Identification of Matrices having a Sparse Representation

Götz E. Pfander *, Holger Rauhut †, Jared Tanner ‡

We consider the problem of recovering a matrix from its action on a known vector in the setting where the matrix can be represented efficiently in a known matrix dictionary. Connections with sparse signal recovery allows for the use of efficient reconstruction techniques such as Basis Pursuit (BP). Of particular interest is the dictionary of time–frequency shift matrices and its role for channel estimation and identification in communications engineering. We present recovery results for BP with the time–frequency shift dictionary and various dictionaries of random matrices.

1. INTRODUCTION

Inferring reliable information from limited data is a key task in the sciences. For example, identifying a channel operator from its response to a limited number of test signals is a crucial step in radar and communications engineering [27, 34, 32, 40]. Here we consider the canonical setting where an operator is approximated by a linear map, that is, by a matrix $\Gamma \in \mathbb{C}^{n\times m}$. While it is clear that Γ is determined by its action on any m vectors that span $\mathbb{C}^m$, significantly fewer measurements may be sufficient if *a-priori* information about the operator is at hand. For instance, one commonly considers the question whether a single test signal h, referred to also as identifier, can be used to identify Γ from Γh. *A priori* information guaranteeing that such an h exists is generally deduced from from physical considerations which may ensure that Γ can be efficiently represented or approximated using relatively few basic matrices from a known matrix dictionary.

In wireless communications or sonar, for example, the transmission channel can generally be well approximated by a linear combination of a small number of time–frequency shift matrices. Signals travel from the source to the receiver along a number of different paths, each of which can be modeled by a time shift (delay dependent on the length of the path traveled) and a frequency shift (Doppler effect caused by the motion of the transmitter, of the receiver, and of reflecting objects) [5]. It is frequently assumed, that the number of relevant (but unknown) paths, that is, the number of involved time–frequency shifts is relatively small when compared to the symbol length. In wireless communications the benefit of recovering the operator at the receiver is clear: knowledge

*School of Engineering and Science, Jacobs University Bremen, 28759 Bremen, Germany, g.pfander@iu-bremen.de

†Numerical Harmonic Analysis Group, Faculty of Mathematics, University of Vienna, Nordbergstrasse 15, A-1090 Vienna, Austria, holger.rauhut@univie.ac.at. H.R. is supported by the European Union's Human Potential Programme under contract MEIF-CT 2006-022811.

‡Department of Mathematics, University of Utah, 155 South 1400 East, Salt Lake City, UT 84112-0090, USA. J.T. would like to thank John E. and Marva M. Warnock for their generous support in the form of an endowed chair. tanner@math.utah.edu

of the operator is necessary to invert it and to recover the information carrying channel input from the channel output.

Complexity regularization has recently seen a resurgence of interest in the signal processing community under the monikers *sparse signal recovery* and *sparse approximation*. In sparse signal recovery, one seeks the solution of an underdetermined system of equations $Ax = b$, $A \in \mathbb{C}^{n\times N}$, $n < N$, with x having the fewest number of non-zero entries from all solutions of $Ax = b$. We show in Section 2 that the identification of a matrix from its action on a single test signal falls into the same setting as sparse signal recovery when the matrix is known to have a sparse representation. This observation allows us to adopt efficient algorithms from sparse signal recovery for the sparse matrix identification question. Examples of applications include the channel identification, estimation, or sounding problem described in part above, which also have been considered in the case of time–invariant channels in [25, 11, 10]. Numerical results based on basis pursuit have been obtained for time–varying channels in [39].

In brief, the content of this paper is organized as follows. In Section 2 we formalize the matrix identification problem for matrices with sparse representations. We establish a connection to the recovery problem of vectors with sparse representations and state the main results that are proven and discussed in greater detail in Section 4 and Section 5. In particular, we consider matrix ensembles of random Gaussian or Bernoulli matrices as well as partial Fourier matrices (Section 2.1 and Section 4).

In Section 2.2 and Section 5 we consider matrix dictionaries of time–frequency shift matrices which are of particular interest due to their efficacy in approximating time–varying transmission channels. We would like to emphasize that the common framework of the identification problem for matrices with a sparse representation and the sparse signal recovery problem implies that the results achieved on the recovery of matrices with a sparse representation in the dictionary of time–frequency shift matrices are at the same time results for the recovery of signals with a sparse representation in Gabor frames.

We conclude with numerical experiments in Section 6. They verify our main results concerning sparse representations with time–frequency shift matrices stated in Theorem 2.5, and show that the precise recoverability thresholds follow those proven for Gaussian random matrices in [21]; that is, for matrices having a $k-$sparse representation we observe Basis Pursuit to successfully recover the matrix from its action on a single vector provided $k \leq n/(2\log n)$.

2. MAIN RESULTS AND CONTEXT

Before comparing the matrix identification problem with sparse signal recovery, we formalize the notion of a matrix having a $k-$sparse representation.

DEFINITION 2.1. *A matrix Γ has a k-sparse representation in the matrix dictionary $\mathbf{\Psi} = \{\Psi_j\}_{j=1}^N$ if*

$$\Gamma = \sum_j x_j \Psi_j \qquad \text{with} \quad \|x\|_0 = k, \tag{1}$$

and $\|x\|_0$ counts the number of non-zero entries in x, that is $\|x\|_0 = |\operatorname{supp} x|$.

The set of elementary matrices comprising $\mathbf{\Psi}$ may form a basis for $\mathbb{C}^{n\times m}$ but it may as well only span a subspace of $\mathbb{C}^{n\times m}$ and/or contain linearly dependent subsets. In Definition 2.1 we place no restriction on the dictionary $\mathbf{\Psi}$.

Identification of matrices having a sparse representation from their action on a single vector (henceforth referred to simply as *sparse matrix identification*, which is not to be confused with the notion of sparse matrices in numerical analysis) can be formulated as sparse signal recovery problem through the simple observation that the action of Γ on a test signal $h \in \mathbb{C}^m$ can be expressed as

$$\Gamma h = \Big(\sum_{j=1}^{N} x_j \Psi_j\Big) h = \sum_{j=1}^{N} x_j \big(\Psi_j h\big) = (\Psi_1 h \,|\, \Psi_2 h \,|\, \ldots \,|\, \Psi_N h)\, x = (\mathbf{\Psi} h) x \tag{2}$$

where $x = (x_1,\ x_2,\ \cdots,\ x_N)^T$ and $(\mathbf{\Psi} h) = (\Psi_1 h \,|\, \Psi_2 h \,|\, \ldots \,|\, \Psi_N h)$.

In classical sparse signal recovery the sparsest vector x satisfying $Ax = b$ is sought given b and A; to identify the matrix Γ, Γh takes the place of b and the j^{th} column of A is $\Psi_j h$ for $j = 1, 2, \ldots, N$.

As mentioned above, we note that in case of the Ψ_j being time–frequency shift matrices, the columns in $A = (\mathbf{\Psi} h)$ form a Gabor system with window h. Consequently, all our identifiability results concerning representations with time–frequency shift matrices are also results for the recovery of signals that are sparse in a Gabor system.

Remark 2.2. Although sparse matrix identification can be cast as sparse signal recovery, two important differences should be noted.

- Sparse signal recovery is only of interest for $k-$sparse vectors with $k < n$ ($A \in \mathbb{C}^{n\times N}$ with $n < N$) as when there is one solution involving exactly n non–zero terms, then there exist infinitely many $n-$term solutions, x. However, such a solution is not catastrophic if that is the minimal sparsity for a given b. In contrast, sparse matrix identification is *only* of interest for $k < n$ as $n-$term solutions to $(\mathbf{\Psi} h)x = \Gamma h$ do not imply that Γ has an $n-$sparse representation in the dictionary $\mathbf{\Psi}$; rather, there always exists infinitely many $n-$sparse matrices Γ' consistent with the observations $\Gamma' h = \Gamma h$. As such, the recovery of an $n-$sparse x in the sparse matrix identification setting does not give any information about the matrix to be identified, Γ.

- In sparse signal recovery the columns of A are used to represent or to approximate b, whereas for sparse matrix identification the matrices Ψ_j are used to represent or approximate Γ; however, the Ψ_j do not appear explicitly when sparse matrix identification is cast as sparse signal recovery (2), but only the action of Ψ_j on the test vector h appears. The test vector $h \in \mathbb{C}^m$ has no analog in traditional sparse signal recovery, and can be exploited in sparse matrix identification to design desirable characteristics in $\Psi_j h$. This design freedom is utilized extensively in our main results concerning the matrix dictionary of time–frequency shifts, Theorem 2.5.

Note that the computational difficulty in sparse signal recovery, sparse approximation, and our formulation of sparse matrix identification arises from the fact that the support set of the non-zero entries in x is unknown. While the direct solution of finding the sparsest representation of Γ in the dictionary $\mathbf{\Psi}$

$$\min \|x'\|_0 \qquad \text{subject to} \quad (\mathbf{\Psi} h)x' = \Gamma h, \tag{3}$$

involves a combinatorial search of the support set and is therefore computationally intractable, a number of computationally efficient algorithms have been shown to recover the sparsest solution if appropriate conditions are met. We concentrate here on recoverability conditions for the canonical sparse signal recovery algorithm Basis Pursuit (BP) where the convex problem

$$\min \|x'\|_1 \qquad \text{subject to} \quad (\mathbf{\Psi} h)x' = \Gamma h, \tag{4}$$

$\|x\|_1 = \sum_j |x_j|$, is solved as a proxy to (3).

The convex program (4) can be solved efficiently using well established optimization algorithms for second-order cone programming and linear programming [6, 15, 26], for complex and real valued systems, respectively. We give theoretical and numerical evidence for conditions where the solution of (4) coincides exactly with that of (3). Many other algorithms may also be used as proxys for (3), including Orthogonal Matching Pursuit (OMP) [42, 22, 29], stagewise orthogonal matching pursuit (StOMP) [13], and an algorithm based upon error correcting codes [2] - to name a few. Our principal technical results in Section 5.1 also give results for OMP, but for conciseness we do not state them here, leaving them to the interested reader.

In practice, the measured vector Γh will be contaminated by noise, and, in addition, the operator Γ will not be strictly sparse, but will instead be well approximated by a sparse representation; in this case the minimization problem (4) will be replaced by its well known variant

$$\min \|x'\|_1 \qquad \text{subject to} \quad \|(\mathbf{\Psi} h)x' - \Gamma h\|_2 \leq \epsilon, \tag{5}$$

where $\|z\|_2 = \sqrt{\sum_j |z_j|^2}$ as usual.

In addition to the case where the operator Γ is applied to a single test signal, we may also consider two or more test signals $h_1, \dots, h_r$. Then the vector of concatenated observations $\Gamma h_1, \dots, \Gamma h_r$ is given as

$$\begin{pmatrix} \Gamma h1 \\ \vdots \\ \Gamma h_r \end{pmatrix} = \begin{pmatrix} \Psi_1 h_1 & \dots & \Psi_N h_1 \\ \vdots & & \vdots \\ \Psi_1 h_r & \dots & \Psi_N h_r \end{pmatrix} x = \begin{pmatrix} \mathbf{\Psi} h_1 \\ \vdots \\ \mathbf{\Psi} h_r \end{pmatrix} x,$$

and our sparse matrix identification task is again reduced to a sparse signal recovery problem. We will, however, not pursue this extension here.

2.1. Dictionaries of random matrices

Many known results in sparse signal recovery, sparse approximations and their companion theory of compressed sensing involve random matrices [9, 4, 12, 21, 37]. Based on these results, we obtain recovery results for matrix dictionaries where all its member matrices are chosen at random. From a practical point of view such random matrix dictionaries do not seem to be useful in the sparse matrix identification setting; nevertheless, the statements give some insight into the sparse matrix identification question as they can be regarded as results for generic matrix dictionaries, that is, they apply to *most* matrix dictionaries.

Theorem 2.3. *Let h be a non-zero vector in $\mathbb{C}^m$.*

(a) Let all entries of the N matrices $\Psi_j \in \mathbb{R}^{n\times m}$, $j=1,\dots,N$ be chosen independently according to a standard normal distribution (Gaussian ensemble); or

(b) let all entries of the N matrices $\Psi_j \in \mathbb{R}^{n\times m}$, $j=1,\dots,N$ be independent Bernoulli ± 1 variables (Bernoulli ensemble).

Then there exists a positive constant c so that for $\varepsilon > 0$,

$$k \le c\,\frac{n}{\log\left(\frac{N}{n\varepsilon}\right)}$$

implies that with probability of at least $1-\varepsilon$ all matrices Γ having a k-sparse representation with respect to $\mathbf{\Psi} = \{\Psi_j\}$ can be recovered from Γh by Basis Pursuit (4).

Using Theorem 3.6, this recovery result can be made stable under perturbation of Γh by noise, and also applies when Γ is not exactly k-sparse, but can be well-approximated by a k-sparse operator.

Precise information on the constant c will be given in Section 4. In case of the Gaussian ensemble Donoho and Tanner [16, 17, 14, 20, 21] obtained sharp thresholds separating regions in the (k/n, n/N) plane where recovery holds or fails with high probability; Section 4.1 recounts these and additional results on Gaussian systems. Theorem 2.3(b) is proven in Section 4.2, and similar results for certain diagonal matrices are proven in Section 4.3.

2.2. The dictionary of time–frequency shift matrices

As outlined in the introduction, the matrix dictionary of time–frequency shifts appears naturally in the channel identification problem in wireless communications [5]. Due to physical considerations wireless channels may indeed be modeled by sparse linear combinations of time–frequency shifts $M_\ell T_p$, where the translation operators T_p and modulation operator M_ℓ on $\mathbb{C}^n$ are given by

$$(T_p h)_q = h_{p+q} \quad \text{and} \quad (M_\ell h)_q = e^{2\pi i \ell q/n} h_q. \tag{6}$$

The system of time–frequency shifts $\boldsymbol{\mathcal{G}} = \{M_\ell T_p : \ell, p = 0,\dots,n-1\}$ forms a basis of $\mathbb{C}^{n\times n}$ and for any non-zero h, the vector dictionary $\boldsymbol{\mathcal{G}}h$ is a Gabor system [24, 30, 28]. Below, we focus on the so-called Alltop window h^A [3, 41] with entries

$$h^A_q \;=\; \frac{1}{\sqrt{n}} e^{2\pi i q^3/n}, \quad q = 0,\dots,n-1, \tag{7}$$

and the randomly generated window h^R with entries

$$h^R_q \;=\; \frac{1}{\sqrt{n}} \epsilon_q, \quad q = 0,\dots,n-1, \tag{8}$$

where the ϵ_q are independent and uniformly distributed on the torus $\{z \in \mathbb{C}, |z| = 1\}$.

Invoking existing recovery results [19, 23, 42, 44] (see Theorems 3.1 and 3.2 below) and our results on the coherence of Gabor systems $\boldsymbol{\mathcal{G}}h^A$ and $\boldsymbol{\mathcal{G}}h^R$ in Section 5.1, we obtain

Theorem 2.4.

(a) Let n be prime and h^A be the Alltop window defined in (7). If $k < \frac{\sqrt{n}+1}{2}$ then BP recovers from Γh^A all matrices Γ having a k-sparse representation with respect to the time–frequency shift dictionary.

(b) Let n be even and choose h^R to be the random unimodular window in (8). Let $t > 0$ and suppose

$$k \leq \frac{1}{4}\sqrt{\frac{n}{C\log(n)+t}} + \frac{1}{2} \tag{9}$$

with $C = 2\log(4) \approx 2.77$. Then with probability of at least $1 - e^{-t}$ BP recovers from Γh^R all matrices $\Gamma \in \mathbb{C}^{n\times n}$ having a k-sparse representation.

A slight variation of part (b) also holds for n odd, but is omitted for conciseness. Further note that Theorem 2.4 also holds with BP literally being replaced by Orthogonal Matching Pursuit [42]. Moreover, Theorem 3.2 shows that recovery is stable under perturbation of Γh^A and Γh^R by noise.

In contrast with Theorem 2.3 for random matrices, where k is allowed to be of order $\mathcal{O}(n/\log(n))$, Theorem 2.4 requires k to be of order $\sqrt{n}$ or $\sqrt{n/\log(n)}$. Substantially larger order $\mathcal{O}(n/\log^\alpha(n))$ thresholds for $\alpha > 1$ are also possible when using h^A or h^R to identify a matrix Γ which is the linear combination of a small number of time–frequency shift matrices. However, this larger regime of successful recovery necessitates passing from a worst case analysis for sparse Γ to an average case analysis in the sense that the coefficient vector x is chosen at random. Theorem 2.5 follows from recent work by Tropp, [43], and our coherence results in Section 5.1.

THEOREM 2.5. *Let $k \geq 3$ and let Λ be chosen uniformly at random among all subsets of $\{0,\ldots,n-1\}^2$ of cardinality k. Suppose further that $x \in \mathbb{C}^n$ has support Λ with random phases $(\mathrm{sgn}(x_{\ell p}))_{(\ell,p)\in\Lambda}$ that are independent and uniformly distributed on the torus $\{z, |z| = 1\}$. Let*

$$\Gamma = \sum_{(\ell,p)\in\Lambda} x_{\ell p} M_\ell T_p.$$

(a) Let n be prime and choose the Alltop window h^A from (7). Assume that

$$k \leq \frac{n}{2^{u+4}\log(n)^{u+1}} \tag{10}$$

for some $u > 0$ satisfying $2^{u+3}\log(n)^{u+1} > 2/c$ with c no smaller than 0.0818. Set

$$\sigma = \left(\frac{c}{2} - \frac{1}{2^{u+4}\log(n)^{u+1}}\right)^2 2^{u+4}\log(n)^u \geq 1. \tag{11}$$

Then with probability at least

$$1 - (2n^{-2u} + 2k^{-\sigma})$$

Basis Pursuit (4) recovers Γ from Γh^A.

(b) Let n be an even number and choose the random window h^R from (8). Choose $\tau > 8$ and assume

$$k \leq \frac{n}{2^{u+4}\tau\log(n)^{u+1}} \tag{12}$$

for some $u > 0$ satisfying $2^{u+4}\tau \log(n)^{u+1} > 2/c$ with c no smaller than 0.0818 and such that

$$\sigma = \left(\frac{c}{2} - \frac{1}{2^{u+4}\tau \log(n)^{u+1}}\right)^2 2^{u+4} \log(n)^{u-1} \geq 1. \tag{13}$$

Then with probability at least

$$1 - (2n^{-2u} + 2k^{-\sigma} + 4n^{-(\tau/4-2)})$$

Basis Pursuit (4) recovers Γ from Γh^R. (A similar result also hold for n odd.)

Roughly speaking the above result tells us that Γ can be recovered from Γh^A or Γh^R with *high probability* provided that the sparsity of Γ satisfies $k \leq cn/\log(n)^{u+1}$ for some positive c, u which are governing the probability of success.

A simple argument from time–frequency analysis gives

Corollary 2.6. *Theorems 2.4, 2.5, and 5.1, also hold with the windows h^A and h^R replaced by their Fourier transforms $\widehat{h^A}$ and $\widehat{h^R}$, with entries defined as $\widehat{h}_j = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} h_q e^{2\pi i jq/n}$.*

3. TOOLS IN THE THEORY OF SPARSE SIGNAL RECOVERY

It was shown in (2) that for any test signal h, we have $\Gamma h = (\mathbf{\Psi} h)x$ where x is the sparse coefficient vector of Γ. This observation links the sparse matrix identification question with sparse signal recovery where one seeks the sparsest solution (3) to the underdetermined system $Ax = b$; in the sparse matrix identification setting $(\mathbf{\Psi} h) = (\Psi_1 h \,|\, \Psi_1 h \,|\, \ldots \,|\, \Psi_N h)$ takes the place of A and Γh the place of b. In contrast to sparse approximation, where the dictionary A is usually fixed, for sparse matrix identification we have the additional freedom of designing the test signal h in order for $(\mathbf{\Psi} h)$ to have desirable properties.

Let us shortly recall known results in sparse signal recovery and sparse approximation that we apply to the sparse matrix identification question. In Section 3.1 we review the notion of coherence (14) and its implications for sparse signal recovery and approximation using Basis Pursuit, (4) and (5), as well as Orthogonal Matching Pursuit. In Section 3.2 we review the restricted isometry property, allowing for improved recoverability results for Basis Pursuit.

3.1. Coherence

The recoverability properties of sparse signal recovery algorithms for an underdetermined system $Ax = b$ is often measured by the coherence of A,

$$\mu = \max_{r \neq s} |\langle a_r, a_s \rangle|, \tag{14}$$

where a_r is the r^{th} column of A and $\|a_r\|_2 = 1$ for all r.

Theorem 3.1 (Tropp [42]; Donoho, Elad [18]). *Let A be a unit norm dictionary with coherence μ. If*

$$(2k - 1)\mu < 1 \tag{15}$$

then BP (as well as OMP) recovers all k-sparse vectors x from $b = Ax$.

Recovery is also stable under perturbation by noise when BP (4) is replaced with (5).

THEOREM 3.2 (DONOHO ET AL. [19], THEOREM 3.1). *Let A, μ be as above and suppose that $(4k-1)\mu < 1$. Assume that x is k-sparse and we have perturbed observations $b = Ax + z$ with $\|z\|_2 \le \epsilon$. Then the solution $x^\#$ of the BP variant*

$$\min \|x'\|_1 \quad \text{subject to} \quad \|Ax' - b\|_2 \le \delta$$

satisfies

$$\|x^\# - x\|_2^2 \le \frac{(\epsilon+\delta)^2}{1-\mu(4k-1)}.$$

Theorems 3.1 and 3.2 ensure that the solutions of (4) and (5) correspond (exactly and approximately, respectively) to the solution of (3) for *all* $k-$sparse x. For a broad class of dictionaries the coherence is of order $\mathcal{O}(1/\sqrt{n})$, see Sections 4 and 5 for random and Gabor dictionaries, respectively. Hence, Theorems 3.1 and 3.2 ensure (stable) recovery provided $k = \mathcal{O}(\sqrt{n})$.

In contrast to these $\mathcal{O}(\sqrt{n})$ thresholds, which are valid for all x, Tropp [43] developed a general framework for the analysis of Basis Pursuit (4), which is still based on the coherence of a general dictionary, but shows that (4) is often successful for substantially larger k than those considered in Theorems 3.1 and 3.2. This comes, however, at the cost of assuming a random model on the sparse signal to be recovered. It allows us to prove the near optimal order $n/(\log n)^{1+u}$ recoverability result for the time–frequency-shift dictionary, Theorem 2.5, where $u > 0$ determines success rates. We state the results of Tropp, where $\|\cdot\|_{2,2}$ denotes the operator norm given by $\|A\|_{2,2} = \sup_{\|x\|_2=1} \|Ax\|_2$, and A_Λ is the restriction of a matrix A to the columns indexed by Λ.

THEOREM 3.3 (TROPP [43], THEOREM B). *Let A be an $n \times N$ vector dictionary with unit norm columns and coherence μ. Suppose that Λ is selected uniformly at random among all subsets of $\{1,\dots,N\}$ of size $k \ge 3$. Let $s \ge 1$. Then*

$$s\sqrt{\mu^2 k \log k} + \frac{k}{N}\|A\|_{2,2}^2 \le c\delta \tag{16}$$

implies

$$\mathbb{P}\left(\|A_\Lambda^* A_\Lambda - \mathrm{Id}\|_{2,2} \ge \delta\right) \le 2k^{-s^2}.$$

The constant c is no smaller than 0.0818.

THEOREM 3.4 (TROPP [43], THEOREM 13). *Let A be an $n \times N$ dictionary with coherence μ. Suppose $\Lambda \subseteq \{1,\dots,N\}$ of cardinality k ($|\Lambda| = k$) is such that*

$$\|A_\Lambda^* A_\Lambda - \mathrm{Id}\|_{2,2} \le 1/2$$

and $8\mu^2 k \le \log^{-(u+1)}(N)$ for some $u > 0$. Suppose that $x \in \mathbb{C}^N$ has support Λ with random phases $\mathrm{sgn}(x_r)$*, $r \in T$, that are independent and uniformly distributed on the torus $\{z, |z| = 1\}$. Then with probability at least $1 - 2N^{-u}$ the sparse vector x can be recovered from $b = Ax$ by Basis Pursuit.*

3.2. Restricted isometry property

Candès, Romberg and Tao introduced the Restricted Isometry Property (RIP) which is an alternative perspective to coherence [9, 8].

DEFINITION 3.5. *Let $A \in \mathbb{R}^{n\times N}$ and $k < n$. The restricted isometry constant $\delta_k = \delta_k(A)$ is the smallest number such that for all subset $\Lambda \subset \{1,\dots,N\}$ of cardinality at most k ($|\Lambda| = k$) it holds that*

$$(1-\delta_k)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta_k)\|x\|_2^2$$

for all x supported on Λ.

A is said to satisfy the restricted isometry property if it has small isometry constants, say $\delta_k < 1/2$; such matrices allow stable sparse recovery by BP.

THEOREM 3.6 (CANDÈS, ROMBERG AND TAO [8]). *Assume that the restricted isometry constants of A satisfy*

$$\delta_{3k} + 3\delta_{4k} < 2.$$

Let $x \in \mathbb{C}^N$ and assume we have noisy data $y = Ax + \eta$ with $\|\eta\|_2 \le \epsilon$. Denote by x^k the truncated vector corresponding to the k largest absolute values of x. Then the solution $x^\#$ of (5) satisfies

$$\|x^\# - x\|_2 \ \le\ C_1\epsilon + C_2\frac{\|x-x^k\|_1}{\sqrt{k}}. \tag{17}$$

The constants C_1 and C_2 depend only on δ_{3k} and δ_{4k}.

Note that for x k-sparse and noise level $\epsilon = 0$, Theorem 3.6 guarantees exact recovery of x by (4).

4. RANDOM MATRICES

Many of the recent results in sparse signal recovery with recoverability thresholds for $k \le Cn/\log(n)$ either assume that A is a random Gaussian or Bernoulli matrix [9, 12, 4, 37], or partial random Fourier matrix [7, 38, 36, 29, 35]. Recoverability results in these cases can be obtained by establishing the restricted isometry property, see Definition 3.5, or through a careful analysis of the geometric structure of the convex hull associated with the columns of A [16, 17, 14, 20, 21]. We apply these results to the matrix identification problem when the matrix has a sparse representation in terms of certain random matrices.

4.1. Gaussian matrix ensemble

Assume all entries of the N matrices $\Psi_j \in \mathbb{R}^{n\times m}$ in $\mathbf{\Psi}$ are independent standard Gaussian random variables and h is an arbitrary non-zero vector in $\mathbb{R}^m$ (independent of $\mathbf{\Psi}$). Then the entries of the dictionary $A = (\mathbf{\Psi}h) \in \mathbb{R}^{n\times N}$ whose columns are given by $\Psi_j h$, $j = 1,\dots,N$, are jointly independent and of the form

$$Z = \sum_{\ell=1}^{n} g_\ell h_\ell$$

where the g_ℓ are independent standard Gaussian random variables. By rotational invariance of the distribution of the Gaussian vector $(g_1,\dots,g_n)$ the random variable Z has the same distribution as

$\|h\|_2 g$ where g is a (scalar-valued) standard Gaussian. Hence, the dictionary $(\mathbf{\Psi} h)$ has the same distribution as

$$\|h\|_2 A \in \mathbb{R}^{n\times N},$$

where A is a random matrix whose entries are independent standard Gaussians. Thus, the existing literature in sparse approximation concerning Gaussian matrices applies, see for instance [4, 9, 12, 37, 21] and additional results discussed in the remainder of this section.

Donoho [16, 17] and Donoho and Tanner [14, 20, 21] supplied precise recoverability thresholds, $\rho(n/N)$, for when the solution of (4) coincides with the solution of (3) provided $k \le n\cdot(\rho(n/N)-\tilde{\epsilon})n$ for $\tilde{\epsilon} > 0$. Thresholds are presented both for finite n, N and user prescribed probabilities of success, as well as for the limiting behavior of these thresholds as $n, N \to \infty$. For any fixed ratio $n/N > 0$, the limiting threshold is bounded away from zero, $\rho(n/N) > 0$, and the probability that the solution of (4) coincides with the solution of (3) grown exponentially in N as k decreases below $n \cdot \rho(n/N)$; that is, for any $k \le n \cdot (\rho(n/N) - \tilde{\epsilon})$ with $\tilde{\epsilon} > 0$ there exists an $\alpha(\tilde{\epsilon}, n/N) > 0$ such that the probability that (4) does not recover the solution of (3) is bounded by $\exp(-\alpha N)$.

In the setting of matrix identification, N is typically much larger than n and the asymptotic behavior of the threshold $\rho(\delta)$ with $\delta \to 0$ is of particular interest. For instance, if $N = n^2$, [21] states that the threshold for recovery of *most* signals behaves as $\rho(1/n) \to 1/(2\log(n))$ as $n \to \infty$, and the threshold for recovery of *all* signals behaves as $\rho(1/n) \to 1/(2e\log(n))$ as $n \to \infty$.

For the noisy setting, the restricted isometry property ensures stable recovery with probability at least $1-\varepsilon$ provided

$$k \le c\frac{n}{\log(\frac{N}{n\varepsilon})}. \tag{18}$$

Hence, by Theorem 3.6 we have stable recovery by (5) in this regime and the statement of Theorem 2.3(a) follows.

Note that although the threshold (18) for the noisy setting appears analogous to the asymptotic thresholds of Donoho and Tanner [21] mentioned earlier, some differences exist, the most important one being the stability associated with the restricted isometry property as stated in Theorem 3.6. Setting the polynomial order proability $\varepsilon = N^{-\beta}$ in (18) yields $k \le c_\beta n/\log(N/n)$, which up to the constant is indeed comparable to the thresholds in [21]. However, Donoho and Tanner achieve a considerable smaller failure probability $\exp(-\alpha N)$ in this regime, whereas setting $\varepsilon = \exp(-\alpha N)$ in (18) gives only the vanishingly small $k \le cn/(\alpha N)$.

When moving from the noiseless setting of (4) to the noisy case (5) it is to be expected that a smaller region and/or probability of successful recovery would result. However, it is unclear whether the lower bound (18) is tight in its dependence on ϵ, n, and N. Alternatively to (18), Wainwright [45] presented the recoverability threshold $k \le n/(2\log(N-k))$ below which (5) recovers the correct support (and sign) of a strictly sparse signal. Unfortunately, lacking a precise estimate of the success probability in [45], a direct comparison with (18) and the results in [21] is not possible.

4.2. Bernoulli matrix ensemble

The recoverability results for Bernoulli matrices in Theorem 2.3(b) are based on establishing the restricted isometry property given in Definition 3.5.

To this end, we assume that the entries of the N matrices $\Psi_j \in \mathbb{R}^{n\times m}$ in $\mathbf{\Psi}$ are selected as independent ± 1 Bernoulli variables, that is, $+1$ or -1 with equal probability, and let h an arbitrary non-zero vector. Then an entry of the dictionary $A = (\mathbf{\Psi} h)$ is given by

$$a_{pq} = \sum_{\ell=1}^{n} \epsilon_\ell^{pq} h_\ell, \quad p = 1, \dots, m, \; q = 1, \dots, N, \tag{19}$$

where the ϵ_ℓ^{pq} are independent Bernoulli variables, that is, the a_{pq} are independent Rademacher series [31]. Theorem 4.1 shows that the matrix A has the restricted isometry property with high probability for sparsities k that are nearly linear in m. Hence, by Theorem 3.6, for an arbitrary non-zero choice of h we can recover any Γ having a k-sparse representation in terms of random Bernoulli matrices from the action of Γh through Basis Pursuit (4).

THEOREM 4.1. *Let $h \in \mathbb{R}^m$ be normalized by $\|h\|_2 = 1/\sqrt{m}$. Let A be the random matrix with entries defined in (19). Assume $\delta \in (0, 1)$ and $t > 0$. If*

$$m \geq C_1 \delta^{-2}(k \log(N/k) + \log(2e + 24e/\delta) + t). \tag{20}$$

Then with probability at least $1 - e^{-t}$ the restricted isometry property is satisfied, that is, for all $\Lambda \subset \{1, \dots, N\}$ of cardinality at most k it holds that

$$(1 - \delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta)\|x\|_2^2$$

for all x supported on Λ. The constant satisfies $C_1 \leq 23.15$.

Proof. Let $v \in \mathbb{R}^N$ be an arbitrary vector. We form the inner product of a row of A with v,

$$X_p = \sum_{q=1}^{n} a_{pq} v_q = \sum_{q=1}^{N} \sum_{\ell=1}^{n} \epsilon_\ell^{pq} h_\ell v_q.$$

By independence of the ϵ_ℓ^{pq}, the X_p are similarly independent. By Khintchine's inequality the even moments of X can be estimated by the moments of a standard Gaussian variable g [31, 33]

$$\mathbb{E}[|X_p|^{2z}] \leq \|v\|_2 \|h\|_2 \frac{(2z)!}{2^z z!} = \|v\|_2 \|h\|_2 \mathbb{E}[|g|^{2z}], \quad z \in \mathbb{N}.$$

Following Lemma 5 and the proof of Lemma 6 in [1] this implies the concentration inequality,

$$\mathbb{P}(|\|Av\|_2^2 - \|v\|_2^2| \geq \epsilon \|v\|_2^2) \leq 2 \exp\left(-\tfrac{n}{2}(\epsilon^2/2 - \epsilon^3/3)\right).$$

By Theorem 2.2 in [37], see also Theorem 5.2 in [4], this implies that the restricted isometry property holds under the stated condition on m. □

Note that for fixed δ and t condition (20) can be rewritten as

$$k \leq cn/\log(N/k)$$

for some constant c.

Combining Theorems 3.6 and 4.1 yields Theorem 2.3(b).

4.3. Diagonal matrices

Diagonal matrices act as multiplication operators on $\mathbb{C}^n$. Using a Fourier expansion of the diagonal, we observe that any diagonal matrix can be expressed as linear combinations of modulation operators $M_\ell \in \mathbb{C}^{n\times n}$, $\ell = 0,\dots,n-1$, defined in (6). We now consider the case that only a small number of components of the output of a diagonal operator Γ can be measured; the assumption that Γ is sparse in the dictionary of modulation operators shall be used to recover Γ from these components.

To this end, let Ω be a subset of $\{0,\dots,n-1\}$ of cardinality m and denote by $M_\ell^\Omega \in \mathbb{C}^{m\times m}$ the submatrix of M_ℓ with columns and rows restricted to the index set Ω. Let

$$\mathbf{\Psi}^\Omega = \{M_\ell^\Omega, \ell = 0,\dots,n-1\}$$

and $h = \mathbf{1} = (1,\dots,1)^T$. If $\Gamma^\Omega = \sum_{\ell=0}^{n-1} x_\ell M_\ell^\Omega$ then $\Gamma^\Omega\mathbf{1}$ coincides with the restriction of $\Gamma\mathbf{1} = \sum_{\ell=0}^{n-1} x_\ell M_\ell \mathbf{1}$ to the indices in Ω.

The matrix A whose columns are the elements of the dictionary $(\mathbf{\Psi}^\Omega\mathbf{1}) = \{M_\ell^\Omega\mathbf{1}, \ell = 0,\dots,n-1\}$ is precisely a row submatrix of the Fourier matrix,

$$A \;=\; A^\Omega \;=\; (e^{2\pi i r\ell})_{r\in\Omega,\ell=0,\dots,n-1} \in \mathbb{C}^{m\times n}.$$

If the subset Ω is chosen uniformly at random among all subsets of size m then A^Ω is a random matrix. This random partial Fourier matrix was studied in [7, 9, 38], see also [36] for a slight variation. Indeed, under the condition

$$k \le c\,\frac{m}{\log^4(n)\log(\varepsilon^{-1})}$$

the restricted isometry property holds with probability at least $1-\varepsilon$ [38] and by Theorem 3.6 we obtain stable recovery of all matrices having a sparse representation in terms of $\mathbf{\Psi}^\Omega$.

5. TIME–FREQUENCY SHIFT DICTIONARIES

In this section we establish coherence results for the dictionary of time–frequency shift matrices and prove Theorems 2.4 and 2.5.

5.1. Coherence for the time–frequency shift dictionary

We apply known recovery results [19, 23, 42, 44, 43] for dictionaries with small coherence (14). Assuming $\|h\|_2 = 1$, then (14) simplifies to

$$\mu \;=\; \max_{(\ell,p)\neq(\ell',p')} |\langle M_\ell T_p h, M_{\ell'}T_{p'}h\rangle| \tag{21}$$

for Gabor systems. Based on results by Alltop in [3], Strohmer and Heath showed in [41] that the coherence (21) of $\mathcal{G}h^A$ given in (7) satisfies

$$\mu \;=\; \frac{1}{\sqrt{n}} \tag{22}$$

for n prime. This is almost optimal since the general lower bound in [41] for the coherence of frames with n^2 elements in $\mathbb{C}^n$ yields $\mu \geq \frac{1}{\sqrt{n+1}}$.

Unfortunately, the coherence (21) of h^A applies only for n prime. For arbitrary n we consider the random window h^R.

THEOREM 5.1. *Let $n \in \mathbb{N}$ and choose a random window h^R with entries*

$$h_q^R = \frac{1}{\sqrt{n}}\epsilon_q, \quad q = 0, \ldots, n-1, \tag{23}$$

where the ϵ_q are independent and uniformly distributed on the torus $\{z \in \mathbb{C}, |z| = 1\}$. Let μ be the coherence of the associated Gabor dictionary (21), then for $\alpha > 0$ and n even,

$$\mathbb{P}\big(\mu \geq \frac{\alpha}{\sqrt{n}}\big) \leq 4n(n-1)e^{-\alpha^2/4}, \tag{24}$$

while for n odd,

$$\mathbb{P}\big(\mu \geq \frac{\alpha}{\sqrt{n}}\big) \leq 2n(n-1)\left(e^{-\frac{n-1}{n}\alpha^2/4} + e^{-\frac{n+1}{n}\alpha^2/4}\right). \tag{25}$$

Up to the constant factor α, the coherence in Theorem 5.1 comes close to the lower bound $\mu \geq \frac{1}{\sqrt{n+1}}$ with high probability. Theorems 2.4 and 2.5 follow from these order $\mathcal{O}(1/\sqrt{n})$ coherence results in this section and the Theorems 3.1 and 3.2 of [19, 23, 42, 44] and Theorems 3.3 and 3.4 of Tropp [43] respectively.

Proof of Theorem 5.1. The technical details for n even and odd are slightly different, for conciseness we only state the proof for n even, and outline the proof for n odd.

A direct computation shows that

$$|\langle M_{\ell'}T_{p'}h^R, M_\ell T_p h^R\rangle| = |\langle M_{\ell-\ell'}T_{p-p'}h^R, h^R\rangle|$$

and, therefore, it suffices to consider $\langle M_\ell T_p h^R, h^R\rangle$, $\ell, p = 0, \ldots, n-1$; furthermore, as $\langle M_\ell h^R, h^R\rangle = \langle M_\ell \mathbf{1}, |h^R|^2\rangle = 0$ for $\ell \neq 0$, we consider only the case $p \neq 0$.

Writing $\epsilon_q = e^{2\pi i y_q}$ with $y_q \in [0, 1)$ we obtain

$$\langle M_\ell T_p h^R, h^R\rangle = \frac{1}{n}\sum_{q=0}^{n-1} e^{2\pi i \frac{q\ell}{n}}\epsilon_{q-p}\overline{\epsilon_q} = \frac{1}{n}\sum_{q=0}^{n-1} e^{2\pi i\left(y_{q-p}-y_q+\frac{q\ell}{n}\right)},$$

where $\epsilon_{q-p} = \epsilon_{n+q-p}$ if $q - p < 0$, that is, the indices are understood modulo n. Set

$$\delta_q^{(p,\ell)} = e^{2\pi i\left(y_{q-p}-y_q+\frac{q\ell}{n}\right)},$$

and note that $\delta_q^{(p,\ell)}$ is uniformly distributed on the torus $\mathbb{T}$. However, the $\delta_q^{(p,\ell)}$, $q = 1, \ldots, n$, are no longer jointly independent. But, if $p = 1$, $p = n-1$, or if neither p nor $n - p$ divide n, then the $n/2$ random variables $\epsilon_0\overline{\epsilon_p}, \epsilon_p\overline{\epsilon_{2p}}, \ldots, \epsilon_{p(n/2-1)}\overline{\epsilon_{pn/2}}$ are jointly independent, as well as the remaining

$n/2$ variables $\epsilon_{pn/2}\overline{\epsilon_{p(n/2+1)}}, \ldots, \epsilon_{p(n-1)}\overline{\epsilon_0}$. The indices are again understood modulo n. If $p \geq 2$ or $n-p \geq 2$ divides n, then we form the p random vectors

$$\begin{aligned}
Y_1 =& (\epsilon_0\overline{\epsilon_p}, \epsilon_p\overline{\epsilon_{2p}}, \ldots, \epsilon_{n-p}\overline{\epsilon_0}), \\
Y_2 =& (\epsilon_1\overline{\epsilon_{p+1}}, \epsilon_{p+1}\overline{\epsilon_{2p+1}}, \ldots, \epsilon_{n-p+1}\overline{\epsilon_1}), \\
&\vdots \\
Y_p =& (\epsilon_{p-1}\overline{\epsilon_{2p-1}}, \epsilon_{2p-1}\overline{\epsilon_{3p-1}}, \ldots, \epsilon_{n-1}\overline{\epsilon_{p-1}}).
\end{aligned}$$

These vectors are jointly independent. Moreover, $p \leq n/2$ allows partitioning the entries of a single vector Y into two groups Λ_p^1 and Λ_p^2 with $|\Lambda_p^1|, |\Lambda_p^2| \geq 1$ and the elements of each group are jointly independent. Indeed, this can be seen by forming subgroups of two *adjacent elements* with possibly a remaining single element group. Then all subgroups are jointly independent and the two elements inside a subgroup are independent as well.

Now by forming unions $\cup_{i=1}^p \Lambda_i^1$ and $\cup_{i=1}^p \Lambda_i^2$ we can always partition the index set $\{0, \ldots, n-1\}$ into two subsets Λ_1, $\Lambda_2 \subset \{0, \ldots, n-1\}$ with $|\Lambda_1| = |\Lambda_2| = n/2$ such that the random variables $\{\delta_q^{(p,\ell)}, q \in \Lambda^i\}$ are jointly independent for both $i = 1, 2$.

In the following, we will use the complex Bernstein inequality, see for example [43, Proposition 15] and [33]. It states that for an independent sequence $\epsilon_q, q = 1, \ldots, n$, of random variables which are uniformly distributed on the torus,

$$\mathbb{P}\left(\left|\sum_{q=1}^{n} \epsilon_q\right| \geq nu\right) \leq 2e^{-nu^2/2}. \tag{26}$$

Using the pigeonhole principle and the inequality (26) we obtain

$$\begin{aligned}
\mathbb{P}\left(|\langle M_\ell T_p h^R, h^R\rangle| \geq t\right) &= \mathbb{P}\big(\big|\sum_{q=0}^{n-1} \delta_q^{(p,\ell)}\big| \geq nt\big) \leq \mathbb{P}\big(\big|\sum_{q\in\Lambda^1} \delta_q^{(p,\ell)}\big| \geq nt/2\big) + \mathbb{P}\big(\big|\sum_{q\in\Lambda^2} \delta_q^{(p,\ell)}\big| \geq nt/2\big) \\
&\leq 4\exp(-nt^2/4).
\end{aligned}$$

Forming the union bound over all possible $(p, \ell) \in \{0, \ldots, n-1\}^2 \setminus \{(0, 0)\}$ and choosing $t = \alpha/\sqrt{n}$ yields the statement of Theorem 5.1 for n even.

The proof of Theorem 5.1 for n odd uses essentially the same technique as for n even, with the difference that the random variables $\delta_k^{(m,\ell)}$ are grouped into sets of unequal cardinality, $|\Lambda^1| = (n-1)/2$ and $|\Lambda^2| = (n+1)/2$. For large n the probability tail bounds are nearly the same for n even (24) and n odd (25). □

5.2. Proof of Theorem 2.4

Part (a) follows directly from Theorem 3.1 and the coherence of $\mathcal{G}h^A$ (22).

Part (b) follows from Theorem 3.1 and Theorem 5.1. In fact, the probability that the condition $\mu < (2k-1)^{-1}$ of Theorem 3.1 does *not* hold for $\mathcal{G}h^R$ is estimated by

$$\mathbb{P}(\mu \geq (2k-1)^{-1}) \leq 4n^2 \exp\left(-\frac{n}{4(2k-1)^2}\right).$$

Requiring that the latter term is less than e^{-t} and solving for k gives (9). □

5.3. Proof of Theorem 2.5

Having established coherence results for $\mathcal{G}h^A$ and $\mathcal{G}h^R$ in Section 5.1, Theorem 2.5 follows from Theorems 3.3 and 3.4 of Tropp [43] as shown below.

(a) Recall from (22) that the coherence for $\mathcal{G}h^A$ satisfies $\mu = n^{-1/2}$. Since $N = n^2$, condition $8\mu^2 k \le (2\log(n))^{-(u+1)} = \log^{-(u+1)}(N)$ in Theorem 3.4 is valid by (10).

Next, observe that h^A unimodular implies that the columns of $\mathcal{G}h^A$ form n orthonormal bases, and, hence, $n = \|(\mathcal{G}h^A)^*\|_{2,2}^2 = \|\mathcal{G}h^A\|_{2,2}^2$. Further, set $s = \sqrt{\sigma}$. Then by (11) and (10) we have

$$s\sqrt{\mu^2 k \log(k)} + \frac{k}{n^2}\|\mathcal{G}h^A\|_{2,2}^2 \;=\; \sqrt{\sigma\frac{k\log k}{n}} + \frac{k}{n} \le c/2.$$

Hence, condition (16) in Theorem 3.3 holds for $A = \mathcal{G}h^A$ and we conclude that $\|A_\Lambda^* A_\Lambda - \mathrm{Id}\|_{2,2} \le 1/2$ with probability at least $1 - 2k^{-s^2} = 1 - 2k^{-\sigma}$.

Now let $\delta = \|A_\Lambda^* A_\Lambda - \mathrm{Id}\|_{2,2}$. Then

$$\mathbb{P}(\text{BP does not recover } \Gamma \text{ from } \Gamma h^A) \le \mathbb{P}(\text{BP does not recover } \Gamma \text{ from } \Gamma h^A | \delta \le 1/2) + \mathbb{P}(\delta > 1/2).$$

Thus by Theorem 3.4 we can lower bound the probability that recovery is successful by

$$1 - (2k^{-\sigma} + 2n^{-2u}).$$

(b) Let μ be the coherence associated with the random Gabor window h^R. We assume that $t = \sqrt{\tau\frac{\log(n)}{n}} \ge \mu$. The probability that this does *not* hold can be estimated by Theorem 5.1, namely

$$\mathbb{P}(\mu > t) \le 4n(n-1)\exp(-nt^2/4) \le 4n^2\exp(-\log(n)\tau/4) \;=\; 4n^{-(\tau/4-2)}. \tag{27}$$

Setting $s = \sqrt{\sigma}$ and using (12), (13), and $\|(\mathcal{G}h^R)\|_{2,2} = n$ to estimate the expression on the left hand side of (16) by

$$\begin{aligned} s\sqrt{\mu^2 k\log(k)} + \frac{k}{n} &\le \sqrt{\sigma\tau\frac{\log(n)}{n}k\log(k)} + \frac{k}{n} \\ &\le \left(\frac{c}{2} - \frac{1}{2^{u+4}\tau\log(n)^{u+1}}\right)\sqrt{2^{u+4}\log(n)^{u-1}\frac{\log(n)}{n}\tau k\log(k)} + \frac{k}{n} \\ &\le \left(\frac{c}{2} - \frac{k}{n}\right)\left(k2^{u+4}\tau\log(n)^{u+1}\frac{1}{n}\right) + \frac{k}{n} \le c/2. \end{aligned}$$

Hence, by Theorem 3.3 we have $\|A_\Lambda^* A_\Lambda - \mathrm{Id}\|_{2,2} > 1/2$ with probability at most $2k^{-s^2} = 2k^{-\sigma}$. Similarly to the proof of part (a), we estimate the probability of successful recovery by

$$\begin{aligned} \mathbb{P}(\text{BP recovers } \Gamma \text{ from } \Gamma h^R) \;&\ge\; 1 - \big(\mathbb{P}(\text{BP does not recover } \Gamma \text{ from } \Gamma h^R | \delta \le 1/2) + \mathbb{P}(\delta > 1/2)\big) \\ &\ge\; 1 - \big(\mathbb{P}(\text{BP does not recover } \Gamma \text{ from } \Gamma h^R | \delta \le 1/2) \\ &\qquad + \mathbb{P}(\delta > 1/2 | \mu \le t) + \mathbb{P}(\mu > t)\big). \end{aligned}$$

By (27) as well as Theorems 3.3 and 3.4, the probability that Γ can be reconstructed from Γh^R by Basis Pursuit (4) exceeds

$$1 - (2n^{-2u} + 2k^{-\sigma} + 4n^{-(\tau/4-2)}).$$

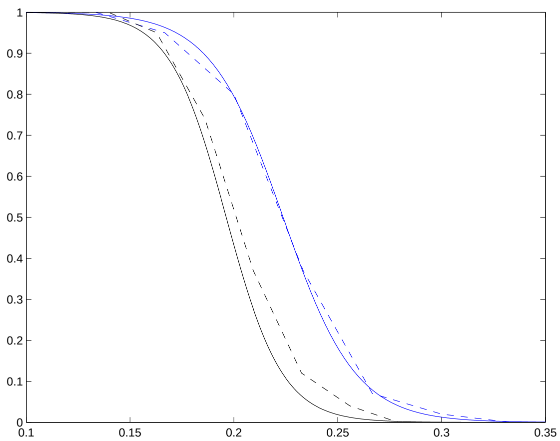

Figure 1. Empirical verification of Theorem 2.5 without noise. The mean response of Y_k^n (dashed) and fitted logistic regression model (solid) for h^A with $n = 43$ (left-black) and h^R with $n = 30$ (right-blue), plotted against the fractional sparsity k/n.

5.4. Proof of Corollary 2.6

Plancherel's theorem and $\widehat{M_\ell T_p h} = T_\ell M_{n-p} \widehat{h} = \sigma M_{n-p} T_\ell \widehat{h}$ with $|\sigma| = 1$ implies that the coherence remains the same under Fourier transform of the window, that is,

$$\begin{aligned} \mu_h &= \sup_{(\ell,p)\neq(\ell',p')} |\langle M_\ell T_p h, M_{\ell'} T_{p'} h\rangle| = \sup_{(\ell,p)\neq(\ell',p')} |\langle \widehat{(M_\ell T_p h)}, \widehat{(M_{\ell'} T_{p'} h)}\rangle| \\ &= \sup_{(\ell,p)\neq(\ell',p')} |\langle M_{n-p} T_\ell \hat{h}, M_{n-p'} T_{\ell'} \hat{h}\rangle| = \mu_{\hat{h}}. \end{aligned}$$

Since all of the results concerning the dictionary of time–frequency shift matrices stated above are based on the coherence this proves the claim.

6. NUMERICAL RESULTS

Theorem 2.5 can be tested empirically for various values of n by trying a number of sparsity levels k and recording the fraction of times (4) recovers the true k-sparse coefficient vector x.

For the Alltop window h^A in (7) we consider the values of n prime from 11 to 59, for the random window h^R in equation (8) we consider the values of n prime from 11 to 59 as well as $n = 10 + 4j$ for $j = 0, 1, \ldots, 12$. Each empirical test consists of generating a random k-sparse $x \in \mathbb{C}^{n^2}$ with non-zero entries x_q selected iid as $r_q \exp(2\pi i \theta_q)$ with r_q drawn from the Gaussian $N(0, 1)$ distribution, and θ_q drawn iid uniformly from $[0, 1)$. For each value of n, 1000 tests are computed per value of $k = 1, 2, \ldots, n-1$. The successful recovery of the coefficient vector x, and, hence, of Γ from Γh^A or Γh^R using Basis Pursuit (4) is recorded in Y_k^n as a 1, and failure to recover as a 0. Following the empirical examination of phase transitions in [13], we approximate the observed probability distribution by fitting the mean response of Y_k^n using the logistic regression model

$$E(Y_k^n) = \frac{\exp(\beta_0(n) + \beta_1(n)k)}{1 + \exp(\beta_0(n) + \beta_1(n)k)}. \tag{28}$$

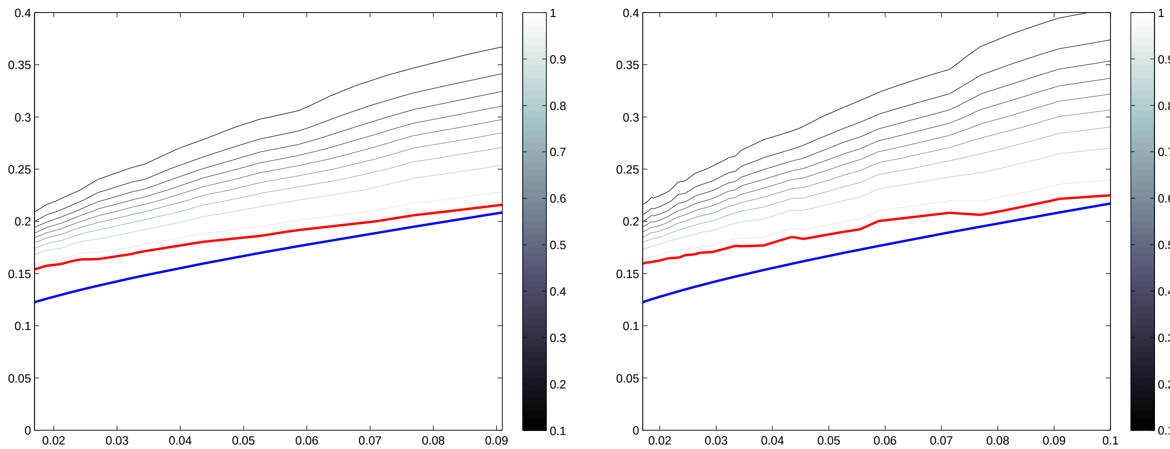

Figure 2. Empirical verification of Theorem 2.5 without noise. Contours of the fitted logistic regression model (gray), the 93% success rate contour (red), and $1/(2\log n)$ (blue). Horizontal axis $1/n$ and vertical axis k/n. Figure 1 shows vertical slices for 1/43 (left) and 1/30 (right).

For illustration purposes, the fitted response for windows h^A with $n = 43$ and h^R with $n = 30$ is shown in Figure 1 along with the mean response of Y_k^n.

The phase transition behaviors are traditionally observed through the fractional sparsity ratio k/n, and the matrix so–called undersampling rate n/N, here $1/n$ for $\mathcal{G}h^A$ and $\mathcal{G}h^R$ [21]. Contours of the fitted logistic regression models for time–frequency shift dictionaries with identifiers h^A and h^R are shown in Figure 2 left and right respectively. Overlaying these contours is the level curve for 93% success rate (red) and $1/(2\log n)$ (blue). The curve $1/(2\log n)$ is known to be the threshold for overwhelming probability of successful recovery in the case of Gaussian random matrices for large n [21]. It is observed in Figure 2 that the curve $1/(2\log n)$ remains below the 93% success rate level curve, indicating consistence of the empirical results with the phase transition $1/(2\log n)$ conjectured for the class of time–frequency shift matrices applied to identifiers h^A and h^R. Moreover, the curve $1/(2\log n)$ increasingly falls below the 93% success rate level curve as n increases, indicating improved agreement in the large n limit. Note that this conjectured phase transition $1/(2\log n)$ is larger than that proven in the main Theorem 2.5, both in order (as $u = 0$ here), as well as in the constant.

As stated earlier, in practice the measurements Γh are observed with noise and although Γ can be well approximated by a $k-$sparse representation, it is rarely strictly $k-$sparse. For both of these reasons, the recovery algorithm (4) is not often used in practice, rather (5) is used to allow for an inexact fit of the measurements.

We empirically test Theorem 2.5 using (5) rather than (4) for the reconstruction algorithm, for the same values of k and n, and the same number of tests as was done to generate Figure 2. The non-zero entries in x are similarly selected from the same distribution as was used to generate Figure 2. Additive noise is simulated at a level of $25db$ signal to noise ratio; that is, η is added to Γh with the entries in η selected iid from the same distribution as the non-zero entries in x, and η is normalized to $\|\eta\|_2 = \|\Gamma h\|_2 \cdot 10^{-5/4}$. The vector x associated with Γ is only considered to have been successfully recovered if the largest k entries of the recovered x' have the same support set Λ as x. The inequality

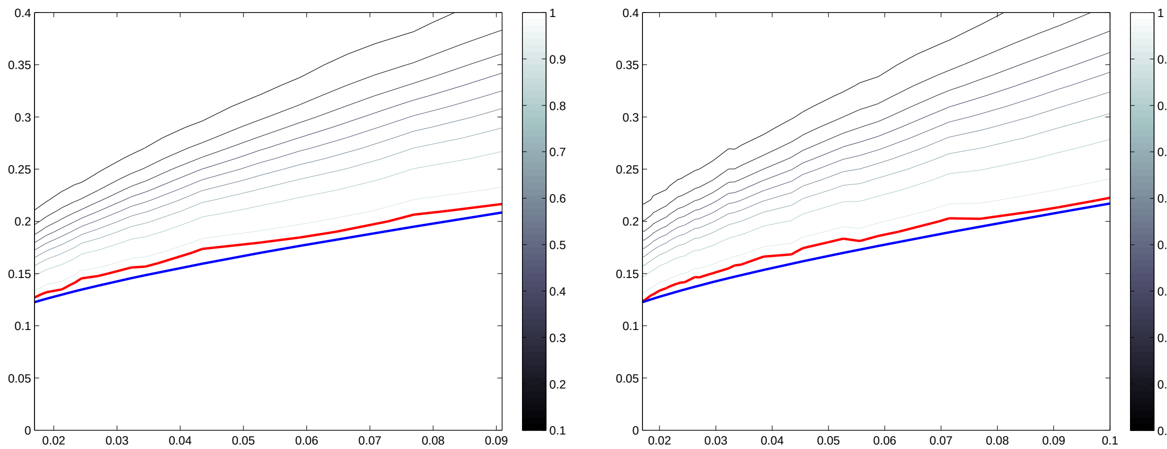

Figure 3. Empirical verification of Theorem 2.5 in the noisy setting, with (4) replaced by (5) and additive noise of 25*db* signal to noise ratio. Contours of the fitted logistic regression model (gray), the 93% success rate contour (red), and $1/(2\log n)$ (blue). Horizontal axis $1/n$ and vertical axis k/n.

fit parameter ϵ in (5) is selected to be at the noise level $10^{-5/4}$. As for Figure 2, we approximate the probability distribution of the empirical observations Y_k^n using the logistic regression model (28). Contours of the fitted logistic regression models for time–frequency shift dictionaries with identifiers h^A and h^R are shown in Figure 3 left and right respectively. Overlaying these contours is the level curve for 93% success rate (red) and $1/(2\log n)$ (blue). Unlike the noiseless case (4), it was shown that the threshold for overwhelming probability of successful recovery in the case of Gaussian random $n \times n^2$ matrices with noise using (5) is $1/(4\log n)$, [45]; however, we observe the $1/(2\log n)$ to fit the empirical data better in this instance.

References

[1] D. Achlioptas. Database-friendly random projections. In *Proc. 20th Annual ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems*, pages 274–281, 2001.

[2] M. Akcakaya and V. Tarokh. Performance bounds on sparse representations using redundant frames. *Preprint*, 2007.

[3] W. O. Alltop. Complex sequences with low periodic correlations. *IEEE Trans. Inf. Theory*, 26(3):350–354, 1980.

[4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, to appear.

[5] P. Bello. Characterization of randomly time-variant linear channels. *IEEE Trans. Commun.*, 11:360–393, 1963.

[6] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge Univ. Press, 2004.

[7] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.

[8] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[9] E. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.

[10] M. Cetin and B. Sadler. Semi–blind sparse channel estimation with constant modulus symbols. In *Proc. IEEE ICASSP 05*, volume 3, pages iii/561– iii/564, Atlanta (GA), 2005.

[11] S. Cotter and B. Rao. Sparse channel estimation via matching pursuit with applications to equalization. 50(3):374–377, 2002.

[12] D. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[13] D. Donoho, I. Drori, J.-L. Starck, and Y. Tsaig. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. *Preprint*, 2006.

[14] D. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Nat. Acad. Sci.*, 102(27):9446–9451, 2005.

[15] D. Donoho and Y. Tsaig. Fast solution of l1-norm minimization problems when the solution may be sparse. *Preprint*, 2006.

[16] D. L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. *Preprint*, 2005.

[17] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.*, 35(4):617–652, 2006.

[18] D. L. Donoho and M. Elad. Optimally sparse representations in general (non-orthogonal) dictionaries via ell^1 minimization. *Proc. Nat. Acad. Sci.*, 100:2197–2202, 2002.

[19] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory*, 52(1):6–18, 2006.

[20] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102(27):9452–9457, 2005.

[21] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *Preprint*, 2006.

[22] A. Gilbert and J. Tropp. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, to appear.

[23] R. Gribonval and P. Vandergheynst. On the exponential convergence of matching pursuits in quasi-incoherent dictionaries. *IEEE Trans. Inform. Theory*, 52(1):255–261, 2006.

[24] K. Gröchenig. *Foundations of Time-Frequency Analysis.* Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, MA, 2001.

[25] D. Han, S.-P. Kim, and J. Principe. Sparse channel estimation with regularization method using convolution inequality for entropy. In *Proc. IJCNN'05, International Joint Conf. on Neural Networks*, volume 4, pages 2359–2362, August 2005.

[26] S. Kim, K. Ksh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale l1-regularized least squares problems with applications in signal processing and statistics. *Preprint*, 2007.

[27] W. Kozek and G. Pfander. Identification of operators with bandlimited symbols. *SIAM J. Math. Anal.*, 37(3):867–888, 2006.

[28] F. Krahmer, G. Pfander, and P. Rashkov. Uncertainty principles for time–frequency representations on finite abelian groups. Preprint, 2006.

[29] S. Kunis and H. Rauhut. Random sampling of sparse trigonometric polynomials II - orthogonal matching pursuit versus basis pursuit. *Found. Comput. Math.*, to appear.

[30] J. Lawrence, G. Pfander, and D. Walnut. Linear independence of Gabor systems in finite dimensional vector spaces. *J. Fourier Anal. Appl.*, 11(6):715–726, 2005.

[31] M. Ledoux and M. Talagrand. *Probability in Banach Spaces. Isoperimetry and Processes.* Springer-Verlag, Berlin, Heidelberg, New York, 1991.

[32] J. Parsons, D. Demery, and A. Turkmani. Sounding techniques for wideband mobile radio channels: a review. *IEE Proceedings-1*, 138(5):437–446, 1991.

[33] G. Pevskir and A. Shiryaev. The Khintchine inequalities and martingale expanding sphere of their action. *Russ. Math. Surv.*, 50(5):849–904, 1995.

[34] G. Pfander and D. Walnut. Measurement of time–variant channels. *IEEE Trans. Info. Theory*, 52(11):4808–4820, 2006.

[35] H. Rauhut. Stability results for random sampling of sparse trigonometric polynomials. *Preprint*, 2006.

[36] H. Rauhut. Random sampling of sparse trigonometric polynomials. *Appl. Comput. Harm. Anal.*, 22(1):16–42, 2007.

[37] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *Preprint*, 2006.

[38] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *Proc. CISS 2006 (40th Annual Conference on Information Sciences and Systems)*, 2006.

[39] S. Sanyal, S. Kukreja, E. Perreault, and D. Westwick. Identification of linear time varying systems using basis pursuit. In *Proceedings of the IEEE EMBS 2005.*

[40] M. Skolnik. *Introduction to Radar Systems*. McGraw-Hill Book Company, New York, 1980.

[41] T. Strohmer and R. W. Heath. Grassmannian frames with applications to coding and communication. *Appl. Comput. Harmon. Anal.*, 14(3):257–275, 2003.

[42] J. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory*, 50(10):2231–2242, 2004.

[43] J. Tropp. Random subdictionairies of general dictionairies. *Preprint*, 2006.

[44] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals. *IEEE Trans. Inf. Theory*, 51(3):1030–1051, 2006.

[45] M. J. Wainwright. Sharp thresholds for noisy and high-dimensional recovery of sparsity using ℓ_1-constrained quadratic programming. *Preprint*, 2006.

Version of June 11, 2007.