} and $\overline{\text{RSS}}$ represent the per-element average of the regulation term and residual sum of squares, respectively.

Proof. Consider the extreme value condition of the free energy density f (51) on \hat{Q} , \hat{m} and $\hat{\chi}$, given by

$$\frac{\partial f}{\partial \hat{Q}} = -Q + \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \frac{1}{2} |\hat{x}_i|^2 \text{D}z_i = 0, \quad (86)$$

$$\frac{\partial f}{\partial \hat{m}} = 2m - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \text{Re} (x_{0,i}^* \hat{x}_i) \text{D}z_i = 0, \quad (87)$$

$$\frac{\partial f}{\partial \hat{\chi}} = \chi - \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \text{Re} \left(\frac{1}{\sqrt{2\hat{\chi}}} z_i^* \hat{x}_i \right) \text{D}z_i = 0, \quad (88)$$

where \hat{x}_i is the same with (58) and (59). The above equations note that

$$\hat{Q}Q = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \frac{\hat{Q}}{2} |\hat{x}_i|^2 \text{D}z_i, \quad (89)$$

$$2\hat{m}m = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \text{Re} ((\hat{m}x_{0,i})^* \hat{x}_i) \text{D}z_i, \quad (90)$$

$$2\hat{\chi}\chi = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \text{Re} \left((\sqrt{2\hat{\chi}}z_i)^* \hat{x}_i \right) \text{D}z_i. \quad (91)$$

Therefore, one can obtain that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \min_{x_i} \left[\frac{\hat{Q}}{2} |x_i|^2 - \text{Re} \left((\hat{m}x_{0,i} + \sqrt{2\hat{\chi}}z_i)^* x_i \right) + \lambda |x_i| \right] \text{D}z_i \\ &= \hat{Q}Q - 2\hat{\chi}\chi - 2\hat{m}m + \bar{r}. \end{aligned} \quad (92)$$

On the other hand, with the close-form expression of \hat{x}_i , we can get that

$$\begin{aligned} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \min_{x_i} \left[\frac{\hat{Q}}{2} |x_i|^2 - \text{Re} \left((\hat{m}x_{0,i} + \sqrt{2\hat{\chi}}z_i)^* x_i \right) + \lambda |x_i| \right] \text{D}z_i \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int -\frac{\hat{Q}}{2} |\hat{x}_i|^2 \text{D}z_i = -\hat{Q}Q. \end{aligned} \quad (93)$$

The expression of the regulation term \bar{r} (83) can then be derived, and the same with $\overline{\text{RSS}}$ (85). \square

Recall the extreme value condition of the free energy density f (51) on Q , m and χ , given by

$$\frac{\partial f}{\partial Q} = G'(-\chi; \mathbf{J}) - \hat{Q}, \quad (94)$$

$$\frac{\partial f}{\partial m} = 2G'(-\chi; \mathbf{J}) - 2\hat{m}, \quad (95)$$

$$\frac{\partial f}{\partial \chi} = -G''(-\chi; \mathbf{J}) \left(Q - 2m + \rho - \frac{\chi}{2}\sigma^2 \right) - \frac{\sigma^2}{2}G'(-\chi; \mathbf{J}) + \hat{\chi} = 0. \quad (96)$$

Together with the expression of $\overline{\text{RSS}}$ (85), the expression of $\hat{\chi}$ (41) can be obtained. The derivatives of function $G(-\chi; \mathbf{J})$ (42) and (43) follow the derivation of (55). Derivatives act on (56) leads to the expressions of $t(-\chi)$ and $t'(-\chi)$.

4 Numerical Experiments

In this section, we mainly provide the numerical simulations to exam the following capabilities.

1. Gaussianity of \mathbf{w} in the case of row-orthogonal matrix design, where $\mathbf{w} = \hat{\mathbf{x}}^d - \mathbf{x}_0$ denotes the difference between the debiased LASSO estimator and the original signal \mathbf{x}_0 .
2. Accuracy of estimating the variance σ_w^2 in both the case of row-orthogonal and Gaussian matrix design.
3. Detection performance of debiased LASSO detector.

4.1 Settings

In all the numerical experiments, we artificially generate the original signal \mathbf{x}_0 , observation matrix \mathbf{A} and the noise $\boldsymbol{\xi}$. The original signal \mathbf{x}_0 is generated from the Bernoulli-Gaussian distribution: $p_x = (1-2\rho)\delta(x) + \frac{2\rho}{\pi\sigma_x^2}e^{-|x|^2/\sigma_x^2}$. For the observation matrix \mathbf{A} , we separately consider Gaussian design and row-orthogonal design. The former is achieved by setting all the entries of \mathbf{A} i.i.d. complex Gaussian variables: $A_{ij} \sim \mathcal{CN}(0, 1/N)$, and the latter is achieved by randomly selecting M rows from a randomly generated $N \times N$ orthogonal matrix. The entries of the noise $\boldsymbol{\xi}$ are i.i.d. complex Gaussian variables: $\xi_i \sim \mathcal{CN}(0, \sigma^2)$. We consider the matched filtering (MF) definition of the signal-to-noise ratio (SNR) [5], such that

$$\text{SNR} = \frac{\gamma\sigma_x^2}{\sigma^2}. \quad (97)$$

4.2 Gaussianity of \mathbf{w} in the case of row-orthogonal matrix design

We investigate the Gaussianity of \mathbf{w} in the case of row-orthogonal matrix design by comparing the empirical cumulative distribution function (ECDF) of its real and imag-

Table 2: p-values of Kolmogorov-Smirnov test of the real and imaginary parts of non-zero entries and zero entries of \mathbf{w}^{CROM} and \mathbf{w}^{CG} .

p-values	$\mathbf{w}^{\text{CROM}} (\mathcal{H}_1)$	$\mathbf{w}^{\text{CROM}} (\mathcal{H}_0)$	$\mathbf{w}^{\text{CG}} (\mathcal{H}_1)$	$\mathbf{w}^{\text{CG}} (\mathcal{H}_0)$
real part	0.7563	0.6702	1.731×10^{-78}	0
imaginary part	0.8259	0.6090	1.729×10^{-70}	0

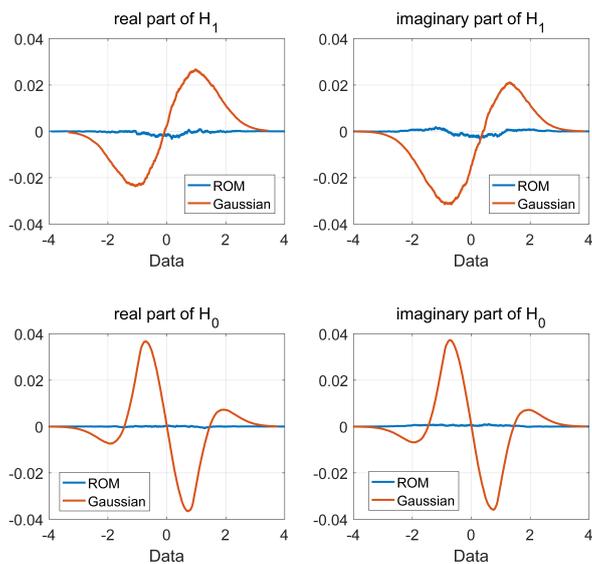


Figure 2: The difference between the Gaussian distribution and the empirical CDF of the real and imaginary parts of non-zero entries and zero entries of \mathbf{w}^{CROM} and \mathbf{w}^{CG} .

inary part with Gaussian distribution. The comparisons are performed for the complex row-orthogonal debiased estimator $\hat{\mathbf{x}}^{\text{d,CROM}}$ (with Λ_{CROM} given by (27)) and the complex Gaussian debiased estimator $\hat{\mathbf{x}}^{\text{d,CG}}$ (with Λ_{CG} given by (24)). Denote by $\mathbf{w}^{\text{CROM}} = \hat{\mathbf{x}}^{\text{d,CROM}} - \mathbf{x}_0$ and $\mathbf{w}^{\text{CG}} = \hat{\mathbf{x}}^{\text{d,CG}} - \mathbf{x}_0$. Let $\tilde{\mathbf{w}}^{\text{CROM}} = \sqrt{2}\mathbf{w}^{\text{CROM}}/\sigma_{w,\text{CROM}}^2$, $\tilde{\mathbf{w}}^{\text{CG}} = \sqrt{2}\mathbf{w}^{\text{CG}}/\sigma_{w,\text{CG}}^2$. Consequently, the ECDF of the real and imaginary part of these two vectors, denoted by $F_{\text{CROM},r}(x)$, $F_{\text{CROM},i}(x)$, $F_{\text{CG},r}(x)$ and $F_{\text{CG},i}(x)$ respectively, are considered to converge to standard Gaussian distribution $\Phi(x)$ weakly. We divide $\tilde{\mathbf{w}}^{\text{CROM}}$ and $\tilde{\mathbf{w}}^{\text{CG}}$ into 4 parts: real and imaginary part of non-zero entries and zero entries, and verify their Gaussianity by demonstrating the differences such as $\Phi(x) - F_{\text{CROM},\mathcal{H}_1,r}(x)$ respectively.

The observation matrix \mathbf{A} is partial Fourier with a size of $M = 768$ and $N = 1024$. The variance σ_x^2 of the non-zero entries in \mathbf{x}_0 is set to be 1 and the regularization parameter λ of LASSO is 0.1. We set the signal density to be $2\rho = 0.1$ and SNR to be 5dB. The empirical laws are obtained by 10^3 Monte-Carol trials.

Fig. 2 shows the difference between Gaussian distribution $\Phi(x)$ and the ECDF of each part of \mathbf{w}^{CROM} and \mathbf{w}^{CG} . The results shows that in the case of partial Fourier matrix design, the ECDF of \mathbf{w}^{CROM} is very close to the Gaussian distribution, while the ECDF of \mathbf{w}^{CG} has a significant difference from the Gaussian distribution. We also employ Kolmogorov-Smirnov (KS) test [29] on them, with the p-values shown in Table 2. The KS test compares the ECDF of the input samples with a certain distribution (here we set it to be Gaussian distribution) and the larger of p-values means the higher probability that the samples come from the given distribution. It is obvious that for each part of \mathbf{w}^{CROM} , KS test verifies that the ECDF well meets the expected theoretical distribution while the other one indicates the opposite result. Our verification of the Gaussianity of \mathbf{w}^{CROM} in turn verifies the correctness of the debiased coefficient Λ_{CROM} in the case of row-orthogonal matrix design.

4.3 Accuracy of estimating σ_w^2

In this section, we investigate the accuracy of the estimation of σ_w^2 by comparing the relative estimation error of different approaches, which is defined as follow

$$\text{REE} = \frac{|\hat{\sigma}_w - \sigma_w|}{\sigma_w}, \quad (98)$$

where $\hat{\sigma}_w$ is the estimated result and the ground truth σ_w is obtained by

$$\sigma_w = \sqrt{\frac{1}{N} \sum_{i=1}^N |w_i|^2}. \quad (99)$$

We compare REE of CROD, CAMP and SDL-test in the case of Gaussian matrix design and row-orthogonal matrix design, respectively. In the following experiments, we set the length of \mathbf{x}_0 to be $N = 256$ and σ_x^2 to be 1. The variance of noise $\sigma^2 = 0.05$, corresponding to a SNR of 13dB. Noting that the estimation approach of SDL-test

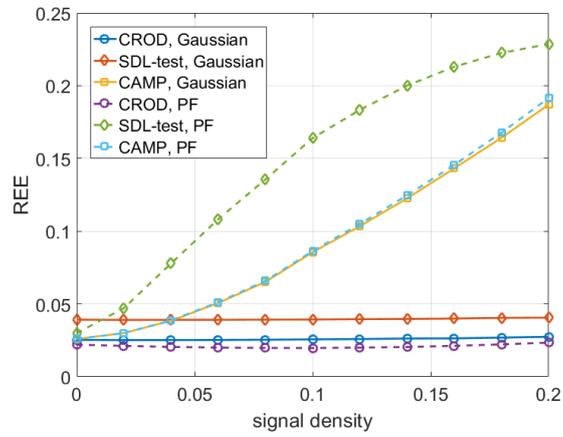


Figure 3: REE of CROD, CAMP and SDL-test under the Gaussian and partial Fourier design model.

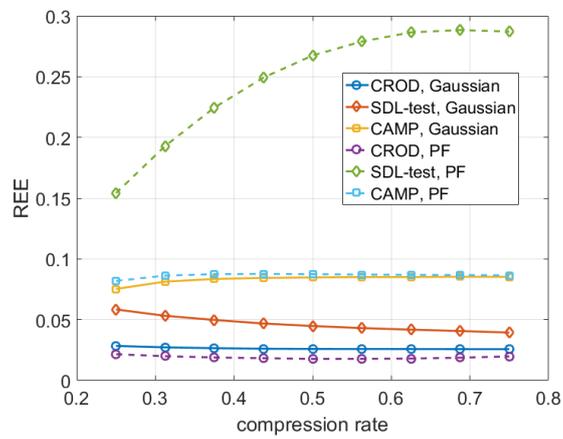


Figure 4: REE of CROD, CAMP and SDL-test under the Gaussian and partial Fourier design model.

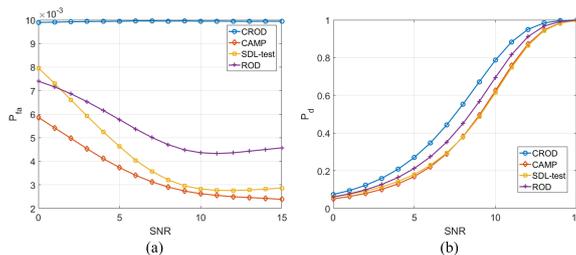


Figure 5: (a) Probability of false alarm and (b) detection of different detectors vary with SNR. Here $2\rho = 0.1$ and $\gamma = 0.5$.

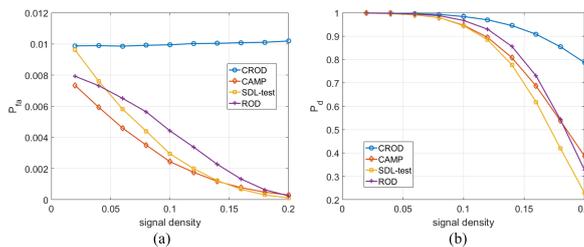


Figure 6: (a) Probability of false alarm and (b) detection of different detectors vary with signal density. Here $\text{SNR} = 13\text{dB}$ and $\gamma = 0.5$.

in [18] is not suitable for complex-valued data, we have modified it to the following form:

$$\hat{\sigma}_w = \frac{\sqrt{\gamma}}{\sqrt{\ln 2}(\gamma - \rho_{CA})} \text{median}(|\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}^{\text{LASSO}}|), \quad (100)$$

The results are obtained by 10^5 Monte-Carlo trials and presented in Fig. 3 and Fig. 4, in which the signal density and compression rate vary, respectively. The results of the numerical experiments illustrate that the variance estimation method we use has the ability to consistently obtain high estimation accuracy under a variety of different conditions.

4.4 Detection performance of debiased LASSO detector

We apply several detectors for solving the sub-Nyquist radar detection problem, in which the observation matrix \mathbf{A} is partial Fourier, and examine their detection performance. Debiased detectors listed in Table 1, including CROD, CAMP, SDL-test and ROD, are compared. In the following experiments, we set the length of \mathbf{x}_0 to $N = 256$ and σ_x^2 to 1. The probability of false alarm is set to 0.01 and all the results are obtained by 10^5 Monte-Carlo trials. In Fig. 5, 6 and 7, we demonstrate how the detection performance of these detectors varies with SNR, signal density, and compression rate, respectively. The results suggest that CROD has the best ability to maintain the false alarm rate

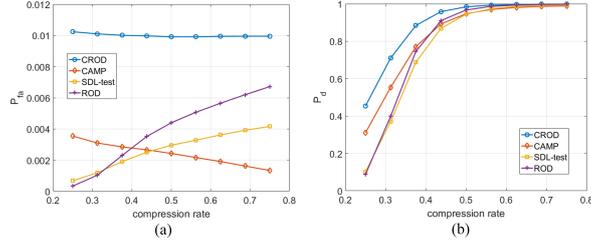


Figure 7: (a) Probability of false alarm and (b) detection of different detectors vary with compression rate. Here $2\rho = 0.1$ and $\text{SNR} = 13\text{dB}$.

under multiple parameter variations, i.e., it can most accurately calculate the threshold (or p-value) for a given false alarm rate.

5 Conclusion

In this paper, we provide the design of debiased LASSO detector solving the detection problem of compressed sensing radar under the row-orthogonal design model. The detection performance of the present detector is theoretically analyzed and proved to be better than LASSO detector. We also compare the proposed approach with other debiased LASSO detectors, simulation results indicate that our approach can provide test statistic and threshold (or p-value) more accurately. Such merit allows precise control of the false alarm rate, resulting in the higher detection rate than other CS detectors.

A Proof of Theorem 2.3

Assume that there exists $\Lambda_2 \neq \Lambda_1$ such that the empirical law of $\mathbf{w}_2(N)$ converges to a Gaussian distribution with mean μ_2 as $N \rightarrow \infty$. According to Lemma B.2, we have the following inequality

$$|w_{2,i} - w_{1,i}| \leq \lambda \left| \frac{1}{\Lambda_2} - \frac{1}{\Lambda_1} \right| \doteq c. \quad (101)$$

We argue that the probability that the variance σ_2^2 of \mathbf{w}_2 is unequal to the variance σ_1^2 of \mathbf{w}_1 is 1. That is, almost surely,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N w_{1,i}^2 \neq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N w_{2,i}^2 - \mu_2^2, \quad (102)$$

where

$$\mu_2 = \frac{1}{N} \sum_{i=1}^N w_{2,i}. \quad (103)$$

Such conclusion holds since the well-known fact that the set of zeros of a nonzero polynomial has measure zero.

Without loss of generality, let $\sigma_2 > \sigma_1$. Denote the empirical distribution of $\mathbf{w}_1(N)$ and $\mathbf{w}_2(N)$ by $F_{1,N}(x)$ and $F_{2,N}(x)$ respectively, given by

$$F_{1,N}(x) = \frac{1}{N} \# \{i \leq N : w_{1,i} \leq x\}, \quad (104)$$

$$F_{2,N}(x) = \frac{1}{N} \# \{i \leq N : w_{2,i} \leq x\}. \quad (105)$$

Therefore, $\forall \delta_1 > 0, \forall \delta_2 > 0, \forall m \in \mathbb{R}, \exists N_1 \in \mathbb{N}, \forall n \geq N_1$, such that

$$|F_{1,n}(m\sigma_1) - F_1(m\sigma_1)| \leq \delta_1, \quad (106)$$

$$|F_{2,n}(m\sigma_2 + \mu_2) - F_2(m\sigma_2 + \mu_2)| \leq \delta_2 \quad (107)$$

hold with probability 1, where

$$F_1(m\sigma_1) = F_2(m\sigma_2 + \mu_2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^m e^{-\frac{t^2}{2}} dt. \quad (108)$$

Thus,

$$|F_{1,n}(m\sigma_1) - F_{2,n}(m\sigma_2 + \mu_2)| \leq \delta_1 + \delta_2. \quad (109)$$

Let $m > \frac{c-\mu_2}{\sigma_2-\sigma_1}$, $m_1 = \frac{m\sigma_2+\mu_2-c}{\sigma_1}$, and

$$\delta_1 = \delta_2 = \frac{1}{\sqrt{18\pi}} (m_1 - m) e^{-\frac{m_1^2}{2}}. \quad (110)$$

Then we have $m\sigma_2 + \mu_2 = m_1\sigma_1 + c$ and $m_1 > m$. Combining with (101) and the definition of the empirical law (104), (105) yields

$$F_{2,n}(m\sigma_2 + \mu_2) \geq F_{1,n}(m_1\sigma_1). \quad (111)$$

On the other hand, (106) and (107) yield

$$F_{1,n}(m_1\sigma_1) \geq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{m_1} e^{-\frac{t^2}{2}} dt - \delta_1, \quad (112)$$

$$F_{2,n}(m\sigma_2 + \mu_2) \leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^m e^{-\frac{t^2}{2}} dt + \delta_2. \quad (113)$$

Then, from a simple math, we can obtain that

$$F_{2,n}(m\sigma_2 + \mu_2) - F_{1,n}(m_1\sigma_1) \leq \delta_1 + \delta_2 - \frac{1}{\sqrt{2\pi}} \int_m^{m_1} e^{-\frac{t^2}{2}} dt < 0, \quad (114)$$

which is in contradiction with (111).

B Proof of Theorem 2.7

For the convenience of readers, we here restate the definition of subdifferential as follow, which will help us prove this theorem.

Definition B.1 ([30]). *The subdifferential of a function $f : \mathcal{I} \rightarrow \mathbb{R}$ at a point $\mathbf{x}_0 \in \mathcal{I}$ is defined to be a set of vectors such that*

$$\partial f(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0} = \left\{ \mathbf{g} \in \mathbb{R}^n : f(\mathbf{x}) - f(\mathbf{x}_0) \geq (\mathbf{x} - \mathbf{x}_0)^T \mathbf{g}, \quad \forall \mathbf{x} \in \mathcal{I} \right\}, \quad (115)$$

where $\mathcal{I} \subset \mathbb{R}^n$ and each \mathbf{g} is called subderivative or subgradient.

It is easy to verify that the subderivative is linear. We then prove the following lemma that simplifies the proof of Theorem 2.7.

Lemma B.2. *The vector $\hat{\mathbf{x}}$ is the solution of the LASSO problem if and only if the following conditions are all satisfied for $i \in \{1, \dots, N\}$:*

$$\mathbf{a}_i^H (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}) = \lambda \frac{\hat{x}_i}{|\hat{x}_i|}, \quad \text{if } \hat{x}_i \neq 0, \quad (116)$$

$$|\mathbf{a}_i^H (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})| \leq \lambda, \quad \text{if } \hat{x}_i = 0, \quad (117)$$

where \mathbf{a}_i is the i -th column of matrix \mathbf{A} and \hat{x}_i is the i -th entry of vector $\hat{\mathbf{x}}$.

Proof. Due to the independent variable \mathbf{x} in LASSO problem is complex-valued, we transform the objective function into a real function $f : \mathbb{R}^{2N} \rightarrow \mathbb{R}$, given by

$$f(\mathbf{x}_r) = \frac{1}{2} \|\mathbf{y}_r - \mathbf{A}_r \mathbf{x}_r\|_2^2 + \lambda \sum_{i=1}^N (x_{r,i}^2 + x_{r,i+N}^2)^{\frac{1}{2}}, \quad (118)$$

where $\mathbf{x}_r = [\text{Re}(\mathbf{x})^T, \text{Im}(\mathbf{x})^T]^T \in \mathbb{R}^{2N}$, $\mathbf{y}_r = [\text{Re}(\mathbf{y})^T, \text{Im}(\mathbf{y})^T]^T \in \mathbb{R}^{2M}$ and

$$\mathbf{A}_r = \begin{bmatrix} \text{Re}(\mathbf{A}) & -\text{Im}(\mathbf{A}) \\ \text{Im}(\mathbf{A}) & \text{Re}(\mathbf{A}) \end{bmatrix} \in \mathbb{R}^{2M \times 2N}.$$

It is easy to verify that $f(\mathbf{x}_r)$ equals to the objective function of LASSO problem. Let $f(\mathbf{x}_r) = g(\mathbf{x}_r) + \lambda h(\mathbf{x}_r)$, in which

$$g(\mathbf{x}_r) = \frac{1}{2} \|\mathbf{y}_r - \mathbf{A}_r \mathbf{x}_r\|_2^2, \quad (119)$$

$$h(\mathbf{x}_r) = \sum_{i=1}^N (x_{r,i}^2 + x_{r,i+N}^2)^{\frac{1}{2}}. \quad (120)$$

The part $g(\mathbf{x}_r)$ is everywhere differentiable, thus

$$\partial g(\mathbf{x}_r) = \left\{ \frac{dg(\mathbf{x}_r)}{d\mathbf{x}_r} \right\} = \{-\mathbf{A}_r (\mathbf{y}_r - \mathbf{A}_r \mathbf{x}_r)\}. \quad (121)$$

The part $h(\mathbf{x}_r)$ could be divided into N parts: $h_i(x_{r,i}, x_{r,i+N}) = (x_{r,i}^2 + x_{r,i+N}^2)^{\frac{1}{2}}$. If $x_{r,i}^2 + x_{r,i+N}^2 > 0$, then $\partial h_i(x_{r,i}, x_{r,i+N}) = \{(x_{r,i}(x_{r,i}^2 + x_{r,i+N}^2)^{-\frac{1}{2}}, x_{r,i+N}(x_{r,i}^2 + x_{r,i+N}^2)^{-\frac{1}{2}})\}$. If $x_{r,i}^2 + x_{r,i+N}^2 = 0$, then $\partial h_i(x_{r,i}, x_{r,i+N}) = \{(g_1, g_2) : (x_{r,i}^2 + x_{r,i+N}^2)^{\frac{1}{2}} \geq g_1 x_{r,i} + g_2 x_{r,i+N}\}$. According to Cauchy inequality, we can easily obtain that $g_1^2 + g_2^2 \leq 1$.

With the fact that $f(\mathbf{x}_r)$ is a convex function, the point $\hat{\mathbf{x}}_r$ is the global minimum of $f(\mathbf{x}_r)$ if and only if $\mathbf{0} \in \partial f(\mathbf{x}_r)|_{\mathbf{x}_r=\hat{\mathbf{x}}_r}$, which indicates that

$$\mathbf{l} \doteq \frac{1}{\lambda} \mathbf{A}_r (\mathbf{y}_r - \mathbf{A}_r \mathbf{x}_r) \in \partial h(\mathbf{x}_r)|_{\mathbf{x}_r=\hat{\mathbf{x}}_r} \quad (122)$$

Therefore,

$$(l_i, l_{i+N}) \in \begin{cases} \left\{ \left(\frac{x_{r,i}}{(x_{r,i}^2 + x_{r,i+N}^2)^{\frac{1}{2}}}, \frac{x_{r,i+N}}{(x_{r,i}^2 + x_{r,i+N}^2)^{\frac{1}{2}}} \right) \right\}, & x_{r,i}^2 + x_{r,i+N}^2 > 0; \\ \{(g_1, g_2) : g_1^2 + g_2^2 \leq 1\}, & x_{r,i}^2 + x_{r,i+N}^2 = 0. \end{cases} \quad (123)$$

Combined with the definition of \mathbf{x}_r , \mathbf{y}_r and \mathbf{A}_r , the previous result yields (116) and (117). \square

Let the threshold of the LASSO detector be κ , such that the probability of false alarm $P_{fa,2}$ satisfies

$$P_{fa,2} = \lim_{N \rightarrow \infty} \frac{1}{N-k} \sum_{i \in S^c} \varphi_{2,i}, \quad (124)$$

where $S = \{i : x_{0,i} \neq 0\}$ is the support set of \mathbf{x}_0 with $|S| = k$, $S^c = \{1, \dots, N\} \setminus S$ and $\varphi_{2,i}$ is a test with

$$\varphi_{2,i} = \begin{cases} 1, & |\hat{x}_i^{\text{LASSO}}| > \kappa; \\ 0, & \text{otherwise.} \end{cases} \quad (125)$$

Then the probability of detection of LASSO detector $P_{d,2}$ has the expression that

$$P_{d,2} = \lim_{N \rightarrow \infty} \frac{1}{k} \sum_{i \in S} \varphi_{2,i}. \quad (126)$$

We now define the test of the debiased LASSO detector as follow:

$$\varphi_{1,i} = \begin{cases} 1, & |\hat{x}_i^{\text{d}}| > \kappa + \frac{\lambda}{\Lambda}; \\ 0, & \text{otherwise.} \end{cases} \quad (127)$$

and the probability of false alarm $P_{fa,1}$, the probability of detection $P_{d,1}$, given by

$$P_{fa,1} = \lim_{N \rightarrow \infty} \frac{1}{N-k} \sum_{i \in S^c} \varphi_{1,i}, \quad (128)$$

$$P_{d,1} = \lim_{N \rightarrow \infty} \frac{1}{k} \sum_{i \in S} \varphi_{1,i}. \quad (129)$$

1) If $\varphi_{2,i} = 0$ and $i \in S^c$, then $|\hat{x}_i^{\text{LASSO}}| \leq \kappa$.

- For i satisfying $\hat{x}_i^{\text{LASSO}} = 0$, we have the following derivation based on Lemma B.2:

$$\begin{aligned} |\hat{x}_i^{\text{d}}| &= \left| \hat{x}_i^{\text{LASSO}} + \frac{1}{\Lambda} \mathbf{a}_i^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}}) \right| \\ &\leq |\hat{x}_i^{\text{LASSO}}| + \frac{1}{\Lambda} |\mathbf{a}_i^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}})| \\ &\leq |\hat{x}_i^{\text{LASSO}}| + \frac{\lambda}{\Lambda}. \end{aligned} \quad (130)$$

- For i satisfying $|\hat{x}_i^{\text{LASSO}}| > 0$, we have the following derivation based on Lemma B.2:

$$\begin{aligned} |\hat{x}_i^{\text{d}}| &= \left| \hat{x}_i^{\text{LASSO}} + \frac{1}{\Lambda} \mathbf{a}_i^H (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}^{\text{LASSO}}) \right| \\ &= \left| \hat{x}_i^{\text{LASSO}} + \frac{\lambda}{\Lambda} \cdot \frac{\hat{x}_i^{\text{LASSO}}}{|\hat{x}_i^{\text{LASSO}}|} \right| \\ &= |\hat{x}_i^{\text{LASSO}}| + \frac{\lambda}{\Lambda}. \end{aligned} \quad (131)$$

Therefore, $|\hat{x}_i^{\text{d}}| \leq \kappa + \frac{\lambda}{\Lambda}$ and $\varphi_{1,i} = 0$. Thus, $\varphi_{1,i} \leq \varphi_{2,i}$ for $i \in S^c$, and

$$\begin{aligned} P_{fa,1} &= \lim_{N \rightarrow \infty} \frac{1}{N-k} \sum_{i \in S^c} \varphi_{1,i} \\ &\leq \lim_{N \rightarrow \infty} \frac{1}{N-k} \sum_{i \in S^c} \varphi_{2,i} = P_{fa,2}. \end{aligned} \quad (132)$$

2) If $\varphi_{2,i} = 1$ and $i \in S$, then $|\hat{x}_i^{\text{LASSO}}| > \kappa$, leading to $|\hat{x}_i^{\text{d}}| > \kappa + \frac{\lambda}{\Lambda}$ and $\varphi_{1,i} = 1$. Thus, $\varphi_{1,i} \geq \varphi_{2,i}$ for $i \in S$, and

$$\begin{aligned} P_{d,1} &= \lim_{N \rightarrow \infty} \frac{1}{k} \sum_{i \in S} \varphi_{1,i} \\ &\geq \lim_{N \rightarrow \infty} \frac{1}{k} \sum_{i \in S} \varphi_{2,i} = P_{d,2}. \end{aligned} \quad (133)$$

C Proof of Claim 3.6

Replica method bases on a simple fact that for any $Z > 0$, we have

$$\ln Z = \lim_{n \rightarrow 0} \frac{Z^n - 1}{n} = \lim_{n \rightarrow 0} \frac{\partial}{\partial n} (Z^n - 1). \quad (134)$$

If Z is a random variable, then we take the average over Z on the both side of (134) yielding

$$\mathbb{E} \ln Z = \lim_{n \rightarrow 0} \frac{\mathbb{E} Z^n - 1}{n} = \lim_{n \rightarrow 0} \frac{1}{n} \ln (\mathbb{E} Z^n) = \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \ln (\mathbb{E} Z^n). \quad (135)$$

Therefore, with the help of replica method, we rewrite the free energy density $f(\lambda)$ as follow

$$f(\lambda) = - \lim_{\beta \rightarrow \infty} \lim_{N \rightarrow \infty} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \frac{1}{\beta N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [Z^n(\mathbf{y}, \mathbf{A}; \lambda, \beta)], \quad (136)$$

where we recall $Z(\mathbf{y}, \mathbf{A}; \lambda, \beta)$ is the partition function and we use Z instead of $Z(\mathbf{y}, \mathbf{A}; \lambda, \beta)$ in all of the following. As most of the work based on replica analysis, we exchange the order of the limits $n \rightarrow 0$ and $N \rightarrow \infty$, thus

$$f(\lambda) = - \lim_{\beta \rightarrow \infty} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \lim_{N \rightarrow \infty} \frac{1}{\beta N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [Z^n]. \quad (137)$$

We will first compute the limit $N \rightarrow \infty$ in Appendix C.1, then we take the other limits to get the final result in Appendix C.2.

C.1 Derivation of the limit $N \rightarrow \infty$

In this subsection, we will introduce the derivation of

$$\frac{1}{N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [Z^n]$$

and take the limit $N \rightarrow \infty$. With the definition of Z , we have

$$\begin{aligned} & \frac{1}{N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [Z^n] \\ &= \frac{1}{N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} \left[\int \prod_{a=1}^n d\mathbf{x}_a \exp \left\{ -\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}_a\|_2^2 - \beta\lambda \|\mathbf{x}_a\|_1 \right\} \right] \\ &= \frac{1}{N} \ln \int \prod_{a=1}^n \exp \left\{ -\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}_a\|_2^2 - \beta\lambda \|\mathbf{x}_a\|_1 \right\} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) p_{\mathbf{A}}(\mathbf{A}) \prod_{a=1}^n d\mathbf{x}_a d\boldsymbol{\xi} d\mathbf{A}, \quad (138) \end{aligned}$$

where $p_{\mathbf{A}}(\mathbf{A})$ and $p_{\boldsymbol{\xi}}(\boldsymbol{\xi})$ are the probability density function of \mathbf{A} and $\boldsymbol{\xi}$, respectively. However, this integral is still difficult to calculate. To solve this problem, we introduce the RS ansatz, which gives a constraint on original signal \mathbf{x}_0 and the replicas \mathbf{x}_a , and we express it as

$$\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) = \mathbf{k}. \quad (139)$$

And we define that

$$f(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0, \mathbf{A}, \boldsymbol{\xi}) \doteq \prod_{a=1}^n \exp \left\{ -\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}_a\|_2^2 \right\} p_{\boldsymbol{\xi}}(\boldsymbol{\xi}) p_{\mathbf{A}}(\mathbf{A}), \quad (140)$$

$$g(\{\mathbf{x}_a\}_{a=1}^n) \doteq \prod_{a=1}^n e^{-\beta\lambda \|\mathbf{x}_a\|_1}. \quad (141)$$

With the conclusion

$$\int f(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0, \mathbf{A}, \boldsymbol{\xi}) d\boldsymbol{\xi} d\mathbf{A} = f_1(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0)), \quad (142)$$

which will be proved in Appendix C.1.1, we can split (138) into two parts under the RS ansatz:

$$\begin{aligned} & \frac{1}{N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [Z^n] \\ &= \frac{1}{N} \ln \int f(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0, \mathbf{A}, \boldsymbol{\xi}) g(\{\mathbf{x}_a\}_{a=1}^n) \delta(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) - \mathbf{k}) \prod_{a=1}^n d\mathbf{x}_a d\boldsymbol{\xi} d\mathbf{A} d\mathbf{k} \\ &= \frac{1}{N} \ln \int f_1(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0)) g(\{\mathbf{x}_a\}_{a=1}^n) \delta(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) - \mathbf{k}) \prod_{a=1}^n d\mathbf{x}_a d\mathbf{k} \\ &= \frac{1}{N} \ln \int f_1(\mathbf{k}) \left(\int g(\{\mathbf{x}_a\}_{a=1}^n) \delta(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) - \mathbf{k}) \prod_{a=1}^n d\mathbf{x}_a \right) d\mathbf{k} \\ &= \frac{1}{N} \ln \int f_1(\mathbf{k}) g_1(\mathbf{k}, \mathbf{x}_0) d\mathbf{k}, \end{aligned} \quad (143)$$

where

$$g_1(\mathbf{k}, \mathbf{x}_0) \doteq \int g(\{\mathbf{x}_a\}_{a=1}^n) \delta(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) - \mathbf{k}) \prod_{a=1}^n d\mathbf{x}_a. \quad (144)$$

Then we employ the following substitution, given by

$$f_1(\mathbf{k}) = \exp\{N\mathcal{T}(\mathbf{k})\}, \quad (145)$$

$$g_1(\mathbf{k}, \mathbf{x}_0) = \exp\{N\mathcal{S}(\mathbf{k}, \mathbf{x}_0)\}. \quad (146)$$

The limit $N \rightarrow \infty$ of (143) can be calculated by saddle point method, which is usually used to approximately calculate the following complex integral when $N \in \mathbb{R}$ is large enough

$$I(N) = \int_{\Gamma_1 \times \Gamma_2 \times \dots \times \Gamma_M} F(\mathbf{z}) e^{Nf(\mathbf{z})} d\mathbf{z}, \quad (147)$$

where $f(\mathbf{z})$ is analytic in complex variable $\mathbf{z} \in \mathbb{C}^M$ and $\{\Gamma_i\}$ converge the integral. Saddle point method claims that for large N , the integral can be approximately given by the function value of the saddle point \mathbf{z}_0 , given by

$$|I(N)| = \left| e^{N(f(\mathbf{z}_0) + O(\frac{1}{N}))} \right|, \quad (148)$$

where \mathbf{z}_0 , the saddle point, satisfies $\left. \frac{\partial f}{\partial \mathbf{z}} \right|_{\mathbf{z}=\mathbf{z}_0} = \mathbf{0}$. Thus,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [Z^n] = \text{extr}_{\mathbf{k}} \{\mathcal{T}(\mathbf{k}) + \mathcal{S}(\mathbf{k}, \mathbf{x}_0)\}. \quad (149)$$

In Appendix C.1.1 and C.1.2, we will provide the derivation of $\mathcal{T}(\mathbf{k})$ and $\mathcal{S}(\mathbf{k}, \mathbf{x}_0)$, respectively.

C.1.1 The derivation of $\mathcal{T}(\mathbf{k})$

In this section, we will introduce the derivation of

$$\int f(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0, \mathbf{A}, \boldsymbol{\xi}) d\boldsymbol{\xi} d\mathbf{A} = \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} \left[\exp \left\{ -\frac{\beta}{2} \sum_{a=1}^n \|\mathbf{y} - \mathbf{A}\mathbf{x}_a\|_2^2 \right\} \right], \quad (150)$$

and further under the RS ansatz

$$\mathcal{T}(\mathbf{k}) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln f_1(\mathbf{k}) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \int f(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0, \mathbf{A}, \boldsymbol{\xi}) d\boldsymbol{\xi} d\mathbf{A}. \quad (151)$$

It is convenient to first take the average with respect to $\boldsymbol{\xi}$:

$$\mathbb{E}_{\boldsymbol{\xi}} \left[\exp \left\{ -\frac{\beta}{2} \sum_{a=1}^n \|\mathbf{A}(\mathbf{x}_0 - \mathbf{x}_a) + \boldsymbol{\xi}\|_2^2 \right\} \right] = \exp \left[\sum_{a=0}^n -\frac{k_a}{2} \mathbf{u}_a^H \mathbf{J} \mathbf{u}_a - N\gamma \ln \left(1 + \frac{\beta n \sigma^2}{2} \right) \right], \quad (152)$$

where we recall that $\mathbf{J} = \mathbf{A}^H \mathbf{A}$, $\gamma = M/N$ and k_a , \mathbf{u}_a are defined as follows:

$$k_0 = -\frac{\beta^2 \sigma^2}{2 + \beta n \sigma^2}, \quad (153)$$

$$\mathbf{u}_0 = \sum_{a=1}^n \mathbf{u}_a, \quad (154)$$

$$k_a = \beta, \quad 1 \leq a \leq n, \quad (155)$$

$$\mathbf{u}_a = \mathbf{x}_a - \mathbf{x}_0, \quad 1 \leq a \leq n. \quad (156)$$

Then we orthogonalize $\{\mathbf{u}_a\}_{a=0}^n$ in order to facilitate the calculation of the average on \mathbf{A} . First, one can rewrite (152) as,

$$\exp \left[\sum_{a=0}^n -\frac{k_a}{2} \mathbf{u}_a^H \mathbf{J} \mathbf{u}_a - N\gamma \ln \left(1 + \frac{\beta n \sigma^2}{2} \right) \right] = \exp \left[\frac{1}{2} \text{Tr}(\mathbf{J} \mathbf{L}) - N\gamma \ln \left(1 + \frac{\beta n \sigma^2}{2} \right) \right], \quad (157)$$

where $\text{Tr}(\cdot)$ denotes the trace of a matrix and $\mathbf{L} \in \mathbb{C}^{N \times N}$ is defined as follow:

$$\mathbf{L} = \frac{\beta^2 \sigma^2}{2 + \beta n \sigma^2} \left(\sum_{a=1}^n \mathbf{u}_a \right) \left(\sum_{a=1}^n \mathbf{u}_a \right)^H - \beta \sum_{a=1}^n \mathbf{u}_a \mathbf{u}_a^H. \quad (158)$$

To obtain a more specific perspective on the matrix \mathbf{L} , we here introduce the RS ansatz. Denote by an overlap matrix $\mathbf{Q} \in \mathbb{C}^{(n+1) \times (n+1)}$, whose elements are given by $Q_{ab} = Q_{ba} = (2N)^{-1} \mathbf{x}_a^H \mathbf{x}_b$ ($0 \leq a \leq n; 0 \leq b \leq n$). Then the RS ansatz restricts the values of the overlap matrix \mathbf{Q} to the following:

$$Q_{ab} = \begin{cases} Q, & (a = b = 1, 2, \dots, n) \\ q, & (a > b = 1, 2, \dots, n; b > a = 1, 2, \dots, n) \\ m, & (a = 0, b = 1, 2, \dots, n) \\ \rho, & (a = b = 0) \end{cases}, \quad (159)$$

where we recall $\rho > 0$ is half of the signal density, Q , m and q are real numbers. Due to the symmetry of the replicas, we assume that $\mathbf{x}_a^H \mathbf{x}_b$ for $a \neq b$ are all real numbers. This indicates that \mathbf{L} has three types of eigenvalues:

$$s_1 = -\frac{4N\beta}{2 + \beta n\sigma^2} ((Q - q) + n(q - 2m + \rho)), \quad (160)$$

$$s_2 = -2N\beta(Q - q), \quad (161)$$

$$s_3 = 0, \quad (162)$$

which can be obtained by

$$\mathbf{L} \sum_{a=1}^n \mathbf{u}_a = -\frac{4N\beta}{2 + \beta n\sigma^2} ((Q - q) + n(q - 2m + \rho)) \sum_{a=1}^n \mathbf{u}_a, \quad (163)$$

$$\mathbf{L} \left(\sum_{a=1}^n \mathbf{u}_a - n\mathbf{u}_b \right) = -2N\beta(Q - q) \left(\sum_{a=1}^n \mathbf{u}_a - n\mathbf{u}_b \right), \quad (164)$$

$$\mathbf{L}\mathbf{v}_k = \mathbf{0}, \quad (165)$$

where \mathbf{v}_k are the vector orthogonal to all \mathbf{u}_a , thus there are at least $N - n$ \mathbf{v}_k 's. Now we get the eigenvalue decomposition of \mathbf{L} , thus

$$\exp \left[\sum_{a=0}^n -\frac{k_a}{2} \mathbf{u}_a^H \mathbf{J} \mathbf{u}_a \right] = \exp \left[\sum_{i=1}^n -\frac{1}{2} \mathbf{v}_i^H \mathbf{J} \mathbf{v}_i \right], \quad (166)$$

where

$$\mathbf{v}_1^H \mathbf{v}_1 = 2N \cdot \frac{2\beta}{2 + \beta n\sigma^2} ((Q - q) + n(q - 2m + \rho)), \quad (167)$$

$$\mathbf{v}_i^H \mathbf{v}_i = 2N \cdot \beta(Q - q), \quad 2 \leq i \leq n. \quad (168)$$

Under such constraint, when N is large enough, the logarithm of the averages on \mathbf{A} can be calculated:

$$\begin{aligned} & \mathcal{T}(\mathbf{k}) \\ & \doteq \mathcal{T}(Q, q, m) \\ & = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \int f(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0, \mathbf{A}, \boldsymbol{\xi}) d\boldsymbol{\xi} d\mathbf{A} \\ & = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathbb{E}_{\mathbf{A}} \left\{ \exp \left[\sum_{i=1}^n -\frac{1}{2} \mathbf{v}_i^H \mathbf{J} \mathbf{v}_i - \frac{N\gamma}{2} \ln(1 + \beta n\sigma^2) \right] \right\} \\ & = G \left(-\frac{2\beta(Q - q + n(q - 2m + \rho))}{2 + \beta n\sigma^2}; \mathbf{J} \right) \\ & \quad + (n - 1)G(-\beta(Q - q); \mathbf{J}) - \gamma \ln \left(1 + \frac{\beta n\sigma^2}{2} \right), \end{aligned} \quad (169)$$

where

$$G(x; \mathbf{J}) \doteq \text{extr}_z \left[-\int \rho_{\mathbf{J}}(s) \ln |z - s| ds + zx \right] - \ln |x| - 1. \quad (170)$$

The derivation of (169) is explained in Appendix D.

C.1.2 The derivation of $\mathcal{S}(\mathbf{k}, \mathbf{x}_0)$

In this section, we will derive

$$\mathcal{S}(\mathbf{k}, \mathbf{x}_0) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \int \prod_{a=1}^n d\mathbf{x}_a e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) - \mathbf{k}). \quad (171)$$

We first express the constraint provided by RS ansatz as follow

$$\begin{aligned} & \delta(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) - \mathbf{k}) \\ \doteq & \frac{1}{(2N)^{n^2-1}} \prod_{a=1}^n \left[\delta\left(\|\mathbf{x}_a\|_2^2 - 2NQ\right) \delta\left(\operatorname{Re}(\mathbf{x}_a^H \mathbf{x}_a) - 2Nm\right) \right] \prod_{a \neq b} \delta(\mathbf{x}_a^H \mathbf{x}_b - 2NQ_{ab}) \end{aligned} \quad (172)$$

We proceed with the derivation of (190) by the Fourier transform of the Dirac's delta function:

$$\delta(\mathbf{x}_a^H \mathbf{x}_b - 2NQ_{ab}) = \frac{1}{4\pi j} \int_{c-j\infty}^{c+j\infty} d\tilde{Q}_{ab} \exp\left\{ \frac{\tilde{Q}_{ab}}{2} (\mathbf{x}_a^H \mathbf{x}_b - 2NQ_{ab}) \right\}, \quad (173)$$

where $c \in \mathbb{R}$ is an arbitrary real number. Therefore,

$$\begin{aligned} & \int \prod_{a=1}^n d\mathbf{x}_a e^{-\beta \|\mathbf{x}_a\|_1} \delta(\mathbf{h}(\{\mathbf{x}_a\}_{a=1}^n, \mathbf{x}_0) - \mathbf{k}) \\ = & \int \prod_{a,b} d\frac{\tilde{Q}_{ab}}{4\pi j} \int \prod_{a=1}^n d\mathbf{x}_a \frac{1}{(2N)^{n^2-1}} \prod_{a,b} \left(e^{\frac{\tilde{Q}_{ab}}{2} \mathbf{x}_a^H \mathbf{x}_b - N\tilde{Q}_{ab}Q_{ab}} \right) \prod_{a=1}^n e^{-\beta \|\mathbf{x}_a\|_1} \\ = & \int \prod_{a,b} d\frac{\tilde{Q}_{ab}}{4\pi j} \exp\left\{ N \left[-\sum_{a,b} \tilde{Q}_{ab}Q_{ab} + \frac{1}{N} \ln \int \prod_{a=1}^n d\mathbf{x}_a \exp\left(\sum_{a,b} \frac{\tilde{Q}_{ab}}{2} \mathbf{x}_a^H \mathbf{x}_b - \sum_{a=1}^n \beta \|\mathbf{x}_a\|_1 \right) \right] \right\} \end{aligned} \quad (174)$$

With the saddle point method, we can approximately calculate the integral of \tilde{Q}_{ab} by the function value of the saddle point \tilde{Q}_{ab}^0 . Under the RS ansatz, we find that the saddle point \tilde{Q}_{ab}^0 have only three types of values, which is given as follow

$$\tilde{Q}_{ab}^0 = \begin{cases} -\tilde{Q}, & (a = b = 1, 2, \dots, n) \\ \tilde{q}, & (a > b = 1, 2, \dots, n; b > a = 1, 2, \dots, n) \\ 2\tilde{m}. & (a = 1, 2, \dots, n, b = 0; a = 0, b = 1, 2, \dots, n) \end{cases} \quad (175)$$

Therefore,

$$\begin{aligned} & \mathcal{S}(\mathbf{k}, \mathbf{x}_0) \\ \doteq & \mathcal{S}(Q, q, m) \\ = & \operatorname{extr}_{\tilde{Q}, \tilde{q}, \tilde{m}} \left(n\tilde{Q}Q - n(n-1)\tilde{q}q - 2n\tilde{m}m + \lim_{N \rightarrow \infty} \frac{1}{N} \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \right), \end{aligned} \quad (176)$$

where

$$\begin{aligned}
& \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \\
& \doteq \ln \int \prod_{a=1}^n d\mathbf{x}_a \exp \left(\sum_{a=1}^n -\frac{\tilde{Q}}{2} \mathbf{x}_a^H \mathbf{x}_a + \sum_{a \neq b} \frac{\tilde{q}}{2} \mathbf{x}_a^H \mathbf{x}_b + \sum_{a=1}^n \tilde{m} \operatorname{Re}(\mathbf{x}_0^H \mathbf{x}_a) - \sum_{a=1}^n \beta \lambda \|\mathbf{x}_a\|_1 \right) \\
& = \sum_{i=1}^N \ln \int \prod_{a=1}^n dx_{a,i} \exp \left(\sum_{a=1}^n -\frac{\tilde{Q}}{2} |x_{a,i}|^2 + \sum_{a \neq b} \frac{\tilde{q}}{2} x_{a,i}^* x_{b,i} + \sum_{a=1}^n \tilde{m} \operatorname{Re}(x_{0,i}^* x_{a,i}) - \sum_{a=1}^n \beta \lambda |x_{a,i}| \right)
\end{aligned}$$

According to the well-known identity

$$\mathbb{E}_z \exp \left(\operatorname{Re}(\sqrt{2}\lambda^* z) \right) = \exp \left(\frac{|\lambda|^2}{2} \right), \quad (178)$$

where z is a standard complex Gaussian variable, one can find that

$$\begin{aligned}
& \exp \left(\sum_{a \neq b} \frac{\tilde{q}}{2} x_{a,i}^* x_{b,i} \right) \\
& = \exp \left(\frac{1}{2} \left| \sum_{a=1}^n \sqrt{\tilde{q}} x_{a,i} \right|^2 - \frac{\tilde{q}}{2} \sum_{a=1}^n |x_{a,i}|^2 \right) \\
& = \int Dz_i \exp \left(\operatorname{Re} \left(\sqrt{2\tilde{q}} z_i^* \sum_{a=1}^n x_{a,i} \right) - \frac{\tilde{q}}{2} \sum_{a=1}^n |x_{a,i}|^2 \right). \quad (179)
\end{aligned}$$

Thus,

$$\begin{aligned}
& \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \\
& = \sum_{i=1}^N \ln \int \prod_{a=1}^n dx_{a,i} Dz_i \exp \left\{ \sum_{a=1}^n \left[-\frac{\tilde{Q} + \tilde{q}}{2} |x_{a,i}|^2 + \operatorname{Re} \left((\sqrt{2\tilde{q}} z_i + \tilde{m} x_{0,i})^* x_{a,i} \right) - \beta \lambda |x_{a,i}| \right] \right\} \\
& = \sum_{i=1}^N \ln \int Dz_i \left(\int dx_i \exp \left\{ -\frac{\tilde{Q} + \tilde{q}}{2} |x_i|^2 + \operatorname{Re} \left((\sqrt{2\tilde{q}} z_i + \tilde{m} x_{0,i})^* x_i \right) - \beta \lambda |x_i| \right\} \right)^n. \quad (180)
\end{aligned}$$

C.2 Final result

We continue working on the result of equation (149). Note that such result holds for $n = 1, 2, \dots$. In the following derivation, we assume the analytic continuation to real n from the expression obtained by evaluating the relevant quantity only for positive

integers n . Thus,

$$\begin{aligned}
& \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathbb{E}_{\mathbf{A}, \boldsymbol{\xi}} [Z^n] \\
&= \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \text{extr}_{Q, q, m} \{ \mathcal{T}(Q, q, m) + \mathcal{S}(Q, q, m) \} \\
&= \text{extr}_{Q, q, m} \left\{ \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{T}(Q, q, m) + \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{S}(Q, q, m) \right\}, \tag{181}
\end{aligned}$$

where

$$\begin{aligned}
& \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{T}(Q, q, m) \\
&= \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \left\{ G \left(-\frac{2\beta(Q - q + n(q - 2m + \rho))}{2 + \beta n \sigma^2}; \mathbf{J} \right) \right. \\
&\quad \left. + (n - 1)G(-\beta(Q - q); \mathbf{J}) - \frac{\gamma}{2} \ln(1 + \beta n \sigma^2) \right\} \\
&= -G'(-\beta(Q - q); \mathbf{J}) \left(\beta(q - 2m + \rho) - \frac{\beta^2 \sigma^2 (Q - q)}{2} \right) \\
&\quad + G(-\beta(Q - q); \mathbf{J}) - \frac{\beta \gamma \sigma^2}{2}, \tag{182}
\end{aligned}$$

and

$$\begin{aligned}
& \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{S}(Q, q, m) \\
&= \text{extr}_{\tilde{Q}, \tilde{q}, \tilde{m}} \left\{ \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \left[n\tilde{Q}Q - n(n - 1)\tilde{q}q - 2n\tilde{m}m + \lim_{N \rightarrow \infty} \frac{1}{N} \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \right] \right\} \\
&= \text{extr}_{\tilde{Q}, \tilde{q}, \tilde{m}} \left\{ \tilde{Q}Q + \tilde{q}q - 2\tilde{m}m + \lim_{N \rightarrow \infty} \frac{1}{N} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \right\}. \tag{183}
\end{aligned}$$

The part of the partial of $\mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m})$ is given by

$$\begin{aligned}
& \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \\
&= \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \sum_{i=1}^N \ln \int \text{D}z_i \left(\int \text{d}x_i \exp \left\{ -\frac{\tilde{Q} + \tilde{q}}{2} |x_i|^2 + \text{Re} \left((\sqrt{2\tilde{q}}z_i + \tilde{m}x_{0,i})^* x_i \right) - \beta \lambda |x_i| \right\} \right)^n \\
&= \sum_{i=1}^N \int \text{D}z_i \ln \left(\int \text{d}x_i \exp \left\{ -\frac{\tilde{Q} + \tilde{q}}{2} |x_i|^2 + \text{Re} \left((\sqrt{2\tilde{q}}z_i + \tilde{m}x_{0,i})^* x_i \right) - \beta \lambda |x_i| \right\} \right). \tag{184}
\end{aligned}$$

When $\beta \rightarrow \infty$, the following substitutions of variables are employed

$$\begin{aligned}
\chi &\doteq \beta(Q - q), \\
\hat{Q} &\doteq \beta^{-1}(\tilde{Q} + \tilde{q}), \\
\hat{\chi} &\doteq \beta^{-2}\tilde{q}, \\
\hat{m} &\doteq \beta^{-1}\tilde{m}.
\end{aligned}$$

With these substitutions, we have

$$\begin{aligned}
& \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{T}(Q, q, m) \\
&= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \left[-G'(-\beta(Q - q); \mathbf{J}) \left(\beta(q - 2m + \rho) - \frac{\beta^2 \sigma^2 (Q - q)}{2} \right) \right. \\
&\quad \left. + G(-\beta(Q - q); \mathbf{J}) - \frac{\beta \gamma \sigma^2}{2} \right] \\
&= G'(-\chi; \mathbf{J})(Q - 2m + \rho - \frac{\chi}{2} \sigma^2) + \frac{\gamma}{2} \sigma^2, \tag{185}
\end{aligned}$$

and

$$\begin{aligned}
& \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{S}_n(Q, q, m) \\
&= \text{extr}_{\tilde{Q}, \tilde{q}, \tilde{m}} \left\{ \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \left[\tilde{Q}Q + \tilde{q}q - 2\tilde{m}m + \lim_{N \rightarrow \infty} \frac{1}{N} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \right] \right\} \\
&= \text{extr}_{\hat{Q}, \hat{q}, \hat{m}} \left\{ -\hat{Q}Q + \hat{\chi}\chi + 2\hat{m}m + \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \lim_{N \rightarrow \infty} \frac{1}{N} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{R}(\hat{Q}, \hat{q}, \hat{m}) \right\}. \tag{186}
\end{aligned}$$

With the saddle point method, one has

$$\begin{aligned}
& \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \lim_{N \rightarrow \infty} \frac{1}{N} \lim_{n \rightarrow 0} \frac{\partial}{\partial n} \mathcal{R}(\tilde{Q}, \tilde{q}, \tilde{m}) \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \sum_{i=1}^N \int \text{D}z_i \ln \left(\int \text{d}x_i \exp \left\{ -\frac{\tilde{Q} + \tilde{q}}{2} |x_i|^2 + \text{Re} \left((\sqrt{2\tilde{q}}z_i + \tilde{m}x_{0,i})^* x_i \right) - \beta \lambda |x_i| \right\} \right) \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \text{D}z_i \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \ln \left(\int \text{d}x_i \exp \left\{ -\beta \left[\frac{\hat{Q}}{2} |x_i|^2 - \text{Re} \left((\sqrt{2\hat{\chi}}z_i + \hat{m}x_{0,i})^* x_i \right) + \lambda |x_i| \right] \right\} \right) \\
&= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \min_{x_i} \left[\frac{\hat{Q}}{2} |x_i|^2 - \text{Re} \left((\hat{m}x_{0,i} + \sqrt{2\hat{\chi}}z_i)^* x_i \right) + \lambda |x_i| \right] \text{D}z_i. \tag{187}
\end{aligned}$$

Finally, associating with the result (185), (186) and (187) yields the expression of the average free energy density in zero-temperature limit

$$\begin{aligned}
f &= \text{extr}_{Q, \hat{Q}, \chi, \hat{\chi}, m, \hat{m}} \left\{ G'(-\chi; \mathbf{J})(Q - 2m + \rho - \frac{\chi}{2} \sigma^2) + \frac{\gamma}{2} \sigma^2 - \hat{Q}Q + \hat{\chi}\chi + 2\hat{m}m \right. \\
&\quad \left. + \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int \min_{x_i} \left[\frac{\hat{Q}}{2} |x_i|^2 - \text{Re} \left((\hat{m}x_{0,i} + \sqrt{2\hat{\chi}}z_i)^* x_i \right) + \lambda |x_i| \right] \text{D}z_i \right\} \tag{188}
\end{aligned}$$

D Derivation of averages with respect to row-orthogonal matrices

We here would like to derive the averages with respect to row-orthogonal matrices $\mathbf{A} \in \mathbb{C}^{M \times N}$ as follow:

$$L_{\text{ROM}}(-\mathbf{x}) \doteq \lim_{N \rightarrow \infty} \frac{1}{N} \ln \mathbb{E}_{\mathbf{A}} \left[\exp \left\{ - \sum_{i=1}^n \frac{1}{2} \mathbf{v}_i^H \mathbf{A}^H \mathbf{A} \mathbf{v}_i \right\} \right], \quad (189)$$

where $\mathbf{v}_i \in \mathbb{C}^{N \times 1}$ ($1 \leq i \leq n$) is a set of orthogonal vectors satisfying $\mathbf{v}_i^H \mathbf{v}_j = 2N\nu_i \delta_{ij}$, δ_{ij} is unity for $i = j$ and vanishes otherwise, $\mathbf{x} = [\nu_1, \nu_2, \dots, \nu_n]^T$ and the average is taken over all the row-orthogonal matrices. Denote by $\mathbf{J} = \mathbf{A}^H \mathbf{A}$, then the Hermitian matrix $\mathbf{J} \in \mathbb{C}^{N \times N}$ can be factorized as $\mathbf{J} = \mathbf{O} \mathbf{D} \mathbf{O}^H$ with an orthogonal matrix $\mathbf{O} \in \mathbb{C}^{N \times N}$ and a diagonal matrix \mathbf{D} , in which its on-diagonal elements are eigenvalues of \mathbf{J} . Due to the fact that \mathbf{A} is row-orthogonal, the eigenvalues of \mathbf{J} are fixed to 1 or 0, that is, \mathbf{D} does not vary with \mathbf{A} . In addition, denote by $\tilde{\mathbf{v}}_i = \mathbf{O}^H \mathbf{v}_i$, the ℓ_2 -norm of $\tilde{\mathbf{v}}_i$ is maintained: $\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i = 2N\nu_i$. Therefore, the average on \mathbf{A} reduces to the average on $\tilde{\mathbf{v}}_i$, which is given by

$$L_{\text{ROM}}(-\mathbf{x}) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left\{ \frac{\int \prod_{i=1}^n d\tilde{\mathbf{v}}_i \exp \left(- \sum_{i=1}^n \frac{1}{2} \tilde{\mathbf{v}}_i^H \mathbf{D} \tilde{\mathbf{v}}_i \right) \prod_{i=1}^n \delta(\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i - 2N\nu_i)}{\int \prod_{i=1}^n d\tilde{\mathbf{v}}_i \prod_{i=1}^n \delta(\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i - 2N\nu_i)} \right\}. \quad (190)$$

Note that here we discard the fact that $\{\mathbf{v}_i\}_{i=1}^n$ are orthogonal to each other, for when the dimension of vectors $N \rightarrow \infty$, the probability that any finite number of vectors are orthogonal to each other is 1.

We proceed with the derivation of (190) by the Fourier transform of the Dirac's delta function:

$$\delta(\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i - 2N\nu_i) = \frac{1}{4\pi j} \int_{c-j\infty}^{c+j\infty} d\Lambda_i \exp \left\{ - \frac{\Lambda_i}{2} (\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i - 2N\nu_i) \right\}, \quad (191)$$

where $c \in \mathbb{R}$ is an arbitrary real number. Thus,

$$\int \prod_{i=1}^n [d\tilde{\mathbf{v}}_i \delta(\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i - 2N\nu_i)] = \frac{(2\pi)^{nN}}{(4\pi j)^n} \int_{c-j\infty}^{c+j\infty} \prod_{i=1}^n d\Lambda_i \exp \left\{ -N \sum_{i=1}^n (\ln \Lambda_i - \Lambda_i \nu_i) \right\}, \quad (192)$$

where we use the complex Gaussian integration formula

$$\frac{1}{(2\pi)^N} \int \exp \left(- \frac{1}{2} \mathbf{z}^H \mathbf{M} \mathbf{z} + \text{Re}(\mathbf{b}^H \mathbf{z}) \right) d\mathbf{z} = \frac{1}{\det \mathbf{M}} \exp \left(\frac{1}{2} \mathbf{b}^H \mathbf{M}^{-1} \mathbf{b} \right), \quad (193)$$

where $\mathbf{b}, \mathbf{z} \in \mathbb{C}^N$ and \mathbf{M} is symmetric positive definite.

Applying saddle-point method, we rewrite (192) as

$$\begin{aligned}
& \int \prod_{i=1}^n [\mathrm{d}\tilde{\mathbf{v}}_i \delta(\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i - 2N\nu_i)] \\
&= (2\pi)^{nN} \exp \left\{ -N \left(\sum_{i=1}^n \operatorname{extr}_{\{\Lambda_i\}} [\ln \Lambda_i - \Lambda_i \nu_i] + O\left(\frac{1}{N}\right) \right) \right\} \\
&= (2\pi)^{nN} \exp \left\{ -N \left(\sum_{i=1}^n (-\ln \nu_i - 1) + O\left(\frac{1}{N}\right) \right) \right\}. \tag{194}
\end{aligned}$$

With similar procedure, the numerator of (190) can be expressed as

$$\begin{aligned}
& \int \prod_{i=1}^n \mathrm{d}\tilde{\mathbf{v}}_i \exp \left(- \sum_{i=1}^n \frac{1}{2} \tilde{\mathbf{v}}_i^H \mathbf{D} \tilde{\mathbf{v}}_i \right) \prod_{i=1}^n \delta(\tilde{\mathbf{v}}_i^H \tilde{\mathbf{v}}_i - 2N\nu_i) \\
&= \frac{1}{(4\pi j)^n} \int_{c-j\infty}^{c+j\infty} \prod_{i=1}^n \mathrm{d}\Lambda_i \int \prod_{i=1}^n \mathrm{d}\tilde{\mathbf{v}}_i \exp \left\{ \sum_{i=1}^n \left(-\frac{1}{2} \tilde{\mathbf{v}}_i^H (\Lambda_i \mathbf{I}_N + \mathbf{D}) \tilde{\mathbf{v}}_i + N\Lambda_i \nu_i \right) \right\} \\
&= \frac{(2\pi)^{nN}}{(4\pi j)^n} \int_{c-j\infty}^{c+j\infty} \prod_{i=1}^n \mathrm{d}\Lambda_i \exp \left\{ \sum_{i=1}^n \left(- \sum_{m=1}^N \ln(\Lambda_i + d_m) + N\Lambda_i \nu_i \right) \right\} \\
&= \frac{(2\pi)^{nN}}{(4\pi j)^n} \int_{c-j\infty}^{c+j\infty} \prod_{i=1}^n \mathrm{d}\Lambda_i \exp \left\{ N \left(\sum_{i=1}^n \left(-\frac{1}{N} \sum_{m=1}^N \ln(\Lambda_i + d_m) + \Lambda_i \nu_i \right) \right) \right\}, \tag{195}
\end{aligned}$$

where d_m ($1 \leq m \leq N$) are the diagonal elements of \mathbf{D} , thus are the eigenvalues of \mathbf{J} . When N are large enough, we have the following approximation:

$$\frac{1}{N} \sum_{m=1}^N \ln(\Lambda_i + d_m) = \int \mathrm{d}\lambda \rho_{\mathbf{J}}(\lambda) \ln(\Lambda_i + \lambda). \tag{196}$$

Therefore, by using such approximation and combining (194) and (195), (190) can be expressed as

$$\begin{aligned}
L_{\text{ROM}}(-\mathbf{x}) &= \lim_{N \rightarrow \infty} \frac{1}{N} \ln \left\{ \frac{1}{(4\pi j)^n} \int_{c-j\infty}^{c+j\infty} \prod_{i=1}^n \mathrm{d}\Lambda_i \cdot \right. \\
&\quad \left. \exp \left[N \left(\sum_{i=1}^n \left(- \int \mathrm{d}\lambda \rho_{\mathbf{J}}(\lambda) \ln(\Lambda_i + \lambda) + \Lambda_i \nu_i - \ln \nu_i - 1 \right) + O\left(\frac{1}{N}\right) \right) \right] \right\} \\
&= \sum_{i=1}^n \left(\operatorname{extr}_{\Lambda_i} \left[- \int \mathrm{d}\lambda \rho_{\mathbf{J}}(\lambda) \ln(\Lambda_i + \lambda) + \Lambda_i \nu_i \right] - \ln \nu_i - 1 \right) \\
&= \sum_{i=1}^n \left(\operatorname{extr}_{\Lambda_i} \left[- \int \mathrm{d}\lambda \rho_{\mathbf{J}}(\lambda) \ln |\Lambda_i - \lambda| - \Lambda_i \nu_i \right] - \ln |-\nu_i| - 1 \right). \tag{197}
\end{aligned}$$

Recall the definition of function G :

$$G(x; \mathbf{J}) \doteq \operatorname{extr}_z \left[- \int \rho_{\mathbf{J}}(s) \ln |z - s| \mathrm{d}s + zx \right] - \ln |x| - 1. \tag{198}$$

We finally obtain:

$$L_{\text{ROM}}(-\mathbf{x}) = \sum_{i=1}^n G(-x_i; \mathbf{J}). \quad (199)$$

E Derivation of χ through linear response argument

Denote by

$$Gi_0(\mathbf{m}, \mathbf{h}, Q, \Lambda) = \text{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \frac{1}{\beta} \ln \int e^{-\frac{\beta}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \beta \text{Re}(\mathbf{h}^H \mathbf{x}) - \frac{\beta}{2} \Lambda \|\mathbf{x}\|_2^2 - \beta \lambda \|\mathbf{x}\|_1} d\mathbf{x}, \quad (200)$$

By recalling the definition of the Boltzmann average $\langle \mathbf{x} \rangle$, one can find that

$$\left. \frac{\partial Gi_0}{\partial \text{Re}(h_i)} \right|_{h_i=h_{0,i}} = \text{Re}(m_i) - \text{Re}(\langle x_i \rangle), \quad (201)$$

$$\left. \frac{\partial Gi_0}{\partial \text{Im}(h_i)} \right|_{h_i=h_{0,i}} = \text{Im}(m_i) - \text{Im}(\langle x_i \rangle), \quad (202)$$

and further

$$\left. \frac{\partial^2 Gi_0}{\partial \text{Re}(h_i)^2} \right|_{h_i=h_{0,i}} + \left. \frac{\partial^2 Gi_0}{\partial \text{Im}(h_i)^2} \right|_{h_i=h_{0,i}} = \beta \left(\langle |x_i|^2 \rangle - |\langle x_i \rangle|^2 \right), \quad (203)$$

where \mathbf{h}_0 is the extreme value of G_0 , h_i and $h_{0,i}$ is the i -th entry of \mathbf{h} and \mathbf{h}_0 , respectively. Therefore,

$$\chi = \frac{1}{2N} \sum_{i=1}^N \left(\left. \frac{\partial^2 Gi_0}{\partial \text{Re}(h_i)^2} \right|_{h_i=h_{0,i}} + \left. \frac{\partial^2 Gi_0}{\partial \text{Im}(h_i)^2} \right|_{h_i=h_{0,i}} \right). \quad (204)$$

We compute χ with the following approximation:

$$\chi = \frac{1}{2N} \sum_{i=1}^N \left(\left. \frac{\partial^2 Gi_1}{\partial \text{Re}(h_i)^2} \right|_{h_i=h_{1,i}} + \left. \frac{\partial^2 Gi_1}{\partial \text{Im}(h_i)^2} \right|_{h_i=h_{1,i}} \right), \quad (205)$$

where

$$\begin{aligned} Gi_1(\mathbf{m}, \mathbf{h}, Q, \Lambda) &= \text{Re}(\mathbf{h}^H \mathbf{m}) - N\Lambda Q - \sum_{i=1}^N \frac{(|h_i| - \lambda)^2}{2\Lambda} \cdot \Theta(|h_i| - \lambda) \\ &\quad - \frac{N}{\beta} G(-\chi; \mathbf{J}) + \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{m}\|_2^2. \end{aligned} \quad (206)$$

and \mathbf{h}_1 is the extreme value of Gi_1 with $h_{1,i}$ denoting the i -th entry of \mathbf{h}_1 . Hence, (76) can be derived from (205).

References

- [1] L. Xiao, Y. Liu, T. Huang, X. Liu, and X. Wang, “Distributed target detection with partial observation,” *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1551–1565, 2018.
- [2] T. Huang, Y. Liu, X. Xu, Y. C. Eldar, and X. Wang, “Analysis of frequency agile radar via compressed sensing,” *IEEE Transactions on Signal Processing*, vol. 66, no. 23, pp. 6228–6240, 2018.
- [3] L. Wang, T. Huang, and Y. Liu, “Randomized stepped frequency radars exploiting block sparsity of extended targets: A theoretical analysis,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 1378–1393, 2021.
- [4] D. Cohen and Y. C. Eldar, “Sub-nyquist radar systems: Temporal, spectral, and spatial compression,” *IEEE Signal Processing Magazine*, vol. 35, no. 6, pp. 35–58, 2018.
- [5] S. Na, K. V. Mishra, Y. Liu, Y. C. Eldar, and X. Wang, “Tendsur: Tensor-based 4d sub-nyquist radar,” *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 237–241, 2018.
- [6] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [7] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [8] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky *et al.*, “A method for large-scale ℓ_1 -regularized least squares,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.
- [9] E. Greenshtein and Y. Ritov, “Persistence in high-dimensional linear predictor selection and the virtue of overparametrization,” *Bernoulli*, vol. 10, no. 6, pp. 971–988, 2004.
- [10] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, “Simultaneous analysis of lasso and dantzig selector,” *The Annals of statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.
- [11] E. Candes and T. Tao, “The dantzig selector: Statistical estimation when p is much larger than n ,” *The Annals of Statistics*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [12] G. Raskutti, M. J. Wainwright, and B. Yu, “Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls,” *IEEE transactions on information theory*, vol. 57, no. 10, pp. 6976–6994, 2011.

- [13] N. Meinshausen and P. Bühlmann, “High-dimensional graphs and variable selection with the lasso,” *The annals of statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.
- [14] P. Zhao and B. Yu, “On model selection consistency of lasso,” *The Journal of Machine Learning Research*, vol. 7, pp. 2541–2563, 2006.
- [15] M. J. Wainwright, “Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso),” *IEEE transactions on information theory*, vol. 55, no. 5, pp. 2183–2202, 2009.
- [16] M. Bayati and A. Montanari, “The dynamics of message passing on dense graphs, with applications to compressed sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, 2011.
- [17] —, “The lasso risk for gaussian matrices,” *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 1997–2017, 2011.
- [18] A. Javanmard and A. Montanari, “Hypothesis testing in high-dimensional regression under the gaussian random design model: Asymptotic theory,” *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6522–6554, 2014.
- [19] L. Anitori, A. Maleki, M. Otten, R. G. Baraniuk, and P. Hoogeboom, “Design and analysis of compressed sensing radar detectors,” *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 813–827, 2012.
- [20] T. Takahashi and Y. Kabashima, “A statistical mechanics approach to de-biasing and uncertainty estimation in lasso for random measurements,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2018, no. 7, p. 073405, 2018.
- [21] T. Tanaka, “A statistical-mechanics approach to large-system analysis of cdma multiuser detectors,” *IEEE Transactions on Information theory*, vol. 48, no. 11, pp. 2888–2910, 2002.
- [22] D. Guo and S. Verdú, “Randomly spread cdma: Asymptotics via statistical physics,” *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.
- [23] T. Tanaka and M. Okada, “Approximate belief propagation, density evolution, and statistical neurodynamics for cdma multiuser detection,” *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 700–706, 2005.
- [24] A. Campo, A. G. i Fabregas, and E. Biglieri, “Large-system analysis of multiuser detection with an unknown number of users: A high-snr approach,” *IEEE transactions on information theory*, vol. 57, no. 6, pp. 3416–3428, 2011.
- [25] Y. Wu and S. Verdú, “Optimal phase transitions in compressed sensing,” *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6241–6263, 2012.

- [26] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, p. 457, 1967.
- [27] M. Opper, O. Winther, and M. J. Jordan, “Expectation consistent approximate inference.” *Journal of Machine Learning Research*, vol. 6, no. 12, 2005.
- [28] M. Opper and D. Saad, *Advanced mean field methods: Theory and practice*. MIT press, 2001.
- [29] F. J. Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [30] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.