

Compressive Sensing with Orthonormal Measurements

Yi Zhou, Huishuai Zhang, and Yingbin Liang^{1,2,3}

Abstract

The Compressive Sensing (CS) approach for recovering sparse signal with orthonormal measurements has been studied under various notions of coherence. However, existing notions of coherence either do not exploit the structure of the underlying signal, or are too complicated to provide an explicit sampling scheme for all orthonormal basis sets. Consequently, there is lack of understanding of key factors that guide the sampling of CS with orthonormal measurements and achieve as low sample complexity as possible. In this paper, we introduce a new notion of π -coherence that exploits both the sparsity structure of the signal and the local coherence. Based on π -coherence, we propose a sampling scheme that is adapted to the underlying true signal and is applicable for CS under all orthonormal basis. Our scheme outperforms (up to a constant factor) existing sampling schemes for orthonormal measurements, and achieves a near-optimal sample complexity (up to certain logarithm factors) for several popular choices of orthonormal basis. Furthermore, we characterize the necessary condition on the sampling schemes for CS with orthonormal measurements. We then propose a practical multi-phase implementation of our sampling scheme, and verify its advantage over existing sampling schemes via application to magnetic resonance imaging (MRI) in medical science.

Index Terms

Compressive Sensing, Orthonormal basis, Coherence, Sample Complexity.

I. INTRODUCTION

With ever growing data size in signal processing applications, taking full number of linear measurements of the signal suffers from both high cost and low efficiency. In the past decade, numerous studies focus on recovering signals with reduced number of linear measurements by exploiting special structures of the signals. Among them,

¹ The material in this paper will be presented in part at the 54th Allerton Conference on Communications, Control, and Computing, Monticello Illinois, October 2016.

² The authors are affiliated with the Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, USA. {yzhou35, hzhang23, yliang06}@syr.edu.

³ This work is supported in part by the grants AFOSR FA9550-16-1-0077 and NSF ECCS 16-09916.

the sparsity of the signal is found to be the key factor to reduce the sample complexity [1], [2], [3], [4]. This leads to the development of the so-called compressive sensing (CS) technique, which has been successfully applied to various applications such as sparse MRI [5], face recognition [6], background subtraction [7] and photo-acoustic tomography [8], just to name a few.

We consider an orthonormal scenario in CS, which consists of a pair of orthonormal basis $\Psi, \Phi \in \mathbb{C}^{n \times n}$ with basis vectors $\Psi := [\psi_1^T; \dots; \psi_n^T]$, $\Phi := [\phi_1^T; \dots; \phi_n^T]$, and an underlying true signal $\mathbf{x}_0 \in \mathbb{C}^n$. The basis Ψ provides a sparse representation of the true signal, i.e., $\boldsymbol{\theta}_0 = \Psi \mathbf{x}_0$ is assumed to be s -sparse. The basis Φ corresponds to the space over which the orthonormal measurements $\{\langle \phi_j, \mathbf{x}_0 \rangle\}_{j=1}^n$ are taken. CS aims to recover the true signal \mathbf{x}_0 by taking as few number of measurements as possible. To be specific, denote $\Omega \subset \{1, \dots, n\}$ as the index set of the sampled orthonormal measurement vectors, and denote Φ_Ω as Φ restricted on the rows with index in Ω . Then CS aims to recover \mathbf{x}_0 via

$$(\mathbf{P}) \quad \min_{\mathbf{x} \in \mathbb{C}^n} \|\Psi \mathbf{x}\|_1 \quad \text{s.t.} \quad \Phi_\Omega \mathbf{x} = \Phi_\Omega \mathbf{x}_0.$$

A fundamental problem is to characterize the number of samples that guarantees that problem (P) uniquely recovers the underlying signal. It has been well understood that sample complexity is highly related to sampling schemes, and three types of sampling schemes have been studied so far.

(a) **Uniform** sampling is considered in [9], in which each orthonormal measurement $\phi_j, j = 1, \dots, n$ is equally likely taken. It has been shown via a RIPless¹ theory that the sample complexity depends on the following notion of coherence

$$(\text{Mutual Coherence}) \quad \mu(\Psi, \Phi) := \max_{1 \leq i, j \leq n} |\langle \psi_i, \phi_j \rangle|^2,$$

and is of the order $\mathcal{O}(\mu(\Psi, \Phi)sn \log n)$. In the low mutual coherence regime (i.e., $\mu(\Psi, \Phi) = \mathcal{O}(1/n)$), this sample complexity is order-wise optimal up to a logarithm factor. However, the sample complexity scales as large as the trivial order $\mathcal{O}(n \log n)$ if any pair of the basis vectors are exactly aligned (i.e., maximally coherent) with each other.

(b) **Measurement-adaptive** sampling scheme was studied in [10], [11], [12], in which each orthonormal measurement ϕ_j is sampled with probability being proportional to the so-called *local coherence*, i.e.,

$$(\text{Local Coherence}) \quad \mu(\Psi, \phi_j) := \max_{1 \leq i \leq n} |\langle \psi_i, \phi_j \rangle|^2. \quad (1)$$

The local coherence is determined by the basis pair and can be viewed as mutual coherence localized at the j th measurement. It was shown in [10] that the sample complexity is of the order $\mathcal{O}(\sum_j \mu(\Psi, \phi_j) s \log^3 s \log n)$ via a

¹RIP stands for the well known restricted isometry property.

RIP argument, and was further improved to $\mathcal{O}(\sum_j \mu(\Psi, \phi_j) s \log s \log n)$ via a RIPless argument in [11], [12].

(c) *Measurement-and-signal adaptive* sampling scheme was proposed in [11]. There, each sample is a subset of measurements (refer to as block measurements), and the sampling is based on the maximum of two notions of coherence that are related to both the support information of the underlying true signal and the basis pair. In a setting where each sample contains only one measurement (refer to as isolated measurement), Ψ being the bivariate Haar wavelet basis and Φ being the discrete Fourier basis, the sampling scheme is given in explicit form with a sample complexity on the order $\mathcal{O}(s \log s \log n)$.

The focus of this paper is to resolve remaining issues for the third type of sampling scheme. The notion of coherence introduced in [11] does not yield a universal sampling scheme for recovering the true signal in all orthonormal basis pairs, and may not be the real underlying quantity that governs the sample complexity. Hence, our goal is to find out the fundamental notion of coherence to guide the design of the sampling scheme, which 1) is adapted to the underlying true signal; 2) is generally applicable to all basis pairs; and 3) yields better sample complexity.

A. Main Contributions

We propose the new notion of π -coherence (see Definition 1) to capture how an orthonormal measurement ϕ_j is coherent with the subspace where the true signal lies. Based on the notion of π -coherence, we propose a Bernoulli sampling scheme that is generally applicable for all pairs of orthonormal basis.

Our sampling scheme reveals the relationship among sample complexity, sparsity structure of the true signal and coherence pattern of the basis pair. We further show that our scheme achieves lower sample complexity (within a constant factor) than uniform sampling [4], measurement-adaptive sampling [10], [12] and measurement-and-signal adaptive sampling [11]. For several popular choices of orthonormal basis pairs, we show that our sampling scheme achieves a near-optimal sample complexity (up to certain logarithm factors). Furthermore, we characterize the necessary condition on the sampling scheme for CS with orthonormal measurements to be successful. Lastly, we show that our sampling scheme maintains robustness in the noisy case.

Our technical proof introduces a weighted infinity norm to control the concentration bounds in a tighter way. Consequently, these new concentration bounds avoid involving complicated notions of coherence as those in [11], and lead to our notion of coherence that has an intuitive physical meaning on the sparsity structure of the signal and the coherence pattern of the orthonormal basis pair.

For practical applications, we propose a multi-phase version of our sampling scheme, which iteratively learns the subspace information of the true signal and updates the sampling scheme in the next phase.

B. Related Work

Various kinds of sampling schemes have been proposed for CS with orthonormal measurements. Some schemes are empirically oriented [13], [14], [15], [16], which were demonstrated to be useful in specific applications. Closely to our work are theoretically oriented schemes such as uniform sampling [4], [9], mixture of full sampling in finite regime and uniform sampling in asymptotic regime [17], measurement-adaptive sampling [10], [11], and measurement-and-signal adaptive sampling [11]. Our work falls into the last type of sampling scheme that is adapted to both the underlying true signal and the basis pair. Differently from [11] that explicitly derives sampling scheme only for a specific basis pair, our scheme is generally applicable to all orthonormal basis pairs and yields a better sample complexity.

There are other series of studies on adaptive CS [18], [19], [20]. There, the focus was on recovering the support of the signal via adapting the distribution from which the random measurements are generated, and do not involve any notion of coherence. In contrast, we are interested in recovering the signal by sampling deterministic orthonormal measurements based on our notion of coherence.

Adaptive sampling has also been considered in matrix-related problems by exploiting the leverage score [21], [22], [23], which is related to the projections of canonical unit vectors onto row/column spaces of the matrix. In particular, [23] proposed a sampling scheme that is adapted to the leverage scores of each entry of the matrix for matrix completion problems.

C. Organization and Notations

This paper proceeds as follows. Section II introduces the signal model and our notion of coherence; In Section III we present our sampling scheme and analyze the performance guarantee of the sampling scheme; In Section IV we propose a multi-phase algorithm to illustrate how to implement our sampling scheme in practice; In Section V we present numerical experiments to demonstrate our theoretical characterization, and finally we conclude the paper in Section VI.

We adopt the following notations in the paper:

\mathbf{x}_0 : the underlying signal to recover;

Ψ, Φ : the representation basis and measurement basis;

Ω : the index set of sampled measurements;

θ_0 : the representation of \mathbf{x}_0 in basis Ψ ;

S : support set of θ_0 ;

Π : the subspace of Ψ where \mathbf{x}_0 lies in.

$\|\cdot\|_p$: Euclidean l_p norm;

$\|\cdot\|$: spectral norm of an operator;

$\mathcal{I}, \mathcal{H}, \mathcal{F}$: the canonical basis, bivariate Haar wavelet basis and discrete Fourier basis, respectively.

II. SIGNAL MODEL AND π -COHERENCE

Suppose that the representation of signal \mathbf{x}_0 in basis Ψ , i.e., $\boldsymbol{\theta}_0 = \Psi\mathbf{x}_0$, is s -sparse. Denote S as the support set of $\boldsymbol{\theta}_0$. Then the following two complementary subspaces are well defined

$$\mathbf{\Pi} : \text{span}\{\boldsymbol{\psi}_j \mid j \in S\}, \quad \mathbf{\Pi}^\perp : \text{span}\{\boldsymbol{\psi}_j \mid j \in S^c\}. \quad (2)$$

It follows that $\mathbf{x}_0 \in \mathbf{\Pi}$, and $\dim(\mathbf{\Pi}) = s$. Following [9], [24], we consider the following random sign model of the underlying signal.

Assumption 1. *The sign of the non-zero entries of $\boldsymbol{\theta}_0$ is distributed independently from their locations as*

$$\forall j \in S, \quad \text{sgn}(\boldsymbol{\theta}_0)_j = \begin{cases} 1, & \text{w.p. } 1/2 \\ -1, & \text{w.p. } 1/2 \end{cases}$$

We also denote $P_{\mathbf{\Pi}}, P_{\mathbf{\Pi}^\perp}$ as the projection operators onto $\mathbf{\Pi}, \mathbf{\Pi}^\perp$, respectively. Now we propose the following notion of coherence, which plays a fundamental role in our design of the sampling scheme and characterization of its performance guarantee.

Definition 1. (π -coherence) *The π -coherence of measurement ϕ_j w.r.t. space $\mathbf{\Pi}$ is defined as*

$$\pi(\mathbf{\Pi}, \phi_j) := \max\{\|P_{\mathbf{\Pi}}\phi_j\|_2^2, \|P_{\mathbf{\Pi}}\phi_j\|_2\|\Psi P_{\mathbf{\Pi}^\perp}\phi_j\|_\infty\}.$$

In the above definition of π -coherence, the first term (i.e., l_2^2) corresponds to the ‘‘energy’’ of each measurement ϕ_j onto $\mathbf{\Pi}$. Intuitively, a larger value of this term implies that measurement ϕ_j is more aligned with the signal subspace $\mathbf{\Pi}$, and hence can retrieve more information of the signal. Moreover, the l_2^2 structure of this term also reveals the sparsity structure of the signal. To be specific, consider *arbitrary* orthogonal decomposition of $\mathbf{\Pi}$ into p subspaces $\{\mathbf{\Pi}_l\}_{l=1}^p$ with corresponding dimensions $\{s_l\}_{l=1}^p$. Then for all j , one has

$$\mathbf{\Pi} = \bigoplus_{l=1}^p \mathbf{\Pi}_l, \quad \|P_{\mathbf{\Pi}}\phi_j\|_2^2 = \sum_{l=1}^p \|P_{\mathbf{\Pi}_l}\phi_j\|_2^2. \quad (3)$$

With regard to the full support $\dim(\mathbf{\Pi}) = s$ of the signal, $\{\dim(\mathbf{\Pi}_l) = s_l\}_{l=1}^p$ can be viewed as a sparsity structure into subspace decompositions. Correspondingly, each term $\|P_{\mathbf{\Pi}_l}\phi_j\|_2^2$ measures the alignment of ϕ_j with subspace $\mathbf{\Pi}_l$, and their summation yields the full coherence.

The second term (i.e., $l_2 \cdot l_\infty$) further contains an l_∞ component, which can be interpreted as (square root of) the local coherence of ϕ_j with regard to $\mathbf{\Pi}^\perp$, i.e.,

$$\|\Psi \mathbf{P}_{\mathbf{\Pi}^\perp} \phi_j\|_\infty = \sqrt{\mu(\mathbf{\Pi}^\perp, \phi_j)} := \max_{i \in S^c} |\langle \psi_i, \phi_j \rangle|. \quad (4)$$

Intuitively, such local coherence is involved because one needs to recover zeros on $\mathbf{\Pi}^\perp$ (recall that $\mathbf{x}_0 \in \mathbf{\Pi}$) to guarantee fully correct recovery of the signal.

In summary, our notion of π -coherence incorporates the general sparsity structure of the signal as in eq. (3) and the local coherence as in eq. (4). This is different from the mutual coherence [4], [9], the asymptotic coherence [17], and the local coherence [10], [11], none of which exploits the sparsity structure of the signal. Moreover, it has a simpler form than the notions of coherence introduced in [11], and is more general in subspace decomposition as eq. (3). In the next section, we demonstrate that our notion of coherence serves more fundamental purpose in characterizing sufficient conditions on sample complexity for performance guarantee.

III. MAIN RESULTS

A. Sampling Scheme and Performance Guarantee

We propose the following Bernoulli sampling, i.e., measurement ϕ_j is taken with probability given by

$$\mathbb{P}(j \in \Omega) \sim \text{Bernoulli}(p_j), \quad \text{for } j = 1, \dots, n. \quad (5)$$

The sampling probability is set based on the π -coherence as we state in the following theorem.

Theorem 1. *Let Assumption 1 hold and fix any pair of orthonormal basis Ψ, Φ . Then \mathbf{x}_0 is the unique minimizer of (\mathbf{P}) with probability at least $1 - s^{-\sqrt{C_0}}$, provided that*

$$p_j \geq \min\{C_0 \pi(\mathbf{\Pi}, \phi_j) \log s \log n, 1\} \quad (6)$$

$$p_j \geq 1/s^{20}. \quad (7)$$

where $C_0 > 1$ is a universal constant.

Proof. The detailed proof is provided in Appendix A. The proof is based on the convex duality argument and the so called golfing scheme originated from matrix completion literature [25]. Specially, novelty of our proof lies in introduction of a weighted infinity norm, which together with the π -coherence allows to control the concentration bounds in a tighter way. We refer to Appendix A for the detailed definitions and proof. \square

In fact, the constraint in eq. (7) is to avoid singularity in the proof, and the power of s can be raised up further by slightly increasing C_0 . Effectively, eq. (6) determines the sampling scheme, which requires to sample the ϕ_j

measurement with probability proportional to the corresponding π -coherence $\pi(\mathbf{\Pi}, \phi_j)$. Thus, inheriting from π -coherence, our sampling scheme exploits the sparsity structure of the signal in eq. (3) and the local coherence in eq. (4). Intuitively, if measurement ϕ_j is more aligned with the signal space $\mathbf{\Pi}$, then such measurement is sampled with higher probability.

It may appear that since the sampling scheme exploits the information of signal space $\mathbf{\Pi}$, the objective of problem **(P)** should also exploit such information for signal recovery. This is not true because we usually do not know the subspace $\mathbf{\Pi}$ accurately in practice. As in the practical multi-phase sampling scheme we provide in Section IV, the subspace information $\mathbf{\Pi}$ is iteratively learned to update the sampling scheme.

We further note that our sampling scheme is generally applicable to all pairs of orthonormal basis. This is in contrast to the same type, but an alternative sampling scheme in [11] that also exploits both signal and basis information based on two complicated notions of coherence. There, the sampling scheme in general do not have closed form. Only for the MRI example (with Haar wavelet basis and Fourier basis), an explicit scheme is obtained, which samples uniformly among the Fourier measurements in different dyadic levels based on the sparsity of the signal in different levels of wavelet basis. In contrast, our sampling scheme exploits the fully decomposable sparsity structure in eq. (3) for all orthonormal basis.

Theorem 1 also characterizes the expected number of measurements, i.e., $\sum_{j=1}^n p_j$, that guarantees correct recovery of the signal. Clearly, the π -coherence plays a central role similar to other notions of coherence in determining sample complexity. Thus, in the following theorem, we compare π -coherence with other notions of coherence, which thus yields comparison of sample complexity among the corresponding sampling schemes.

Theorem 2. *For $j = 1, \dots, n$, the following inequality that compares various notions of coherence holds:*

$$\pi(\mathbf{\Pi}, \phi_j) \leq \|\Psi P_{\mathbf{\Pi}} \phi_j\|_1 \sqrt{\mu(\Psi, \phi_j)} \leq s\mu(\Psi, \phi_j) \leq s\mu(\Psi, \Phi).$$

Proof. See Appendix B. □

The sample complexities of the schemes in [10], [9] are proportional to $s \sum_j \mu(\Psi, \phi_j)$ and $sn\mu(\Psi, \Phi)$, respectively. Hence, by summing over $j = 1, \dots, n$, the last two inequalities in Theorem 2 implies that our scheme achieves a lower sample complexity (up to a logarithm factor). Also, [11] achieves a sample complexity being proportional to $\max\{\Theta, \Gamma\}$, where Θ, Γ are two notions of coherence that both depend on the sampling scheme. There is no general closed form for the optimal sampling scheme that minimizes the two coherences. However, the coherence Θ has a minimizing sampling scheme $p_j \propto \|\Psi P_{\mathbf{\Pi}} \phi_j\|_1 \sqrt{\mu(\Psi, \phi_j)}$ in the case of isolated measurement². Hence by the first inequality of Theorem 2, our scheme achieves a lower sample complexity.

²Recall that isolated measurement corresponds to a single measurement.

B. Complexity, Sparsity and Coherence Pattern

In this subsection, we understand further the sample complexity of our sampling scheme. Due to eq. (6), the sample complexity (i.e., $\sum_{j=1}^n p_j$) is upper bounded by the summation of the two terms involved in π -coherence. The summation of the first l_2^2 term provides a clear view of the sparsity structure of the signal. Specifically, consider the general orthogonal decomposition in eq. (3), one has

$$\sum_{j=1}^n \|\mathbf{P}_{\mathbf{\Pi}} \phi_j\|_2^2 = \sum_{l=1}^p \sum_{j=1}^n \|\mathbf{P}_{\mathbf{\Pi}_l} \phi_j\|_2^2 = \sum_{l=1}^p s_l = s. \quad (8)$$

That is, the summation of the first term is a collection of the sparsity of the signal in each subspace $\mathbf{\Pi}_l$, and in total contributes the whole sparsity of the signal to the sample complexity. The second $l_2 \cdot l_\infty$ term is related to both the sparsity structure and the local coherence, and thus its summation depends on both the sparsity structure and the pattern of the local coherence.

We next demonstrate that the above characterization of the sample complexity for all basis pairs $\mathbf{\Psi}$ and $\mathbf{\Phi}$ reduces to the state-of-the-art results in specific examples. In particular, we wish to understand more explicitly how interaction of the two terms in π -coherence yields the sample complexity via these specific examples. In the sequel, $\mathcal{I}, \mathcal{H}, \mathcal{F}$ denote the canonical basis, bivariate Haar wavelet basis and discrete Fourier basis, respectively.

Example 1. (*maximally incoherent*)

$\mathbf{\Psi} = \mathcal{I}, \mathbf{\Phi} = \mathcal{F}$, and $\mathcal{I}\mathbf{x}_0$ is s -sparse.

This example arises in applications that aims to recover sparse signals via a number of Fourier measurements [26]. It is well known that this pair of basis have an incoherent local coherence pattern, i.e., for all $i, j \in \{1, \dots, n\}$

$$\mu(\psi_i, \phi_j) = 1/n. \quad (9)$$

Then a simple calculation yields that $\forall j, \pi(\mathbf{\Pi}, \phi_j) \equiv s/n$, and consequently our sampling scheme becomes the optimal uniform sampling scheme for this example with sample complexity being of the order $\mathcal{O}(s \log s \log n)$.

Example 2. (*highly coherent*)

$\mathbf{\Psi} = \mathcal{H}, \mathbf{\Phi} = \mathcal{F}$, and $\mathcal{H}\mathbf{x}_0$ is s -sparse.

This example arises in the popular magnetic resonance imaging (MRI) application [27], [11], [10]. There, \mathbf{x}_0 corresponds to medical image that can be sparsely represented in the Haar wavelet basis, and the physical device takes Fourier measurements of the signal. While the mutual coherence is as high as $\mu(\mathcal{H}, \mathcal{F}) = 1$, this pair of basis have a well behaved local coherence pattern. To be specific, we assume without of loss of generality that $n = 2^p$ for some interger p and introduce a dyadic partition of the index set $\{1, \dots, n\}$ into $p + 1$ levels, i.e.,

$L_0 = \{1\}, L_1 = \{2\}, L_2 = \{3, 4\}, \dots, L_p = \{n/2 + 1, \dots, n\}$. For any $j \in \{1, \dots, n\}$ we also define $k(j)$ as the dyadic level that j belongs to, i.e., $j \in L_{k(j)}$. Then for all $i, j \in \{1, \dots, n\}$, the following pattern of local coherence follows from Lemma D.1 in [11]

$$\mu(\psi_i, \phi_j) \leq 2^{-k(j)} 2^{-|k(j)-k(i)|}. \quad (10)$$

Intuitively, the pair of basis are more incoherent in the asymptotic region (i.e., large i, j), and this observation motivates the introduction of the asymptotic coherence in [17]. By exploiting the coherence pattern in eq. (10), we can control the summation of the second term in π -coherence, and characterize the following sample complexity for our sampling scheme.

Proposition 1. *With the coherence pattern in eq. (10), one yields that*

$$\sum_{j=1}^n \|\mathbf{P}_{\mathbf{\Pi}} \phi_j\|_2 \|\Psi \mathbf{P}_{\mathbf{\Pi}^\perp} \phi_j\|_\infty \lesssim \sum_{l=0}^p \sqrt{s_l}, \quad (11)$$

where $\{s_l\}_{l=0}^p$ are sparsity of the signal in different levels of wavelet basis. Consequently, the sample complexity of eq. (6) for Example 2 is of the order $\mathcal{O}(s \log s \log n)$.

Proof. See Appendix C. □

In comparison, our sampling scheme in eq. (6) achieves better sample complexity than [10] for Example 2 (which is on the order $\mathcal{O}(s \log^3 s \log^2 n)$), and recovers the same order-level sample complexity as [11] for Example 2. However, the first inequality in Theorem 2 implies that our sample complexity is in fact lower than that in [11] for sampling with isolated measurement, although they are on the same order. One possible reason is due to the fact that sampling scheme in [11] samples uniformly among the Fourier measurements over the same dyadic levels, leading to a block structure of the sampled Fourier frequencies which does not fit the spectrum of natural images. The advantage of our sampling scheme turns out to be more prominent in experiments in Section V.

Example 3. *(maximally coherent)*

$\Psi = \Phi$, and $\Psi \mathbf{x}_0$ is s -sparse.

That is, the basis pair are exactly aligned with

$$\mu(\psi_i, \phi_j) = \mathbb{I}\{i = j\}, \quad (12)$$

where $\mathbb{I}\{\cdot\}$ is the binary indicator. Thus, π -coherence satisfies $\pi(\mathbf{\Pi}, \phi_j) = \mathbb{I}\{\phi_j \in \mathbf{\Pi}\}$, and Theorem 1 implies that with $C_0 \log s \log n > 1$, $p_j = 1$ if $\phi_j \in \mathbf{\Pi}$, and $p_j = 0$ otherwise. Namely, sample only the measurements that are in $\mathbf{\Pi}$. This is intuitive because only if each basis vector in the signal space $\mathbf{\Pi}$ is sampled, the signal component

over that basis vector can be recovered. Clearly, the sample complexity is $\mathcal{O}(s)$, which is in contrast to the basis adaptive sampling scheme in [10] that samples all n measurements for this example. This extreme example justifies the advantage of incorporating support structure of the signal into our sampling scheme.

The above three examples represent different types of coherence patterns. Our sampling scheme based on π -coherence for general basis pairs achieves near optimal sample complexity when specialized to these cases.

C. Necessary Condition for Orthonormal Sensing

In this subsection, we present a necessary condition on sample complexity to guarantee success of sampling schemes for orthonormal sensing.

Theorem 3. Fix a pair of orthonormal basis Ψ, Φ . If for all $j = 1, \dots, n$

$$p_j = \|\mathbf{P}_{\Pi}\phi_j\|_2^2 - \delta_j, \quad (13)$$

where $\{\delta_j\}_j$ are positive numbers satisfying $\delta_j < \|\mathbf{P}_{\Pi}\phi_j\|_2^2$ and $\sum_j \delta_j \geq 0.1s + 2$. Then there are infinitely many solutions other than \mathbf{x}_0 achieving the same objective value of problem (P) with probability at least 0.1.

Proof. See Appendix D. □

An intuitive explanation for the above necessary condition is that the number of measurements sampled by eq. (13) can be less than its expectation s , and hence the linear system in problem (P) becomes under-determinant for uniquely identifying an s -sparse signal. One can see that the quantities in eq. (13) and eq. (6) have a gap, i.e., the additional $l_2 \cdot l_\infty$ term in the π -coherence (excluding the logarithm terms). This might be an artifact of the duality argument and the use of weighted norm in the proof, but fortunately it does not affect the order of sample complexity of the three popular examples discussed in Section III-B. We note that the necessary condition does not rule out the case that eq. (13) holds only for a subset of $j = 1, \dots, n$, and may still recover the signal.

D. Robustness to Noise

In this subsection, we further characterize the performance of our sampling scheme when the measurements of the signal are corrupted by noise. Consider the following noisy observation model

$$\mathbf{y} = \Phi_{\Omega}\mathbf{x}_0 + \Phi_{\Omega}\mathbf{n}$$

where \mathbf{n} denotes the noise with bounded energy, i.e., $\|\Phi_{\Omega}\mathbf{n}\|_2 \leq \delta$. We consider the following problem (Q) to recover the signal under the noisy model

$$\text{(Q)} \quad \min_{\mathbf{x} \in \mathbb{C}^n} \|\Psi\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\Phi_{\Omega}(\mathbf{x} - \mathbf{y})\|_2 \leq \delta. \quad (14)$$

Then the following theorem states that our sampling scheme guarantees problem **(Q)** to recover the signal with certain bounded error.

Theorem 4. *Let the assumptions in Theorem 1 hold. Then with the same sampling scheme and the same probability guarantee in Theorem 1, the minimizer $\hat{\mathbf{x}}$ of problem **(Q)** satisfies*

$$\sqrt{\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2} \leq C\sqrt{n}\delta. \quad (15)$$

Furthermore, if $\Psi = \mathcal{I}$, then the above result can be improved to

$$\sqrt{\mathbb{E}\|\hat{\mathbf{x}} - \mathbf{x}_0\|_2^2} \leq C\sqrt{|\Omega|}\delta, \quad (16)$$

where C is a universal positive constant.

Proof. See Appendix E. □

Theorem 4 implies that our sampling scheme eq. (6) maintains robustness in the noisy case, and the improved MSE in eq. (16) is comparable to the result in [4, eq.(IV.2)] for recovering s -sparse signal when the number of measurements $|\Omega|$ is of the order s . We note that there is a critical difference between the special case $\Psi = \mathcal{I}$ and other cases $\Psi \neq \mathcal{I}$. The noisy recovery problem **(Q)** with $\Psi = \mathcal{I}$ is similar to the Dantzig selector in [4, eq.(I.14)], and we can conveniently tighten one key step of our analysis as $\|\Psi(\Phi_\Omega \mathbf{h})\|_1 = \|\Phi_\Omega \mathbf{h}\|_1 \leq \sqrt{|\Omega|}\|\Phi_\Omega \mathbf{h}\|_2$. However, for a general orthonormal transformation Ψ , the vector $\Psi(\Phi_\Omega \mathbf{h})$ is no longer restricted on Ω , and the upper bound can be at the level $\sqrt{n}\|\Phi_\Omega \mathbf{h}\|_2$. Intuitively speaking, the noise occurring in the Φ_Ω orthonormal measurements can become dense after transforming into basis Ψ for recovering the underlying signal.

IV. MULTI-PHASE SAMPLING SCHEME

Our adaptive sampling scheme requires the knowledge of the signal space Π , which is typically unknown a priori. We now propose a multi-phase implementation of our sampling scheme, where the signal space is estimated via previous phases, and is then exploited in subsequent phases for adaptive sampling. More specifically, suppose the multi-phase sampling scheme consists of P phases, and each phase takes m measurements. Clearly, we require $mP \leq n$. Then the main steps in each phase $p = 1, \dots, P$ are described in Algorithm 1.

We note that one can of course use other sampling schemes in the initial phase than the uniform sampling scheme. We use uniform sampling scheme to emphasize the advantage of our adaptive sampling scheme in subsequent phases. Moreover, for the noisy case we solve problem **(Q)** instead of problem **(P)** for the corresponding steps.

Trim the Π space: In the multi-phase sampling scheme, estimation of Π, Π^\perp from the recovered signal \mathbf{x}_p is based on the definition in eq. (2). However, in practice, all entries of the representation vector $\theta_p := \Psi \mathbf{x}_p$

Algorithm 1 Multi-phase Sampling for CS

Initial phase $p = 1$:

- 1). Uniform sampling scheme $\xrightarrow{\text{get}} \Omega_1$ with m measurements.
- 2). Solve problem **(P)** with $\Omega \stackrel{\text{set}}{=} \Omega_1$ for the solution \mathbf{x}_1 , and evaluate its $\mathbf{\Pi}, \mathbf{\Pi}^\perp$ according to eq. (2).

Phase $p = 2, \dots, P$:

- 1). Evaluate p_j in eq. (6) with the $\mathbf{\Pi}, \mathbf{\Pi}^\perp$ evaluated in previous phase for unsampled $j \notin \cup_{k=1}^{p-1} \Omega_k$; Set $p_j = 0$ for the sampled measurements.
- 2). Normalize $\{p_j\}$ to a probability distribution, according to which sample m measurements sequentially to get Ω_p . After each sample j we set $p_j = 0$ and renormalize $\{p_j\}$.
- 3). Solve problem **(P)** with $\Omega \stackrel{\text{set}}{=} \cup_{k=1}^p \Omega_k$ for the solution \mathbf{x}_p , and evaluate its $\mathbf{\Pi}, \mathbf{\Pi}^\perp$ according to eq. (2).

Output \mathbf{x}_P .

can be strictly non-zero with majority components having very small magnitudes. Such phenomenon yields $\mathbf{\Pi}$ that span over the entire space, and the sampling scheme in eq. (6) reduces to the uniform sampling scheme and is hence not signal adaptive. To overcome this issue, we relax the criterion in eq. (2) by removing the entries with small magnitudes. To be specific, we set a threshold of the magnitude $\tilde{\theta} > 0$ with which $\mathbf{\Pi}$ is set to be $\text{span}\{\psi_j, |(\theta_p)_j| > \tilde{\theta}\}$. The threshold is searched among all magnitudes of entries to satisfy the following criterion

$$\|\mathbf{P}_{\mathbf{\Pi}}\mathbf{x}_p\|_2^2 = 0.95\|\mathbf{x}_p\|_2^2. \quad (17)$$

That is, $\tilde{\theta}$ thresholds out those entries with small magnitudes and keeps 95% of the energy of \mathbf{x}_p . The resulting $\mathbf{\Pi}$ is then spanned by a small set of basis vectors with coefficients of high magnitudes, providing accurate approximation with a much smaller dimension. We note that one can of course change 95% to other quantities, and it controls the approximation error via $\|\mathbf{P}_{\mathbf{\Pi}}\mathbf{x}_p - \mathbf{x}_p\|_2^2 = 5\%$.

V. NUMERICAL EXPERIMENTS

In this section, we study the performance of our basis-and-signal adaptive sampling scheme defined in eq. (6), and compare it with the uniform sampling [4], basis-adaptive sampling [10], and uniform by level sampling (with isolated measurement) [11]. Since our scheme and the uniform by level sampling [11] are signal-dependent, we implement them by the multi-phase scheme described in Algorithm 1. Since some of the above sampling schemes are proposed only for special cases, we use the common Example 2 as the test example. All experiments are repeated five times and the average is reported as the final result.

A. Total Budget v.s. Relative Error

We first consider the noise-free setting, and apply the sampling schemes of interest to recover a 256×256 MRI brain image. The budget of measurements is set to be 20%, 25%, \dots , 70% of the total dimensions, respectively. The number P of phases for the multi-phase sampling schemes is set to be two – the lowest number we found that gives a good performance empirically. Denote \mathbf{x}^* as the original image and $\hat{\mathbf{x}}$ as the recovered image. We report the relative error $\frac{\|\mathbf{x}^* - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}^*\|_2}$. Figure 1 plots how the relative error reduces as the total budget increases for four sampling schemes. It can be seen that the uniform sampling [4] and the uniform by level scheme [11] suffer from high relative error and instability, while the basis-adaptive scheme in [10] and our scheme provide more accurate and stable recovery. In particular, our scheme outperforms all other three schemes. It is superior over the uniform and basis adaptive sampling due to the fact that it adapts to both the local coherence of basis and the signal space Π . Although the uniform by level sampling also adapts to the signal information, adaption is limited only across levels, not over the entire basis vectors as in our scheme.

To provide a better illustration, we present in Figure 3 the recovered images and their corresponding sampling masks of the sampled Fourier frequencies for the setting with 30% budget of total dimensions.

One can see that the image recovered by our sampling scheme has the best quality. Note that we have shifted the Fourier basis to center the spectrum and the sampling masks. One can see that the sampling mask of the uniform by level scheme in [11] has a block structure, which does not fit the spectrum of the original signal well. The sampling mask of the basis-adaptive scheme in [10] is a good approximation of the spectrum of practical smooth images, but is generated by a fixed probability distribution that is not adapted to the underlying signal. Our sampling mask is the mixture of a uniform mask (in the initial phase) and a mask generated by adaptive sampling scheme (in the second phase). Despite its uniform part of the mask, one can see that the mask of our scheme is well adapted to the spectrum of the original signal.

B. Robustness to Noise

We further consider the noisy case by adding a Gaussian noise $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$, and the standard deviation is set to be $\sigma = 1, 2, 5, 10$, respectively. Then a standard probability argument [28] shows that with high probability $\delta^2 \leq (|\Omega| + \sqrt{8|\Omega|})\sigma^2$, where $|\Omega| = mp$ in phase $p + 1$. We then set the budget to be 30% measurements of the total dimensions, and solve problem **(Q)** for four sampling schemes of interest. Figure 2 plots the relationship between σ and the relative error. One can see that our scheme is most robust to noise.

VI. CONCLUSION

We propose a Bernoulli sampling scheme for CS with orthonormal basis, which is applicable to all orthonormal basis pairs. The scheme is based on π -coherence that exploits both the support structure of the signal and the local

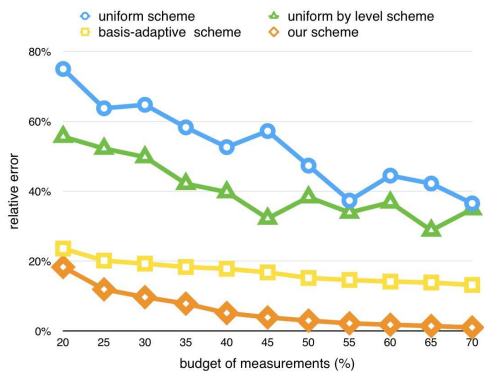


Fig. 1. Comparison of relative errors of four sampling schemes to recover a medical image as the number of measurements increases.

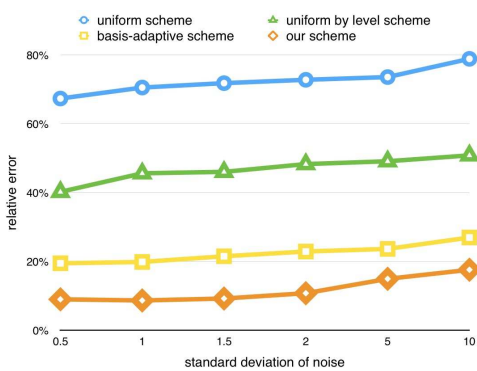


Fig. 2. Comparison of relative errors of four sampling schemes to recover a noisy image as the variance of the noise increases.

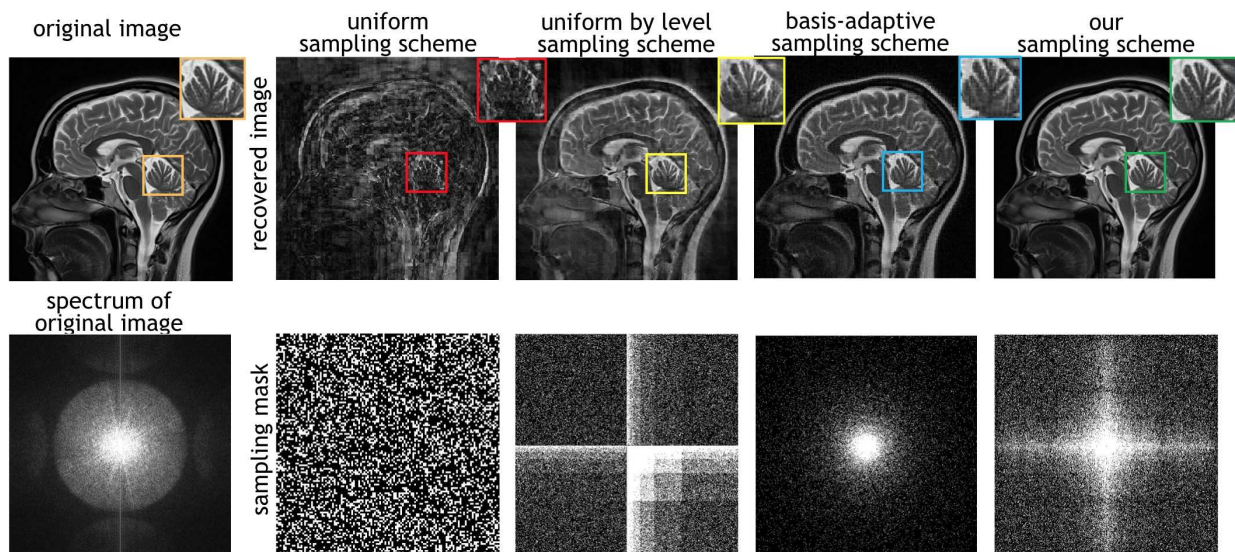


Fig. 3. Comparison of recovered images and sampling masks of four sampling schemes with 30% measurements of total dimensions.

coherence. Our sampling scheme reveals the relationship between sample complexity and sparsity of the signal as well as coherence pattern of the basis, and achieves lower sample complexity (within a constant factor) than that of existing basis and signal adaptive sampling schemes. Furthermore, we characterize the necessary conditions to guarantee the success of sampling scheme for orthonormal measurements. We also propose a practical multi-phase implementation of our sampling scheme, and demonstrate its advantage over other existing sampling schemes via experiments. We anticipate that such signal-dependent adaptive sampling together with multi-phase implementation can be useful for other high dimensional problems with limited budget of measurements.

APPENDIX A

PROOF OF THEOREM 1

We first characterize the sub-differential of $\|\Psi \mathbf{x}_0\|_1$ via the following proposition for convenience.

Proposition 2. The sub-differential of $\|\Psi \mathbf{x}_0\|_1$ w.r.t. \mathbf{x}_0 can be expressed as $\partial_{\mathbf{x}_0} \|\Psi \mathbf{x}_0\|_1 = \gamma_S + \gamma_{S^c}$, where

$$\gamma_S = \sum_{j \in S} \text{sgn}(\boldsymbol{\theta})_j \boldsymbol{\psi}_j, \quad \gamma_{S^c} := \left\{ \sum_{j \in S^c} \mathbf{f}_j \boldsymbol{\psi}_j \mid \|\mathbf{f}\|_\infty \leq 1 \right\}.$$

Moreover, $\gamma_S \in \Pi$ and $\gamma_{S^c} \in \Pi^\perp$.

Proof. Consider an arbitrary vector \mathbf{x} with its support set denoted as S . It is well known that

$$\partial \|\mathbf{x}\|_1 := \{\text{sgn}(\mathbf{x}) + \mathbf{f} \mid \text{supp}(\mathbf{f}) = S^c, \|\mathbf{f}\|_\infty \leq 1\}. \quad (18)$$

Then for the vector $\boldsymbol{\theta} = \Psi \mathbf{x}_0$ with the support set S , the chain rule gives

$$\partial_{\mathbf{x}_0} \|\Psi \mathbf{x}_0\|_1 = \Psi^T \partial \|\Psi \mathbf{x}_0\|_1 \stackrel{(i)}{=} \Psi^T (\text{sgn}(\boldsymbol{\theta}) + \mathbf{f}) \stackrel{(ii)}{=} \underbrace{\sum_{j \in S} \text{sgn}(\boldsymbol{\theta})_j \boldsymbol{\psi}_j}_{\gamma_S} + \underbrace{\sum_{j \in S^c} \mathbf{f}_j \boldsymbol{\psi}_j}_{\gamma_{S^c}}, \quad (19)$$

where (i) follows from the sub-differential in eq. (18), and (ii) follows from the fact that $\text{supp}(\boldsymbol{\theta}) = S$, and $\text{supp}(\mathbf{f}) = S^c$. Clearly, $\gamma_S \in \Pi, \gamma_{S^c} \in \Pi^\perp$. \square

Next we introduce the random variable $\delta_j \sim \text{Bernoulli}(p_j)$, and define the weighted sampling operators

$$R_\Omega(\mathbf{x}) := \sum_{j=1}^n \frac{\delta_j}{p_j} \langle \mathbf{x}, \boldsymbol{\phi}_j \rangle \boldsymbol{\phi}_j, \quad R_\Omega^{1/2}(\mathbf{x}) = \sum_{j=1}^n \frac{\delta_j}{\sqrt{p_j}} \langle \mathbf{x}, \boldsymbol{\phi}_j \rangle \boldsymbol{\phi}_j.$$

Then we present below a set of conditions that guarantee \mathbf{x}_0 to be the unique minimizer of problem **(P)**.

Lemma 1. If there exists a vector $\boldsymbol{\nu}$ such that the following conditions hold

$$\Phi_{\Omega^c} \boldsymbol{\nu} = 0; \quad (20)$$

$$\|\Psi_{S^c} P_{\Pi^\perp} \boldsymbol{\nu}\|_\infty \leq 1/4; \quad (21)$$

$$\|P_\Pi (\gamma_S - \boldsymbol{\nu})\|_2 \leq 1/2s^{10}; \quad (22)$$

$$\|P_\Pi R_\Omega P_\Pi - P_\Pi\| \leq 1/2, \quad (23)$$

then problem **(P)** has the unique minimizer \mathbf{x}_0 .

Proof. Our goal is to show that any feasible perturbation of the signal, i.e., $\mathbf{x}_0 + \Delta \mathbf{x}$ with $\Phi_\Omega \Delta \mathbf{x} = 0$, strictly

increases the objective function of problem **(P)** under the above set of conditions. By convexity we have

$$\begin{aligned}
\|\Psi(\mathbf{x}_0 + \Delta\mathbf{x})\|_1 &\geq \|\Psi\mathbf{x}_0\|_1 + \langle \Delta\mathbf{x}, \partial_{\mathbf{x}_0}\|\Psi\mathbf{x}_0\|_1 \rangle \\
&\stackrel{(i)}{=} \|\Psi\mathbf{x}_0\|_1 + \langle \Delta\mathbf{x}, \gamma_S + \gamma_{S^c} \rangle \\
&= \|\Psi\mathbf{x}_0\|_1 + \langle \Delta\mathbf{x}, \gamma_S - \boldsymbol{\nu} \rangle + \langle \Delta\mathbf{x}, \boldsymbol{\nu} \rangle + \langle \Delta\mathbf{x}, \gamma_{S^c} \rangle \\
&\stackrel{(ii)}{=} \|\Psi\mathbf{x}_0\|_1 + \langle \Delta\mathbf{x}, \gamma_S - \boldsymbol{\nu} \rangle + \langle \Delta\mathbf{x}, \gamma_{S^c} \rangle \\
&\stackrel{(iii)}{=} \|\Psi\mathbf{x}_0\|_1 + \langle P_{\Pi}\Delta\mathbf{x}, P_{\Pi}(\gamma_S - \boldsymbol{\nu}) \rangle \\
&\quad + \langle P_{\Pi^\perp}\Delta\mathbf{x}, -P_{\Pi^\perp}\boldsymbol{\nu} \rangle + \langle P_{\Pi^\perp}\Delta\mathbf{x}, \gamma_{S^c} \rangle \\
&\stackrel{(iv)}{=} \|\Psi\mathbf{x}_0\|_1 + \langle P_{\Pi}\Delta\mathbf{x}, P_{\Pi}(\gamma_S - \boldsymbol{\nu}) \rangle \\
&\quad + \langle \Psi P_{\Pi^\perp}\Delta\mathbf{x}, -\Psi P_{\Pi^\perp}\boldsymbol{\nu} \rangle + \langle \Psi P_{\Pi^\perp}\Delta\mathbf{x}, \mathbf{f} \rangle \\
&\stackrel{(v)}{=} \|\Psi\mathbf{x}_0\|_1 + \langle P_{\Pi}\Delta\mathbf{x}, P_{\Pi}(\gamma_S - \boldsymbol{\nu}) \rangle \\
&\quad + \langle \Psi P_{\Pi^\perp}\Delta\mathbf{x}, -\Psi_{S^c}P_{\Pi^\perp}\boldsymbol{\nu} \rangle + \langle \Psi P_{\Pi^\perp}\Delta\mathbf{x}, \mathbf{f} \rangle,
\end{aligned}$$

where (i) follows from Proposition 2, (ii) is due to eq. (20) and the fact that $\Phi_{\Omega}\Delta\mathbf{x} = 0$, (iii) follows from the facts that $\gamma_S \in \Pi, \gamma_{S^c} \in \Pi^\perp$, (iv) follows from the fact that $\Psi\gamma_{S^c} = \Psi\Psi^T\mathbf{f} = \mathbf{f}$, and (v) follows because $\Psi P_{\Pi^\perp} = \Psi_{S^c}P_{\Pi^\perp}$.

Since $\|\mathbf{f}\|_\infty \leq 1$, the duality between l_1 and l_∞ guarantees the existence of an \mathbf{f} such that $\langle \Psi P_{\Pi^\perp}\Delta\mathbf{x}, \mathbf{f} \rangle = \|\Psi P_{\Pi^\perp}\Delta\mathbf{x}\|_1$. Then the last equation above further gives

$$\begin{aligned}
\|\Psi(\mathbf{x}_0 + \Delta\mathbf{x})\|_1 &\geq \|\Psi\mathbf{x}_0\|_1 + \langle P_{\Pi}\Delta\mathbf{x}, P_{\Pi}(\gamma_S - \boldsymbol{\nu}) \rangle \\
&\quad + \langle \Psi P_{\Pi^\perp}\Delta\mathbf{x}, -\Psi_{S^c}P_{\Pi^\perp}\boldsymbol{\nu} \rangle + \|\Psi P_{\Pi^\perp}\Delta\mathbf{x}\|_1 \\
&\geq \|\Psi\mathbf{x}_0\|_1 - \|P_{\Pi}\Delta\mathbf{x}\|_2 \|P_{\Pi}(\gamma_S - \boldsymbol{\nu})\|_2 \\
&\quad - \|\Psi P_{\Pi^\perp}\Delta\mathbf{x}\|_1 \|\Psi_{S^c}P_{\Pi^\perp}\boldsymbol{\nu}\|_\infty + \|\Psi P_{\Pi^\perp}\Delta\mathbf{x}\|_1 \\
&\stackrel{(i)}{\geq} \|\Psi\mathbf{x}_0\|_1 - \|P_{\Pi}\Delta\mathbf{x}\|_2 \|P_{\Pi}(\gamma_S - \boldsymbol{\nu})\|_2 \\
&\quad + \|P_{\Pi^\perp}\Delta\mathbf{x}\|_2 (1 - \|\Psi_{S^c}P_{\Pi^\perp}\boldsymbol{\nu}\|_\infty) \\
&\stackrel{(ii)}{\geq} \|\Psi\mathbf{x}_0\|_1 - \frac{1}{2s^{10}} \|P_{\Pi}\Delta\mathbf{x}\|_2 + \frac{3}{4} \|P_{\Pi^\perp}\Delta\mathbf{x}\|_2, \tag{24}
\end{aligned}$$

where (i) uses the fact that $\|\Psi P_{\Pi^\perp}\Delta\mathbf{x}\|_1 \geq \|\Psi P_{\Pi^\perp}\Delta\mathbf{x}\|_2 = \|P_{\Pi^\perp}\Delta\mathbf{x}\|_2$, and (ii) follows from eq. (21) and eq. (22). Note that the feasible perturbation condition $\Phi_{\Omega}\Delta\mathbf{x} = 0$ implies that $\|\mathbf{R}_{\Omega}^{1/2}\Delta\mathbf{x}\|_2 = 0$, which further leads

to

$$\|\mathbf{R}_\Omega^{1/2}\mathbf{P}_\Pi\Delta\mathbf{x}\|_2 = \|\mathbf{R}_\Omega^{1/2}\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2. \quad (25)$$

On the other hand, the quantity $\|\mathbf{R}_\Omega^{1/2}\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2$ satisfies

$$\begin{aligned} \|\mathbf{R}_\Omega^{1/2}\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2 &= \langle \mathbf{R}_\Omega^{1/2}\mathbf{P}_\Pi\Delta\mathbf{x}, \mathbf{R}_\Omega^{1/2}\mathbf{P}_\Pi\Delta\mathbf{x} \rangle \\ &\stackrel{(i)}{=} \langle \mathbf{P}_\Pi\Delta\mathbf{x}, \mathbf{P}_\Pi\mathbf{R}_\Omega\mathbf{P}_\Pi\Delta\mathbf{x} \rangle \\ &= \langle \mathbf{P}_\Pi\Delta\mathbf{x}, (\mathbf{P}_\Pi\mathbf{R}_\Omega\mathbf{P}_\Pi - \mathbf{P}_\Pi)\Delta\mathbf{x} \rangle + \|\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2 \\ &\geq -\|\mathbf{P}_\Pi\mathbf{R}_\Omega\mathbf{P}_\Pi - \mathbf{P}_\Pi\| \|\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2 + \|\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2 \\ &\stackrel{(ii)}{\geq} \frac{1}{2}\|\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2, \end{aligned} \quad (26)$$

where (i) follows from the fact that \mathbf{P}_Π , $\mathbf{R}_\Omega^{1/2}$ are self adjoint, and (ii) is due to eq. (23). Then combining eq. (26) with eq. (25), we obtain $\|\mathbf{R}_\Omega^{1/2}\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2^2 \geq \frac{1}{2}\|\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2$. Also observe that $\|\mathbf{R}_\Omega^{1/2}\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2^2 \leq \|\mathbf{R}_\Omega^{1/2}\|^2\|\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2^2 = \frac{1}{\min_j p_j}\|\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2^2$. We then further obtain

$$\|\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2^2 \geq \frac{\min_j p_j}{2}\|\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2 \stackrel{\text{eq. (7)}}{\geq} 1/2s^{20}\|\mathbf{P}_\Pi\Delta\mathbf{x}\|_2^2. \quad (27)$$

With the above inequality, eq. (24) further becomes

$$\begin{aligned} \|\Psi(\mathbf{x}_0 + \Delta\mathbf{x})\|_1 &\geq \|\Psi\mathbf{x}_0\|_1 - \frac{\sqrt{2}s^{10}}{2s^{10}}\|\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2 + \frac{3}{4}\|\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2, \\ &\geq \|\Psi\mathbf{x}_0\|_1 + \left(\frac{3}{4} - \frac{\sqrt{2}}{2}\right)\|\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2. \end{aligned} \quad (28)$$

Now we argue by contradiction to prove that $\|\mathbf{P}_{\Pi^\perp}\Delta\mathbf{x}\|_2 > 0$. Assume this fails to hold, i.e., $\mathbf{P}_\Pi\Delta\mathbf{x} = \Delta\mathbf{x}$. Then feasibility condition $\Phi_\Omega\Delta\mathbf{x} = 0$ implies that $\mathbf{R}_\Omega\mathbf{P}_\Pi\Delta\mathbf{x} = 0$. Putting all these together, we obtain $\|(\mathbf{P}_\Pi\mathbf{R}_\Omega\mathbf{P}_\Pi - \mathbf{P}_\Pi)\Delta\mathbf{x}\|_2 = \|\Delta\mathbf{x}\|_2$, which contradicts the fact that $\|\mathbf{P}_\Pi\mathbf{R}_\Omega\mathbf{P}_\Pi - \mathbf{P}_\Pi\| \leq 1/2$. Hence, we have shown that any feasible perturbation of the signal strictly increases the objective function. \square

It then suffices to show that the set of conditions hold with high probability for our sampling scheme. We first show the bound in eq. (23) in the following lemma.

Lemma 2. *With the sampling scheme in Theorem 1, one has*

$$\|\mathbf{P}_\Pi\mathbf{R}_\Omega\mathbf{P}_\Pi - \mathbf{P}_\Pi\| \leq 1/2 \quad (29)$$

with probability at least $1 - n^{-C_0}$.

Proof. See Appendix F. □

Next, we construct a vector $\boldsymbol{\nu}$ that satisfies eq. (20), eq. (21) and eq. (22) via the ‘‘golfing scheme’’. First we decompose the Bernoulli sampling model $\text{Bernoulli}(p_j)$ into k_0 independent Bernoulli schemes, i.e., $\{\text{Bernoulli}(q_j^k)\}_{k=1}^{k_0}$. Then to be consistent we must have $(1 - p_j) = \prod_{k=1}^{k_0} (1 - q_j^k)$. We also define the corresponding weighted sampling operator $\{R_{\Omega_k}\}_{k=1}^{k_0}$ for each decomposed Bernoulli sampling scheme. Then, we set $k_0 = 11 \log s + 2$, and for all j we set $q_j^1 = q_j^2 = p_j/6$; $q_j^k \equiv q_j$, $k \geq 3$. Then a non-overlapping argument shows that $q_j \geq \frac{2p_j}{3(k_0-2)} \geq C_0 \pi(\boldsymbol{\Pi}, \phi_j) \log n$. Now we start with $\boldsymbol{\nu}_0 = 0$, and construct $\boldsymbol{\nu} := \boldsymbol{\nu}_{k_0}$ via the following iterative scheme

$$\boldsymbol{\nu}_{k+1} = \boldsymbol{\nu}_k - R_{\Omega_{k+1}}(\mathbf{P}_{\boldsymbol{\Pi}}\boldsymbol{\nu}_k - \gamma_S), k = 0, 1, \dots, k_0 - 1.$$

By introducing the variables $\mathbf{w}_0 = -\gamma_S$, $\mathbf{w}_k = \mathbf{P}_{\boldsymbol{\Pi}}\boldsymbol{\nu}_k - \gamma_S$, the above iterative scheme can be expressed as

$$\boldsymbol{\nu}_{k_0} = - \sum_{k=1}^{k_0} R_{\Omega_k} \mathbf{w}_{k-1}, \quad (30)$$

$$\mathbf{w}_k = -(\mathbf{P}_{\boldsymbol{\Pi}}R_{\Omega_k}\mathbf{P}_{\boldsymbol{\Pi}} - \mathbf{P}_{\boldsymbol{\Pi}})\mathbf{w}_{k-1}, k = 1, \dots, k_0, \quad (31)$$

which brings more convenience for subsequent analysis. Before proving the conditions on the constructed $\boldsymbol{\nu}_{k_0}$, we first present the following concentration bound for convenience.

Lemma 3. Consider the $\{\Omega_k\}_{k=1}^{k_0}$ specified in the golfing scheme and any vector \mathbf{w} , one has

$$\|(\mathbf{P}_{\boldsymbol{\Pi}}R_{\Omega_k}\mathbf{P}_{\boldsymbol{\Pi}} - \mathbf{P}_{\boldsymbol{\Pi}})\mathbf{w}\|_2 \leq \frac{1}{2}\|\mathbf{w}\|_2, k = 1, 2, \dots, k_0 \quad (32)$$

with probability at least $1 - n^{-C_0}$.

Proof. See Appendix F. □

Also, note that the infinity norm induced by basis $\boldsymbol{\Phi}$ is $\max_j |[\boldsymbol{\Phi}\mathbf{w}]_j| = \max_j |\langle \mathbf{w}, \phi_j \rangle|$. By scaling this norm with the scalar $\|\mathbf{P}_{\boldsymbol{\Pi}}\phi_j\|_2$, we define the following weighted norm $\|\cdot\|_{\boldsymbol{\Phi}, \infty}$

$$\|\mathbf{w}\|_{\boldsymbol{\Phi}, \infty} := \max_j \frac{|\langle \mathbf{w}, \phi_j \rangle|}{\|\mathbf{P}_{\boldsymbol{\Pi}}\phi_j\|_2}.$$

Equipped with this weighted norm, the following bounds hold.

Lemma 4. Consider the $\{\Omega_k\}_{k=1}^{k_0}$ specified in the golfing scheme and any vector \mathbf{w} , one has

$$\|\boldsymbol{\Psi}_{S^c}(R_{\Omega_k} - I)\mathbf{w}\|_{\infty} \leq \frac{1}{\sqrt{C_0}}\|\mathbf{w}\|_{\boldsymbol{\Phi}, \infty}, \quad k = 1, 2, \dots, k_0$$

with probability at least $1 - n^{-\sqrt{C_0}}$.

Proof. See Appendix F. □

Lemma 5. Consider the $\{\Omega_k\}_{k=1}^{k_0}$ specified in the golfing scheme and any vector $\mathbf{w} \in \Pi$, one has

$$\|(\mathbf{P}_\Pi \mathbf{R}_{\Omega_k} \mathbf{P}_\Pi - \mathbf{P}_\Pi) \mathbf{w}\|_{\Phi, \infty} \leq \begin{cases} \|\mathbf{w}\|_{\Phi, \infty} / \sqrt{\log s}, & k \in \{1, 2\} \\ \|\mathbf{w}\|_{\Phi, \infty} / 2, & k = 3, 4, \dots, k_0 \end{cases}$$

with probability at least $1 - n^{-C_0}$.

Proof. See Appendix F. □

Lemma 6. Under Assumption 1, one has

$$\|\gamma_S\|_{\Phi, \infty} \leq \frac{\sqrt{C_0} \log s}{4} \quad (33)$$

with probability at least $1 - s^{-\sqrt{C_0}/4}$.

Proof. See Appendix F. □

Now equipped with all technical lemmas above, we show that the constructed ν_{k_0} satisfies the set of conditions. First, by the decomposition of the Bernoulli sampling model, eq. (30) directly implies eq. (21). Next we verify eq. (22) via

$$\begin{aligned} \|\mathbf{P}_\Pi(\gamma_S - \nu_{k_0})\|_2 &= \|\mathbf{w}_{k_0}\|_2 \\ &\stackrel{(i)}{=} \|(\mathbf{P}_\Pi \mathbf{R}_{\Omega_{k_0}} \mathbf{P}_\Pi - \mathbf{P}_\Pi) \mathbf{w}_{k_0-1}\|_2 \\ &\stackrel{(ii)}{\leq} \frac{1}{2} \|\mathbf{w}_{k_0-1}\|_2 \leq \dots \leq \frac{1}{2^{11 \log s + 2}} \|\gamma_S\|_2 \\ &\leq \frac{\sqrt{s}}{4s^{11}} \leq \frac{1}{2s^{10}}, \end{aligned}$$

where (i) is due to eq. (31), and (ii) follows from Lemma 3. Finally, we verify eq. (21) via

$$\begin{aligned} \|\Psi_{S^c} \mathbf{P}_{\Pi^\perp} \nu_{k_0}\|_\infty &= \|\Psi_{S^c} \sum_{k=1}^{k_0} \mathbf{P}_{\Pi^\perp} \mathbf{R}_{\Omega_k} \mathbf{w}_{k-1}\|_\infty \\ &\stackrel{(i)}{=} \|\Psi_{S^c} \sum_{k=1}^{k_0} \mathbf{P}_{\Pi^\perp} (\mathbf{R}_{\Omega_k} - I) \mathbf{w}_{k-1}\|_\infty \\ &\leq \sum_{k=1}^{k_0} \|\Psi_{S^c} \mathbf{P}_{\Pi^\perp} (\mathbf{R}_{\Omega_k} - I) \mathbf{w}_{k-1}\|_\infty \\ &\stackrel{(ii)}{=} \sum_{k=1}^{k_0} \|\Psi_{S^c} (\mathbf{R}_{\Omega_k} - I) \mathbf{w}_{k-1}\|_\infty \\ &\stackrel{(iii)}{\leq} \sum_{k=1}^{k_0} \frac{1}{\sqrt{C_0}} \|\mathbf{w}_{k-1}\|_{\Phi, \infty} \end{aligned}$$

$$\begin{aligned}
&\stackrel{\text{(iv)}}{\leq} \sum_{k=1}^{k_0} \frac{1}{\sqrt{C_0}} \frac{1}{\log s} \frac{1}{2^{k-1}} \|\mathbf{w}_0\|_{\Phi, \infty} \\
&\stackrel{\text{(v)}}{\leq} \frac{\|\gamma_S\|_{\Phi, \infty}}{\sqrt{C_0} \log s} \\
&\stackrel{\text{(vi)}}{\leq} \frac{1}{4},
\end{aligned}$$

where (i) is due to $\mathbf{w}_k \in \Pi$, (ii) follows from $\Psi_{S^c} P_{\Pi} = 0$, (iii) is due to Lemma 4, (iv) follows from Lemma 5, (v) is due to $\mathbf{w}_0 = -\gamma_S$ and (vi) follows from Lemma 6. To this end we have shown that the constructed ν_{k_0} satisfies all conditions, and the succeed probability is lower bounded by $1 - s^{-\sqrt{C_0}/4}$.

APPENDIX B

PROOF OF THEOREM 2

We start with the first inequality. For the two terms involved in π -coherence, we have

$$\begin{aligned}
\|P_{\Pi} \phi_j\|_2^2 &= \langle \Psi P_{\Pi} \phi_j, \Psi P_{\Pi} \phi_j \rangle \\
&\stackrel{\text{(i)}}{\leq} \|\Psi P_{\Pi} \phi_j\|_1 \|\Psi P_{\Pi} \phi_j\|_{\infty} \\
&\leq \|\Psi P_{\Pi} \phi_j\|_1 \|\Psi \phi_j\|_{\infty} \\
&= \|\Psi P_{\Pi} \phi_j\|_1 \sqrt{\mu(\Psi, \phi_j)},
\end{aligned}$$

where (i) utilize the duality between l_1 and l_{∞} . For the other term, we have

$$\begin{aligned}
\|P_{\Pi} \phi_j\|_2 \|\Psi P_{\Pi^{\perp}} \phi_j\|_{\infty} &\leq \|\Psi P_{\Pi} \phi_j\|_2 \|\Psi \phi_j\|_{\infty} \\
&\stackrel{\text{(i)}}{\leq} \|\Psi P_{\Pi} \phi_j\|_1 \|\Psi \phi_j\|_{\infty} \\
&= \|\Psi P_{\Pi} \phi_j\|_1 \sqrt{\mu(\Psi, \phi_j)},
\end{aligned}$$

where (i) is due to $l_2 \leq l_1$. Hence, the first inequality holds. For the second inequality, we begin with

$$\begin{aligned}
\|\Psi P_{\Pi} \phi_j\|_1 \sqrt{\mu(\Psi, \phi_j)} &= \|\Psi P_{\Pi} \phi_j\|_1 \|\Psi \phi_j\|_{\infty} \\
&\stackrel{\text{(i)}}{\leq} s \|\Psi P_{\Pi} \phi_j\|_{\infty} \|\Psi \phi_j\|_{\infty} \\
&\leq s \|\Psi \phi_j\|_{\infty}^2 \\
&= s \mu(\Psi, \phi_j),
\end{aligned}$$

where (i) follows from $\dim(\Pi) = s$. The third inequality holds straightforwardly.

APPENDIX C

PROOF OF PROPOSITION 1

Adopting the definition of local coherence, we define

$$\|\Psi P_{\mathbf{\Pi}_l} \phi_j\|_\infty = \sqrt{\mu(\mathbf{\Pi}_l, \phi_j)} \quad (34)$$

$$\|\Psi P_{\mathbf{\Pi}^\perp} \phi_j\|_\infty = \sqrt{\mu(\mathbf{\Pi}^\perp, \phi_j)}. \quad (35)$$

Since $\dim(\mathbf{\Pi}_l) = s_l$, we obtain by the relationship between l_2 and l_∞ that

$$\|P_{\mathbf{\Pi}_l} \phi_j\|_2^2 \leq s_l \|\Psi P_{\mathbf{\Pi}_l} \phi_j\|_\infty^2 \stackrel{\text{eq. (34)}}{=} s_l \mu(\mathbf{\Pi}_l, \phi_j). \quad (36)$$

Thus, we obtain the upper bound

$$\|P_{\mathbf{\Pi}} \phi_j\|_2 \|\Psi P_{\mathbf{\Pi}^\perp} \phi_j\|_\infty \stackrel{(i)}{=} \sqrt{\sum_{l=1}^p \|P_{\mathbf{\Pi}_l} \phi_j\|_2^2} \sqrt{\mu(\mathbf{\Pi}^\perp, \phi_j)} \stackrel{(ii)}{\leq} \sqrt{\sum_{l=1}^p s_l \mu(\mathbf{\Pi}_l, \phi_j) \mu(\mathbf{\Pi}^\perp, \phi_j)}, \quad (37)$$

where (i) is due to eq. (3) and eq. (35), and (ii) follows from eq. (36). Next we decompose $\mathbf{\Pi}$ with regard to different dyadic levels. Specifically, we denote $\mathbf{\Pi}_l := \mathbf{\Pi} \cap \text{span}\{\psi_j \mid j \in L_l\}$, i.e., the part of $\mathbf{\Pi}$ in the subspace spanned by the wavelet basis in level l , and it follows that $\mathbf{\Pi} = \bigoplus_{l=0}^p \mathbf{\Pi}_l$. Now we have

$$\begin{aligned} & \sum_{j=1}^n \|P_{\mathbf{\Pi}} \phi_j\|_2 \|\Psi P_{\mathbf{\Pi}^\perp} \phi_j\|_\infty \\ & \stackrel{(i)}{\leq} \sum_{j=1}^n \sqrt{\sum_{l=0}^p s_l \mu(\mathbf{\Pi}_l, \phi_j) \mu(\mathbf{\Pi}^\perp, \phi_j)} \\ & \leq \sum_{j=1}^n \sqrt{\sum_{l=0}^p s_l \mu(\mathbf{\Pi}_l, \phi_j) \mu(\Psi, \phi_j)} \\ & \leq \sum_{j=1}^n \sqrt{\sum_{l=0}^p s_l \max_{i \in L_l} \mu(\psi_i, \phi_j) \max_i \mu(\psi_i, \phi_j)} \\ & \stackrel{(ii)}{\leq} \sum_{j=1}^n \sqrt{\sum_{l=0}^p s_l 2^{-k(j)} 2^{-|k(j)-l|} 2^{-k(j)}} \\ & \leq \sum_{j=1}^n \sum_{l=0}^p \sqrt{s_l} 2^{-k(j)} 2^{-|k(j)-l|/2} \\ & \stackrel{(iii)}{\leq} \sum_{l=0}^p \sqrt{s_l} \sum_{k(j)=0}^p 2^{-|k(j)-l|/2} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{2\sqrt{2}}{\sqrt{2}-1} \sum_{l=0}^p \sqrt{s_l} \\
&\leq \frac{2\sqrt{2}}{\sqrt{2}-1} s,
\end{aligned}$$

where (i) is due to eq. (37), (ii) follows from eq. (10), and (iii) is due to $|L_{k(j)}| \leq 2^{k(j)}$. Since we have shown that $\sum_{j=1}^n \|\mathbf{P}_{\Pi} \phi_j\|_2^2 = s$, the overall sample complexity of eq. (6) is upper bounded by $\mathcal{O}(s \log s \log n)$.

APPENDIX D

PROOF OF THEOREM 3

For the scheme in eq. (13), the number of sampled measurements $|\Omega|$ satisfies

$$\mathbb{E}|\Omega| = \sum_{j=1}^n (\|\mathbf{P}_{\Pi} \phi_j\|_2^2 - \delta_j) \leq 0.9s - 2. \quad (38)$$

Hence, Markov inequality gives that $|\Omega| < s - 2$ with probability at least 0.1.

Now let us identify a set of candidates $\hat{\mathbf{x}}$ that achieve the same objective value as \mathbf{x}_0 . Recall that $\boldsymbol{\theta} := \boldsymbol{\Psi} \mathbf{x}_0$ with support set S , and we construct $\hat{\mathbf{x}}$ as follows

$$\hat{\mathbf{x}} := \mathbf{x}_0 - \boldsymbol{\Psi}^T(\boldsymbol{\lambda} \cdot \text{sgn}(\boldsymbol{\theta})), \quad (39)$$

where ‘ \cdot ’ denotes the element-wise product, and $\boldsymbol{\lambda}$ is any vector that satisfies

$$|\boldsymbol{\lambda}| \preceq |\boldsymbol{\theta}|, \quad (40)$$

$$\langle \boldsymbol{\lambda}, \mathbf{1}_S \rangle = 0, \quad (41)$$

$$\boldsymbol{\Phi}_{\Omega} \boldsymbol{\Psi}^T(\boldsymbol{\lambda} \cdot \text{sgn}(\boldsymbol{\theta})) = 0. \quad (42)$$

We use ‘ \preceq ’ to denote that ‘ \leq ’ holds element-wise, and denote $\mathbf{1}_S$ as a vector that has 1s on the support set S . By eq. (41) and eq. (42), we only need to identify the entries of $\boldsymbol{\lambda}$ that are supported on S , and the total number of linear constraints is $1 + |\Omega| < s$. Thus, we can identify infinite many $\boldsymbol{\lambda}$ s that satisfy eqs. (41) and (42). Finally, to ensure eq. (40), we simply multiply each $\boldsymbol{\lambda}$ by $\frac{\min_i |\theta_i|}{\|\boldsymbol{\lambda}\|_{\infty}}$.

We complete the proof by showing that every $\hat{\mathbf{x}}$ that satisfies eqs. (40) to (42) achieves the same objective value of problem **(P)** as \mathbf{x}_0 .

$$\begin{aligned}
\|\boldsymbol{\Psi} \hat{\mathbf{x}}\|_1 &= \|\boldsymbol{\theta} - \boldsymbol{\lambda} \cdot \text{sgn}(\boldsymbol{\theta})\|_1 = \sum_{i \in S} |\theta_i - \lambda_i \text{sgn}(\theta_i)| \\
&\stackrel{(i)}{=} \sum_{i \in S} |\theta_i| - \lambda_i = \|\boldsymbol{\Psi} \mathbf{x}_0\|_1 - \langle \boldsymbol{\lambda}, \mathbf{1}_S \rangle \stackrel{(ii)}{=} \|\boldsymbol{\Psi} \mathbf{x}_0\|_1,
\end{aligned}$$

where (i) follows from eq. (40), and (ii) is due to eq. (41). Moreover, eq. (42) guarantees the feasibility of $\hat{\mathbf{x}}$.

APPENDIX E
PROOF OF THEOREM 4

Denote $\hat{\mathbf{x}}$ as the minimizer of problem **(Q)**, and \mathbf{x}_0 as the underlying true signal. Then we must have

$$\|\Psi \hat{\mathbf{x}}\|_1 \leq \|\Psi \mathbf{x}_0\|_1.$$

On the other hand, triangle inequality gives

$$\|\Phi_\Omega(\hat{\mathbf{x}} - \mathbf{x}_0)\|_2 \leq \|\Phi_\Omega(\hat{\mathbf{x}} - \mathbf{y})\|_2 + \|\Phi_\Omega(\mathbf{x}_0 - \mathbf{y})\|_2 \leq 2\delta, \quad (43)$$

where we have used the fact that $\hat{\mathbf{x}}, \mathbf{x}_0$ are both feasible and hence satisfy the constraint. Define $\mathbf{h} = \hat{\mathbf{x}} - \mathbf{x}_0$. With the above two bounds we can bound the l_2 norm of \mathbf{h} as follows

$$\begin{aligned} \|\mathbf{h}\|_2^2 &= \|\Phi_\Omega \mathbf{h}\|_2^2 + \|\Phi_{\Omega^c} \mathbf{h}\|_2^2 \\ &\stackrel{(i)}{\leq} 4\delta^2 + \|\Phi_{\Omega^c} \mathbf{h}\|_2^2 \\ &= 4\delta^2 + \|\mathbf{P}_\Pi \Phi_{\Omega^c} \mathbf{h}\|_2^2 + \|\mathbf{P}_{\Pi^\perp} \Phi_{\Omega^c} \mathbf{h}\|_2^2, \end{aligned} \quad (44)$$

where (i) is due to eq. (43). In the sequel, we bound $\|\mathbf{P}_\Pi \Phi_{\Omega^c} \mathbf{h}\|_2^2$ and $\|\mathbf{P}_{\Pi^\perp} \Phi_{\Omega^c} \mathbf{h}\|_2^2$, respectively.

Bound $\|\mathbf{P}_{\Pi^\perp} \Phi_{\Omega^c} \mathbf{h}\|_2$. Following from the proof of Theorem 1, for any $\Phi_\Omega \Delta \mathbf{x} = 0$, one has

$$\begin{aligned} \|\Psi(\mathbf{x}_0 + \Delta \mathbf{x})\|_1 &\stackrel{(i)}{\geq} \|\Psi \mathbf{x}_0\|_1 + \left(\frac{3}{4} - \frac{\sqrt{2}}{2}\right) \|\mathbf{P}_{\Pi^\perp} \Delta \mathbf{x}\|_2 \\ &\stackrel{(ii)}{\geq} \|\Psi \mathbf{x}_0\|_1 + \left(\frac{3}{4} - \frac{\sqrt{2}}{2}\right) \|\mathbf{P}_{\Pi^\perp} \Phi_{\Omega^c} \mathbf{h}\|_2 \\ &\geq \|\Psi \mathbf{x}_0\|_1 + C \|\mathbf{P}_{\Pi^\perp} \Phi_{\Omega^c} \mathbf{h}\|_2, \end{aligned} \quad (45)$$

where (i) is due to eq. (28), (ii) has set $\Delta \mathbf{x} = \Phi_{\Omega^c} \mathbf{h}$, and C is a universal constant that may vary at different steps.

With the above bound, we obtain

$$\begin{aligned} \|\Psi(\mathbf{x}_0 + \mathbf{h})\|_1 &\geq \|\Psi \mathbf{x}_0 + \Phi_{\Omega^c} \Psi \mathbf{h}\|_1 - \|\Phi_\Omega \Psi \mathbf{h}\|_1 \\ &\stackrel{(\text{eq. (45)})}{\geq} \|\Psi \mathbf{x}_0\|_1 + C \|\mathbf{P}_{\Pi^\perp} \Phi_{\Omega^c} \mathbf{h}\|_2 - \|\Psi(\Phi_\Omega \mathbf{h})\|_1, \end{aligned}$$

Then by $\|\Psi(\mathbf{x}_0 + \mathbf{h})\|_1 = \|\Psi \hat{\mathbf{x}}\|_1 \leq \|\Psi \mathbf{x}_0\|_1$, we further obtain

$$\|\mathbf{P}_{\Pi^\perp} \Phi_{\Omega^c} \mathbf{h}\|_2 \leq C \|\Psi(\Phi_\Omega \mathbf{h})\|_1 \leq C \sqrt{n} \|\Phi_\Omega \mathbf{h}\|_2 \leq C \sqrt{n} \delta. \quad (46)$$

Clearly, when $\Psi = \mathcal{I}$, the above estimate can be tightened as $C \|\Phi_\Omega \mathbf{h}\|_1 \leq C \sqrt{|\Omega|} \delta$.

Bound $\|\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}\|_2$. We start with

$$\begin{aligned}\|\mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}\|_2^2 &= \langle \mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}, \mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h} \rangle \\ &= \langle \mathbf{P}_{\Pi}\mathbf{R}_{\Omega}\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}, \mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h} \rangle \\ &\stackrel{\text{(eq. (23))}}{\geq} \frac{1}{2}\|\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}\|_2^2.\end{aligned}\tag{47}$$

Also note that

$$\mathbf{R}_{\Omega}^{1/2}\Phi_{\Omega^c}\mathbf{h} = 0 = \mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h} + \mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h},$$

which further implies that

$$\|\mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}\|_2 = \|\mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}\|_2.\tag{48}$$

Combining eq. (48) with eq. (47), we obtain

$$\|\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}\|_2^2 \leq 2\|\mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi}\Phi_{\Omega^c}\mathbf{h}\|_2^2 = 2\|\mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}\|_2^2.\tag{49}$$

With eqs. (46) and (49), eq. (44) can be upper bounded as

$$\|\mathbf{h}\|_2^2 \leq 4\delta^2 + 2\|\mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}\|_2^2 + \|\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}\|_2^2.\tag{50}$$

Then taking expectation with respect to Ω and using the fact that $\mathbb{E}\|\mathbf{R}_{\Omega}^{1/2}\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}\|_2^2 = \langle \mathbb{E}\mathbf{R}_{\Omega}\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}, \mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h} \rangle = \|\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}\|_2^2$, we finally obtain

$$\mathbb{E}\|\mathbf{h}\|_2^2 \leq 4\delta^2 + 3\|\mathbf{P}_{\Pi^{\perp}}\Phi_{\Omega^c}\mathbf{h}\|_2^2 \leq 4\delta^2 + Cn\delta^2.\tag{51}$$

Clearly, for the special case $\Psi = \mathcal{I}$, the above bound reduces to $C|\Omega|\delta^2$.

APPENDIX F

PROOF OF TECHNICAL LEMMAS

Proof of Lemma 2. The operator can be equivalently expressed as:

$$(\mathbf{P}_{\Pi}\mathbf{R}_{\Omega}\mathbf{P}_{\Pi} - \mathbf{P}_{\Pi})(\cdot) = \sum_{j=1}^n \left(\frac{\delta_j}{p_j} - 1 \right) \langle \phi_j, \mathbf{P}_{\Pi}(\cdot) \rangle \mathbf{P}_{\Pi}\phi_j \stackrel{\text{def}}{=} \sum_{j=1}^n \mathcal{X}_j(\cdot).$$

Then for each operator $\mathcal{X}_j(\cdot)$, its operator norm can be bounded as

$$\|\mathcal{X}_j(\cdot)\| = \max_{\mathbf{z}} \frac{\|\mathcal{X}_j(\mathbf{z})\|_2}{\|\mathbf{z}\|_2} \leq \max_{\mathbf{z}} \frac{\|\mathbf{P}_{\Pi}\phi_j\|_2^2 \|\mathbf{z}\|_2}{p_j \|\mathbf{z}\|_2} \stackrel{\text{eq. (6)}}{\leq} \frac{1}{C_0 \log n}.$$

Moreover, the operator norm of its variance $\|\sum_{j=1}^n \mathbb{E} \mathcal{X}_j^2(\cdot)\|$ can be bounded as

$$\begin{aligned}
\left\| \sum_{j=1}^n \mathbb{E} \mathcal{X}_j^2(\cdot) \right\| &= \max_{\mathbf{z}} \frac{\|\sum_{j=1}^n \mathbb{E} \mathcal{X}_j^2(\mathbf{z})\|_2}{\|\mathbf{z}\|_2} \\
&= \max_{\mathbf{z}} \frac{1}{\|\mathbf{z}\|_2} \left\| \sum_{j=1}^n \mathbb{E} \left(\frac{\delta_j}{p_j} - 1 \right)^2 \langle \phi_j, \mathbf{P}_{\Pi} \mathbf{z} \rangle \langle \phi_j, \mathbf{P}_{\Pi} \phi_j \rangle \mathbf{P}_{\Pi} \phi_j \right\|_2 \\
&= \max_{\mathbf{z}} \frac{1}{\|\mathbf{z}\|_2} \left\| \sum_{j=1}^n \frac{(1-p_j) \|\mathbf{P}_{\Pi} \phi_j\|_2^2}{p_j} \langle \phi_j, \mathbf{P}_{\Pi} \mathbf{z} \rangle \mathbf{P}_{\Pi} \phi_j \right\|_2 \\
&\leq \max_{\mathbf{z}} \frac{1}{\|\mathbf{z}\|_2} \left\{ \max_j \frac{(1-p_j) \|\mathbf{P}_{\Pi} \phi_j\|_2^2}{p_j} \right\} \left\| \sum_{j=1}^n \langle \phi_j, \mathbf{P}_{\Pi} \mathbf{z} \rangle \phi_j \right\|_2 \\
&\leq \max_{\mathbf{z}} \left\{ \max_j \frac{\|\mathbf{P}_{\Pi} \phi_j\|_2^2}{p_j} \right\} \frac{\|\mathbf{P}_{\Pi} \mathbf{z}\|_2}{\|\mathbf{z}\|_2} \\
&\stackrel{(i)}{\leq} \frac{1}{C_0 \log n},
\end{aligned}$$

where (i) follows from eq. (6). Then applying the matrix Bernstein's inequality (See Theorem 7 in Appendix G) yields the result. \square

Proof of Lemma 3. For all $k = 1, \dots, k_0$, we expand $(\mathbf{P}_{\Pi} \mathbf{R}_{\Omega_k} \mathbf{P}_{\Pi} - \mathbf{P}_{\Pi}) \mathbf{w}$ as

$$(\mathbf{P}_{\Pi} \mathbf{R}_{\Omega_k} \mathbf{P}_{\Pi} - \mathbf{P}_{\Pi}) \mathbf{w} = \sum_{j=1}^n \left(\frac{\delta_j^k}{q_j^k} - 1 \right) \langle \phi_j, \mathbf{P}_{\Pi} \mathbf{w} \rangle \mathbf{P}_{\Pi} \phi_j \stackrel{\text{def}}{=} \sum_{j=1}^n \mathcal{X}_j^k.$$

Then for all j we have

$$\begin{aligned}
\|\mathcal{X}_j^k\|_2 &= \left\| \left(\frac{\delta_j^k}{q_j^k} - 1 \right) \langle \mathbf{P}_{\Pi} \phi_j, \mathbf{w} \rangle \mathbf{P}_{\Pi} \phi_j \right\|_2 \\
&\leq \frac{\|\mathbf{P}_{\Pi} \phi_j\|_2^2}{q_j^k} \|\mathbf{w}\|_2 \\
&\stackrel{(i)}{\leq} \frac{\|\mathbf{w}\|_2}{C_0 \log n}, k = 1, 2, \dots, k_0
\end{aligned}$$

where (i) is due to the specification of q_j^k in the golfing scheme. Moreover, for the variance, we have

$$\begin{aligned}
\sum_{j=1}^n \mathbb{E} \|\mathcal{X}_j^k\|_2^2 &= \sum_{j=1}^n \mathbb{E} \left(\frac{\delta_j^k}{q_j^k} - 1 \right)^2 |\langle \phi_j, \mathbf{P}_{\Pi} \mathbf{w} \rangle|^2 \|\mathbf{P}_{\Pi} \phi_j\|_2^2 \\
&\leq \sum_{j=1}^n \frac{\|\mathbf{P}_{\Pi} \phi_j\|_2^2}{q_j^k} |\langle \phi_j, \mathbf{P}_{\Pi} \mathbf{w} \rangle|^2 \\
&\leq \max_j \left(\frac{\|\mathbf{P}_{\Pi} \phi_j\|_2^2}{q_j^k} \right) \|\mathbf{P}_{\Pi} \mathbf{w}\|_2^2
\end{aligned}$$

$$\stackrel{(i)}{\leq} \frac{\|\mathbf{w}\|_2^2}{C_0 \log n}, k = 1, 2 \dots k_0$$

where (i) follows from the specification of q_j^k in the golfing scheme. The desired result follows from the vector Bernstein's inequality (See Theorem 6 in Appendix G). \square

Proof of Lemma 4. Since Ψ_{S^c} is restricted on the rows ψ_i^T with $i \in S^c$, $\Psi_{S^c}(R_{\Omega_k} - I)\mathbf{w}$ is supported on S^c . The i -th entry in S^c of vector $\Psi_{S^c}(R_{\Omega_k} - I)\mathbf{w}$ can be alternatively expressed as

$$[\Psi_{S^c}(R_{\Omega_k} - I)\mathbf{w}]_i = \sum_{j=1}^n \left(\frac{\delta_j^k}{q_j^k} - 1 \right) \langle \phi_j, \mathbf{w} \rangle \langle \phi_j, \psi_i \rangle \stackrel{\text{def}}{=} \sum_{j=1}^n \mathcal{X}_j^k, \quad i \in S^c.$$

By the specifications in the golfing scheme, we have for all k that $q_j^k \geq C_0 \|\mathbf{P}_{\Pi} \phi_j\|_2 \|\Psi \mathbf{P}_{\Pi^\perp} \phi_j\|_\infty \log n$. We then obtain

$$\begin{aligned} |\mathcal{X}_j^k| &= \left| \left(\frac{\delta_j^k}{q_j^k} - 1 \right) \langle \phi_j, \mathbf{w} \rangle \langle \phi_j, \psi_i \rangle \right| \\ &\leq \left| \frac{\langle \phi_j, \mathbf{w} \rangle \langle \phi_j, \psi_i \rangle}{C_0 \|\mathbf{P}_{\Pi} \phi_j\|_2 \|\Psi \mathbf{P}_{\Pi^\perp} \phi_j\|_\infty \log n} \right| \\ &\stackrel{(i)}{\leq} \frac{\|\mathbf{w}\|_{\Phi, \infty}}{C_0 \log n}, \end{aligned}$$

where (i) follows from the fact that $|\langle \phi_j, \psi_i \rangle| \leq \|\Psi \mathbf{P}_{\Pi^\perp} \phi_j\|_\infty$ for all $i \in S^c$. Moreover, we also have $q_j^k \geq C_0 \|\mathbf{P}_{\Pi} \phi_j\|_2^2 \log n$. Then we obtain

$$\begin{aligned} \sum_{j=1}^n \mathbb{E} |\mathcal{X}_j^k|^2 &= \sum_{j=1}^n \mathbb{E} \left(\frac{\delta_j^k}{q_j^k} - 1 \right)^2 |\langle \phi_j, \mathbf{w} \rangle \langle \phi_j, \psi_i \rangle|^2 \\ &\leq \sum_{j=1}^n \frac{|\langle \phi_j, \mathbf{w} \rangle \langle \phi_j, \psi_i \rangle|^2}{q_j^k} \\ &\leq \max_j \frac{|\langle \phi_j, \mathbf{w} \rangle|^2}{q_j^k} \sum_{j=1}^n |\langle \phi_j, \psi_i \rangle|^2 \\ &\leq \max_j \frac{|\langle \phi_j, \mathbf{w} \rangle|^2}{C_0 \|\mathbf{P}_{\Pi} \phi_j\|_2^2 \log n} \\ &\leq \frac{\|\mathbf{w}\|_{\Phi, \infty}^2}{C_0 \log n}. \end{aligned}$$

The desired result then follows from the scalar Bernstein's inequality (See Theorem 5 in Appendix G) and a union bound. \square

Proof of Lemma 5. By the definition of the weighted norm, we have

$$\begin{aligned} \|(\mathbf{P}_{\Pi} \mathbf{R}_{\Omega_k} \mathbf{P}_{\Pi} - \mathbf{P}_{\Pi}) \mathbf{w}\|_{\Phi, \infty} &= \max_j \frac{|\langle (\mathbf{P}_{\Pi} \mathbf{R}_{\Omega_k} \mathbf{P}_{\Pi} - \mathbf{P}_{\Pi}) \mathbf{w}, \phi_j \rangle|}{\|\mathbf{P}_{\Pi} \phi_j\|_2} \\ &= \max_j \left| \frac{\sum_{i=1}^n \left(\frac{\delta_i^k}{q_i^k} - 1 \right) \langle \phi_i, \mathbf{P}_{\Pi} \mathbf{w} \rangle \langle \mathbf{P}_{\Pi} \phi_i, \phi_j \rangle}{\|\mathbf{P}_{\Pi} \phi_j\|_2} \right|. \end{aligned}$$

Now consider the j -th term, i.e.,

$$\left| \sum_{i=1}^n \frac{\left(\frac{\delta_i^k}{q_i^k} - 1 \right) \langle \phi_i, \mathbf{P}_{\Pi} \mathbf{w} \rangle \langle \mathbf{P}_{\Pi} \phi_i, \phi_j \rangle}{\|\mathbf{P}_{\Pi} \phi_j\|_2} \right| \stackrel{\text{def}}{=} \left| \sum_{i=1}^n \mathcal{X}_i^k \right|$$

we have that

$$\begin{aligned} |\mathcal{X}_i^k| &= \left| \frac{\left(\frac{\delta_i^k}{q_i^k} - 1 \right) \langle \phi_i, \mathbf{P}_{\Pi} \mathbf{w} \rangle \langle \mathbf{P}_{\Pi} \phi_i, \phi_j \rangle}{\|\mathbf{P}_{\Pi} \phi_j\|_2} \right| \\ &\stackrel{(i)}{\leq} \left| \frac{\langle \phi_i, \mathbf{w} \rangle \|\mathbf{P}_{\Pi} \phi_i\|_2}{q_i^k} \right| \\ &\leq \begin{cases} [C_0 \log n \log s]^{-1} \|\mathbf{w}\|_{\Phi, \infty}, & k \in \{1, 2\} \\ [C_0 \log n]^{-1} \|\mathbf{w}\|_{\Phi, \infty}, & k = 3, 4 \dots k_0 \end{cases} \end{aligned}$$

where (i) follows from Cauchy-Swartz inequality and the fact that $\mathbf{w} \in \Pi$. Moreover, we further derive

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} |\mathcal{X}_i^k|^2 &= \sum_{i=1}^n \mathbb{E} \left(\frac{\delta_i^k}{q_i^k} - 1 \right)^2 \left| \frac{|\langle \phi_i, \mathbf{P}_{\Pi} \mathbf{w} \rangle|^2 |\langle \mathbf{P}_{\Pi} \phi_i, \phi_j \rangle|^2}{\|\mathbf{P}_{\Pi} \phi_j\|_2^2} \right| \\ &\leq \frac{1}{\|\mathbf{P}_{\Pi} \phi_j\|_2^2} \sum_{i=1}^n \left| \frac{|\langle \phi_i, \mathbf{P}_{\Pi} \mathbf{w} \rangle|^2 |\langle \phi_i, \mathbf{P}_{\Pi} \phi_j \rangle|^2}{q_i^k} \right| \\ &\stackrel{(i)}{\leq} \max_i \frac{|\langle \phi_i, \mathbf{w} \rangle|^2}{q_i^k} \frac{1}{\|\mathbf{P}_{\Pi} \phi_j\|_2^2} \sum_{i=1}^n |\langle \phi_i, \mathbf{P}_{\Pi} \phi_j \rangle|^2 \\ &= \max_i \frac{|\langle \phi_i, \mathbf{w} \rangle|^2}{q_i^k} \\ &\stackrel{(ii)}{\leq} \begin{cases} [C_0 \log n \log s]^{-1} \|\mathbf{w}\|_{\Phi, \infty}^2, & k \in \{1, 2\} \\ [C_0 \log n]^{-1} \|\mathbf{w}\|_{\Phi, \infty}^2, & k = 3, 4 \dots k_0 \end{cases} \end{aligned}$$

where (i) follows from the fact that $\mathbf{w} \in \Pi$, and (ii) utilizes the specifications of q_i^k in the golfing scheme. The desired result follows from the scalar Bernstein's inequality (See Theorem 5 in Appendix G) and a union bound. \square

Proof of Lemma 6. Recall that $\gamma_S = \sum_{i \in S} \text{sgn}(\boldsymbol{\theta})_i \boldsymbol{\psi}_i$, and by definition

$$\begin{aligned} \|\gamma_S\|_{\Phi, \infty} &= \max_j \frac{|\langle \boldsymbol{\phi}_j, \gamma_S \rangle|}{\|\mathbf{P}_{\Pi} \boldsymbol{\phi}_j\|_2} \\ &= \max_j \frac{|\sum_{i \in S} \text{sgn}(\boldsymbol{\theta})_i \langle \boldsymbol{\phi}_j, \boldsymbol{\psi}_i \rangle|}{\|\mathbf{P}_{\Pi} \boldsymbol{\phi}_j\|_2} \\ &\stackrel{\text{def}}{=} \max_j \frac{|\sum_{i \in S} \mathcal{X}_i|}{\|\mathbf{P}_{\Pi} \boldsymbol{\phi}_j\|_2}. \end{aligned}$$

Now for each j , to bound the term $|\sum_{i \in S} \mathcal{X}_i|$, we derive

$$|\mathcal{X}_i| = |\text{sgn}(\boldsymbol{\theta})_i \langle \boldsymbol{\phi}_j, \boldsymbol{\psi}_i \rangle| \leq \|\mathbf{P}_{\Pi} \boldsymbol{\phi}_j\|_2 = \frac{\sqrt{C_0} \|\mathbf{P}_{\Pi} \boldsymbol{\phi}_j\|_2 \log s/4}{\sqrt{C_0} \log s/4},$$

where we have used the fact that $\boldsymbol{\psi}_i \in \Pi$ for $i \in S$. Moreover, we have for $C_0 > 1$

$$\sum_{i \in S} \mathbb{E} |\mathcal{X}_i|^2 \leq \sum_{i \in S} |\langle \boldsymbol{\phi}_j, \boldsymbol{\psi}_i \rangle|^2 = \|\mathbf{P}_{\Pi} \boldsymbol{\phi}_j\|_2^2 \leq \frac{C_0 \|\mathbf{P}_{\Pi} \boldsymbol{\phi}_j\|_2^2 \log^2 s/16}{\sqrt{C_0} \log s/4}.$$

Then scalar Bernstein's inequality (See Theorem 5 in Appendix G) and a union bound gives the desired result. \square

APPENDIX G

BERNSTEIN INEQUALITIES

Theorem 5 (Scalar Bernstein inequality). *Let $x_1, \dots, x_m \in \mathbb{R}$ be independent, zero mean random variables. If for all $i = 1, \dots, m$, it holds that $|x_i| \leq B$ and $\sum_{i=1}^m \mathbb{E} |x_i|^2 \leq \sigma^2$, then*

$$\mathbf{P} \left(\left| \sum_{i=1}^m x_i \right| > t \right) \leq 2 \exp \left(-\frac{t^2/2}{\sigma^2 + Bt/3} \right). \quad (52)$$

Theorem 6 (Vector Bernstein inequality). *Let $\mathbf{x}_1, \dots, \mathbf{x}_m \in \mathbb{R}^n$ be independent, zero mean random vectors. If for all $i = 1, \dots, m$, it holds that $\|\mathbf{x}_i\|_2 \leq B$ and $\sum_{i=1}^m \mathbb{E} \|\mathbf{x}_i\|_2^2 \leq \sigma^2$, then for all $0 < t < \sigma^2/B$*

$$\mathbf{P} \left(\left\| \sum_{i=1}^m \mathbf{x}_i \right\|_2 > t \right) \leq \exp \left(-\frac{t^2}{8\sigma^2} + \frac{1}{4} \right). \quad (53)$$

Theorem 7 (Matrix Bernstein inequality). *Let $\mathbf{X}_1, \dots, \mathbf{X}_m \in \mathbb{R}^{n \times n}$ be independent, zero mean random matrices. If for all $i = 1, \dots, m$, it holds that $\|\mathbf{X}_i\| \leq B$ and $\max\{\|\sum_{i=1}^m \mathbb{E} \mathbf{X}_i^T \mathbf{X}_i\|, \|\sum_{i=1}^m \mathbb{E} \mathbf{X}_i \mathbf{X}_i^T\|\} \leq \sigma^2$, then for all $c > 0$*

$$\left\| \sum_{i=1}^m \mathbf{X}_i \right\| \leq 2\sqrt{c\sigma^2 \log 2n} + cB \log 2n \quad (54)$$

with probability at least $1 - 2n^{-(c-1)}$.

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [2] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [3] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [4] E. Candès and Y. Plan, "A probabilistic and riplless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 11, pp. 7235–7254, 2011.
- [5] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing mri," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [6] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [7] V. Cevher, A. Sankaranarayanan, M. Duarte, D. Reddy, and R. Baraniuk, "Compressive sensing for background subtraction," in *European Conference on Computer Vision*, 2008, pp. 155–168.
- [8] J. Provost and F. Lesage, "The application of compressed sensing for photo-acoustic tomography," *IEEE Transactions on Medical Imaging*, vol. 28, no. 4, pp. 585–594, 2009.
- [9] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, p. 969, 2007.
- [10] F. Krahermer and R. Ward, "Stable and robust sampling strategies for compressive imaging," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 612–622, 2014.
- [11] C. Boyer, J. Bigot, and P. Weiss, "Compressed sensing with structured sparsity and structured acquisition," *arxiv*, 2015. [Online]. Available: <http://arxiv.org/abs/1505.01619>
- [12] J. Bigot, C. Boyer, and P. Weiss, "An analysis of block sampling strategies in compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 2125–2139, April 2016.
- [13] Z. Wang and G. Arce, "Variable density compressed image sampling," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 264–270, 2010.
- [14] C. Schroder, P. Bornert, and B. Aldefeld, "Spatial excitation using variable-density spiral trajectories," *Journal of Magnetic Resonance Imaging*, vol. 18, no. 1, pp. 136–141, 2003.
- [15] G. Puy, J. Marques, R. Gruetter, J. Thiran, D. Ville, P. Vandergheynst, and Y. Wiaux, "Spread spectrum magnetic resonance imaging," *arxiv*, 2012. [Online]. Available: <https://arxiv.org/pdf/1203.2205>
- [16] N. Chauffert, P. Ciuciu, and P. Weiss, "Variable density compressed sensing in MRI. Theoretical vs heuristic sampling strategies," in *International Symposium on Biomedical Imaging*, 2013.
- [17] B. Adcock, P. Univ, A. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing," Tech. Rep., 2013.
- [18] L. Malloy and D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4001–4012, 2013.
- [19] M. Iwen and A. H. Tewfik, "Adaptive compressed sensing for sparse signals in noise," in *Conference on Circuits, Systems & Computers*, 2011, pp. 1240–1244.
- [20] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in *IEEE Statistical Signal Processing Workshop*, 2012, pp. 401–404.

- [21] C. Boutsidis, M. Mahoney, and P. Drineas, “An improved approximation algorithm for the column subset selection problem,” *Computing Research Repository*, vol. abs/0812.4, 2008.
- [22] M. W. Mahoney, “Randomized algorithms for matrices and data,” *Foundations and Trends in Machine Learning*, vol. 3, no. 2, pp. 123–224, 2011.
- [23] Y. Chen, S. Bhojanapalli, S. Sanghavi, and R. Ward, “Coherent matrix completion,” in *Proceedings of the 31st International Conference on Machine Learning*, 2014, pp. 674–682.
- [24] H. Rauhut, “Compressive sensing and structured random matrices,” *Radon Series Comp. Appl. Math*, 2010.
- [25] D. Gross, “Recovering low-rank matrices from few coefficients in any basis,” *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.
- [26] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb 2006.
- [27] M. Guerquin-Kern, M. Haberland, K. Pruessmann, and M. Unser, “A fast wavelet-based reconstruction method for magnetic resonance imaging,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, Sept 2011.
- [28] E. J. Candès and Y. Plan, “Matrix completion with noise,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.