

On the Effective Measure of Dimension in the Analysis Cosparse Model

Raja Giryes, Yaniv Plan, and Roman Vershynin

Abstract—Many applications have benefited remarkably from low-dimensional models in the recent decade. The fact that many signals, though high dimensional, are intrinsically low dimensional has given the possibility to recover them stably from a relatively small number of their measurements. For example, in *compressed sensing* with the standard (synthesis) sparsity prior and in *matrix completion*, the number of measurements needed is proportional (up to a logarithmic factor) to the signal’s manifold dimension.

Recently, a new natural low-dimensional signal model has been proposed: the cosparse analysis prior. In the noiseless case, it is possible to recover signals from this model, using a combinatorial search, from a number of measurements proportional to the signal’s manifold dimension. However, if we ask for stability to noise or an efficient (polynomial complexity) solver, all the existing results demand a number of measurements which is far removed from the manifold dimension, sometimes far greater. Thus, it is natural to ask whether this gap is a deficiency of the theory and the solvers, or if there exists a real barrier in recovering the cosparse signals by relying only on their manifold dimension. Is there an algorithm which, in the presence of noise, can accurately recover a cosparse signal from a number of measurements proportional to the manifold dimension? In this work, we prove that there is no such algorithm. Further, we show through numerical simulations that even in the noiseless case convex relaxations fail when the number of measurements is comparable to the manifold dimension. This gives a practical counter-example to the growing literature on compressed acquisition of signals based on manifold dimension.

R. Giryes is with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708 USA (e-mail: raja.giryes@duke.edu).

Y. Plan is with the Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada (e-mail: yaniv@math.ubc.ca).

R. Vershynin is with the Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, U.S.A (e-mail: romanv@umich.edu).

Manuscript received October 3, 2014. RG is partially supported by AFOSR. YP is partially supported by an NSF Postdoctoral Research Fellowship under award No. 1103909. RV is partially supported by NSF under grant DMS 1265782, by USAF under grant FA9550-14-1-0009, and by Humboldt Research Foundation.

I. INTRODUCTION

Low-dimensional signal models have played an important role in many signal processing applications in the recent decade, where in many cases the use of these models has provided state of the art results [1], [2], [3], [4]. All have relied on the fact that the treated signals, which have high ambient dimension, reside in a low dimensional manifold [5], [6], [7], [8], [9], [10], or a union of manifolds, e.g., a union of subspaces.

The fact that a signal belongs to a low dimensional manifold may make it possible to recover it from few measurements. This is exactly the essence of the compressed sensing [11], [12] and matrix completion [13] problems.

Our core problem is to recover an unknown signal $x \in \mathbb{R}^d$ from a given set of its linear measurements

$$y = Ax + z, \quad (\text{I.1})$$

where $A \in \mathbb{R}^{m \times d}$ is the measurement matrix with $m \ll d$ and z is an additive noise which can be either adversarial with bounded energy [12], [14], [15] or random with a certain given distribution, e.g., Gaussian [16].

In the standard compressed sensing problem the low dimensionality of x is modeled using the synthesis sparsity model. The signal x is assumed to be either explicitly sparse, i.e., has a small number of non-zeros, or with a sparse representation under a given dictionary $D \in \mathbb{R}^{d \times n}$.

If every group of $2k$ columns in AD (if x is explicitly sparse then $D = \text{Id}$) are independent, where k is the sparsity of the signal, then x can be uniquely recovered from y in the noiseless case ($z = 0$) [17], [18] using a combinatorial search.

The above observation shows that it is possible to recover the signal with a number of measurements of the order of its manifold dimension k . However, the combinatorial search is not feasible for any practical size of d [19]. Therefore several relaxation techniques have been proposed [12], [20], [21], [22], [23], [24], [25], [26]. It has been shown that if AD is a subgaussian

random matrix or a partial Fourier matrix then it is possible using these practical methods to recover \mathbf{x} from \mathbf{y} using only $O(k \log^c(n))$ measurements (c is a given constant) [15], [27]. Note that up to the log factor, the number of needed measurements is of the order of the manifold dimension of the signal.

Noise is easily incorporated into these results. The same number of measurements guarantees robustness to adversarial noise. Further, in the case of random white Gaussian noise the recovery error turns to be of the order of $k\sigma^2 \log(n)$ [16], [28], [29], where $k\sigma^2$ is roughly the energy of the noise in the low dimensional subspace the signal resides in.

Similar results hold outside of the standard (synthesis) sparsity model of compressed sensing. Indeed, in the *matrix completion* problem, in which the signal is a low-rank matrix, once again one may reconstruct the matrix from a number of measurements proportional to the manifold dimension [13], [30]. Further, there exists a body of literature giving an abstract theory of signal reconstruction based on manifold dimension [10], [31], [32], [33], [34], [35].

A. The Cosparse Analysis Model

Recently, a new signal sparsity framework has been proposed: the cosparse analysis model [36], [37] that looks at the behavior of the signal after applying a given operator $\Omega \in \mathbb{R}^{p \times d}$ on it. We introduce it with an important example: the vertical and horizontal finite difference operator (2D-DIF). The motivation for the usage of this operator is that in many cases the image is not sparse but its gradient is, so the signal “becomes sparse” only after the application of 2D-DIF. Therefore it is common to represent the structure of an image through its behavior after the application of the 2D-DIF operator. Note that 2D-DIF, when used with ℓ_1 -minimization (See (III.1) hereafter), corresponds to the anisotropic two dimensional total variation (2D-TV) [38].

For simplicity, we suppose that d is a square number and consider an $\sqrt{d} \times \sqrt{d}$ signal matrix X , e.g., an image. Then the horizontal differences operator, \mathbf{H} , and the vertical finite differences operator, \mathbf{V} , are defined as follows:

$$\mathbf{H}(X)_{i,j} := X_{i,j} - X_{i,j+1} \quad (\text{I.2})$$

$$\mathbf{V}(X)_{i,j} := X_{i,j} - X_{i+1,j} \quad (\text{I.3})$$

with addition of indices being done mod \sqrt{d} . One may then unfold the signal X into vector form, and correspondingly represent \mathbf{H} and \mathbf{V} as $d \times d$ matrices. Thus, we take $\Omega \in \mathbb{R}^{2d \times d}$, in the 2D-DIF model, to be

the horizontal differences matrix stacked on top of the vertical differences matrix .

In the general case, for a given operator $\Omega \in \mathbb{R}^{p \times d}$, the cosparse analysis model assumes that $\Omega \mathbf{x}$ should be sparse. The subspace in which the signal resides is characterized by the zeros in $\Omega \mathbf{x}$. The number of zeros in $\Omega \mathbf{x}$ is denoted as the cosparsity of the signal. Each zero entry characterizes a row in Ω to which the signal is orthogonal. Denoting by T the unknown support (the locations of the non-zero entries) of $\Omega \mathbf{x}$ and by Ω_T the submatrix of Ω restricted to the rows in T we have that the subspace of \mathbf{x} is the one orthogonal to the subspace spanned by the rows of Ω_{T^c} . In what follows, define

$$K_T := \{\mathbf{x} \in \mathbb{R}^d : \Omega_{T^c} \mathbf{x} = 0\} \quad (\text{I.4})$$

to be such a subspace. Notice that the dimension of K_T , which we denote by b , equals $d - \text{rank}(\Omega_{T^c})$. In general, T is not known, and so it is natural to assume a signal structure of the following form

$$K_b := \bigcup_{\dim(K_T)=b} K_T. \quad (\text{I.5})$$

Note that this is a finite union of b -dimensional subspaces. As each signal in K_b belongs to a subspace of dimension b , we say that K_b has a *manifold dimension* b .¹

Recent literature [38], [40], [41], [42], [43], [44] shows that \mathbf{x} may be reconstructed efficiently and stably from $O(|T| \log(p))$ random linear measurements. How does this compare to the manifold dimension of the signal? Assume $|T|$ is fixed and Ω is in general position, e.g., each entry is taken from a Gaussian ensemble. Then $\mathbf{x} \in K_b$, where $b = (d - p + |T|)_+$ (as \mathbf{x} is orthogonal to $p - |T|$ rows in Ω). In particular, if $|T| = p - d + 1$ then $b = 1$ and therefore \mathbf{x} resides in a 1-dimensional subspace. Surprisingly, modern theory requires more than $|T| = p - d + 1$ measurements for recovering the signal, i.e., more measurements than the ambient dimension in the case $p \geq 2d$.

This behavior is not unique only to the case of Ω in general position. For example, consider the 2D-DIF model with $b = 2$. In this case K_2 consists of images

¹See [39] for a definition of manifolds and manifold dimension. We note that K_b is technically not a manifold, but this can be easily remedied. Consider the slightly smaller set: $K'_b = K_b \setminus \bigcup K_T \cap K_{T'}$ in which we take the union over all $T \neq T'$ such that $\dim(K_T) = \dim(K_{T'}) = b$. K'_b is a manifold, and none of the proofs would change under this alternative definition of the signal set. Nevertheless, we define K_b as in Equation (I.5) for simplicity of presentation.

with only two connected components.² However, also in this case the current theory requires the number of measurements to be proportional to $|T|$, the number of edges in these images, which might be much larger than 2. Notice that for the same manifold dimension, we may have different number of edges in different images. For example, in Fig. 1a the number of edges is roughly proportional to \sqrt{d} and in Fig. 1c it is roughly proportional to d .

Following the above two examples, it is natural to ask whether the state of the art theory may be improved. Indeed, let $\mathbf{A} \in \mathbb{R}^{m \times d}$ be a Gaussian matrix and observe that by solving

$$\min_{\mathbf{x}'} \|\Omega \mathbf{x}'\|_0 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A} \mathbf{x}', \quad (\text{I.6})$$

where $\|\cdot\|_0$ is the ℓ_0 pseudo-norm that counts the number of non-zeros in a vector, one may recover any $\mathbf{x} \in K_1$ using only two measurements. In the general case there is a need for $2b$ measurements to recover a signal in K_b [37]. However, solving (I.6) is NP-hard and requires a combinatorial search. Thus, there is a large gap between the theory for tractable, stable signal recovery and what can be done by combinatorial search with noiseless measurements.

B. Our Contribution

With these observations before us, it is natural to ask whether the gap between the required number of measurements and the manifold dimension is a deficiency of the utilized approximation strategies and the used proof techniques, or whether there exists a real barrier with recovering the cosparse signals by relying only on the manifold dimension. Is there any algorithm that can robustly reconstruct a cosparse signal from a number of measurements proportional to the manifold dimension b ?

This paper addresses this question by considering the effect of Gaussian noise. We show that unless the number of measurements is much larger than the manifold dimension, there is no estimator that can stably reconstruct the signal. We show this for two different analysis dictionaries: 1) the vertical and horizontal finite difference operator and 2) a random Gaussian matrix. We show that when $m < d < p$, the error must be exponentially larger than the noise no matter what estimator is used.

²All images in the subspace K_T have the same pattern of edges, defined by the indices of T (two adjacent pixels have an edge when they may take different values). The edges of T separate the image into connected components. If there are only two connected components it is enough to use only two numbers, which set the grey value of each component, to define each image in this subspace. In this case, K_T is a two-dimensional subspace.

We state our two main theorems below.

Theorem I.1. *Let Ω be the 2D-DIF operator and let K_2 be the union of 2-dimensional subspaces generated by this matrix. Suppose $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$ for some $\mathbf{x} \in K_2$, $\mathbf{A} \in \mathbb{R}^{m \times d}$ with $\|\mathbf{A}\| \leq 1$, and $\mathbf{z} \sim \mathcal{N}(0, \sigma^2 \cdot \text{Id})$. Then for any estimator $\hat{\mathbf{x}}(\mathbf{y})$ we have*

$$\max_{\mathbf{x} \in K_2} \mathbb{E} \|\hat{\mathbf{x}} - \mathbf{x}\|_2 \geq C\sigma \exp(cd/m).$$

Theorem I.2. *Let $\Omega \in \mathbb{R}^{p \times d}$ be a matrix with independent standard normal entries and let K_1 be the union of 1-dimensional subspaces generated by this matrix. Suppose $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$ for some $\mathbf{x} \in K_1$, $\mathbf{A} \in \mathbb{R}^{m \times d}$ with $\|\mathbf{A}\| \leq 1$, and $\mathbf{z} \sim \mathcal{N}(0, \sigma^2 \cdot \text{Id})$. Then for any estimator $\hat{\mathbf{x}}(\mathbf{y})$ we have*

$$\max_{\mathbf{x} \in K_1} \mathbb{E} \|\hat{\mathbf{x}} - \mathbf{x}\|_2 \geq C\sigma \exp\left(C \frac{d-1}{m} \left(1 - \frac{d-2}{p}\right)\right).$$

The theorems are proven in Section II by providing packings for the sets K_1 and K_2 , and combining these with a hypothesis testing argument.

We can say that both theorems show that $O(d)$ measurements are needed for any algorithm to get signal reconstruction without incurring a huge error if $p \geq 2d$; the latter assumption is implicit in Theorem I.1 where $p = 2d$. Notice that in Theorem I.2, if $d < p < 2d$ then we need $m = O(p - d)$, which is still remarkably larger than the manifold dimension $b = 1$. Remarkably, both theorems lower bound the efficacy of any estimator, tractable or not, and thus show that even the performance of ℓ_0 minimization is not characterized well by the manifold dimension of the signal. Therefore we may conclude that the fact that the needed number of measurements is of the order of $|T|$ is not a result of a flaw in the existing reconstruction guarantees or a problem with the studied methods. We perform several experiments to demonstrate this fact and show that indeed, the size of $|T|$ (determined by the cosparsity of the signal, the number of zeros in $\Omega\mathbf{x}$) is a better measure for its compressibility than its manifold dimension.

C. Organization

The paper is organized as follows. Section II gives the proofs of Theorems I.1 and I.2. In Section III we demonstrate the lower bounds we have developed through several experiments that use the ℓ_1 -minimization technique with a Gaussian matrix and the 2D-DIF operator as the analysis dictionary. In Section IV we discuss the implications of the derived results and conclude the work.

II. PROOFS OF THEOREMS I.1 AND I.2

Both of our main theorems are proven by construction of random packings, followed by a hypothesis testing argument. Recall that a packing $\mathcal{X} \subset K$ with ℓ_2 balls of radius δ is a set satisfying $\|x - y\|_2 \geq \delta$ for any $x, y \in \mathcal{X}$ with $x \neq y$. We denote by $P(K, \delta)$ the maximal cardinality of such a set, i.e., the *packing number*.

We now gather supporting lemmas and then put them together in Section II-D. We will construct a packing when Ω is the vertical and horizontal differences operator and when Ω is Gaussian. In both cases, we will make a random construction using the following observation.

Lemma II.1 (Random packing). *Let F be a distribution supported on some set $K \in \mathbb{R}^n$. Let x, x' be independently chosen from F . Suppose that*

$$\mathbb{P}\{\|x - x'\|_2 < \delta\} \leq \eta$$

for some $\eta, \delta > 0$. Then,

$$P(K, \delta) \geq \eta^{-1/2}.$$

Proof: Let q be $\eta^{-1/2}$ rounded up to the next nearest integer. Pick q points independently from K . Then, by considering each of the $\binom{q}{2}$ pairs and using the union bound we have

$$\begin{aligned} \mathbb{P}\left\{\min_{x \neq x'} \|x - x'\|_2 < \delta\right\} &\leq \binom{q}{2} \eta \\ &< \frac{(\eta^{-1/2} + 1)\eta^{-1/2}}{2} \cdot \eta < 1 \end{aligned} \quad (\text{II.1})$$

so long as $\eta < 1$. The minimum above is taken over all $x \neq x'$ in the q random points. Thus, with probability greater than 0, the q points satisfy

$$\|x - x'\|_2 \geq \delta \quad \text{for } x \neq x'.$$

Thus, there must exist at least one such arrangement of points, making the requisite packing. ■

A. Packing when Ω is the vertical and horizontal differences operator

We construct a packing for the set $K := K_2 \cap S^{d-1}$ when Ω is the vertical and horizontal finite differences operator.

Lemma II.2. *Suppose that $d \geq 64$ is a square number. Then*

$$P(K, 1/2) \geq \exp(d/64).$$

Proof: Ω is most simply visualized acting on images. Thus, set $n := \sqrt{d}$ to be the number of vertical or horizontal pixels in an image. We make the corresponding abuse of notation and take $\Omega : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^p$

and $K \subset \mathbb{R}^{n \times n}$. Each zero entry of Ωx forces two adjacent entries of x to be equal. Thus, the set K is precisely the set of normalized images x composed of two connected components, x_1 and x_2 , with x constant on each component (See Figure 1a). Our question reduces to constructing a packing for pairs of blobs in $\mathbb{R}^{n \times n}$. We restrict our attention to blobs with the pictorial representation of Figure 1b. Naturally, our packing is "non-exhaustive", as it aims at enumerating a large number of images (and not all images) that reside in K .

Note that each entry of our blobs has magnitude $1/n$. Each of the $q := n(n-2)/3$ question marks can take the value $1/n$ or $-1/n$. Thus, our problem reduces to constructing a packing of the Hamming cube

$$\tilde{K} := \frac{1}{n} \{+1, -1\}^q.$$

Such packings are well known. We use a random construction.

We pack \tilde{K} by randomly picking a number of points in it and showing that with nonzero probability each pairwise distance is at least $1/2$. Thus, let x and x' be two points picked uniformly at random from \tilde{K} . Note that $\|n \cdot x - n \cdot x'\|_2^2 \sim 4 \cdot \text{Binomial}(q, 1/2)$. It follows from Hoeffding's inequality that

$$\begin{aligned} \mathbb{P}\left\{\|x - x'\|_2^2 < \frac{q}{d}\right\} &= \mathbb{P}\left\{\|n \cdot x - n \cdot x'\|_2^2 < q\right\} \\ &\leq \exp(-q/8). \end{aligned} \quad (\text{II.2})$$

We conclude the construction of the packing by applying Lemma II.1. In using the lemma, note that we assumed $d \geq 64$, and thus $q = n(n-2)/3 \geq d/4$. ■

B. Gaussian Ω

We now construct a packing when Ω is Gaussian.

In this section we take $\Omega \in \mathbb{R}^{p \times d}$ to have *i.i.d.* standard normal entries. Note that with probability 1 the rows of Ω are in general position and thus

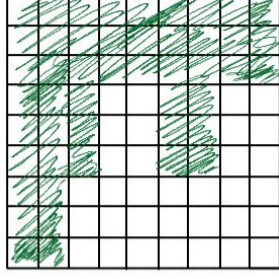
$$K_b := \bigcup_{|T|=p-(d-b)} K_T.$$

Let $K := K_1 \cap S^{d-1}$. We construct a packing for K .

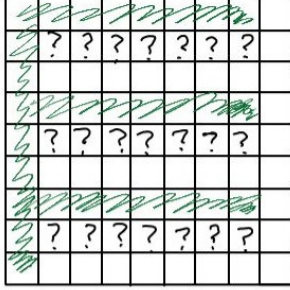
Lemma II.3.

$$P(K, 1/2) \geq \frac{1}{\sqrt{3}} \exp\left(\frac{d-1}{8} \cdot \left(1 - \frac{d-2}{p}\right)\right).$$

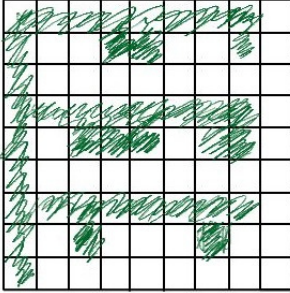
Proof: As above, we randomly construct our packing. We will pick a series of points in K as follows. First, pick $d-1$ rows of Ω uniformly at random indexed by the set $\Lambda \in [p]$. Take x which satisfies $\Omega_\Lambda = 0$. This



(a) A point in K_2 . All green squares have one value and all white squares have another.



(b) Visualization of packing patterns. Green corresponds with $1/n$, white corresponds with $-1/n$, and ? can be either.



(c) One possible point in the packing.

Fig. 1

restricts \mathbf{x} to a 1-dimensional space. Since \mathbf{x} must be on the unit sphere, there are only two possible points. Pick one at random.

Take \mathbf{x} and \mathbf{x}' to be two randomly generated points; we must show that they are far apart with high probability. Let Λ and Λ' be the corresponding rows of Ω , so that $\Omega_{\Lambda}\mathbf{x} = 0$ and $\Omega_{\Lambda'}\mathbf{x}' = 0$. Note that \mathbf{x} and \mathbf{x}' are both drawn uniformly from the sphere, but they are not independent because Λ and Λ' may have some intersection. The proof will follow by controlling their

dependence. First, we note that Λ and Λ' do not have an overly large intersection (with high probability).

Lemma II.4 (Bounding $|\Lambda \cap \Lambda'|$).

$$\mathbb{P}\left\{|\Lambda \cap \Lambda'| \geq (d-1) \cdot \left(\frac{d+p}{2p}\right)\right\} \leq \exp\left(-\frac{(d-1)(p-d+2)}{2p}\right) \quad (\text{II.3})$$

Proof: An application of Corollary 1.1 in [45] (an extension of Hoeffding's inequality for sampling without replacement [46]) gives

$$\mathbb{P}\left\{|\Lambda \cap \Lambda'| \geq \frac{(d-1)^2}{p} + t\right\} \leq \exp\left(-\frac{2t^2}{(d-1)(1 - \frac{d-2}{p})}\right). \quad (\text{II.4})$$

The lemma follows by taking $t = (d-1)(p-d+2)/(2p)$. ■

Now condition on Λ, Λ' , and $\Omega_{\Lambda \cap \Lambda'}$. Both \mathbf{x} and \mathbf{x}' must satisfy $\Omega_{\Lambda \cap \Lambda'}\mathbf{x} = 0$, thus reducing the dimension of the space they live in to $d - |\Lambda \cap \Lambda'|$. Set $T := \Lambda \setminus \Lambda'$ and $T' := \Lambda' \setminus \Lambda$. Note that Ω_T and $\Omega_{T'}$ are independent. Thus, by the rotational invariance of the Gaussian distribution, \mathbf{x} and \mathbf{x}' are distributed uniformly at random in the orthogonal complement to $\text{span}(\Omega_{\Lambda \cap \Lambda'})$. The distance between \mathbf{x} and \mathbf{x}' is equal in distribution to the distance between two points chosen uniformly at random on S^{q-1} where $q := d - |\Lambda \cap \Lambda'|$. Let $z := \|\mathbf{x} - \mathbf{x}'\|_2$. Note that the geodesic distance between \mathbf{x} and \mathbf{x}' is equal to $2 \arcsin(z/2)$. The distribution of z does not change if we fix \mathbf{x}' , and thus, the probability that $z \leq 1/2$ is precisely the normalized measure of a spherical cap with geodesic radius $2 \arcsin(1/2) \leq 0.52$. Bounds for this quantity are well known (see [47][Theorem 1.1]) giving

$$\mathbb{P}\left\{\|\mathbf{x} - \mathbf{x}'\|_2 \leq \frac{1}{2}\right\} \leq 2 \exp\left(-\frac{q}{2}\right). \quad (\text{II.5})$$

We only need to bound q , but this is done in Lemma II.4. Let E be the good event that $|\Lambda \cap \Lambda'| \leq (d-1) \cdot \left(\frac{d+p}{2p}\right)$. By Lemma II.4, $\mathbb{P}\{E\} \geq 1 - \exp\left(-\frac{(d-1)(p-d+2)}{2p}\right)$. On the event E , we have $q \geq d - (d-1) \cdot \left(\frac{d+p}{2p}\right) = 1 + (d-1) \cdot \left(\frac{p-d}{2p}\right)$.

Putting these pieces together, we get

$$\begin{aligned}
& \mathbb{P} \left\{ \|\mathbf{x} - \mathbf{x}'\|_2 \leq \frac{1}{2} \right\} \\
& \leq \mathbb{P} \left\{ \|\mathbf{x} - \mathbf{x}'\|_2 \leq \frac{1}{2} \mid E \right\} + \mathbb{P} \{E^c\} \\
& \leq 2 \exp \left(-\frac{1}{2} - \frac{(d-1)(p-d)}{4p} \right) \\
& \quad + \exp \left(-\frac{(d-1)(p-d+2)}{2p} \right) \\
& \leq 3 \exp \left(-\frac{(d-1)(p-d+2)}{4p} \right),
\end{aligned}$$

where in the last line we have used the fact that $\frac{2(d-1)}{4p} \leq \frac{1}{2}$.

Complete the proof by applying Lemma II.1 with $\delta = 1/2$ and η as in the last line of the above equation. ■

C. Implications of Set Packing

Consider a vector \mathbf{x} which is known to reside in a set K . We show that if K admits a large packing, then \mathbf{x} cannot be robustly reconstructed from few linear measurements by any method. The proof proceeds by showing that the distance between some pair of points in the packing will be reduced immensely when subsampling, and thus the corresponding two points in K will be nearly indistinguishable amid noise.

We need the following lemma in the proof. This lemma is a classical result about packing numbers. In this lemma and the ones to follow, we denote by B^n ($\subset \mathbb{R}^n$) the ℓ_2 ball of radius 1 centered at the origin.

Lemma II.5 (Minimum distance in a packing). *Let $\mathcal{X} \subset B^m$ be a finite set of points. Then*

$$\min_{\substack{\mathbf{x} \neq \mathbf{y} \\ \mathbf{x}, \mathbf{y} \in \mathcal{X}}} \|\mathbf{x} - \mathbf{y}\|_2 \leq \frac{4}{|\mathcal{X}|^{1/m}}.$$

The proof is a simple and classical volumetric argument, see [48].

The following lemma begins to address the problem of signal estimation.

Lemma II.6. *Let $\mathcal{X} \subset B^d$ be a finite set of points. Suppose $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$ for some $\mathbf{x} \in \mathcal{X}$, $\mathbf{A} \in \mathbb{R}^{m \times d}$ with $\|\mathbf{A}\| \leq 1$, and $\mathbf{z} \sim \mathcal{N}(0, \sigma^2 \cdot \text{Id})$. Assume*

$$\frac{4}{|\mathcal{X}|^{1/m}} \leq \sigma.$$

Let $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{y})$ be any estimator of \mathbf{x} . Then,

$$\min_{\mathbf{x} \in \mathcal{X}} \mathbb{P} \{ \hat{\mathbf{x}} = \mathbf{x} \} \leq \frac{3}{4}. \quad (\text{II.6})$$

Proof: We will lower-bound the worst-case probability of error by the probability of error under a suitably unfavorable prior. (This reduction from minimax to Bayesian is a standard trick, see, for example, [49, Equation 2]).

First, apply the packing bound (Lemma II.5) to $\mathcal{A}\mathcal{X}$ to show that there are some two points $\mathbf{A}\mathbf{x}_1, \mathbf{A}\mathbf{x}_2 \in \mathcal{A}\mathcal{X}$ satisfying

$$\|\mathbf{A}\mathbf{x}_1 - \mathbf{A}\mathbf{x}_2\|_2 := \varepsilon \leq \frac{4}{|\mathcal{X}|^{1/m}} \leq \sigma.$$

Consider the prior distribution which picks \mathbf{x}_1 with probability $1/2$ and \mathbf{x}_2 with probability $1/2$. For any given estimator, the worst-case probability of error (the left-hand side of Equation (II.6)) is lower-bounded by the probability of error under this prior. This is further minimized by the *Bayes Estimator* which chooses \mathbf{x}_1 or \mathbf{x}_2 based on which has the highest posterior probability conditional on \mathbf{y} . The Bayes estimator simply takes

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}_1, \mathbf{x}_2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2.$$

It is straightforward to show that this estimator satisfies,

$$\begin{aligned}
\mathbb{P} \{ \hat{\mathbf{x}} = \mathbf{x} \} &= \mathbb{P} \left\{ \mathcal{N}(0, 1) \leq \frac{\varepsilon}{2\sigma} \right\} \\
&\leq \mathbb{P} \left\{ \mathcal{N}(0, 1) \leq \frac{1}{2} \right\} \leq \frac{3}{4}.
\end{aligned} \quad (\text{II.7})$$

The following proposition is the synthesized tool relating packings to minimax error that we will use to prove our main theorems.

Proposition II.7. *Let $K \subset \mathbb{R}^d$ be a cone. Let \mathcal{X} be a δ -packing of $K \cap B^d$. Suppose $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$ for $\mathbf{x} \in K$, $\mathbf{A} \in \mathbb{R}^{m \times d}$ with $\|\mathbf{A}\| \leq 1$, and $\mathbf{z} \sim \mathcal{N}(0, \sigma^2 \cdot \text{Id})$. Then for any estimator $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{y})$, we have*

$$\sup_{\mathbf{x} \in K} \mathbb{E} \|\hat{\mathbf{x}} - \mathbf{x}\|_2 \geq \frac{\delta \sigma |\mathcal{X}|^{1/m}}{32}.$$

Proof: We begin by rescaling the problem so that the noise level is just large enough that a signal in \mathcal{X} will be hard to recover, i.e., so that we may use Lemma II.6. Let

$$\lambda := \frac{4}{|\mathcal{X}|^{1/m} \sigma}$$

and set $\tilde{\mathbf{y}} = \lambda \mathbf{y}$, $\tilde{\mathbf{x}} = \lambda \mathbf{x}$, and $\tilde{\mathbf{z}} = \lambda \mathbf{z}$. Thus, $\tilde{\mathbf{y}} = \mathbf{A}\tilde{\mathbf{x}} + \tilde{\mathbf{z}}$ and $\tilde{\mathbf{x}}, \tilde{\mathcal{X}}$ satisfy the conditions of Lemma II.6. Note also that $\tilde{\mathbf{x}} \in \lambda K = K$. We further restrict $\tilde{\mathbf{x}}$ to lie in \mathcal{X} .

Now, by Lemma II.6, for any estimator $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\mathbf{y})$,

$$\min_{\tilde{\mathbf{x}} \in \mathcal{X}} \mathbb{P} \{ \hat{\mathbf{x}}(\mathbf{y}) = \tilde{\mathbf{x}} \} \leq 3/4.$$

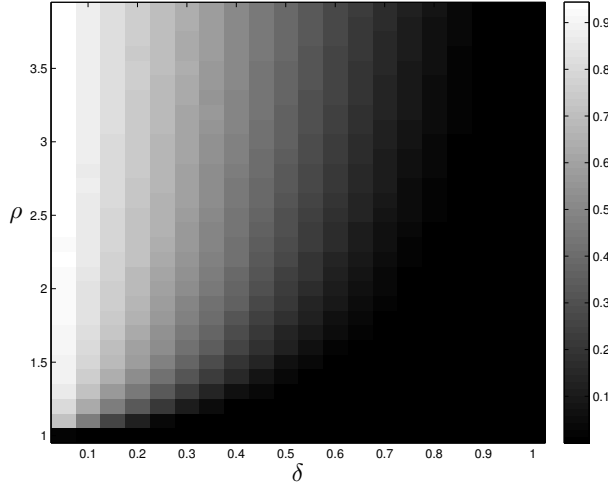


Fig. 2: Recovery error of ℓ_1 -minimization with Gaussian analysis dictionary in the noiseless case. Experiment setup: $d = 200$, $\delta = \frac{m}{d}$ and $\rho = \frac{p}{d}$. For each configuration we average over 500 realizations. Color attribute: Mean Squared Error.

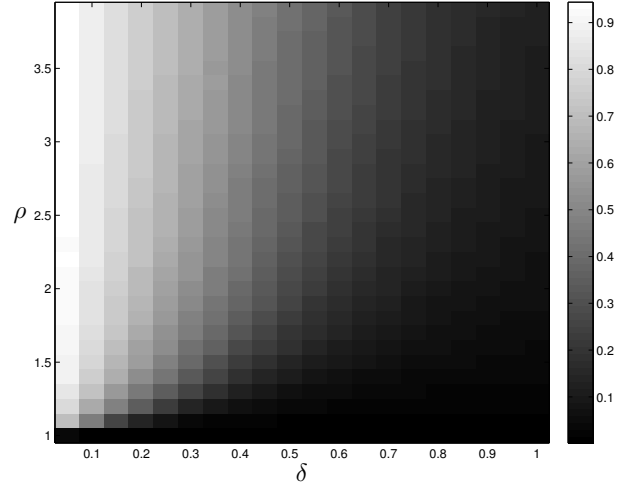


Fig. 3: Recovery error of ℓ_1 -minimization with Gaussian analysis dictionary in the noisy case with $\sigma = 0.01$. Experiment setup: $d = 200$, $\delta = \frac{m}{d}$ and $\rho = \frac{p}{d}$. For each configuration we average over 500 realizations. Color attribute: Mean Squared Error.

Since no estimator can reliably guess \tilde{x} on a δ -packing, it follows that no estimator can estimate \tilde{x} to accuracy better than $\delta/2$ with high probability. Otherwise, such an estimator could be projected onto \mathcal{X} to make a reliable guess. In other words,

$$\min_{\tilde{x} \in \mathcal{X}} \mathbb{P} \{ \|\hat{x}(y) - \tilde{x}\|_2 < \delta/2 \} \leq 3/4.$$

This implies that

$$\max_{\tilde{x} \in \mathcal{X}} \mathbb{E} \|\hat{x} - \tilde{x}\|_2 \geq \frac{\delta}{8}.$$

Divide both sides of the equation by λ to undo the scaling and finish the proof. ■

Remark II.8 (Interpretation in terms of complexity, or metric dimension, of K). *While the manifold dimension of a signal set determines signal recoverability in the noiseless case [31], it can fail to characterize the noisy case. Instead, a classical metric notion of dimension, following ideas of Kolmogorov and Le Cam, provides a more apt characterization. Indeed, set $D(K) := \log(P(K \cap B^d, 1/2))$. Le Cam [50] showed that $D(K)$ is a effective metric characterisation of the complexity (or dimension) of the set K in regards to many point estimation problems. As a simple example, if K is a q -dimensional subspace it is well-known that $D(K)$ is proportional to q , just like the manifold dimension. However, in contrast to the manifold dimension, this metric definition takes into account the geometry of the set, thus*

allowing characterization of signal recoverability amid noise. The above proposition states that signal recovery error is at least proportional to $\exp(D(K)/m)$, i.e., if the number of measurements is below the effective dimension, the error amid noise blows up exponentially fast as a function of the ratio.

D. Putting it together

Theorem I.1 now follows by combining the packing number of Lemma II.2 with Proposition II.7. Theorem I.2 follows by combining Lemma II.3 with Proposition II.7.

III. EXPERIMENTS

To demonstrate the results of the theorems, in this section we look at the performance of analysis ℓ_1 -minimization

$$\min_{x'} \|\Omega x'\|_1 \quad \text{s.t.} \quad \|y - Ax'\|_2 \leq \sqrt{m}\sigma, \quad (\text{III.1})$$

in recovering signals with low dimensionality and different cosparsity levels. In all the experiments the measurement matrix A is a random Gaussian matrix with normalized columns.

In the first experiment, we select Ω to be a random Gaussian matrix and the signal x to be a Gaussian random vector projected to a one dimensional subspace orthogonal to randomly selected $d - 1$ rows from Ω . In Fig. 2 we present the recovery performance in the

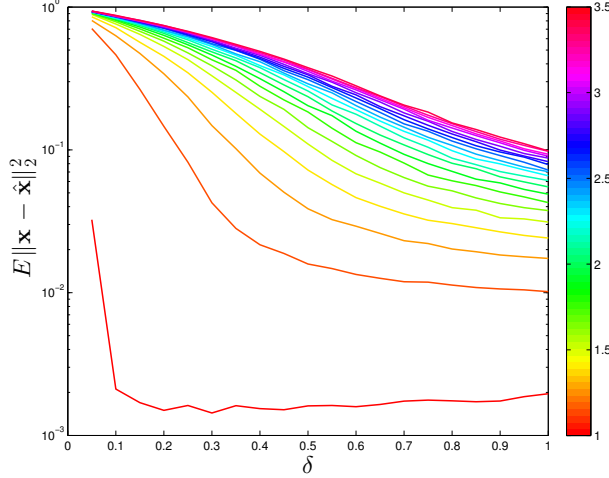


Fig. 4: Recovery error of ℓ_1 -minimization with Gaussian analysis dictionary in the noisy case with $\sigma = 0.01$ as a function of $\delta = \frac{m}{d}$ for different selections of $\rho = \frac{p}{d}$. The signal dimension is $d = 200$ and we average over 500 realizations. Color attribute: The color of each graph corresponds to the value of ρ . The bottom graph corresponds to $\rho = 1$ and the upper one to $\rho = 3.5$.

noiseless case for a fixed signal ambient dimension $d = 200$ and different combinations of the sampling rate $\delta = \frac{m}{d}$ and the redundancy ratio $\rho = \frac{p}{d}$. Interestingly, we observe empirically that the theoretical instability to noise (Theorem I.2) also implies instability to ℓ_1 relaxation. Indeed, notice that though the manifold dimension of the signal is equal to 1 in all the experiments, the success in recovery heavily depends on p , which changes only the cosparsity of the signal. As expected from the theory, as soon as p increases to be slightly larger than d , the number of measurements needed to reconstruct the signal increases enormously.

In Fig 3 we present the reconstruction error in the noisy case when an additive random white Gaussian noise with standard deviation $\sigma = 0.01$ is added to the measurements. Notice that the error is saturated by 1, the signal energy, as when the noise is very large the best estimator is the zero estimator, for which the error equals to the signal energy.

As predicted from the theorem, when $\rho > 1$ the recovery error grows exponentially as m , the number of measurements, decreases. This becomes clearer for larger values of ρ . To show it more clearly we present in Fig. 4 the recovery error as a function of the sampling rate, which is a function of m . The bottom graph corresponds to $\rho = 1$, where the manifold dimension equals d minus the cosparsity, for which we get a good recovery

almost for all values of m . This is not surprising because when $d = m$, Ω is invertible and the analysis cosparsity model may be recast into the standard synthesis model, in which one expects $O(\log(p))$ measurements to suffice. Indeed, we see that already when we have ten percent of the measurements (corresponds to $20 > 2 \log(p) \simeq 10.6$ measurements) we get a very good recovery. However, the behavior is quite different as soon as ρ increases to slightly greater than 1. The rest of the graphs are above it and ordered according to increasing values of ρ . As ρ becomes larger the error increases and its behavior as a function of m becomes more and more exponential.

In the second experiment we consider the 2D-DIF operator. Notice that for this operator we may have different cosparsity levels for the same manifold dimension. We consider vectors of size $d = 144$ that represent two dimensional images of size 12×12 with manifold dimension two. An image of dimension two is an image with two connected components, each with a different gray value. We generate randomly such images with different cosparsity levels. The values in the first and second connected components are selected randomly from the ranges $[0, 1]$ and $[-1, 0]$ respectively. Note that the cosparsity level defines the length of the edge in the image. The images are generated by setting all the pixels in the image to a value from the range $[0, 1]$ and then picking one pixel at random and starting a random walk (from its location) that assigns to all the pixels in its path a value from the range $[-1, 0]$ (the same value). The random walk stops once it gets to a pixel that it has visited before. With high probability the resulted image will be with only two connected components. We generate many images like that and sort them according to their cosparsity (eliminating images that have more than two connected components). Note that the larger the cosparsity the shorter the edge.

We test the reconstruction performance for different cosparsity levels and sampling rates. The recovery error in the noiseless and noisy cases are presented in Figs. 5 and 6. It can be clearly seen that the recovery performance in both cases is determined by the cosparsity level and not the manifold dimension of the signal which is fixed in all the experiments. As predicted by Theorem I.1, if we rely only on the manifold dimension we get a very unstable recovery. However, if we take into account also the cosparsity level we can have a better prediction of our success rate. As we have seen in the previous experiment, also here we can see that instability in the noisy case also implies instability to relaxation.

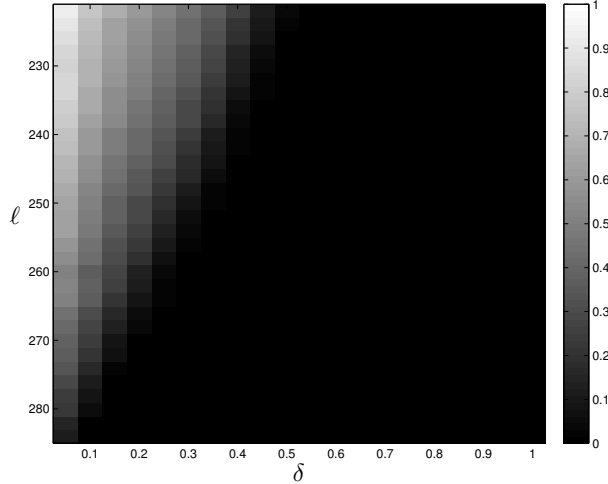


Fig. 5: Recovery error of ℓ_1 -minimization with the 2D-DIF analysis dictionary in the noiseless case. Experiment setup: $d = 144$, $\delta = \frac{m}{d}$ and ℓ is the cosparsity level. For each configuration we average over 500 realizations. Color attribute: Mean Squared Error.

IV. CONCLUSION

In this work we have inquired whether it is possible to provide recovery guarantees for compressed sensing with signals from the cosparsity analysis framework by only having information about their manifold dimension. Though the answer for this question is positive for standard compressed sensing (with the standard sparsity model) and the matrix completion problem, we have shown that this is not the case here. We have demonstrated this with two analysis dictionaries, the Gaussian matrix and the 2D-DIF operator, both in theory and simulations. Our conclusion is that in the cosparsity analysis framework the “correct” measure to use for predicting the recovery success of any tractable method is the cosparsity of the signal (number of zeros) and not the manifold dimension.

It would be interesting to check whether it is possible to carry over the results in this paper to non-subspace models such as curved manifolds. Notice that some theoretical guarantees have already been given for the latter case [32], [33], showing that it is possible to recover a b -dimensional submanifold from a number of samples proportional to b . However, these results also depend on a number of other quantities, such as the volume and condition number of the manifold. Therefore, another direction would be to see whether adding similar assumptions to the analysis model would allow sampling in the manifold dimension.

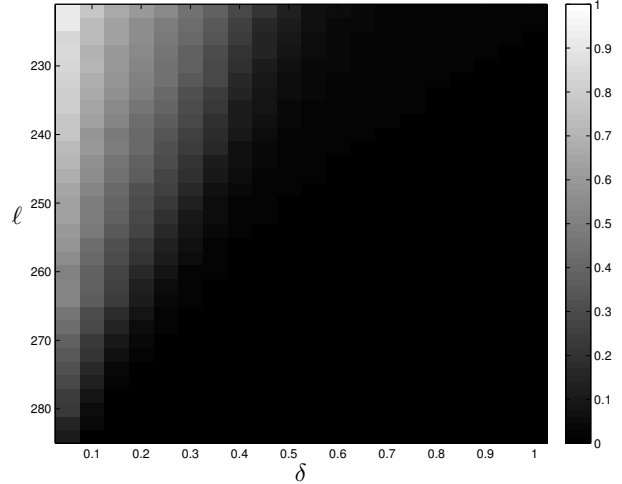


Fig. 6: Recovery error of ℓ_1 -minimization with the 2D-DIF dictionary in the noisy case with $\sigma = 0.01$. Experiment setup: $d = 144$, $\delta = \frac{m}{d}$ and ℓ is the cosparsity level. For each configuration we average over 500 realizations. Color attribute: Mean Squared Error.

An additional open question raised by this work is whether instability in the noisy case leads to instability to relaxation. Indeed, we have observed this phenomenon empirically in the experiments and therefore believe that there is a room to prove such a result.

ACKNOWLEDGMENT

The authors thank the reviewers of the manuscript for their suggestions which improved the paper.

REFERENCES

- [1] A. M. Bruckstein, D. L. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.
- [2] J. Romberg, “Imaging via compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 14–20, March 2008.
- [3] R. M. Willett, R. F. Marcia, and J. M. Nichols, “Compressed sensing for practical optical imaging systems: a tutorial,” *Optical Engineering*, vol. 50, no. 7, pp. 072 601–072 601–13, 2011.
- [4] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer Publishing Company, Incorporated, 2010.
- [5] Y. Lu and M. Do, “A theory for sampling signals from a union of subspaces,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2334–2345, Jun. 2008.
- [6] T. Blumensath and M. Davies, “Sampling theorems for signals from the union of finite-dimensional linear subspaces,” *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, April 2009.
- [7] Y. C. Eldar and G. Kutyniok, *Compressed Sensing Theory and Applications*, 1st ed. Cambridge University Press, 2012.
- [8] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, 1st ed. Springer Publishing Company, Incorporated, 2013.

- [9] Y. Plan, R. Vershynin, and E. Yudovina, "High-dimensional estimation with geometric constraints," *CoRR*, vol. abs/1404.3749, 2014.
- [10] R. Vershynin, "Estimation in high dimensions: a geometric perspective," *CoRR*, vol. abs/1405.5103, 2014.
- [11] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [12] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [13] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.
- [14] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [15] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, dec. 2005.
- [16] —, "The Dantzig selector: Statistical estimation when p is much larger than n ," *Annals Of Statistics*, vol. 35, p. 2313, 2007.
- [17] D. L. Donoho and M. Elad, "Optimal sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Science*, vol. 100, pp. 2197–2202, Mar 2003.
- [18] R. Giryes and M. Elad, "Can we allow linear dependencies in the dictionary in the synthesis framework?" in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 5459 – 5463.
- [19] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, pp. 227–234, 1995.
- [20] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [21] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, pp. 3397–3415, 1993.
- [22] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, Apr. 2010.
- [23] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301 – 321, May 2009.
- [24] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [25] T. Blumensath and M. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265 – 274, 2009.
- [26] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM J. Numer. Anal.*, vol. 49, no. 6, pp. 2543–2563, 2011.
- [27] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and gaussian measurements," in *Information Sciences and Systems, 2006 40th Annual Conference on*, 22-24 2006, pp. 207–212.
- [28] P. Bickel, Y. Ritov, and A. Tsybakov, "Simultaneous analysis of lasso and dantzig selector," *Annals of Statistics*, vol. 37, no. 4, pp. 1705–1732, 2009.
- [29] R. Giryes and M. Elad, "RIP-based near-oracle performance guarantees for SP, CoSaMP, and IHT," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1465–1468, March 2012.
- [30] E. Candès and Y. Plan, "Matrix completion with noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, June 2010.
- [31] Y. Eldar, D. Needell, and Y. Plan, "Uniqueness conditions for low-rank matrix recovery," *Applied and Computational Harmonic Analysis*, vol. 33, no. 2, pp. 309 – 314, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1063520312000528>
- [32] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of computational mathematics*, vol. 9, no. 1, pp. 51–77, 2009.
- [33] M. B. Wakin, "Manifold-based signal recovery and parameter estimation from compressive measurements," *arXiv preprint arXiv:1002.1247*, 2010.
- [34] H. L. Yap, M. B. Wakin, and C. J. Rozell, "Stable manifold embeddings with operators satisfying the restricted isometry property," in *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*. IEEE, 2011, pp. 1–6.
- [35] A. Eftekhar and M. B. Wakin, "New analysis of manifold embeddings and signal recovery from compressive measurements," *arXiv preprint arXiv:1306.4748*, 2013.
- [36] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, June 2007.
- [37] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The cosparse analysis model and algorithms," *Appl. Comput. Harmon. Anal.*, vol. 34, no. 1, pp. 30 – 56, 2013.
- [38] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM Journal on Imaging Sciences*, vol. 6, no. 2, pp. 1035–1058, 2013.
- [39] J. M. Lee, *Smooth manifolds*. Springer, 2003.
- [40] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59 – 73, 2011.
- [41] Y. Liu, T. Mi, and S. Li, "Compressed sensing with general frames via optimal-dual-based ℓ_1 -analysis," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4201–4214, 2012.
- [42] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. Davies, "Greedy-like algorithms for the cosparse analysis model," *Linear Algebra and its Applications*, vol. 441, no. 0, pp. 22 – 60, 2014, special Issue on Sparse Approximate Solution of Linear Systems.
- [43] M. Kabanava and H. Rauhut, "Analysis ℓ_1 -recovery with frames and Gaussian measurements," *Preprint*, 2013.
- [44] R. Giryes, "A greedy algorithm for the analysis transform domain," to appear in *Neurocomputing*, 2015.
- [45] R. J. Serfling, "Probability inequalities for the sum in sampling without replacement," *Annals Of Statistics*, vol. 2, no. 1, pp. 39–48, 1974.
- [46] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 13–30, 1963.
- [47] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*. Springer-Verlag, 1991.
- [48] G. Pisier, *The volume of convex bodies and Banach space geometry*. Cambridge University Press, 1999, vol. 94.
- [49] A. Guntuboyina, "Lower bounds for the minimax risk using divergences, and applications," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2386–2399, 2011.
- [50] L. Le Cam, *Asymptotic methods in statistical decision theory*. Springer New York, 1986.