Progress In Electromagnetics Research, Vol. 142, 559–578, 2013

LASSO BASED PERFORMANCE EVALUATION FOR SPARSE ONE-DIMENSIONAL RADAR PROBLEM UNDER RANDOM SUB-SAMPLING AND GAUSSIAN NOISE

Yin Xiang*, **Bingchen Zhang, and Wen Hong**

Institute of Electronics, Chinese Academy of Sciences, Beijing, China

Abstract—Sparse microwave imaging is the combination of microwave imaging and sparse signal processing, which aims to extract physical and geometry information of sparse or transformed sparse scene from least number of radar measurements. As a primary investigation on its performance, this paper focuses on the performance guarantee for a one-dimensional radar, which detects delays of several point targets located at a sparse scene via randomly sub-sampling of radar returns. Based on the Lasso framework, the quantity relationship among three important factors is discussed, including the sub-sampling ratio ρ_M, sparse ratio ρ_K and signal-to-noise ratio (SNR), where ρ_M is the ratio of number of random sub-sampling to that of Nyquist's sampling, and ρ_K is the ratio of sparsity to the number of unknowns. In particular, to ensure correct delay detection and accurate back scattering coefficient reconstruction for each target, one needs ρ_M to be greater than $C(\rho_K)\rho_K \log N$ and the input SNR be of order $\log N$, where N is the number of range cells in scene.

1. INTRODUCTION

Sparse microwave imaging is a new concept that combines microwave imaging and sparse signal processing. It seeks to efficiently acquire the target's geometry and physical characteristics, which could be sparsely represented, from limit radar observations or measurements in temporal, frequency, spatial or spectrum domain [1]. In all features of sparse microwave imaging, one of the sparse signal-processing tools named as compressive sensing (CS) plays an important role. CS was developed by Donoho [2], and Candes and Wakin [3] in 2004.

Received 14 March 2013, Accepted 29 July 2013, Scheduled 22 September 2013
* Corresponding author: Yin Xiang (xy_overlimit@sina.cn).

It combines the processing of sampling and data compression, which allows one to measure the sparse of compressed signal via very small number of incoherent linear projections, then recover it accurately via non-linear optimization methods. To date, to capture a signal vector $\mathbf{x}$ of dimension N with K $(K \ll N)$ freedom, one only needs $M \sim \mathbf{O}(K \log N)$ measurements. It means that the sufficient number of samples measured by the sparse microwave imaging radar is proportional to the sparsity of scene under observation, but not to the bandwidth of radar system, when CS is applied. Accordingly, compared with existing radar systems, the sparse microwave imaging radar may have the potential ability to reduce the system data amount and complexity [4].

In recent years, plenty of works have investigated the theory and application of sparse microwave imaging radar, and a few of them are mentioned here. In [5], Baraniuk and Steeghs suggest to apply CS to radar imaging. After that, the CS based SAR imaging [6, 7], ISAR imaging [8, 9] and other applications of microwave imaging [10–12] are extensively discussed. A more detailed and comprehensive account can be found in [4, 13, 14]. In particular, the principle of sparse microwave imaging is fundamentally established in [4].

The works mentioned above show big progress of sparse microwave imaging, but more efforts need to be made. If sparse microwave imaging radar is going to be practical, one fatal question that we have to face is how to evaluate the system performance. Namely, how many samples will be sufficient against the given sparsity of scene and the signal-to-noise-ratio (SNR) of radar measurement.

In this paper, we primarily attempt to answer the above question. We focus on the fundamental radar microwave imaging problem. In this problem, the radar transmits band-limit pulse to a one-dimensional sparse scene, and the echo is randomly sub-sampled to form compressed measurements. The term 'sub-sample' means that the average sampling rate is lower than the Nyquist's rate and that it is randomly generated to form a proper CS measurement matrix, which mainly determines the performance of system.

In the CS theory, several criteria are developed to investigate whether the measurement matrix has a good performance for the recovering of the sparse vector, such as mutual coherence [15], restricted isometric property (RIP) [16], restricted orthogonality property (ROP) [17], exact reconstruction criteria (ERC) [18]. In general, mutual coherence is relatively easy to calculate, but the results based on mutual coherence are sub-optimal compared with the results on RIP, ROP, and ERC. RIP, ROP, and ERC based performance guarantees are better, but still too strict to get a numerical constant

C in the performance guarantee formed as M $(M > C \cdot K \log N)$.

In this paper, we would like to derive the performance guarantee based on the Lasso framework. The sparse recovery property has already been extensively demonstrated in [18]. With the help of the existed theoretical tools, a quantity bound related to the aforementioned three factors is given by considering the particular measurement matrix of one-dimensional sparse microwave imaging radar. The rest of the paper is organized as follows. Section 2 briefly introduces the model of radar problem and Lasso framework. Section 3 discusses the performance guarantee for noiseless case. Section 4 is about the same task for the noisy case. Section 5 makes simple simulations to support our analysis, and the last section concludes the whole paper.

2. PROBLEM SETUP

2.1. Radar Problem Based on Sparse Signal Processing

In this model, one-dimensional sparse scene, which contains only a small number of point scatters, is measured by randomly sub-sampled radar echo. The radar return equation is written as a linear convolution form

$$s(t) = \int h(t-\tau)x(\tau)\mathrm{d}\tau + n(t) \tag{1}$$

where, s denotes the echo, and h denotes the transmitted pulse and is considered as linear modulation signal written as

$$h(\tau) = \mathrm{rect}\left(\frac{\tau}{T}\right)\exp\left(\o\pi K\tau^2\right), \quad \tau \in \left[-\frac{T}{2}, \frac{T}{2}\right) \tag{2}$$

where, T, K are the time duration and linear rate of the linear modulation signal, respectively, and $\mathrm{rect}\,(x) = \begin{cases} 1, & |x| \le 0.5 \\ 0, & \text{else} \end{cases}$, $x(\tau) = \sum_{k=1}^{K} x_k \delta(\tau - \tau_k)$. x_k stands for the back-scattering coefficients of the target k, n the additive Gaussian noise, t the fast time, and τ the delay of each target.

Let $\mathbf{y} = \mathbf{\Theta} \circ s = [s(t_1), \ldots, s(t_M)]^T$ be a set of non-uniform samples randomly sampled from $s(t)$, where $\mathbf{\Theta}$ denotes the linear operation for sub-sampling. Together with (1), each sample of $s(t)$ can be represented as

$$y_m = \int h(t_m - \tau)x(\tau)\mathrm{d}\tau + n(t_m), \quad m = 1, \ldots, M \tag{3}$$

where, M is the number of sub-samplers.

To solve Equation (3) via sparse signal processing methods, $x(\tau)$ should be discreted to the vector form as $\mathbf{x} = [x(\Delta\tau), \ldots, x(n\Delta\tau), \ldots]$, $n = 1, \ldots, N$, where $\Delta\tau$ is selected to be $\Delta\tau = 1/B$ so as to preserve the stability of sparse recovery due to [19], and N is the number of the range cells, i.e., the number of unknowns. Then (3) is changed as

$$\mathbf{y} = \mathbf{\Theta}\mathbf{H}\mathbf{x} + \mathbf{n} = \mathbf{\Phi}\mathbf{x} + \mathbf{n} \tag{4}$$

where, $\mathbf{\Phi} = \mathbf{\Theta}\mathbf{H} \in C^{M\times N}$ is the measurement matrix of the sparse microwave radar. Without losing any generality, assume that $\mathbf{y}$ is randomly selected from the Nyquist's samples of $s(t)$, then $\mathbf{H}$ can be expressed as a discrete linear convolution matrix (see Fig. 1).

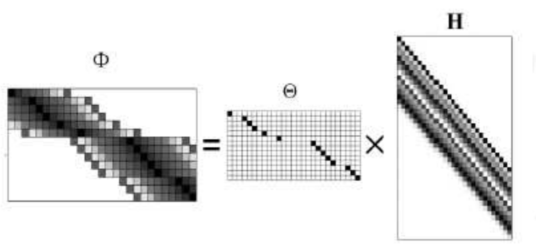

Figure 1. Construction of measurement matrix of radar with randomly sub-samplers, Zhang et al. [4].

Let N_S, N_H be the number of Nyquist's samples in $s(t)$, $h(t)$, respectively. The sub-sampling ratio is defined as

$$\rho_M = \frac{M}{N_s} \tag{5}$$

Assume that $\mathbf{x}$ is sparse and includes up to K non-zero entries, all of which should not be constrained in small regions; otherwise some local region will be too dense to be correctly recovered. For instance, let an extremely large scene include N ($N \gg N_H$) range cells, but contain one N_H-length continuous region that is full filled. Though other parts of the scene can be empty, it is obvious that the full filled region can never be reconstructed from sub-Nyquist's samples/measurements. As a result, the sparse ratio ρ_K is needed to be defined as a value such that any N_h-length continues range cells contain no more than $\rho_K N_h$ non-zero entries, that is

$$\rho_K = \sup_k \frac{\|\mathbf{x}_\Xi\|_0}{N_h}, \quad \Xi = \{k, k+1, \ldots, k+N_h-1\} \tag{6}$$

Assume that the complex noise $\mathbf{n}$ satisfies $n[i] \overset{i.i.d.}{\sim} \mathrm{N}(0, \sigma_0^2)$, $i = 1, \ldots, M$, where N() stands for the normal distribution, and σ_0^2 is the variance of noise and also means its average power. Then the SNR of radar can be defined as the ratio of signal average power to σ_0^2 [20]. For instance, if there is only one point target located at the scene, the SNR equals $|x|^2/\sigma_0^2$, where $|x|$ is the amplitude of the target echo. However, for a scene that contains $K \gg 1$ targets, it is hard to express the particular average signal power. As a result, the SNR used in this paper is given as

$$\mathrm{SNR} = \frac{\sum_{k=1}^{K} |x_k|^2 / K}{\sigma_0^2} = \frac{\|\mathbf{x}\|_2^2}{K\sigma_0^2} \tag{7}$$

which is a simple generalization of the one-target case.

As the signal vector, $\mathbf{x}$ is sparse in its own domain and can be recovered by solving the following convex optimization problem [3]

$$\hat{\mathbf{x}} = \arg_{\mathbf{x}} \min \|\mathbf{x}\|_1 \; s.t. \; \|\mathbf{y} - \Phi\mathbf{x}\|_2 \leq \varepsilon \tag{8}$$

where $\|\|_1$ stands for the L_1-norm of vector and $\varepsilon = \|\mathbf{n}\|_2$. This L_1-constrained form is known by the name of Basis Pursuit De-Noising (BPDN), which can be transformed into the flowing L_1-penalized optimization problem, also called as Lasso, by the statistic community

$$\hat{\mathbf{x}} = \arg_{\mathbf{x}} \min L(\mathbf{x}); \quad L(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \gamma \|\mathbf{x}\|_1 \tag{9}$$

where, $\gamma > 0$ is regularization parameter, which is chosen due to the level of back-projection noise $\mathbf{\Phi}^*\mathbf{n}$, and plays as a role of de-noising threshold clearing all recovery noise outside a finite support. In the following paragraphs, we will make our analysis about the performance evaluation under the Lasso framework.

2.2. A General Sparse Recovery Property of Lasso

Lasso (least absolute shrinkage and selection operator), is used by the statistic community to demonstrate a particular regularization version of least squares (see (9) for its definition). From the Bayesian aspect, Lasso performs a zero-mean Laplace priori on the under-test signal $\mathbf{x}$. As regularization goes stronger (γ becomes larger), Lasso pushes more and more entries of the recovered solution $\hat{\mathbf{x}}$ to be zeros. Accordingly, Lasso can be used to find the sparse solution of given under-determined linear problems formed as (4).

In a huge number of research papers about Lasso, the general sparse recovery property is extensively studied, and no limitation on the kind of the measurement matrix is made in [18]. One can make the following statements from [18]:

Let $\mathbf{x}$, $\hat{\mathbf{x}}$ be the original and the recovered sparse signal via Lasso, respectively, $\Lambda = \text{supp}(\mathbf{x})$ denotes support of $\mathbf{x}$, then if and only if

$$\left\| \boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}^c} \mathbf{P}^{\perp}_{V_\Lambda} \mathbf{n} + \gamma \boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}^c} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \left(\boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \right)^{-\mathbf{1}} \mathbf{g}_{\boldsymbol{\Lambda}} \right\|_\infty < \gamma \tag{10}$$

$\hat{\mathbf{x}}$ is also supported in Λ, i.e., $\text{supp}(\hat{\mathbf{x}}) \subseteq \text{supp}(\mathbf{x})$, and

$$\mathbf{e}_\Lambda := \hat{\mathbf{x}}_{\boldsymbol{\Lambda}} - \mathbf{x}_{\boldsymbol{\Lambda}} = \left(\boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \right)^{-\mathbf{1}} \boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \mathbf{n} - \gamma \left(\boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \right)^{-\mathbf{1}} \mathbf{g}_{\boldsymbol{\Lambda}} \tag{11}$$

where, Λ^C is the complement set of Λ, $\boldsymbol{\Phi}_\Lambda$ the partial matrix consisting of $\boldsymbol{\Phi}$'s columns index by Λ, and so as for $\boldsymbol{\Phi}_{\Lambda^C}$, $\mathbf{P}^{\perp}_{V_\Lambda}$ denotes the orthogonal projection to the space V_Λ, which is expended by column vector of $\boldsymbol{\Phi}_\Lambda$, vector $\mathbf{g}$ satisfies $\mathbf{g}[i] = \begin{cases} \text{sgn}(\hat{\mathbf{x}}[i]); & \hat{\mathbf{x}}[i] \neq 0 \\ c[i], |c[i]| \leq 1; & \text{else} \end{cases}$, where $\text{sgn}()$ is the complex sign function, $c[i]$ a proper value that makes Equation (11) hold, and $\mathbf{g}_{\boldsymbol{\Lambda}}$ the row restriction of $\mathbf{g}$ indexed by Λ.

Condition (10) and formula (11) demonstrate two functions of Lasso for sparse recovery:

a) Preserve the support of original signal, and clean noise outside its support;
b) Preserve a stable recovery on the support.

The relationship of the three fatal factors used to evaluate the radar performance, sparsity K, sufficient measurement number M, and SNR is buried in (10) and (11). If a particular measurement matrix is taken into consideration, the relationship will be discovered.

3. NOISELESS CASE

Firstly, we would like to study the case when the noise is absent, so that the affection of factor SNR is not considered in this section. The following analysis discovers the basic relationship between ρ_K and ρ_M.

As $\mathbf{n} = \mathbf{0}$ in noiseless case, (10) and (11) become

$$\left\| \boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}^c} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \left(\boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \right)^{-\mathbf{1}} \mathbf{g}_{\boldsymbol{\Lambda}} \right\|_\infty < 1 \tag{12}$$

$$\mathbf{e}_\Lambda = -\gamma \left(\boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \right)^{-\mathbf{1}} \mathbf{g}_{\boldsymbol{\Lambda}} \tag{13}$$

Condition (12) can be expressed as

$$\left| \left\langle \left(\boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \boldsymbol{\Phi}_{\boldsymbol{\Lambda}} \right)^{-\mathbf{1}} \boldsymbol{\Phi}^*_{\boldsymbol{\Lambda}} \phi_j, \mathbf{g}_{\boldsymbol{\Lambda}} \right\rangle \right| < 1, \quad \forall j \notin \Lambda \tag{14}$$

where, ϕ_j is the j-th column of $\boldsymbol{\Phi}$. Whether it would hold or not has no relationship with the regularization constant γ, so that γ can be chosen as small as possible.

Indeed, since $\mathbf{g}_{\boldsymbol{\Lambda}}$ is a bounded but unknown vector, (14) holds if $\|(\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}^*\boldsymbol{\Phi}_{\boldsymbol{\Lambda}})^{-\mathbf{1}}\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}^*\phi_j\|_2$ always takes small values for any j outside Λ, which depends on two properties of $\boldsymbol{\Phi}$, the *on-support isometric* and *of-support incoherence*.

3.1. On-support Isometric

In the context of CS, the isometric property demonstrates how close between $\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}^*\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}$ and the identity matrix [21]. $\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}^*\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}$ is well-posed, and $\|(\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}^*\boldsymbol{\Phi}_{\boldsymbol{\Lambda}})^{-\mathbf{1}}\|_2$ is bounded by a limited value, when the isometric property of $\boldsymbol{\Phi}_{\boldsymbol{\Lambda}}$ is good.

As mentioned before, $\boldsymbol{\Phi}$ consists of random row selections from $\mathbf{H}$ according to the sub-sampling operator $\boldsymbol{\Theta}$. ρ_M can seem as the probability of the random selection. Then, the Gram matrix of $\boldsymbol{\Phi}$ can be written as the sum of N_S random matrix as

$$\boldsymbol{\Phi}^*\boldsymbol{\Phi} = \sum_{n=1}^{N_S} \zeta_n h_{n,\bullet}^* h_{n,\bullet} = \rho_M \mathbf{H}^*\mathbf{H} + \sum_{n=1}^{N_S} \varsigma_n h_{n,\bullet}^* h_{n,\bullet} \tag{15}$$

where, $\zeta_n,\ n = 1, \ldots, N$ is an i.i.d. Bernoulli series with $\mathrm{P}(\zeta = 1) = \rho_M,\ \mathrm{P}(\zeta = 0) = 1 - \rho_M$, and $\varsigma_n = \zeta_n - \mathrm{E}\zeta_n,\ n = 1, \ldots, N_s$, and $h_{n,\bullet}$ is the n-th row of $\mathbf{H}$.

Then the Gram matrix of $\boldsymbol{\Phi}_\Lambda$ can be writhen as

$$\boldsymbol{\Phi}_\Lambda^*\boldsymbol{\Phi}_\Lambda = \rho_M \mathbf{H}_\Lambda^*\mathbf{H}_\Lambda + \sum_{n=1}^{N_S} \varsigma_n h_{n,\Lambda}^* h_{n,\Lambda} \tag{16}$$

and its maximum and minimum eigenvalues can be bound by

$$\lambda_{\max}(\boldsymbol{\Phi}_\Lambda^*\boldsymbol{\Phi}_\Lambda) \le \rho_M \lambda_{\max}(\mathbf{H}_\Lambda^*\mathbf{H}_\Lambda) + \left\|\sum_{n=1}^{N_S} \varsigma_n h_{n,\Lambda}^* h_{n,\Lambda}\right\|_{2-2} \tag{17}$$

$$\lambda_{\min}(\boldsymbol{\Phi}_\Lambda^*\boldsymbol{\Phi}_\Lambda) \ge \rho_M \lambda_{\min}(\mathbf{H}_\Lambda^*\mathbf{H}_\Lambda) - \left\|\sum_{n=1}^{N_S} \varsigma_n h_{n,\Lambda}^* h_{n,\Lambda}\right\|_{2-2} \tag{18}$$

where, $\lambda_{\max}(\cdot)$, $\lambda_{\min}(\cdot)$ stands for the maximum and minimum eigenvalue of some matrix, and $h_{n,\Lambda}$ is a column restriction of $h_{n,\bullet}$ indexed by Λ.

Firstly, the range of eigenvalue of $\mathbf{H}_\Lambda^*\mathbf{H}_\Lambda$ can be estimated according to the gerschgorin's circle theorem [22],

$$\lambda_{\max}(\mathbf{H}_\Lambda^*\mathbf{H}_\Lambda) \le \|h_k\|_2 + \sum_{i\in\Lambda, i\neq k} |\langle h_k, h_i\rangle| \tag{19}$$

$$\lambda_{\min}(\mathbf{H}_\Lambda^*\mathbf{H}_\Lambda) \ge \|h_k\|_2 - \sum_{i\in\Lambda, i\neq k} |\langle h_k, h_i\rangle| \tag{20}$$

hold for any $k \in \Lambda$. From the definition of $\mathbf{H}$ we can state that h_k, h_i, the two discrete forms of chirp signal with delay of $k\Delta\tau$, $i\Delta\tau$, respectively, share the common support if and only if $i \in [k - \frac{N_H}{2}, k + \frac{N_H}{2})$. Denote $\Lambda_k = \Lambda \cap [k - \frac{N_H}{2}, k + \frac{N_H}{2})$ as a local support of $\mathbf{x}$ in a N_H-length continues region of $[k - \frac{N_H}{2}, k + \frac{N_H}{2})$. According to (6), we have $\|\Lambda_k\|_0 \leq \rho_K N_H$, where $\|\Lambda_k\|_0$ denotes the number of candidates of Λ_k.

Define the mutual coherence of $\mathbf{H}$ as

$$\mu_H = \frac{\sup\limits_{i\neq k} |\langle h_i, h_k\rangle|}{\|h_i\|_2 \, \|h_k\|_2} = \frac{1}{N_H} \sup_{i\neq k} |\langle h_i, h_k\rangle| \tag{21}$$

Then,

$$\sum_{i\in\Lambda, i\neq k} |\langle h_k, h_i\rangle| = \sum_{i\in\Lambda_k, i\neq k} |\langle h_k, h_i\rangle| \leq \|\Lambda_k\|_0 \cdot \mu_H N_H \leq \rho_k N_H \cdot \mu_H N_H$$

Substituting the above inequality to (19) and (20) gives

$$\lambda_{\max}\left(\mathbf{H}_\Lambda^* \mathbf{H}_\Lambda\right) \leq N_H \left(1 + \rho_k N_H \mu_H\right) \tag{22}$$

$$\lambda_{\min}\left(\mathbf{H}_\Lambda^* \mathbf{H}_\Lambda\right) \geq N_H \left(1 - \rho_k N_H \mu_H\right) \tag{23}$$

Secondly, the spectrum bound for $\sum\limits_{n=1}^{N_S} \varsigma_n h_{\Lambda,n}^* h_{\Lambda,n}$, the sum of independent random matrix, can be estimated according to the Bornstein's inequality [23], which states

$$\mathrm{P}\left(\left\|\sum_{n=1}^{N_S} \mathbf{X}_n\right\|_{2-2} \geq t\right) \leq (d_1 + d_2) \exp\left(-\frac{t^2/2}{\sigma^2 + Bt/3}\right) \tag{24}$$

where, $\mathbf{X}_n \in C^{d_1 \times d_2}$, $n = 1, \ldots, N_S$ is a series of independent random distributed matrix such that $\mathrm{E}(\mathbf{X}_n) = 0$, $\sup_n \|\mathbf{X}_n\|_{2-2} \leq B$, and $\max(\|\mathrm{E}(\sum\limits_{n=1}^{N_S} \mathbf{X}_n^* \mathbf{X}_n)\|_{2-2}, \|\mathrm{E}(\sum\limits_{n=1}^{N_S} \mathbf{X}_n \mathbf{X}_n^*)\|_{2-2}) \leq \sigma^2$.

Let $\mathbf{X}_n = \varsigma_n h_{n,\Lambda}^* h_{n,\Lambda}$, $n = 1, \ldots, N_S$ be random *K-by-K* matrix, then it satisfies $\mathrm{E}(\mathbf{X}_n) = 0$, and $\|\mathbf{X}_n\|_{2-2} = |\varsigma_n| \cdot \|h_{n,\Lambda}^* h_{n,\Lambda}\|_{2-2} \leq \|h_{n,\Lambda}\|_2^2$.

Since the row vector $h_{\bullet,n}$ is also a discrete form of the chirp signal and has a limited duration of $[n - \frac{N_H}{2}, n + \frac{N_H}{2})$, its restricted part, $h_{\Lambda,n}$, is supported in $\Lambda_n = \Lambda \cap [n - \frac{N_H}{2}, n + \frac{N_H}{2})$ with $\|\Lambda_n\|_0 \leq \rho_K N_H$ as the same reason mentioned before. As a result,

$$\|X_n\|_{2-2} \leq \|h_{n,\Lambda}\|_2^2 \leq \|h_{n,\Lambda}\|_\infty^2 \cdot \|\Lambda_n\|_0 \leq \rho_K N_H = B \tag{25}$$

Since $\mathbf{X}_n = \mathbf{X}_n^*$, we get

$$\left\| \mathrm{E}\left(\sum_{n=1}^{N_S} \mathbf{X}_n^* \mathbf{X}_n\right) \right\|_{2-2} = \left\| \mathrm{E}\left(\sum_{n=1}^{N_S} \mathbf{X}_n \mathbf{X}_n^*\right) \right\|_{2-2}$$

$$= \mathrm{E}\left(\sum_{n=1}^{N_S} \mathbf{X}_n \mathbf{X}_n^*\right) = \rho_M \left(1-\rho_M\right) \cdot \left\| \sum_{n=1}^{N_S} \|h_{n,\Lambda}\|_2^2 h_{n,\Lambda}^* h_{n,\Lambda} \right\|_{2-2}$$

$$\leq \rho_M \cdot \rho_K N_H \cdot \left\| \sum_{n=1}^{N_S} h_{n,\Lambda}^* h_{n,\Lambda} \right\|_{2-2} = \rho_M \cdot \rho_K N_H \cdot \|H_\Lambda^* H_\Lambda\|_{2-2}$$

$$\leq \rho_M \cdot \rho_K N_H \cdot N_H \left(1+\rho_K N_H \mu_H\right) = \sigma^2 \tag{26}$$

where, the first inequality holds for $\|h_{n,\Lambda}\|_2^2 \leq \rho_K N_H$ (see (23)) and for $h_{n,\Lambda}^* h_{n,\Lambda}$ being a non-negative definite matrix, and the last inequality holds according to (16).

Substituting (25)–(26) into (24) gives

$$\mathrm{P}\left(\left\| \sum_{n=1}^{N_S} \varsigma_n h_{n,\Lambda}^* h_{n,\Lambda} \right\|_{2-2} \geq \rho_M N_H \left(1+\rho_K N_H \mu_H\right) \cdot \delta \right)$$

$$\leq 2K \exp\left(-\frac{\rho_M \left(1+\rho_K N_H \mu_H\right)}{\rho_K} \frac{\delta^2/2}{1+\delta/3} \right) \tag{27}$$

where δ is an arbitrary small positive value.

Together with (22)–(23) and (27), the following formula can be derived

$$\mathrm{P}\left(\left\| \mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}^- \rho_M N_H \mathbf{I} \right\|_{2-2} \geq \rho_M N_H \left(\rho_K N_H \mu_H + (1+\rho_K N_H \mu_H) \cdot \delta\right) \right)$$

$$\leq 2K \exp\left(-\frac{\rho_M \left(1+\rho_K N_H \mu_H\right)}{\rho_K} \frac{\delta^2/2}{1+\delta/3} \right) \tag{28}$$

where, $\mathbf{I}$ is the *K-by-K* identity matrix. Formula (28) demonstrates that $\mathbf{\Phi}_{\mathbf{\Lambda}}$ has a good isometric property with high probability, because both μ_H and δ are of small value.

3.2. Of-support Incoherence

In CS theory, the mutual coherence of $\mathbf{\Phi}$ in (4) should be designed as smaller as possible. A sufficient small coherence, i.e., incoherence, also bounds $\mathbf{\Phi}_{\mathbf{\Lambda}}^* \phi_j$, $\forall j \notin \Lambda$.

Let $\mathbf{v}$ be an arbitrary bounded vector. Due to (15),

$$\langle \mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j, \mathbf{v}\rangle = \rho_M h_j^{*}\mathbf{H}_{\Lambda}\mathbf{v} + \sum_{n=1}^{N_S} \varsigma_n h_{n,j}^{*} h_{n,\Lambda}\mathbf{v} \tag{29}$$

Firstly, let $\Lambda_j = \Lambda \cap \left[j - \frac{N_H}{2}, j + \frac{N_H}{2}\right)$, then

$$\left|h_j^{*}\mathbf{H}_{\Lambda}\mathbf{v}\right| \leq \left\|h_j^{*}\mathbf{H}_{\Lambda}\right\|_2 \|\mathbf{v}\|_2 = \sqrt{\sum_{i\in\Lambda} |\langle h_j, h_i\rangle|^2}\, \|\mathbf{v}\|_2 = \sqrt{\sum_{i\in\Lambda_j} |\langle h_j, h_i\rangle|^2}\, \|\mathbf{v}\|_2$$

$$\leq \sqrt{\|\Lambda_j\|_0} \cdot \mu_H N_H \|\mathbf{v}\|_2 \leq \sqrt{\rho_K N_H} \cdot N_H \mu_H \|\mathbf{v}\|_2 \tag{30}$$

Secondly, let $\mathbf{X}_n = \varsigma_n h_{n,j}^{*} h_{n,\Lambda}\mathbf{v}$, $n = 1, \ldots, N_S$ be random valuable, then $\mathrm{E}(\mathbf{X}_n) = 0$, and

$$\|\mathbf{X}_n\|_{2-2} = |\varsigma_n| \cdot \left\|h_{n,j}^{*} h_{n,\Lambda}\mathbf{v}\right\|_2 \leq \|h_{n,\Lambda}\|_2 \|\mathbf{v}\|_2 \leq \sqrt{\rho_K N_N}\, \|\mathbf{v}\|_2 = B \tag{31}$$

$$\left\|\mathrm{E}\left(\sum_{n=1}^{N_S} \mathbf{X}_n^{*}\mathbf{X}_n\right)\right\|_{2-2} = \left\|\mathrm{E}\left(\sum_{n=1}^{N_S} \mathbf{X}_n\mathbf{X}_n^{*}\right)\right\|_{2-2} = \mathrm{E}\left(\sum_{n=1}^{N_S} |\mathbf{X}_n|^2\right)$$

$$= \rho_M (1-\rho_M) \cdot \sum_{n=1}^{N_S} \left|h_{j,n}^{*}\right|^2 |h_{\Lambda,n}\mathbf{v}|^2 \leq \rho_M \cdot \|\mathbf{H}_{\Lambda}\mathbf{v}\|_2^2$$

$$\leq \rho_M N_H (1 + \rho_k N_H \mu_H) \cdot \|\mathbf{v}\|_2^2 = \sigma^2 \tag{32}$$

Substitute (31)–(32) into (24), it gives

$$\mathrm{P}\left(\left\|\sum_{n=1}^{N_S} \varsigma_n h_{j,n}^{*} h_{\Lambda,n}\mathbf{v}\right\|_{2-2} \geq \frac{\rho_M N_H (1 + \rho_K N_H \mu_H)}{\sqrt{\rho_K N_N}} \cdot \delta \|\mathbf{v}\|_2\right)$$

$$\leq 2\exp\left(-\frac{\rho_M (1 + \rho_K N_H \mu_H)}{\rho_K} \frac{\delta^2/2}{1+\delta/3}\right) \tag{33}$$

Combine (30) and (33), and consider about all $j \notin \Lambda$, then gives

$$\mathrm{P}\left(|\langle \mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j, \mathbf{v}\rangle| \geq \frac{\rho_M N_H (\rho_K N_H \mu_H + (1 + \rho_K N_H \mu_H) \cdot \delta)}{\sqrt{\rho_K N_N}} \|\mathbf{v}\|_2\right)$$

$$\leq 2(N-K)\exp\left(-\frac{\rho_M (1 + \rho_K N_H \mu_H)}{\rho_K} \frac{\delta^2/2}{1+\delta/3}\right) \tag{34}$$

Then $\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j$ are well bounded for high probability.

3.3. Support Preserving Condition

The support preserving condition of noiseless case is demonstrated by (12), i.e., (14). Formulas (28) and (34) derived in the last two subsections play a key role in determining whether the condition holds or not.

Firstly, we state that $(\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-\mathbf{1}}\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j$ is supported in $\Lambda_j = \Lambda \cap [j - \frac{N_H}{2}, j + \frac{N_H}{2})$ and then

$$\left\langle (\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-\mathbf{1}} \mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j, \mathbf{g}_{\mathbf{\Lambda}} \right\rangle = \left\langle (\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-\mathbf{1}} \mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j, \hat{\mathbf{g}}_{\mathbf{\Lambda}} \right\rangle \tag{35}$$

where, $\hat{\mathbf{g}}_{\mathbf{\Lambda}}[i] = \begin{cases} \mathbf{g}_{\mathbf{\Lambda}}[i]; & i \in \Lambda_j \\ 0; & \text{else} \end{cases}$, i.e., $\hat{\mathbf{g}}_{\mathbf{\Lambda}}$ eliminates all non-zeros of $\mathbf{g}_{\mathbf{\Lambda}}$ outside Λ, and

$$\|\hat{\mathbf{g}}_{\mathbf{\Lambda}}\|_2 \leq \sqrt{\|\Lambda_j\|_0}\, \|\hat{\mathbf{g}}_{\mathbf{\Lambda}}\|_\infty \leq \sqrt{\rho_K N_H} \tag{36}$$

The proof is obvious. According to (15), $\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j = \sum_{n=1}^{N_S} \zeta_n h_{n,\mathbf{\Lambda}}^{*} h_{n,j}$. For a given j, $h_{n,j} = 0$ if $j \notin [j - \frac{N_H}{2}, j + \frac{N_H}{2})$, so that the vector $h_{n,\mathbf{\Lambda}}^{*} h_{n,j}$ is supported in Λ_j. Namely, $\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j$ is also supported in Λ_j. Let $\mathbf{w}_\Lambda$ be a vector defined on Λ and be supported in Λ_j such that $\mathbf{w}_{\mathbf{\Lambda}_j} = (\mathbf{\Phi}_{\mathbf{\Lambda}_j}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}_j})^{-\mathbf{1}}\mathbf{\Phi}_{\mathbf{\Lambda}_j}^{*}\phi_j$, then $\mathbf{w}_\Lambda$ satisfies $(\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})\mathbf{w}_\Lambda = \mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j$. Due to the uniqueness of well-posed matrix inversion, $\mathbf{w}_\Lambda = (\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-1}\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j$ is supported in Λ_j.

Secondly, we would like to derive the bound for $\langle(\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-\mathbf{1}}\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j, \mathbf{g}_{\mathbf{\Lambda}}\rangle$. Assume $\mathbf{v} = (\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-1}\hat{\mathbf{g}}_{\mathbf{\Lambda}}$. From (36) it gives

$$\|\mathbf{v}\|_2 = \left\| (\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-1} \right\|_{2-2} \cdot \sqrt{\rho_K N_H} \tag{37}$$

Together with (28), (34)–(35), the following formula can be derived

$$\begin{aligned} &\left| \left\langle (\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-\mathbf{1}} \mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j, \mathbf{g}_{\mathbf{\Lambda}} \right\rangle \right| = |\langle \mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\phi_j, v\rangle| \\ &< \frac{\rho_M N_H \left(\rho_K N_H \mu_H + (1 + \rho_K N_H \mu_H) \cdot \delta\right)}{\sqrt{\rho_K N_N}} \|\mathbf{v}\|_2 \\ &\leq \frac{\rho_M N_H (\rho_K N_H \mu_H + (1 + \rho_K N_H \mu_H) \cdot \delta)}{\sqrt{\rho_K N_N}} \left\| (\mathbf{\Phi}_{\mathbf{\Lambda}}^{*}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-1} \right\|_{2-2} \sqrt{\rho_K N_H} \\ &< \frac{\rho_M N_H \left(\rho_K N_H \mu_H + (1 + \rho_K N_H \mu_H) \cdot \delta\right)}{\rho_M N_H - \rho_M N_H \left(\rho_K N_H \mu_H + (1 + \rho_K N_H \mu_H) \cdot \delta\right)} \\ &= \frac{\left(\rho_K N_H \mu_H + (1 + \rho_K N_H \mu_H) \cdot \delta\right)}{1 - \left(\rho_K N_H \mu_H + (1 + \rho_K N_H \mu_H) \cdot \delta\right)} \end{aligned} \tag{38}$$

holds for any $j \notin \Lambda$ with a probability exceeding

$$1 - 2N \exp\left(-\frac{\rho_M \left(1 + \rho_K N_H \mu_H\right)}{\rho_K} \frac{\delta^2/2}{1 + \delta/3}\right) \tag{39}$$

According to condition (14), letting the right side of (38) equal to "1", gives

$$\delta = \frac{\frac{1}{2} - \rho_K N_H \mu_H}{1 + \rho_K N_H \mu_H} \tag{40}$$

Finally, substituting (40) to (39) and together with (38), we conclude that:

Statement 1: In the noiseless sparse delay detection problem, if ρ_M, ρ_K satisfies

$$\rho_M \geq \rho_K \log N \cdot \frac{1}{2} \frac{7 + 4\rho_K N_H \mu_H}{\left(\frac{1}{2} - \rho_K N_H \mu_H\right)^2} = C\left(\rho_K\right) \rho_K \log N \tag{41}$$

the sparse signal's support is preserved with a high probability exceeding $1 - 2/\sqrt{N}$.

In the above formula, $C(\rho_K)$ is not a numerical value independent of ρ_K but is increased while ρ_K is enlarged. The constant of $N_H \mu_H$, which demonstrates the mutual incoherence between the radar waveform and its time delays, dominates how $C(\rho_K)$ changes with ρ_K. For instance, let the bandwidth and time duration of waveform given by (2) be chosen as 150 MHz and 6 µs, respectively, then $N_H \mu_H \cong 14.81$. Assume $\rho_K \leq 1/10 N_H \mu_H = 0.6753\%$, then it can be calculated from (41) that $C(\rho_K)$ takes value in $(14, 23.125]$.

4. NOISY CASE

4.1. Support Preserving Condition for Noisy Case

Comparing the noisy and noiseless condition expressed by (12) and (14), it can be found that the existence of measurement noise is bad for the support preserving. In a noisy case, the regularization constant γ should be sufficient large so that

$$\left\| \mathbf{\Phi}_{\mathbf{\Lambda}^c}^* \mathbf{P}_{V_\Lambda}^{\perp} \mathbf{n} \right\|_\infty < \gamma \left(1 - \left\| \mathbf{\Phi}_{\mathbf{\Lambda}^c}^* \mathbf{\Phi}_{\mathbf{\Lambda}} \left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right)^{-\mathbf{1}} \mathbf{g}_{\mathbf{\Lambda}} \right\|_\infty\right) \tag{42}$$

Formula (42) demonstrates a sharp sufficient support preserving condition for the noisy case.

Assume the right side of (38) equal to positive parameter $\alpha < 1$, i.e., $\left\| \mathbf{\Phi}_{\mathbf{\Lambda}^c}^* \mathbf{\Phi}_{\mathbf{\Lambda}} \left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right)^{-\mathbf{1}} \mathbf{g}_{\mathbf{\Lambda}} \right\|_\infty < \alpha$, and

$$\delta = \frac{\frac{\alpha}{1+\alpha} - \rho_K N_H \mu_H}{1 + \rho_K N_H \mu_H} \tag{43}$$

Substituting (43) to (39) shows that in a noisy case if

$$\rho_M \geq \rho_K \log N \cdot \frac{3+\frac{\alpha}{1+\alpha}+2\rho_K N_H \mu_H}{\left(\frac{\alpha}{1+\alpha}-\rho_K N_H \mu_H\right)^2} = C'\left(\rho_K,\alpha\right)\cdot\rho_K \log N \quad (44)$$

$\|\mathbf{\Phi}^*_{\mathbf{\Lambda}^c}\mathbf{\Phi}_{\mathbf{\Lambda}}(\mathbf{\Phi}^*_{\mathbf{\Lambda}}\mathbf{\Phi}_{\mathbf{\Lambda}})^{-\mathbf{1}}\mathbf{g}_{\mathbf{\Lambda}}\|_\infty < \alpha$ will hold with a probability exceeding $1-2/\sqrt{N}$. Together with (42), inequality can be derived

$$\gamma \geq \frac{1}{1-\alpha}\left\|\mathbf{\Phi}^*_{\mathbf{\Lambda}^c}\mathbf{P}^{\perp}_{V_\Lambda}\mathbf{n}\right\|_\infty \quad (45)$$

The following derivations will show how $\|\mathbf{\Phi}^*_{\mathbf{\Lambda}^c}\mathbf{P}^{\perp}_{V_\Lambda}\mathbf{n}\|_\infty$ is bounded.

Consider a concentration property of real Gaussian variable $n \sim \mathrm{N}(0,\sigma_0^2)$, it is easy to prove that

$$\begin{aligned}\mathrm{P}\left(|n|\geq t\right) &= \frac{2}{\sqrt{2\pi}\sigma_0}\int_t^{+\infty}\exp\left(-\frac{x^2}{2\sigma_0^2}\right)dx \\ &\leq \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_0}\int_t^{+\infty}\frac{x}{t}\exp\left(-\frac{x^2}{2\sigma_0^2}\right)dx = \sqrt{\frac{2}{\pi}}\frac{\sigma_0}{t}\exp\left(-\frac{t^2}{2\sigma_0^2}\right) \quad (46)\end{aligned}$$

The conclusion of (46) can be extended for complex Gaussian variable.

Selecting some $j \notin \Lambda$, from (46) we can derive that

$$\begin{aligned}\mathrm{P}\left(\left|\phi_j^*\mathbf{P}^{\perp}_{V_\Lambda}\mathbf{n}\right|\geq t\right) &\leq \sqrt{\frac{2}{\pi}}\frac{\sigma_0\left\|\mathbf{P}^{\perp}_{V_\Lambda}\phi_j\right\|_2}{t}\exp\left(-\frac{t^2}{2\sigma_0^2\left\|\mathbf{P}^{\perp}_{V_\Lambda}\phi_j\right\|_2^2}\right) \\ &\leq \sqrt{\frac{2}{\pi}}\frac{\sigma_0\|\phi_j\|_2}{t}\exp\left(-\frac{t^2}{2\sigma_0^2\|\phi_j\|_2^2}\right) \quad (47)\end{aligned}$$

The second inequality in (47) holds because $\|\mathbf{P}^{\perp}_{V_\Lambda}\phi_j\|_2 \leq \|\phi_j\|_2$.

Consider $\|\phi_j\|_2^2 = \rho_M N_H + \sum_{n=1}^{N_S}\varsigma_n = \rho_M N_H + \sum_{n\in[j-\frac{N_H}{2},j+\frac{N_H}{2})}\varsigma_n$, then we get for some j that

$$\mathrm{P}\left(\left|\|\phi_j\|_2^2-\rho_M N_H\right|\geq\delta\rho_M N_H\right)\leq 2\exp\left(-\rho_M N_H\frac{\delta^2/2}{1+\delta/3}\right) \quad (48)$$

due to the Bornstein's inequality.

Choosing a δ that satisfies $\delta = 1$ and substituting it into (48), we get

$$\mathrm{P}\left(\left|\|\phi_j\|_2^2-\rho_M N_H\right|\geq\rho_M N_H\right)\leq 2\exp\left(-\frac{3}{8}\rho_M N_H\right) \quad (49)$$

Since $\rho_M N_H \gg 1$, (49) states that $\|\phi_j\|_2^2 < 2\rho_M N_H$ almost surely holds. As a result, combining (47) and (49) and considering $j \notin \Lambda$, we can derive that

$$\mathrm{P}\left(\left\|\mathbf{\Phi}_{\mathbf{\Lambda}^c}^* \mathbf{P}_{V_\Lambda}^\perp \mathbf{n}\right\|_\infty \geq t\right) \leq 2(N-K)\sqrt{\frac{\rho_M N_H}{\pi}}\frac{\sigma_0}{t}\exp\left(-\frac{t^2}{4\sigma_0^2 \rho_M N_H}\right) \quad (50)$$

Let $t = 2\sigma_0\sqrt{\rho_M N_H \log N}$, then (50) becomes

$$\mathrm{P}\left(\left\|\mathbf{\Phi}_{\mathbf{\Lambda}^c}^* \mathbf{P}_{V_\Lambda}^\perp \mathbf{n}\right\|_\infty \geq 2\sigma_0\sqrt{\rho_M N_H \log N}\right) \leq \left(1 - \frac{K}{N}\right)\frac{1}{\sqrt{\pi \log N}} \quad (51)$$

It can be concluded from (44)–(45) and (51) that if ρ_M, ρ_K satisfies (44) and

$$\gamma = \frac{1}{1-\alpha} 2\sigma_0\sqrt{\rho_M N_H \log N} \quad (52)$$

the support preserving will almost ensure via Lasso with a probability exceeding $1 - 2/\sqrt{N} - \frac{1-K/N}{\sqrt{\pi \log N}}$.

4.2. On Support Reconstruction Error and SNR

As the signal support is picked out, it can be derived from (11) that

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 = \|\hat{\mathbf{x}}_{\mathbf{\Lambda}} - \mathbf{x}_{\mathbf{\Lambda}}\|_2 \leq \left\|\left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right)^{-1} \mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{n}\right\|_2 + \gamma\left\|\left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right)^{-1} \mathbf{g}_{\mathbf{\Lambda}}\right\|_2 \quad (53)$$

On the left side of (50), the first term is proportional to noise and the second term proportional to γ. Indeed, it is shown by (52) that γ is proportional to noise, too. In the following, we will demonstrate how measurement noise affects the reconstruction error.

Firstly, we can be derive that

$$\left\|\left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right)^{-1} \mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{n}\right\|_2 \leq \sqrt{K}\left\|\left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right)^{-1} \mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{n}\right\|_\infty \quad (54)$$

Let $\mathbf{u}$ be any column of $\mathbf{\Phi}_{\mathbf{\Lambda}}(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}})^{-1}$, then $\|(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}})^{-1} \mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{n}\|_\infty$ $= \sup_{\mathbf{u}} |\mathbf{u}^* \mathbf{n}|$ and $\|\mathbf{u}\|_2 \leq 1/\sqrt{\lambda_{\min}(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}})}$. While (42)–(43) hold, from (28) we can get

$$\lambda_{\min}\left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right) > \frac{\rho_M N_H}{1+\alpha} \quad (55)$$

By the same means to derive (51), we get

$$\mathrm{P}\left(\|\mathbf{u}^* \mathbf{n}\|_\infty \geq \frac{\sigma_0}{\sqrt{\lambda_{\min}\left(\mathbf{\Phi}_{\mathbf{\Lambda}}^* \mathbf{\Phi}_{\mathbf{\Lambda}}\right)}}\sqrt{2\log N}\right) \leq \frac{K}{N}\frac{1}{\sqrt{\pi \log N}} \quad (56)$$

Substituting (55)–(56) to (54),

$$\left\| \left(\mathbf{\Phi}_{\mathbf{\Lambda}}^{*} \mathbf{\Phi}_{\mathbf{\Lambda}} \right)^{-\mathbf{1}} \mathbf{\Phi}_{\mathbf{\Lambda}}^{*} \mathbf{n} \right\|_{2} < \sigma_0 \sqrt{\frac{\log N}{\rho_M N_H}} \cdot \sqrt{2\left(1+\alpha\right)} \cdot \sqrt{K} \tag{57}$$

holds with a probability exceeding $1 - \frac{K/N}{\sqrt{\pi \log N}}$.

Secondly, it can be derived that

$$\begin{aligned} \left\| \left(\mathbf{\Phi}_{\mathbf{\Lambda}}^{*} \mathbf{\Phi}_{\mathbf{\Lambda}} \right)^{-\mathbf{1}} \mathbf{g}_{\mathbf{\Lambda}} \right\|_{2} &\leq \left\| \left(\mathbf{\Phi}_{\mathbf{\Lambda}}^{*} \mathbf{\Phi}_{\mathbf{\Lambda}} \right)^{-\mathbf{1}} \right\|_{2-2} \left\| \mathbf{g}_{\mathbf{\Lambda}} \right\|_{2} \\ &= \frac{1}{\lambda_{\min}\left(\mathbf{\Phi}_{\mathbf{\Lambda}}^{*} \mathbf{\Phi}_{\mathbf{\Lambda}} \right)} \cdot \sqrt{K} < \frac{1+\alpha}{\rho_M N_H} \cdot \sqrt{K} \end{aligned} \tag{58}$$

Then substituting (52), (57)–(58) to (53), we have

$$\left\| \hat{\mathbf{x}} - \mathbf{x} \right\|_2 < \left(\sqrt{2\left(1+\alpha\right)} + \frac{2\left(1+\alpha\right)}{1-\alpha} \right) \sqrt{K} \cdot \sigma_0 \sqrt{\frac{\log N}{\rho_M N_H}} \tag{59}$$

Formula (59) can be used to relate $\left\| \hat{\mathbf{x}} - \mathbf{x} \right\|_2$ and the SNR defined by (7). According to (7), (59) becomes

$$\frac{\left\| \hat{\mathbf{x}} - \mathbf{x} \right\|_2}{\left\| \mathbf{x} \right\|_2} < \frac{\sqrt{2\left(1+\alpha\right)}\left(1-\alpha\right) + 2\left(1+\alpha\right)}{1-\alpha} \cdot \sqrt{\frac{\log N}{\mathrm{SNR} \cdot \rho_M N_H}} \tag{60}$$

As $\sqrt{2\left(1+\alpha\right)}\left(1-\alpha\right) + 2\left(1+\alpha\right) \leq 4$ holds for $\alpha \in (0, 1)$, it can be derived from (60) that

$$\frac{\left\| \hat{\mathbf{x}} - \mathbf{x} \right\|_2}{\left\| \mathbf{x} \right\|_2} < \frac{4}{1-\alpha} \cdot \sqrt{\frac{\log N}{\mathrm{SNR} \cdot \rho_M N_H}} \tag{61}$$

The relative mean square error (RMSE) can be defined as

$$\mathrm{RMSE} = \frac{\left\| \hat{\mathbf{x}} - \mathbf{x} \right\|_2^2}{\left\| \mathbf{x} \right\|_2^2} \tag{62}$$

Together with (52), (57), (61)–(62), we learn that

$$\mathrm{RMSE} < \left(\frac{4}{1-\alpha} \right)^2 \frac{\log N}{\mathrm{SNR} \cdot \rho_M N_H} \tag{63}$$

holds for a probability exceeding $1 - 2/\sqrt{N} - 1/\sqrt{\pi \log N}$.

It can be found from (63) that RMSE is proportional to 1/SNR, and the smaller α is, the more RMSE can be gained from SNR. However, according to (44), a too small α will result in extremely large sub-sampling ratio ρ_M for the same ρ_K. Accordingly, the parameter α balances the demand for a more efficiently measurement radar system and the demand for a more recovery accuracy. Moreover, $\rho_M N_H$ in the multiplication can seem as a gain from partial pulse compression due to the sub-sampling.

Finally, combining the results of (44), (52), and (65), we can make a conclusion that:

Statement 2: In the noisy sparse delay detection problem, if ρ_M, ρ_K, SNR given by (5)–(7) satisfy

$$\rho_M \geq C'(\rho_K, \alpha) \cdot \rho_K \log N$$

where $C'(\rho_K, \alpha) = \frac{3+\frac{\alpha}{1+\alpha}+2\rho_K N_H \mu_H}{\left(\frac{\alpha}{1+\alpha}-\rho_K N_H \mu_H\right)^2}$, and

$$\text{SNR} \geq \left(\frac{4}{1-\alpha}\right)^2 \frac{\log N}{\rho_M N_H} \frac{1}{\text{RMSE}_{\max}}$$

choose as Lasso parameter as

$$\gamma = \frac{1}{1-\alpha} 2\sigma_0 \sqrt{\rho_M N_H \log N}$$

then the Lasso will preserve the support of targets in scene, and achieve an upper bound for RMSE denoted by $\text{RMSE}_{\max}$, with a probability exceeding $1 - 2/\sqrt{N} - 1/\sqrt{\pi \log N}$.

5. SIMULATIONS AND DISSUASIONS

5.1. Noiseless Case

To validate our first statement in Section 3, we assume that a radar transmits a chirp signal demonstrated by (2) with $T = 6\,\mu\text{s}$, $K = 25\,\text{MHz}/\mu\text{s}$ towards an unknown sparse scene, which means that the bandwidth of the chirp is $B = 150\,\text{MHz}$ and that the range resolution of radar is 1 m. Let there be $N = 10,000$ range cells with 1 m length in the scene, and several point targets with equal backscattering coefficients, i.e., "1", are randomly put into different cells. Once the sparse scene is fixed, the echo is randomly sampled below the rate that the Nyquist's theory demands. The noiseless sub-sampled data is then used to recover the scene via Lasso to validate whether the signal support is preserved, that is to check whether $\text{supp}(\hat{\mathbf{x}}) \subseteq \text{supp}(\mathbf{x})$.

We make a large number of simulations with the above setting, and then draw the ρ_M-ρ_K curve, which is also called phase transitions by Donoho and Stodden [24], to discover the relationship between ρ_M and ρ_K in noiseless case.

Firstly, let ρ_M vary in the interval $[10\%, 20\%]$, and for each ρ_M, search for a proper ρ_K that is the maximum sparse ratio which ensures that in over $1 - 2/\sqrt{N} = 98\%$ simulations the signal's supports are preserved. We use the bisearch method to accelerate the searching. In each round of bisearching, there are 100 simulations of TwIST [25] with Lasso parameter of $\gamma = 0.1\rho_M N_H$ to be performed to calculate the successful rate of support preserving. After several rounds, the searching process converges to the proper ρ_K.

Secondly, we calculate the proper ρ_K for a given ρ_M according to the statement 1, and compare it with the simulation results. Due to the aforementioned assumption, we have $N_H = BT = 900$ and $\mu_H = 0.0164$. The compared results are shown by the following figure.

Figure 2 depicts the theoretical phase transitions predicated by statement one and the empirical phase transitions given by simulations. One can see that empirical phase transition matches closely with the statement one prediction, which means that statement one gives a sharp predication for the phase transitions of the noiseless radar delay detection problem.

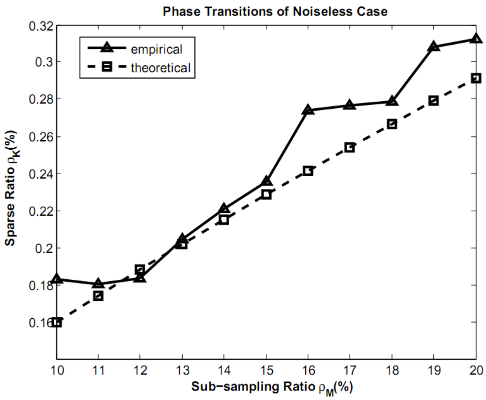

Figure 2. Comparing the simulation result and the result given by statement one.

5.2. Noisy Case

To verify the second statement given in Section 4, we follow the same assumption from the last sub-section except for adding some noise in sub-sampled data. As the back scattering coefficient of the point targets is set to "1", the SNR can be calculated by $\mathrm{SNR} = 1/\sigma_0^2$.

Firstly, let ρ_M be one of the three values $\{10\%, 15\%, 20\%\}$, and for each choice of ρ_M, let ρ_K vary in the interval $[0, \rho_{K,noisefree}]$, where $\rho_{K,noiseless}$ denotes the largest ρ_K which makes (41) hold. Once ρ_K is given, a bisearch method is performed to find the smallest

SNR, which ensures the support of original scene is preserved and the RMSE $\leq$ RMSE$_{\max} = 0.1$ in over $1 - 2/\sqrt{N} - 1/\sqrt{\pi \log N} = 80\%$ of simulations. In each round of bisearch, the simulations are performed 100 times.

Secondly, once ρ_M and ρ_K are given, α can be calculated from (44). Then theoretical demands for SNR which ensure support preserving, and SNR$_{Lasso} \geq 10\,\mathrm{dB}$ is predicated by statement two.

Figure 3 compares the minimal demands for input SNR for different ρ_M and ρ_K given by theoretical predication and simulations. It is shown in the figure that the predication given by statement two of noisy case is not so sharp, especially for small ρ_K. However, the tendencies of the theoretical and empirical curves are still matched.

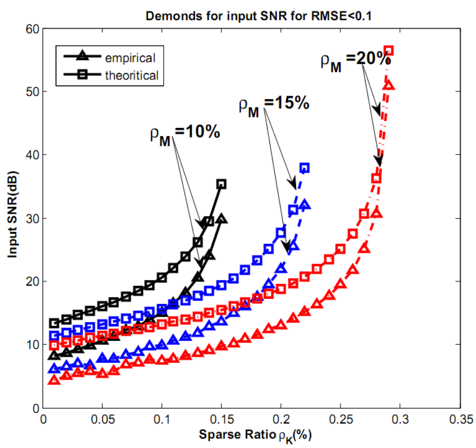

Figure 3. Comparing the simulation result and the result given by statement two.

6. SUMMARY

In this paper, we investigate the Lasso based performance guarantee for one-directional sparse microwave imaging problem. The problem refers to recovery of a sparse scene from random samples of radar returns via Lasso methods. The quantity relationship among three important factors, the sufficient sub-sampling number ratio ρ_M, sparse ratio ρ_K, and input signal-to-noise ratio SNR is discussed, which explicitly demonstrates the performance guarantee of the mentioned sparse microwave imaging radar. Although the theoretical bound is not closed to the optimal value, numerical results support the claim that the derived guarantee is at least reasonably well behaved.

ACKNOWLEDGMENT

This work has been supported by the National Statement tasks and project plan of 973 program of China under Grants No. 2010CB731905.

REFERENCES

1. The Institute of Electronics, Chinese Academy of Sciences, "Statement tasks and project plan of 973 program: Studies on theory, system, and methodology of sparse microwave imaging," 2009.
2. Donoho, D. L., "Compressive sensing," *IEEE Transactions on Information Theory*, Vol. 52, No. 4, 1289–1306, 2006.
3. Candes, E. J. and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, Vol. 25, No. 2, 21–30, 2008.
4. Wu, Y., B. Zhang, and W. Hong, "Sparse microwave imaging: Principles and applications," *China Science (F)*, Vol. 55, 1722–1754, 2012.
5. Baraniuk, R. and P. Steeghs, "Compressive radar imaging," *IEEE Radar Conference*, 128–133, Waltham, Massachusetts, 2007.
6. Wei, S. J., X.-L. Zhang, J. Shi, and G. Xiang, "Sparse reconstruction for SAR imaging based on compressive sensing," *Progress In Electromagnetics Research*, Vol. 109, 63–81, 2010.
7. Ren, X. Z., Y. F. Li, and R. Yang, "Four-dimensional SAR imaging scheme based on compressive sensing," *Progress In Electromagnetics Research B*, Vol. 39, 225–239, 2012.
8. Luo, Y., Q. Zhang, Y.-Q. Bai, and Y.-L. Duan, "High-resolution ISAR imaging with sparse-spectrum OFDM-LFM waveform," *PIERS Proceedings*, 230–234, Kuala Lumpur, Malaysia, Mar. 27–30, 2012.
9. Liu, J., X. Li, S. Xu, and Z. Zhuang, "ISAR imaging of non-uniform rotation targets with limited pulse via compressed sensing," *Progress In Electromagnetics Research B*, Vol. 41, 285–305, 2012.
10. Gurbuz, A., J. McClellan, and W. Scott, Jr., "Compressive sensing for subsurface imaging using ground penetrating radar," *Signal Processing*, Vol. 89, No. 10, 1959–1972, 2009.
11. Stojanovic, I., M. Çetin, and W. Karl, "Joint space aspect reconstruction of wide-angle sar exploiting sparsity," *SPIE*, Vol. 6970, 2008.

12. Chen, C. and P. Vaidyanathan, "Compressive sensing in MIMO radar," *2008 42nd Asilomar Conference on Signals, Systems and Computers*, 41–44, IEEE, 2008.
13. Ender, J., "On compressive sensing applied to radar," *Signal Processing*, Vol. 90, No. 5, 1402–1414, 2010.
14. Patel, V., G. Easley, D. Healy, and R. Chellappa, "Compressed synthetic aperture radar," *IEEE Journal of Selected Topics, in Signal Processing*, Vol. 4, No. 2, 244–254, 2010.
15. Ben-Haim, Z., Y. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Trans. on Signal Processing*, Vol. 58, No. 10, 5030–5043, 2010.
16. Candes, E., J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, Vol. 59, No. 8, 1207–1223, 2006.
17. Candes, E. and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, Vol. 51, No. 12, 4203–4215, 2005.
18. Tropp, J., "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. on Information Theory*, Vol. 50, No. 10, 2231–2242, 2004.
19. Candès, E. J. and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, 2013.
20. Eaves, J. L. and E. K. Reedy, *Principles of Modern Radar*, Van Nostrand Reinhold Company, New York, 1987.
21. Candès, E. J. and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, Vol. 51, No. 13, 4203–4215, 2005.
22. Golub, G. H. and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1996.
23. Mackey, L., M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp, "Matrix concentration inequalities via the method of exchangeable Pairs," 29 pages, 2012.
24. Donoho, D. and V. Stodden, "Breakdown point of model selection when the number of variables exceeds the number of observations," *International Joint Conference on Neural Networks, IJCNN'06*, 1916–1921, 2006.
25. Bioucas-Dias, J. and M. Figueiredo, "Two-step algorithms for linear inverse problems with non-quadratic regularization," *IEEE International Conference on Image Processing, ICIP 2007*, Vol. 1, I-105–I-108, IEEE, 2007.