

Analysis ℓ_1 -recovery with frames and Gaussian measurements

Holger Rauhut*, Maryia Kabanava†

November 3, 2014

Abstract

This paper provides novel results for the recovery of signals from undersampled measurements based on analysis ℓ_1 -minimization, when the analysis operator is given by a frame. We both provide so-called uniform and nonuniform recovery guarantees for cosparsely (analysis-sparse) signals using Gaussian random measurement matrices. The nonuniform result relies on a recovery condition via tangent cones and the uniform recovery guarantee is based on an analysis version of the null space property. Examining these conditions for Gaussian random matrices leads to precise bounds on the number of measurements required for successful recovery. In the special case of standard sparsity, our result improves a bound due to Rudelson and Vershynin concerning the exact reconstruction of sparse signals from Gaussian measurements with respect to the constant and extends it to stability under passing to approximately sparse signals and to robustness under noise on the measurements.

Keywords: compressive sensing, ℓ_1 -minimization, analysis regularization, frames, Gaussian random matrices.

1 Introduction

Compressive sensing [12, 6, 17] is a recent field that has seen enormous research activity in the past years. It predicts that certain signals (vectors) can be recovered from what was previously believed to be incomplete information using efficient reconstruction methods. Applications of this principle range from magnetic resonance imaging over radar and remote sensing to astronomical signal processing and more. The key assumption of the theory is that the signal to be recovered is sparse or can at least be well-approximated by a sparse one. Most research activity so far has been dedicated to the synthesis sparsity model where one assumes that the signal can be written as a linear combination of only a small number of elements from a basis, or more generally an overcomplete frame. In certain situations, however, it turns out to be more efficient to work with an analysis-based sparsity model. Here, one rather assumes that the application of a linear map yields a vector with a large number of zero entries. While the synthesis and the analysis model are equivalent in special cases, they are very distinct in an overcomplete case. By now, comparably few investigations have been dedicated to the analysis sparsity model and its rigorous understanding is still in its infancy.

*RWTH Aachen University, Lehrstuhl C für Mathematik (Analysis), Templergraben 55, 52062 Aachen Germany, rauhut@mathc.rwth-aachen.de

†RWTH Aachen University, Lehrstuhl C für Mathematik (Analysis), Templergraben 55, 52062 Aachen Germany, rauhut@mathc.rwth-aachen.de

The analysis based sparsity model and corresponding reconstruction methods were introduced systematically in recent work of Nam et al. [29]. Nevertheless we note that it appeared also in earlier works, see e.g. [7]. In particular, the popular method of total variation minimization [8, 30] in image processing is closely related to analysis based sparsity with respect to a difference operator. An estimate of the number of Gaussian measurements for successful recovery via total variation minimization has been recently obtained in [22].

In this paper we consider the analysis based sparsity model for the important case that the analysis transform is given by inner products with respect to a possibly redundant frame. As reconstruction method we study a corresponding analysis ℓ_1 -minimization approach. Furthermore, we assume that the linear measurements are obtained via an application of a Gaussian random matrix. The main results of this paper provide precise estimates of the number of measurements required for the reconstruction of a signal whose analysis representation has a given number of zero elements. Moreover, stability estimates are given. An alternative bound on the number of measurements can be found in [22].

1.1 Problem statement and main results

We consider the task of reconstructing a signal $\mathbf{x} \in \mathbb{R}^d$ from incomplete and possibly corrupted measurements given by

$$\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{m \times d}$ with $m \ll d$ is the measurement matrix and \mathbf{w} corresponds to noise. Since this system is underdetermined it is impossible to recover \mathbf{x} from \mathbf{y} without additional information, even when $\mathbf{w} = \mathbf{0}$.

As already mentioned, the underlying assumption in compressive sensing is sparsity. The *synthesis sparsity prior* assumes that \mathbf{x} can be represented as a linear combination of a small number of elements of a dictionary $\mathbf{D} \in \mathbb{R}^{d \times n}$, i.e.,

$$\mathbf{x} = \mathbf{D}\alpha, \quad \alpha \in \mathbb{R}^n,$$

where the number of non-zero elements of α , denoted by $\|\alpha\|_0$, is considerably less than n . Often \mathbf{D} is chosen as a unitary matrix, which refers to sparsity of \mathbf{x} in an orthonormal basis. Unfortunately, the approach to recover α , or \mathbf{x} respectively, from $\mathbf{y} = \mathbf{M}\mathbf{x} = \mathbf{M}\mathbf{D}\alpha$ (assuming the noiseless case for simplicity) via ℓ_0 -minimization, i.e.,

$$\min_{\alpha \in \mathbb{R}^n} \|\alpha\|_0 \quad \text{subject to } \mathbf{M}\mathbf{D}\alpha = \mathbf{y},$$

is NP-hard in general. A by-now well-studied tractable alternative is the ℓ_1 -minimization approach of finding the minimizer α^* of

$$\min_{\alpha \in \mathbb{R}^n} \|\alpha\|_1 \quad \text{subject to } \mathbf{M}\mathbf{D}\alpha = \mathbf{y} \tag{2}$$

The restored signal is then given by $\mathbf{x}^* = \mathbf{D}\alpha^*$. This optimization problem is referred to as basis pursuit [10]. In the noisy case, one passes to

$$\min_{\alpha \in \mathbb{R}^n} \|\alpha\|_1 \quad \text{subject to } \|\mathbf{M}\mathbf{D}\alpha - \mathbf{y}\|_2 \leq \eta, \tag{3}$$

where η corresponds to an estimate of the noise level.

The *analysis sparsity prior* assumes that \mathbf{x} is sparse in some transform domain, that is, given a matrix $\mathbf{\Omega} \in \mathbb{R}^{p \times d}$ – the so-called *analysis operator* – the vector $\mathbf{\Omega}\mathbf{x}$ is sparse. For instance, such operators can be generated by the discrete Fourier transform, the finite difference operator (related to total variation), wavelet [26, 32, 34], curvelet [5] or Gabor transforms [21].

Analogously to (2), a possible strategy for the reconstruction of analysis-sparse vectors (or cosparsity vectors, see below) is to solve the analysis ℓ_1 -minimization problem

$$\min_{\mathbf{z} \in \mathbb{R}^d} \|\mathbf{\Omega}\mathbf{z}\|_1 \quad \text{subject to} \quad \mathbf{M}\mathbf{z} = \mathbf{y}, \quad (4)$$

or, in the noisy case,

$$\min_{\mathbf{z} \in \mathbb{R}^d} \|\mathbf{\Omega}\mathbf{z}\|_1 \quad \text{subject to} \quad \|\mathbf{M}\mathbf{z} - \mathbf{y}\|_2 \leq \eta. \quad (5)$$

Both optimization problems can be solved efficiently using convex optimization techniques, see e.g. [3]. If $\mathbf{\Omega}$ is an invertible matrix, then these analysis ℓ_1 -minimization problems are equivalent to (2) and (3). However, in general the analysis ℓ_1 -minimization problems cannot be reduced to the standard ℓ_1 -minimization problems.

We note, that one may also pursue greedy or other iterative approaches for recovery, see e.g. [17] for an overview in the standard synthesis sparsity case and see e.g. [29] for the analysis sparsity case. However, we will concentrate on the above optimization approaches here.

In the remainder of this paper, we assume that the analysis operator is given by a frame. Put formally, let $\{\omega_i\}_{i=1}^p$, $\omega_i \in \mathbb{R}^d$, $p \geq d$, be a frame, i.e., there exist positive constants $A, B > 0$ such that for all $\mathbf{x} \in \mathbb{R}^d$

$$A\|\mathbf{x}\|_2^2 \leq \sum_{i=1}^p |\langle \omega_i, \mathbf{x} \rangle|^2 \leq B\|\mathbf{x}\|_2^2.$$

Its elements are collected as rows of the matrix $\mathbf{\Omega} \in \mathbb{R}^{p \times d}$. The analysis representation of a signal \mathbf{x} is given by the vector $\mathbf{\Omega}\mathbf{x} = \{\langle \omega_i, \mathbf{x} \rangle\}_{i=1}^p \in \mathbb{R}^p$. (We note that in the literature it is often common to collect the elements of a frame rather as columns of a matrix. However, for our purposes it is more convenient to collect them as rows.) The frame is called tight if the frame bounds coincide, i.e., $A = B$.

Cosparsity is now defined as follows.

Definition 1. Given an analysis operator $\mathbf{\Omega} \in \mathbb{R}^{p \times d}$, the cosparsity of $\mathbf{x} \in \mathbb{R}^d$ is defined as

$$l := p - \|\mathbf{\Omega}\mathbf{x}\|_0.$$

The index set of the zero entries of $\mathbf{\Omega}\mathbf{x}$ is called the cosupport of \mathbf{x} . If \mathbf{x} is l -cosparsity, then $\mathbf{\Omega}\mathbf{x}$ is s -sparse with $s = p - l$.

The motivation to work with the cosupport rather than the support in the context of analysis sparsity is that in contrast to synthesis sparsity, it is the location of the *zero*-elements which define a corresponding subspace. In fact, if Λ is the cosupport of \mathbf{x} , then it follows from Definition 1 that

$$\langle \omega_j, \mathbf{x} \rangle = 0, \quad \text{for all } j \in \Lambda.$$

Hence, the set of l -cosparsity signals can be written as

$$\bigcup_{\Lambda \subset [p]: \#\Lambda=l} W_\Lambda,$$

where W_Λ denotes the orthogonal complement of the linear span of $\{\omega_j : j \in \Lambda\}$.

In contrast to standard sparsity, there are often certain restrictions on the values that the cosparsity can take. In fact, in the generic case that the frame elements ω_j are in general position in \mathbb{R}^d , then every set of d rows of $\mathbf{\Omega}$ are linearly independent. Then the maximal number of zeros that can be achieved for a nontrivial vector \mathbf{x} in the analysis representation $\mathbf{\Omega}\mathbf{x}$ is less than d , since otherwise $\mathbf{x} = 0$. Thus, for the cosparsity l of any non-zero vector \mathbf{x} it holds $l < d$ in this case. Nevertheless, if there are linear dependencies among the frame elements ω_j , then larger values of l are possible. This applies to certain redundant frames as well as to the difference operator (related to total variation).

Our main results concern the minimal number m of measurements that are necessary to recover an l -cosparse vector \mathbf{x} from $\mathbf{y} = \mathbf{M}\mathbf{x}$ with $\mathbf{M} \in \mathbb{R}^{m \times d}$. As it is hard to come up with theoretical guarantees for deterministic matrices \mathbf{M} , we pass to random matrices. As common in compressive sensing, we work with Gaussian random matrices, that is, with matrices having independent standard normal distributed entries. Gaussian random matrices have already proven to provide accurate theoretical guarantees in the context of standard synthesis sparsity, see e.g. [9, 13]. Moreover, empirical tests indicate that also other types of random matrices behave very similar to Gaussian random matrices in terms of recovery performance [14], although a theoretical justification may be much harder than for Gaussian matrices.

We both provide so-called nonuniform and uniform recovery guarantees. The nonuniform result states that a given fixed cosparse vector \mathbf{x} is recovered via analysis ℓ_1 -minimization from $\mathbf{y} = \mathbf{M}\mathbf{x}$ with high probability using a random choice of a Gaussian measurement matrix \mathbf{M} under a suitable condition on the number of measurements. In contrast, the uniform result states that a single random draw of a Gaussian matrix \mathbf{M} is able to recover *all* cosparse signals \mathbf{x} simultaneously with high probability. Clearly, the uniform statement is stronger than the nonuniform one, however, as we will see, the uniform statement requires more measurements.

We start with the nonuniform guarantee for recovery of cosparse signals with respect to frames using Gaussian measurement matrices.

Theorem 1. *Let $\mathbf{\Omega} \in \mathbb{R}^{p \times d}$ be a frame with frame bounds $A, B > 0$ and let \mathbf{x} be l -cosparse, that is, $\mathbf{\Omega}\mathbf{x}$ is s -sparse with $s = p - l$. For a Gaussian random matrix $\mathbf{M} \in \mathbb{R}^{m \times d}$ and $0 < \varepsilon < 1$, if*

$$\frac{m^2}{m+1} \geq \frac{2Bs}{A} \left(\sqrt{\ln\left(\frac{ep}{s}\right)} + \sqrt{\frac{A \ln(\varepsilon^{-1})}{Bs}} \right)^2, \quad (6)$$

then with probability at least $1 - \varepsilon$, the vector \mathbf{x} is the unique minimizer of $\|\mathbf{\Omega}\mathbf{z}\|_1$ subject to $\mathbf{M}\mathbf{z} = \mathbf{M}\mathbf{x}$.

Roughly speaking, that is, ignoring terms of lower order, a fixed l -cosparse vector is recovered with high probability from

$$m > 2(B/A)s \ln(ep/s)$$

Gaussian measurements where $s = p - l$. Note that the number of measurements increases with increasing frame ratio B/A , and the optimal behavior occurs for tight frames. For $\mathbf{\Omega} = \mathbf{Id}$, this bound slightly strengthens the main result for sparse recovery in [9]. We will also show stability of the reconstruction with respect to noise on the measurements, see Theorem 6 below. The proof of the above result (given in Section 2) relies on a characterization of the minimizer via tangent cones (Theorems 3 and 4) which is similar to corresponding conditions stated in [9, 28]. Moreover, our proof uses an extension of the Gordon's escape through a mesh theorem (Theorem 5).

We now pass to the uniform recovery result which additionally takes into account that in practice the signals are often only approximately cosparse. The quantity

$$\sigma_s(\mathbf{\Omega}\mathbf{x})_1 := \inf \{ \|\mathbf{\Omega}\mathbf{x} - \mathbf{u}\|_1 : \mathbf{u} \text{ is } s\text{-sparse} \}$$

describes the ℓ_1 -best approximation error to $\mathbf{\Omega}\mathbf{x}$ by s -sparse vectors.

Theorem 2. *Let $\mathbf{M} \in \mathbb{R}^{m \times d}$ be a Gaussian random matrix, $0 < \rho < 1$ and $0 < \varepsilon < 1$. If*

$$\frac{m^2}{m+1} \geq \frac{2Bs(1+(1+\rho^{-1})^2)}{A} \left(\sqrt{\ln \frac{ep}{s}} + \frac{1}{\sqrt{2}} + \sqrt{\frac{A \ln(\varepsilon^{-1})}{Bs(1+(1+\rho^{-1})^2)}} \right)^2, \quad (7)$$

then with probability at least $1 - \varepsilon$ for every vector $\mathbf{x} \in \mathbb{R}^d$ a minimizer $\hat{\mathbf{x}}$ of $\|\mathbf{\Omega}\mathbf{z}\|_1$ subject to $\mathbf{M}\mathbf{z} = \mathbf{M}\mathbf{x}$ approximates \mathbf{x} with ℓ_2 -error

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2(1+\rho)^2}{\sqrt{A}(1-\rho)} \frac{\sigma_s(\mathbf{\Omega}\mathbf{x})_1}{\sqrt{s}}.$$

Roughly speaking, with high probability every l -cosparse vector can be recovered via analysis ℓ_1 -minimization using a single random draw of a Gaussian matrix if

$$m > 10(B/A)s \ln(ep/s). \quad (8)$$

Moreover, the recovery is stable under passing to approximately cosparse vectors when adding slightly more measurements. The proof of this theorem relies on an extension of the null space property, which is well known in the synthesis sparsity case [11, 17, 20] and was adapted to the analysis sparsity setting in [1, 16]. In fact, for the standard case $\mathbf{\Omega} = \mathbf{Id}$, we improve a result of Rudelson and Vershynin [33] (also relying on the null space property) with respect to the constants in (7) and add stability in ℓ_2 . We further show that recovery is robust under perturbations of the measurements in Theorem 11. We note, that in the standard exact sparse case with no noise, the constant 8 in (8) can be replaced by $2e$, see the contribution by Donoho and Tanner in [13]. Their methods, however, are completely different to ours, and it is not clear, whether they can be extended to analysis sparsity.

1.2 Related work

Let us discuss briefly related theoretical studies on recovery of analysis sparse vectors and compare them with our main results. An earlier version of Theorem 2 was shown by Candès and Needell in [7]. However, they were only able to treat the case that the analysis operator is given by a tight frame, that is, when $A = B$. Moreover, their analysis is based on a version of the restricted isometry property and does not provide explicit constants in the corresponding bound on the required number of measurements. To be fair, we note, however, that their analysis applies to general subgaussian random matrices. The results of [7] were extended to the case of non-tight frames and Weibull matrices in the work of Foucart in [16]. The analysis in [16] incorporates the robust null space property, the verification of which for the Weibull matrices relies on a variant of the classical restricted isometry property. In our work we prove that Gaussian random matrices satisfy the robust null space property by referring to a modification of the Gordon's escape through a mesh theorem.

A recent contribution by Needell and Ward [30] provides theoretical recovery guarantees for the special case of total variation minimization, which corresponds to analysis ℓ_1 -minimization with a certain difference operator. Unfortunately, we cannot cover this situation with our main results because the difference operator is not a frame. Nevertheless, it would be interesting to pursue theoretical recovery guarantees for total variation minimization and Gaussian random matrices using the approach of this paper.

Nam et al.'s work [29] provides a systematic introduction of the analysis sparsity model and treats also greedy recovery methods, see also [18]. Further contributions are contained in [25, 35].

Our nonuniform recovery guarantees rely on a geometric characterization of the successful recovery. We obtain quantitative estimates by bounding a certain Gaussian width which can be thought as an intrinsic complexity measure. The authors of [2] exploit the geometry of optimality conditions to study phase transition phenomena in random linear inverse problems and random demixing problems. They express their results in terms of the statistical dimension which is essentially equivalent to the Gaussian width, see Section 10.3 of [2] for further details.

Also, we note that the optimization problems (4) and (5) often appear in image processing [4, 8].

1.3 Notation

We use the notation $\mathbf{\Omega}_\Lambda$ to refer to a submatrix of $\mathbf{\Omega}$ with the rows indexed by Λ ; $(\mathbf{\Omega}\mathbf{x})_S$ stands for the vector whose entries indexed by S coincide with the entries of $\mathbf{\Omega}\mathbf{x}$ and the rest are filled by zeros. As we have already mentioned, the ℓ_0 -norm $\|\cdot\|_0$ of a vector corresponds to the number of non-zero elements in it. The unit ball in \mathbb{R}^d with respect to the ℓ_q -norm is denoted by B_q^d . The operator norm of a matrix \mathbf{A} is defined by $\|\mathbf{A}\|_{2\rightarrow 2} := \sup_{\|x\|_2 \leq 1} \|\mathbf{A}\mathbf{x}\|_2$ and the Frobenius norm is given by

$$\|\mathbf{A}\|_F := \left(\sum_{i=1}^m \sum_{j=1}^d |A_{ij}|^2 \right)^{1/2}.$$

It is well-known that Frobenius norm dominates the operator norm, $\|\mathbf{A}\|_{2\rightarrow 2} \leq \|\mathbf{A}\|_F$. Finally, $[p]$ is the set of all natural numbers not exceeding p , i.e., $[p] = \{1, 2, \dots, p\}$.

2 Nonuniform recovery

In this section we prove Theorem 1 and extend it to robust recovery in Theorem 6. The proof strategy is similar as in [9]. We rely on conditions on the measurement matrix \mathbf{M} involving tangent cones, which should be of independent interest. In order to check these conditions for a Gaussian random matrix we rely on an extension of the Gordon's escape through a mesh theorem. (In contrast to [9], the standard version of the Gordon's result is not sufficient for our purposes.)

2.1 Recovery via tangent cones

Our conditions for successful recovery of cosparsely signals are formulated via tangent cones. For fixed $\mathbf{x} \in \mathbb{R}^d$ we define the convex cone

$$T(\mathbf{x}) = \text{cone}\{\mathbf{z} - \mathbf{x} : \mathbf{z} \in \mathbb{R}^d, \|\mathbf{\Omega}\mathbf{z}\|_1 \leq \|\mathbf{\Omega}\mathbf{x}\|_1\},$$

where the notation ‘‘cone’’ stands for the conic hull of the indicated set. The following result is analogous to Proposition 2.1 in [9].

Theorem 3. *Let $\mathbf{M} \in \mathbb{R}^{m \times d}$. A vector $\mathbf{x} \in \mathbb{R}^d$ is the unique minimizer of $\|\boldsymbol{\Omega}\mathbf{z}\|_1$ subject to $\mathbf{M}\mathbf{z} = \mathbf{M}\mathbf{x}$ if and only if $\ker \mathbf{M} \cap T(\mathbf{x}) = \{\mathbf{0}\}$.*

Proof. First assume that $\ker \mathbf{M} \cap T(\mathbf{x}) = \{\mathbf{0}\}$. Let $\mathbf{z} \in \mathbb{R}^d$ be a vector that satisfies $\mathbf{M}\mathbf{z} = \mathbf{M}\mathbf{x}$ and $\|\boldsymbol{\Omega}\mathbf{z}\|_1 \leq \|\boldsymbol{\Omega}\mathbf{x}\|_1$. This means that $\mathbf{z} - \mathbf{x} \in T(\mathbf{x})$ and $\mathbf{z} - \mathbf{x} \in \ker \mathbf{M}$. According to our assumption we conclude that $\mathbf{z} - \mathbf{x} = \mathbf{0}$, so that \mathbf{x} is the unique minimizer.

The other direction is proved by contradiction. Let \mathbf{x} be the unique minimizer of (4). Take any $\mathbf{v} \in T(\mathbf{x}) \setminus \{\mathbf{0}\}$. Then \mathbf{v} can be written as

$$\mathbf{v} = \sum_j t_j (\mathbf{z}_j - \mathbf{x}), \quad t_j \geq 0, \quad \|\boldsymbol{\Omega}\mathbf{z}_j\|_1 \leq \|\boldsymbol{\Omega}\mathbf{x}\|_1.$$

Since $\mathbf{v} \neq \mathbf{0}$, it holds $\sum_j t_j > 0$ and we can define $t'_j := \frac{t_j}{\sum_j t_j}$. Suppose $\mathbf{v} \in \ker \mathbf{M}$. Then

$$\mathbf{0} = \mathbf{M} \begin{pmatrix} \mathbf{v} \\ \sum_j t_j \end{pmatrix} = \mathbf{M} \begin{pmatrix} \sum_j t'_j \mathbf{z}_j \\ \sum_j t_j \end{pmatrix} - \mathbf{M}\mathbf{x},$$

so that $\mathbf{M} \left(\sum_j t'_j \mathbf{z}_j \right) = \mathbf{M}\mathbf{x}$. Together with the estimate

$$\left\| \sum_j t'_j \mathbf{z}_j \right\|_1 \leq \sum_j t'_j \|\mathbf{z}_j\| \leq \|\mathbf{x}\|_1$$

and uniqueness of the minimizer, this implies $\sum_j t'_j \mathbf{z}_j = \mathbf{x}$. Hence $\mathbf{v} = \mathbf{0}$, which leads to a contradiction. Thus, $\ker \mathbf{M} \cap T(\mathbf{x}) = \{\mathbf{0}\}$. \square

When the measurements are noisy, we use the following condition for successful recovery [9].

Theorem 4. *Let $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{M} \in \mathbb{R}^{m \times d}$ and $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{w}$ with $\|\mathbf{w}\|_2 \leq \eta$. If*

$$\inf_{\substack{\mathbf{v} \in T(\mathbf{x}) \\ \|\mathbf{v}\|_2=1}} \|\mathbf{M}\mathbf{v}\|_2 \geq \tau \tag{9}$$

for some τ , then a minimizer $\hat{\mathbf{x}}$ of (5) satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2\eta}{\tau}.$$

Proof. Since $\hat{\mathbf{x}}$ is a minimizer of (5), we have $\|\boldsymbol{\Omega}\hat{\mathbf{x}}\|_1 \leq \|\boldsymbol{\Omega}\mathbf{x}\|_1$ and $\hat{\mathbf{x}} - \mathbf{x} \in T(\mathbf{x})$. Our assumption (9) implies

$$\|\mathbf{M}(\hat{\mathbf{x}} - \mathbf{x})\|_2 \geq \tau \|\hat{\mathbf{x}} - \mathbf{x}\|_2. \tag{10}$$

On the other hand, we can give an upper bound for $\|\mathbf{M}\hat{\mathbf{x}} - \mathbf{M}\mathbf{x}\|_2$ by

$$\|\mathbf{M}\hat{\mathbf{x}} - \mathbf{M}\mathbf{x}\|_2 \leq \|\mathbf{M}\hat{\mathbf{x}} - \mathbf{y}\|_2 + \|\mathbf{M}\mathbf{x} - \mathbf{y}\|_2 \leq 2\eta. \tag{11}$$

Combining (10) and (11) we get the desired estimate. \square

2.2 Nonuniform recovery with Gaussian measurements

To prove the non-uniform recovery result for Gaussian random measurements (Theorem 1) we rely on the condition stated in Theorem 3, which requires that the null space of the measurement matrix \mathbf{M} misses the set $T(\mathbf{x})$. The next ingredient of the proof is a variation of the Gordon's escape through a mesh theorem, which was first used in the context of compressed sensing in [33]. To state this theorem, we introduce some notation and formulate auxiliary lemmas.

Let $\mathbf{g} \in \mathbb{R}^m$ be a standard Gaussian random vector, that is, a vector of independent mean zero, variance one normal distributed random variables. Then for

$$E_m := \mathbb{E} \|\mathbf{g}\|_2 = \sqrt{2} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)}$$

we have

$$\frac{m}{\sqrt{m+1}} \leq E_m \leq \sqrt{m},$$

see [19, 17]. For a set $T \subset \mathbb{R}^d$ we define its Gaussian width by

$$\ell(T) := \mathbb{E} \sup_{\mathbf{x} \in T} \langle \mathbf{x}, \mathbf{g} \rangle,$$

where $\mathbf{g} \in \mathbb{R}^d$ is a standard Gaussian random vector.

Lemma 1 (Gordon [19]). *Let $X_{i,j}$ and $Y_{i,j}$, $i = 1 \dots, m$, $j = 1, \dots, n$ be two mean-zero Gaussian random variables. If*

$$\begin{aligned} \mathbb{E} |X_{i,j} - X_{k,l}|^2 &\leq \mathbb{E} |Y_{i,j} - Y_{k,l}|^2, \text{ for all } i \neq k, \text{ and } j, l \\ \mathbb{E} |X_{i,j} - X_{i,l}|^2 &\geq \mathbb{E} |Y_{i,j} - Y_{i,l}|^2 \text{ for } i, j, l, \end{aligned}$$

then

$$\mathbb{E} \min_i \max_j X_{i,j} \geq \mathbb{E} \min_i \max_j Y_{i,j}.$$

Remark 1. Gordon's lemma extends to the case of Gaussian processes indexed by possibly infinite sets where the expected maxima or minima are replaced by corresponding lattice suprema or infima, see for instance [17, Remark 8.28] or [24].

We further exploit the concentration of measure phenomenon, which asserts that Lipschitz functions concentrate well around their expectation [23, 27].

Lemma 2 (Concentration of measure). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be an L -Lipschitz function:*

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L \|\mathbf{x} - \mathbf{y}\|_2, \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Let $\mathbf{g} = (g_1, g_2, \dots, g_n)$ be a vector of independent standard normal random variables. Then, for all $t > 0$,

$$\mathbb{P}(\mathbb{E}[f(\mathbf{g})] - f(\mathbf{g}) \geq t) \leq e^{-\frac{t^2}{2L^2}}.$$

Next, we state our modification of the Gordon's escape through a mesh theorem, see [19] for the original version. Below, $\Omega(T)$ corresponds to the set of elements produced by applying Ω to the elements of T .

Theorem 5. Let $\Omega \in \mathbb{R}^{p \times d}$ be a frame with constants $A, B > 0$. Let $\mathbf{M} \in \mathbb{R}^{m \times d}$ be a Gaussian random matrix and T be a subset of the unit sphere $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 = 1\}$. Then, for $t > 0$, it holds

$$\mathbb{P} \left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 > E_m - \frac{1}{\sqrt{A}} \ell(\Omega(T)) - t \right) \geq 1 - e^{-\frac{t^2}{2}}. \quad (12)$$

Proof. Recall that

$$\|\mathbf{M}\mathbf{x}\|_2 = \max_{\mathbf{y} \in S^{m-1}} \langle \mathbf{M}\mathbf{x}, \mathbf{y} \rangle.$$

For $\mathbf{x} \in T$ and $\mathbf{y} \in S^{m-1}$ we compare the two Gaussian processes

$$X_{\mathbf{x},\mathbf{y}} := \langle \mathbf{M}\mathbf{x}, \mathbf{y} \rangle \quad \text{and} \quad Y_{\mathbf{x},\mathbf{y}} := \frac{1}{\sqrt{A}} \langle \mathbf{g}, \Omega\mathbf{x} \rangle + \langle \mathbf{h}, \mathbf{y} \rangle,$$

where $\mathbf{g} \in \mathbb{R}^p$ and $\mathbf{h} \in \mathbb{R}^m$ are independent standard Gaussian random vectors. Let $\mathbf{x}, \mathbf{x}' \in \mathbb{S}^{d-1}$ and $\mathbf{y}, \mathbf{y}' \in S^{m-1}$. Since M_{ij} are independent with $\mathbb{E} M_{ij} = 0$, $\mathbb{E} M_{ij}^2 = 1$, we have

$$\begin{aligned} \mathbb{E} |X_{\mathbf{x},\mathbf{y}} - X_{\mathbf{x}',\mathbf{y}'}|^2 &= \mathbb{E} \left| \sum_{i=1}^m \sum_{j=1}^d M_{ij} (x_j y_i - x'_j y'_i) \right|^2 = \sum_{i=1}^m \sum_{j=1}^d (x_j y_i - x'_j y'_i)^2 \\ &= \sum_{i=1}^m \sum_{j=1}^d (x_j^2 y_i^2 + x'^2_j y'^2_i - 2x_j x'_j y_i y'_i) = \|\mathbf{x}\|_2^2 \|\mathbf{y}\|_2^2 + \|\mathbf{x}'\|_2^2 \|\mathbf{y}'\|_2^2 - 2\langle \mathbf{x}, \mathbf{x}' \rangle \langle \mathbf{y}, \mathbf{y}' \rangle \\ &= 2 - 2\langle \mathbf{x}, \mathbf{x}' \rangle \langle \mathbf{y}, \mathbf{y}' \rangle. \end{aligned} \quad (13)$$

Independence and the isotropicity of the Gaussian vectors \mathbf{g} and \mathbf{h} together with the fact that Ω is a frame with lower frame bound A imply

$$\begin{aligned} \mathbb{E} |Y_{\mathbf{x},\mathbf{y}} - Y_{\mathbf{x}',\mathbf{y}'}|^2 &= \mathbb{E} \left| \frac{1}{\sqrt{A}} \langle \mathbf{g}, \Omega\mathbf{x} - \Omega\mathbf{x}' \rangle \right|^2 + \mathbb{E} |\langle \mathbf{h}, \mathbf{y} - \mathbf{y}' \rangle|^2 \\ &= \frac{1}{A} \|\Omega\mathbf{x} - \Omega\mathbf{x}'\|_2^2 + \|\mathbf{y} - \mathbf{y}'\|_2^2 \geq \|\mathbf{x} - \mathbf{x}'\|_2^2 + \|\mathbf{y} - \mathbf{y}'\|_2^2 \\ &= \|\mathbf{x}\|_2^2 + \|\mathbf{x}'\|_2^2 - 2\langle \mathbf{x}, \mathbf{x}' \rangle + \|\mathbf{y}\|_2^2 + \|\mathbf{y}'\|_2^2 - 2\langle \mathbf{y}, \mathbf{y}' \rangle \\ &= 4 - 2\langle \mathbf{x}, \mathbf{x}' \rangle - 2\langle \mathbf{y}, \mathbf{y}' \rangle. \end{aligned} \quad (14)$$

When $\mathbf{x} = \mathbf{x}'$, we have

$$\mathbb{E} |Y_{\mathbf{x},\mathbf{y}} - Y_{\mathbf{x},\mathbf{y}'}|^2 = \|\mathbf{y} - \mathbf{y}'\|_2^2 = 2 - 2\langle \mathbf{y}, \mathbf{y}' \rangle.$$

Combining (13) and (14), we obtain

$$\mathbb{E} |Y_{\mathbf{x},\mathbf{y}} - Y_{\mathbf{x}',\mathbf{y}'}|^2 - \mathbb{E} |X_{\mathbf{x},\mathbf{y}} - X_{\mathbf{x}',\mathbf{y}'}|^2 \geq 2(1 - \langle \mathbf{x}, \mathbf{x}' \rangle)(1 - \langle \mathbf{y}, \mathbf{y}' \rangle)$$

and since $\langle \mathbf{x}, \mathbf{x}' \rangle \leq \|\mathbf{x}\|_2 \|\mathbf{x}'\|_2 \leq 1$ and similarly for \mathbf{y}, \mathbf{y}' , it follows that

$$\mathbb{E} |Y_{\mathbf{x},\mathbf{y}} - Y_{\mathbf{x}',\mathbf{y}'}|^2 - \mathbb{E} |X_{\mathbf{x},\mathbf{y}} - X_{\mathbf{x}',\mathbf{y}'}|^2 \geq 0.$$

Moreover, we have

$$\mathbb{E} |Y_{\mathbf{x},\mathbf{y}} - Y_{\mathbf{x},\mathbf{y}'}|^2 = \mathbb{E} |X_{\mathbf{x},\mathbf{y}} - X_{\mathbf{x},\mathbf{y}'}|^2.$$

Due to Gordon's lemma (Lemma 1) and Remark 1,

$$\begin{aligned}
\mathbb{E} \inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 &= \mathbb{E} \inf_{\mathbf{x} \in T} \max_{\mathbf{y} \in S^{m-1}} X_{\mathbf{x}, \mathbf{y}} \geq \mathbb{E} \inf_{\mathbf{x} \in T} \max_{\mathbf{y} \in S^{m-1}} Y_{\mathbf{x}, \mathbf{y}} \\
&= \mathbb{E} \inf_{\mathbf{x} \in T} \max_{\mathbf{y} \in S^{m-1}} \left\{ \frac{1}{\sqrt{A}} \langle \mathbf{g}, \boldsymbol{\Omega}\mathbf{x} \rangle + \langle \mathbf{h}, \mathbf{y} \rangle \right\} = \mathbb{E} \inf_{\mathbf{x} \in T} \left\{ \frac{1}{\sqrt{A}} \langle \mathbf{g}, \boldsymbol{\Omega}\mathbf{x} \rangle + \|\mathbf{h}\|_2 \right\} \\
&= \mathbb{E} \|\mathbf{h}\|_2 - \frac{1}{\sqrt{A}} \mathbb{E} \sup_{\mathbf{x} \in T} \langle \mathbf{g}, \boldsymbol{\Omega}\mathbf{x} \rangle = E_m - \frac{1}{\sqrt{A}} \mathbb{E} \sup_{\mathbf{z} \in \boldsymbol{\Omega}(T)} \langle \mathbf{g}, \mathbf{z} \rangle = E_m - \frac{1}{\sqrt{A}} \ell(\boldsymbol{\Omega}(T)). \tag{15}
\end{aligned}$$

Let $F(\mathbf{M}) := \inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2$. For any $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times d}$

$$\begin{aligned}
\inf_{\mathbf{x} \in T} \|\mathbf{A}\mathbf{x}\|_2 &\leq \inf_{\mathbf{x} \in T} (\|\mathbf{B}\mathbf{x}\|_2 + \|(\mathbf{A} - \mathbf{B})\mathbf{x}\|_2) \\
&\leq \inf_{\mathbf{x} \in T} \|\mathbf{B}\mathbf{x}\|_2 + \|\mathbf{A} - \mathbf{B}\|_{2 \rightarrow 2} \leq \inf_{\mathbf{x} \in T} \|\mathbf{B}\mathbf{x}\|_2 + \|\mathbf{A} - \mathbf{B}\|_F.
\end{aligned}$$

The second inequality follows from the fact that $T \subset \mathbb{S}^{d-1}$. By interchanging \mathbf{A} and \mathbf{B} we conclude that

$$|F(\mathbf{A}) - F(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|_F.$$

This means that F is 1-Lipschitz with respect to the Frobenius norm (which corresponds to the ℓ_2 -norm when interpreting a matrix as a vector) and due to concentration of measure (Lemma 2)

$$\mathbb{P} \left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 \leq \mathbb{E} \inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 - t \right) \leq e^{-t^2/2}.$$

Applying the estimate (15) to the previous inequality gives

$$\mathbb{P} \left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 \leq E_m - \frac{\ell(\boldsymbol{\Omega}(T))}{\sqrt{A}} - t \right) \leq \mathbb{P} \left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 \leq \mathbb{E} \inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 - t \right) \leq e^{-t^2/2},$$

which concludes the proof. \square

The previous result suggests to estimate the Gaussian width of $\boldsymbol{\Omega}(T)$ with $T := T(\mathbf{x}) \cap \mathbb{S}^{d-1}$. Since $\boldsymbol{\Omega}$ is a frame with upper frame constant B , we have

$$\boldsymbol{\Omega}(T) \subset \boldsymbol{\Omega}(T(\mathbf{x})) \cap \boldsymbol{\Omega}(\mathbb{S}^{d-1}) \subset K(\boldsymbol{\Omega}\mathbf{x}) \cap (\sqrt{B}\mathbb{B}_2^p),$$

where

$$K(\boldsymbol{\Omega}\mathbf{x}) = \text{cone} \{ \mathbf{y} - \boldsymbol{\Omega}\mathbf{x} : \mathbf{y} \in \mathbb{R}^p, \|\mathbf{y}\|_1 \leq \|\boldsymbol{\Omega}\mathbf{x}\|_1 \}.$$

The supremum over a larger set can only increase, hence

$$\ell(\boldsymbol{\Omega}(T)) \leq \sqrt{B} \ell(K(\boldsymbol{\Omega}\mathbf{x}) \cap \mathbb{B}_2^p). \tag{16}$$

We next recall an upper bound for the Gaussian width $\ell(K(\boldsymbol{\Omega}\mathbf{x}) \cap \mathbb{B}_2^p)$ from [9] involving the polar cone $\mathcal{N}(\boldsymbol{\Omega}\mathbf{x}) = K(\boldsymbol{\Omega}\mathbf{x})^\circ$ defined by

$$\mathcal{N}(\boldsymbol{\Omega}\mathbf{x}) = \{ \mathbf{z} \in \mathbb{R}^p : \langle \mathbf{z}, \mathbf{y} - \boldsymbol{\Omega}\mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{y} \in \mathbb{R}^p \text{ such that } \|\mathbf{y}\|_1 \leq \|\boldsymbol{\Omega}\mathbf{x}\|_1 \}.$$

Proposition 1. Let $\mathbf{g} \in \mathbb{R}^p$ be a standard Gaussian random vector. Then

$$\ell(K(\boldsymbol{\Omega}\mathbf{x}) \cap \mathbb{B}_2^p) \leq \mathbb{E} \min_{\mathbf{z} \in \mathcal{N}(\boldsymbol{\Omega}\mathbf{x})} \|\mathbf{g} - \mathbf{z}\|_2.$$

The proof relies on tools from convex analysis, see [9, 17], [3, Ch. 5.9].

Proposition 2. Let s be the sparsity of the vector $\boldsymbol{\Omega}\mathbf{x} \in \mathbb{R}^p$. Then

$$\ell(K(\boldsymbol{\Omega}\mathbf{x}) \cap \mathbb{B}_2^p)^2 \leq 2s \ln \frac{ep}{s}. \quad (17)$$

Proof. By Proposition 1 and Hölder's inequality

$$\ell(K(\boldsymbol{\Omega}\mathbf{x}) \cap \mathbb{B}_2^p)^2 \leq \left(\mathbb{E} \min_{\mathbf{z} \in \mathcal{N}(\boldsymbol{\Omega}\mathbf{x})} \|\mathbf{g} - \mathbf{z}\|_2 \right)^2 \leq \mathbb{E} \min_{\mathbf{z} \in \mathcal{N}(\boldsymbol{\Omega}\mathbf{x})} \|\mathbf{g} - \mathbf{z}\|_2^2. \quad (18)$$

Let S denote the support of $\boldsymbol{\Omega}\mathbf{x}$. Then one can verify that

$$\mathcal{N}(\boldsymbol{\Omega}\mathbf{x}) = \bigcup_{t \geq 0} \{ \mathbf{z} \in \mathbb{R}^p : z_i = t \operatorname{sgn}(\boldsymbol{\Omega}\mathbf{x})_i, i \in S, |z_i| \leq t, i \in S^c \}, \quad (19)$$

see [17, Lemma 9.23] for a proof. To proceed, we fix t , minimize $\|\mathbf{g} - \mathbf{z}\|_2^2$ over all possible entries z_j , take the expectation of the obtained expression and finally optimize over t . According to (19), we have

$$\begin{aligned} \min_{\mathbf{z} \in \mathcal{N}(\boldsymbol{\Omega}\mathbf{x})} \|\mathbf{g} - \mathbf{z}\|_2^2 &= \min_{\substack{t \geq 0 \\ |z_i| \leq t, i \in S^c}} \sum_{i \in S} (g_i - t \operatorname{sgn}(\boldsymbol{\Omega}\mathbf{x})_i)^2 + \sum_{i \in S^c} (g_i - z_i)^2 \\ &= \min_{t \geq 0} \sum_{i \in S} (g_i - t \operatorname{sgn}(\boldsymbol{\Omega}\mathbf{x})_i)^2 + \sum_{i \in S^c} S_t(g_i)^2, \end{aligned}$$

where S_t is the soft-thresholding operator given by

$$S_t(x) = \begin{cases} x + t, & x < -t, \\ 0, & -t \leq x \leq t, \\ x - t, & x > t. \end{cases}$$

Taking expectation we arrive at

$$\begin{aligned} \mathbb{E} \min_{\mathbf{z} \in \mathcal{N}(\boldsymbol{\Omega}\mathbf{x})} \|\mathbf{g} - \mathbf{z}\|_2^2 &\leq \mathbb{E} \left[\sum_{i \in S} (g_i - t \operatorname{sgn}(\boldsymbol{\Omega}\mathbf{x})_i)^2 \right] + \mathbb{E} \left[\sum_{i \in S^c} S_t(g_i)^2 \right] \\ &= s(1 + t^2) + (p - s) \mathbb{E} S_t(g)^2, \end{aligned} \quad (20)$$

where g is a univariate standard Gaussian random variable. To calculate the expectation of $S_t(g)^2$, we apply the direct integration

$$\begin{aligned} \mathbb{E} S_t(g)^2 &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{-t} (x + t)^2 e^{-\frac{x^2}{2}} dx + \int_t^{\infty} (x - t)^2 e^{-\frac{x^2}{2}} dx \right] \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-\frac{(x+t)^2}{2}} dx = \frac{2e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-\frac{x^2}{2}} e^{-xt} dx \\ &\leq e^{-\frac{t^2}{2}} \sqrt{\frac{2}{\pi}} \int_0^{\infty} x^2 e^{-\frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}. \end{aligned} \quad (21)$$

Substituting the estimate (21) into (20) gives

$$\mathbb{E} \min_{\mathbf{z} \in \mathcal{N}(\boldsymbol{\Omega}\mathbf{x})} \|\mathbf{g} - \mathbf{z}\|_2^2 \leq s(1+t^2) + (p-s)e^{-\frac{t^2}{2}}.$$

Setting $t = \sqrt{2\ln(p/s)}$ finally leads to

$$\ell(K(\boldsymbol{\Omega}\mathbf{x}) \cap \mathbb{B}_2^p)^2 \leq s(1+2\ln(p/s)) + s = 2s\ln(ep/s).$$

This concludes the proof. \square

By combining inequalities (16) and (17) we obtain

$$\ell(\boldsymbol{\Omega}(T))^2 \leq 2Bs \ln \frac{ep}{s}.$$

of Theorem 1. Set $t = \sqrt{2\ln(\varepsilon^{-1})}$. The fact that $E_m \geq m/\sqrt{m+1}$ along with condition (6) yields

$$E_m \geq \frac{1}{\sqrt{A}}\ell(\boldsymbol{\Omega}(T)) + t.$$

Theorem 5 implies

$$\mathbb{P}\left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 > 0\right) \geq \mathbb{P}\left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 > E_m - \frac{1}{\sqrt{A}}\ell(\boldsymbol{\Omega}(T)) - t\right) \geq 1 - e^{-\frac{t^2}{2}} = 1 - \varepsilon,$$

which guarantees that $\ker \mathbf{M} \cap T(\mathbf{x}) = \{\mathbf{0}\}$ with probability at least $1 - \varepsilon$. As the final step we apply Theorem 3. \square

We now extend Theorem 1 to robust recovery.

Theorem 6. *Let $\boldsymbol{\Omega} \in \mathbb{R}^{p \times d}$ be a frame with frame bounds $A, B > 0$ and let \mathbf{x} be l -cosparse and $s = p - l$. For a random draw $\mathbf{M} \in \mathbb{R}^{m \times d}$ of a Gaussian random matrix, let noisy measurements $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{w}$ be given with $\|\mathbf{w}\|_2 \leq \eta$. If for $0 < \varepsilon < 1$ and some $\tau > 0$*

$$\frac{m^2}{m+1} \geq \frac{2Bs}{A} \left(\sqrt{\ln \frac{ep}{s}} + \sqrt{\frac{A \ln(\varepsilon^{-1})}{Bs}} + \tau \sqrt{\frac{A}{2sB}} \right)^2, \quad (22)$$

then with probability at least $1 - \varepsilon$, any minimizer $\hat{\mathbf{x}}$ of (5) satisfies

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2\eta}{\tau}.$$

Proof. We use the recovery condition stated in Theorem 4. Set $t = \sqrt{2\ln(\varepsilon^{-1})}$. Our previous considerations and the choice of m in (22) guarantee that

$$E_m - \frac{1}{\sqrt{A}}\ell(\boldsymbol{\Omega}(T)) - t \geq \frac{m}{\sqrt{m+1}} - \sqrt{\frac{2Bs}{A} \ln \frac{ep}{s}} - \sqrt{2\ln(\varepsilon^{-1})} \geq \tau.$$

The monotonicity of probability and Theorem 5 yield

$$\mathbb{P}\left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 \geq \tau\right) \geq \mathbb{P}\left(\inf_{\mathbf{x} \in T} \|\mathbf{M}\mathbf{x}\|_2 \geq E_m - \frac{1}{\sqrt{A}}\ell(\boldsymbol{\Omega}(T)) - t\right) \geq 1 - \varepsilon.$$

\square

3 Uniform recovery

This section is dedicated to the proof of the uniform recovery result in Theorem 2. It relies on the Ω -null space property, which extends the null space property known from the standard synthesis sparsity case, see e.g. [11, 17, 20]. We analyze this property directly for Gaussian random matrices with similar techniques as used in the previous section.

3.1 Ω -null space property

Let us start with the Ω -null space property which is a sufficient condition for the exact reconstruction of every cosparse vector.

Definition 2. A matrix $\mathbf{M} \in \mathbb{R}^{m \times d}$ is said to satisfy the Ω -null space property of order s with constant $0 < \rho < 1$, if for any set $\Lambda \subset [p]$ with $\#\Lambda \geq p - s$ it holds

$$\|\Omega_{\Lambda^c} \mathbf{v}\|_1 \leq \rho \|\Omega_{\Lambda} \mathbf{v}\|_1 \quad \text{for all } \mathbf{v} \in \ker \mathbf{M}. \quad (23)$$

If Ω is the identity map $\mathbf{Id} : \mathbb{R}^d \rightarrow \mathbb{R}^d$, then (23) is the standard null space property. We start with a result on exact recovery of cosparse vectors.

Theorem 7. *If $\mathbf{M} \in \mathbb{R}^{m \times d}$ satisfies the Ω -null space property of order s with $0 < \rho < 1$, then every l -cosparse vector $\mathbf{x} \in \mathbb{R}^d$ with $l = p - s$ is the unique solution of (4) with $\mathbf{y} = \mathbf{M}\mathbf{x}$.*

This theorem follows immediately from the next result, which also implies a certain stability estimate in ℓ_1 .

Theorem 8. *Let $\mathbf{x} \in \mathbb{R}^d$ be an arbitrary vector and $\hat{\mathbf{x}}$ be a solution of (4) with $\mathbf{y} = \mathbf{M}\mathbf{x}$, where $\mathbf{M} \in \mathbb{R}^{m \times d}$ satisfies the Ω -null space property of order s with constant $\rho \in (0, 1)$. Then*

$$\|\Omega(\mathbf{x} - \hat{\mathbf{x}})\|_1 \leq \frac{2(1 + \rho)}{1 - \rho} \sigma_s(\Omega \mathbf{x})_1. \quad (24)$$

Proof. Since $\hat{\mathbf{x}}$ is the solution of (4), we must have $\|\Omega \hat{\mathbf{x}}\|_1 \leq \|\Omega \mathbf{x}\|_1$. Take any $\Lambda \subset [p]$ with $\#\Lambda \geq p - s$. Then

$$\|\Omega_{\Lambda^c} \hat{\mathbf{x}}\|_1 + \|\Omega_{\Lambda} \hat{\mathbf{x}}\|_1 \leq \|\Omega_{\Lambda^c} \mathbf{x}\|_1 + \|\Omega_{\Lambda} \mathbf{x}\|_1.$$

By the triangle inequality, the vector $\mathbf{v} := \mathbf{x} - \hat{\mathbf{x}}$ satisfies

$$\|\Omega_{\Lambda^c} \mathbf{x}\|_1 - \|\Omega_{\Lambda^c} \mathbf{v}\|_1 + \|\Omega_{\Lambda} \mathbf{v}\|_1 - \|\Omega_{\Lambda} \mathbf{x}\|_1 \leq \|\Omega_{\Lambda^c} \mathbf{x}\|_1 + \|\Omega_{\Lambda} \mathbf{x}\|_1,$$

which implies

$$\|\Omega_{\Lambda} \mathbf{v}\|_1 \leq \|\Omega_{\Lambda^c} \mathbf{v}\|_1 + 2\|\Omega_{\Lambda} \mathbf{x}\|_1 \leq \rho \|\Omega_{\Lambda} \mathbf{v}\|_1 + 2\|\Omega_{\Lambda} \mathbf{x}\|_1.$$

Hereby, we have applied the Ω -null space property (23). Rearranging and choosing a set Λ of size $p - s$ which minimizes $\|\Omega_{\Lambda} \mathbf{x}\|_1$ yields

$$\|\Omega_{\Lambda} \mathbf{v}\|_1 \leq \frac{2}{1 - \rho} \sigma_s(\Omega \mathbf{x})_1.$$

Furthermore, another application of the Ω -null space property gives

$$\|\Omega \mathbf{v}\|_1 = \|\Omega_{\Lambda} \mathbf{v}\|_1 + \|\Omega_{\Lambda^c} \mathbf{v}\|_1 \leq (1 + \rho) \|\Omega_{\Lambda} \mathbf{v}\|_1 \leq \frac{2(1 + \rho)}{1 - \rho} \sigma_s(\Omega \mathbf{x})_1.$$

This completes the proof. □

In order to provide a suitable stability estimate in ℓ_2 we require a slightly stronger version of the Ω -null space property.

Definition 3. A matrix $\mathbf{M} \in \mathbb{R}^{m \times d}$ is said to satisfy the ℓ_2 -stable Ω -null space property of order s with constant $0 < \rho < 1$, if, for any set $\Lambda \subset [p]$ with $\#\Lambda \geq p - s$, it holds

$$\|\Omega_{\Lambda^c} \mathbf{v}\|_2 \leq \frac{\rho}{\sqrt{s}} \|\Omega_{\Lambda} \mathbf{v}\|_1 \quad \text{for all } \mathbf{v} \in \ker \mathbf{M}. \quad (25)$$

Remark 2. The Hölder's inequality implies $\|\Omega_{\Lambda^c} \mathbf{v}\|_1 \leq \sqrt{s} \|\Omega_{\Lambda^c} \mathbf{v}\|_2$ for any set $\Lambda \subset [p]$ with $\#(\Lambda^c) = s$. This means that if $\mathbf{M} \in \mathbb{R}^{m \times d}$ satisfies the ℓ_2 -stable Ω -null space property of order s with constant $0 < \rho < 1$, then it satisfies the Ω -null space property of the same order and with the same constant.

Theorem 9. Let $\mathbf{M} \in \mathbb{R}^{m \times d}$ satisfy the ℓ_2 -stable Ω -null space property of order s with constant $0 < \rho < 1$. Then for any $\mathbf{x} \in \mathbb{R}^d$ the solution $\hat{\mathbf{x}}$ of (4) with $\mathbf{y} = \mathbf{M}\mathbf{x}$ approximates the vector \mathbf{x} with ℓ_2 -error

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2(1 + \rho)^2}{\sqrt{A}(1 - \rho)} \frac{\sigma_s(\Omega \mathbf{x})_1}{\sqrt{s}}. \quad (26)$$

Inequality (26) means that l -cosparse vectors are exactly recovered by (4) and vectors $\mathbf{x} \in \mathbb{R}^d$, such that $\Omega \mathbf{x}$ is close to an s -sparse vector in ℓ_1 , can be well approximated in ℓ_2 by a solution of (4). The proof goes along the same lines as in the standard case in [17]. The novelty here is that we exploit the sparsity not of the signal itself, but of its analysis representation. So first we extend the ℓ_1 -error estimate above to an ℓ_2 -error estimate for $\Omega \mathbf{x}$ and use the fact that Ω is a frame to bound the ℓ_2 -error $\|\mathbf{x} - \hat{\mathbf{x}}\|_2$. The statement of Theorem 9 was generalized to the setting of a perturbed frame and imprecise knowledge of the measurement matrix in [1, Theorem 3.1].

of Theorem 9. We define the vector $\mathbf{v} := \hat{\mathbf{x}} - \mathbf{x} \in \ker \mathbf{M}$ and denote by $S_0 \subset [p]$ an index set of s largest absolute entries of $\Omega \mathbf{v}$. Since $\#S_0^c = p - s$ and $\mathbf{M} \in \mathbb{R}^{m \times d}$ satisfies the ℓ_2 -stable Ω -null space property, it follows

$$\|(\Omega \mathbf{v})_{S_0}\|_2 \leq \frac{\rho}{\sqrt{s}} \|\Omega_{S_0^c} \mathbf{v}\|_1 \leq \frac{\rho}{\sqrt{s}} \|\Omega \mathbf{v}\|_1. \quad (27)$$

We partition the indices of S_0^c into subsets S_1, S_2, \dots of size s in order of decreasing magnitude of $(\Omega \mathbf{v})_i$. Then for each $k \in S_{i+1}$, $i \geq 0$,

$$|(\Omega \mathbf{v})_k| \leq \frac{1}{s} \sum_{j \in S_i} |(\Omega \mathbf{v})_j| \quad \text{and} \quad \|(\Omega \mathbf{v})_{S_{i+1}}\|_2 \leq \frac{1}{\sqrt{s}} \|(\Omega \mathbf{v})_{S_i}\|_1.$$

Along with the triangle inequality this gives

$$\|(\Omega \mathbf{v})_{S_0^c}\|_2 \leq \sum_{i \geq 1} \|(\Omega \mathbf{v})_{S_i}\|_2 \leq \frac{1}{\sqrt{s}} \sum_{i \geq 0} \|(\Omega \mathbf{v})_{S_i}\|_1 = \frac{1}{\sqrt{s}} \|\Omega \mathbf{v}\|_1. \quad (28)$$

Inequalities (27) and (28) together with Remark 2 and Theorem 8 yield

$$\|\Omega \mathbf{v}\|_2 \leq \|(\Omega \mathbf{v})_{S_0}\|_2 + \|(\Omega \mathbf{v})_{S_0^c}\|_2 \leq \frac{1 + \rho}{\sqrt{s}} \|\Omega \mathbf{v}\|_1 \leq \frac{2(1 + \rho)^2}{(1 - \rho)\sqrt{s}} \sigma_s(\Omega \mathbf{x})_1. \quad (29)$$

Finally, we use that $\mathbf{\Omega}$ is a frame with lower frame bound A to conclude that

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{1}{\sqrt{A}} \|\mathbf{\Omega}\mathbf{x} - \mathbf{\Omega}\hat{\mathbf{x}}\|_2 \leq \frac{2(1+\rho)^2}{\sqrt{A}(1-\rho)\sqrt{s}} \sigma_s(\mathbf{\Omega}\mathbf{x})_1.$$

This completes the proof. \square

When the measurements are given with some error, the author in [16] introduced the following extension of the $\mathbf{\Omega}$ -null space property in order to guarantee robustness of the recovery.

Definition 4. A matrix $\mathbf{M} \in \mathbb{R}^{m \times d}$ is said to satisfy the robust ℓ_2 -stable $\mathbf{\Omega}$ -null space property of order s with constant $0 < \rho < 1$ and $\tau > 0$, if for any set $\Lambda \subset [p]$ with $\#\Lambda \geq p - s$ it holds

$$\|\mathbf{\Omega}_{\Lambda^c}\mathbf{v}\|_2 \leq \frac{\rho}{\sqrt{s}} \|\mathbf{\Omega}_{\Lambda}\mathbf{v}\|_1 + \tau \|\mathbf{M}\mathbf{v}\|_2 \quad \text{for all } \mathbf{v} \in \mathbb{R}^d. \quad (30)$$

If $\mathbf{v} \in \ker \mathbf{M}$, the term $\|\mathbf{M}\mathbf{v}\|_2$ vanishes, and we see that the robust ℓ_2 -stable $\mathbf{\Omega}$ -null space property implies the ℓ_2 -stable $\mathbf{\Omega}$ -null space property. The robust ℓ_2 -stable null space property guarantees the stability and robustness of the ℓ_1 -minimization (5).

Theorem 10. Let $\mathbf{M} \in \mathbb{R}^{m \times d}$ satisfy the robust ℓ_2 -stable $\mathbf{\Omega}$ -null space property of order s with constants $0 < \rho < 1$ and $\tau > 0$. Then for any $\mathbf{x} \in \mathbb{R}^d$ the solution $\hat{\mathbf{x}}$ of (5) with $\mathbf{y} = \mathbf{M}\mathbf{x} + \mathbf{w}$, $\|\mathbf{w}\|_2 \leq \eta$, approximates the vector \mathbf{x} with ℓ_2 -error

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2(1+\rho)^2}{\sqrt{A}(1-\rho)} \frac{\sigma_s(\mathbf{\Omega}\mathbf{x})_1}{\sqrt{s}} + \frac{2\tau(3+\rho)}{\sqrt{A}(1-\rho)} \eta. \quad (31)$$

Proof. Theorem 5 in [16] with $q = p = 2$ provides the bound for $\|\mathbf{\Omega}(\mathbf{x} - \hat{\mathbf{x}})\|_2$. Taking into account that $\mathbf{\Omega}$ is a frame with lower frame constant A , we obtain estimate (31). \square

3.2 Uniform recovery from Gaussian measurements

We now show Theorem 2 by establishing the ℓ_2 -stable $\mathbf{\Omega}$ -null space property of order s for a Gaussian measurement matrix \mathbf{M} by following a similar strategy as in Section 2. To this end we introduce the set

$$W_{\rho,s} := \left\{ \mathbf{w} \in \mathbb{R}^d : \|\mathbf{\Omega}_{\Lambda^c}\mathbf{w}\|_2 > \rho/\sqrt{s} \|\mathbf{\Omega}_{\Lambda}\mathbf{w}\|_1 \text{ for some } \Lambda \subset [p], \#\Lambda = p - s \right\}.$$

In fact, if

$$\inf \left\{ \|\mathbf{M}\mathbf{w}\|_2 : \mathbf{w} \in W_{\rho,s} \cap \mathbb{S}^{d-1} \right\} > 0, \quad (32)$$

then for all $\mathbf{w} \in \ker \mathbf{M}$ and any $\Lambda \subset [p]$ with $\#\Lambda = p - s$ we have

$$\|\mathbf{\Omega}_{\Lambda^c}\mathbf{w}\|_2 \leq \frac{\rho}{\sqrt{s}} \|\mathbf{\Omega}_{\Lambda}\mathbf{w}\|_1,$$

which means that \mathbf{M} satisfies the ℓ_2 -stable $\mathbf{\Omega}$ -null space property of order s . To show (32) we apply Theorem 5, which requires to study the Gaussian width of the set $\mathbf{\Omega}(W_{\rho,s} \cap \mathbb{S}^{d-1})$. Since $\mathbf{\Omega}$ is a frame with upper frame bound B , we have

$$\mathbf{\Omega}(W_{\rho,s} \cap \mathbb{S}^{d-1}) \subset \mathbf{\Omega}(W_{\rho,s}) \cap \left(\sqrt{B}\mathbb{B}_2^p \right) \subset T_{\rho,s} \cap \left(\sqrt{B}\mathbb{B}_2^p \right) = \sqrt{B}(T_{\rho,s} \cap \mathbb{B}_2^p), \quad (33)$$

with

$$T_{\rho,s} = \{ \mathbf{u} \in \mathbb{R}^p : \|\mathbf{u}_S\|_2 \geq \rho/\sqrt{s} \|\mathbf{u}_{S^c}\|_1 \text{ for some } S \subset [p], \#S = s \}.$$

Then

$$T_{\rho,s} \cap \mathbb{B}_2^p = \bigcup_{\#S=s} \left\{ \mathbf{u} \in \mathbb{R}^p : \|\mathbf{u}\|_2 \leq 1, \|\mathbf{u}_S\|_2 > \frac{\rho}{\sqrt{s}} \|\mathbf{u}_{S^c}\|_1 \right\}.$$

Lemma 3. *Let D be the set defined by*

$$D := \text{conv} \{ \mathbf{x} \in \mathbb{S}^{p-1} : \#\text{supp } \mathbf{x} \leq s \}. \quad (34)$$

(a) *Then D is the unit ball with respect to the norm*

$$\|\mathbf{x}\|_D := \sum_{l=1}^L \left[\sum_{i \in I_l} (x_i^*)^2 \right]^{1/2},$$

where $L = \lceil \frac{p}{s} \rceil$,

$$I_l = \begin{cases} \{s(l-1)+1, \dots, sl\}, & l = 1, \dots, L-1, \\ \{s(L-1)+1, \dots, p\}, & l = L, \end{cases}$$

and \mathbf{x}^* is the non-increasing rearrangement of \mathbf{x} .

(b) *It holds*

$$T_{\rho,s} \cap \mathbb{B}_2^p \subset \sqrt{1 + (1 + \rho^{-1})^2} D. \quad (35)$$

A similar result was stated as Lemma 4.5 in [33]. For the sake of completeness we present the proof.

Proof. (a) Suppose $\mathbf{x} \in D$. It can be represented as $\mathbf{x} = \sum_i \alpha_i \mathbf{x}_i$ with $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$ and $\mathbf{x}_i \in \mathbb{S}^{p-1}$, $\#\text{supp } \mathbf{x}_i \leq s$. Then $\|\mathbf{x}_i\|_D = \|\mathbf{x}_i\|_2 = 1$. By the triangle inequality

$$\|\mathbf{x}\|_D \leq \sum_i \alpha_i \|\mathbf{x}_i\|_D = \sum_i \alpha_i = 1.$$

This proves that D is a subset of the unit ball with respect to the $\|\cdot\|_D$ -norm.

On the other hand, let $\|\mathbf{x}\|_D \leq 1$. We partition the index set $[p]$ into subsets S_1, S_2, \dots of size s in order of decreasing magnitude of entries x_k . Set $\alpha_i = \|\mathbf{x}_{S_i}\|_2$. Then \mathbf{x} can be written as

$$\mathbf{x} = \sum_{i:\alpha_i \neq 0} \alpha_i \left(\frac{1}{\alpha_i} \mathbf{x}_{S_i} \right), \quad \text{where} \quad \sum_{i:\alpha_i \neq 0} \alpha_i = \sum_i \|\mathbf{x}_{S_i}\|_2 = \|\mathbf{x}\|_D \leq 1$$

and, for $\alpha_i \neq 0$, $\|\frac{1}{\alpha_i} \mathbf{x}_{S_i}\|_2 = \frac{1}{\alpha_i} \|\mathbf{x}_{S_i}\|_2 = 1$. Thus $\mathbf{x} \in D$.

(b) Take an arbitrary $\mathbf{x} \in T_{\rho,s} \cap \mathbb{B}_2^p$. To show (35) we estimate $\|\mathbf{x}\|_D$. According to the definition of $\|\cdot\|_D$ in Lemma 3 (a),

$$\begin{aligned} \|\mathbf{x}\|_D &= \sum_{l=1}^L \left[\sum_{i \in I_l} (x_i^*)^2 \right]^{\frac{1}{2}} \\ &= \left[\sum_{i=1}^s (x_i^*)^2 \right]^{\frac{1}{2}} + \left[\sum_{i=s+1}^{2s} (x_i^*)^2 \right]^{\frac{1}{2}} + \sum_{l \geq 3}^L \left[\sum_{i \in I_l} (x_i^*)^2 \right]^{\frac{1}{2}}. \end{aligned} \quad (36)$$

To bound the last term in the inequality above, we first note that for each $i \in I_l$, $l \geq 3$,

$$x_i^* \leq \frac{1}{s} \sum_{j \in I_{l-1}} x_j^* \quad \text{and} \quad \left[\sum_{i \in I_l} (x_i^*)^2 \right]^{1/2} \leq \frac{1}{\sqrt{s}} \sum_{j \in I_{l-1}} x_j^*.$$

Summing up over $l \geq 3$ yields

$$\sum_{l \geq 3}^L \left[\sum_{i \in I_l} (x_i^*)^2 \right]^{1/2} \leq \frac{1}{\sqrt{s}} \sum_{l \geq 2} \sum_{j \in I_l} x_j^*.$$

Since $\mathbf{x} \in T_{\rho,s} \cap \mathbb{B}_2^p$, it holds $\|\mathbf{x}\|_2 \leq 1$ and there is $S \subset [p]$, $\#S = s$, such that $\|\mathbf{x}_S\|_2 > \rho/\sqrt{s} \|\mathbf{x}_{S^c}\|_1$. Then

$$\sum_{l \geq 2} \sum_{i \in I_l} x_i^* \leq \|\mathbf{x}_{S^c}\|_1 < \frac{\sqrt{s}}{\rho} \|\mathbf{x}_S\|_2 \leq \frac{\sqrt{s}}{\rho} \left[\sum_{i=1}^s (x_i^*)^2 \right]^{1/2}$$

and

$$\sum_{l \geq 3}^L \left[\sum_{i \in I_l} (x_i^*)^2 \right]^{1/2} \leq \rho^{-1} \left[\sum_{i=1}^s (x_i^*)^2 \right]^{1/2}.$$

Applying the last estimate to (36) and taking into account that $\|x\|_2^2 \leq 1$ we derive that

$$\begin{aligned} \|\mathbf{x}\|_D &\leq (1 + \rho^{-1}) \left[\sum_{i=1}^s (x_i^*)^2 \right]^{1/2} + \left[\sum_{i=s+1}^{2s} (x_i^*)^2 \right]^{1/2} \\ &\leq (1 + \rho^{-1}) \left[\sum_{i=1}^s (x_i^*)^2 \right]^{1/2} + \left[1 - \sum_{i=1}^s (x_i^*)^2 \right]^{1/2}. \end{aligned}$$

Set $a = \left[\sum_{i=1}^s (x_i^*)^2 \right]^{1/2}$. The maximum of the function

$$f(a) := (1 + \rho^{-1})a + \sqrt{1 - a^2}, \quad 0 \leq a \leq 1,$$

is attained at the point

$$a = \frac{1 + \rho^{-1}}{\sqrt{1 + (1 + \rho^{-1})^2}}$$

and is equal to $\sqrt{1 + (1 + \rho^{-1})^2}$. Thus for any $\mathbf{x} \in W$ it holds

$$\|\mathbf{x}\|_D \leq \sqrt{1 + (1 + \rho^{-1})^2},$$

which proves (35). □

Lemma 3 (b) implies

$$\ell(T_{\rho,s} \cap \mathbb{B}_2^p) \leq \sqrt{1 + (1 + \rho^{-1})^2} \ell(D). \quad (37)$$

Lemma 4. *The Gaussian width of the set D defined by (34) satisfies*

$$\ell(D) \leq \sqrt{2s \ln \frac{ep}{s}} + \sqrt{s}.$$

Proof. The supremum of the linear functional $\langle \mathbf{g}, \mathbf{x} \rangle$ over D is achieved at an extreme point, i.e., at an $\mathbf{x} \in S^{p-1}$ with $\#\text{supp } \mathbf{x} \leq s$. Hence, by Hölder's inequality

$$\ell(D) = \mathbb{E} \sup_{\mathbf{x} \in D} \langle \mathbf{g}, \mathbf{x} \rangle = \mathbb{E} \sup_{\substack{\|\mathbf{x}\|_2=1, \\ \#\text{supp } \mathbf{x} \leq s}} \langle \mathbf{g}, \mathbf{x} \rangle \leq \mathbb{E} \max_{S \subset [p], \#S=s} \|\mathbf{g}_S\|_2.$$

An estimate on the maximum squared ℓ_2 -norm of a sequence of standard Gaussian random vectors (see e.g. [31, Lemma 3.2] or [17, Proposition 8.2]) gives

$$\ell(D) \leq \sqrt{\mathbb{E} \max_{S \subset [p], \#S=s} \|\mathbf{g}_S\|_2^2} \leq \sqrt{2 \ln \binom{p}{s}} + \sqrt{s} \leq \sqrt{2s \ln \frac{ep}{s}} + \sqrt{s}.$$

The last inequality follows from the fact that $\binom{p}{s} \leq \left(\frac{ep}{s}\right)^s$, see e.g. [17, Lemma C.5]. \square

of Theorem 2. Expressions (33), (37) and Lemma 4 show that

$$\begin{aligned} \ell\left(\Omega\left(W_{\rho,s} \cap \mathbb{S}^{d-1}\right)\right) &\leq \sqrt{B[1+(1+\rho^{-1})^2]} \ell(D) \\ &\leq \sqrt{B[1+(1+\rho^{-1})^2]} \left(\sqrt{2s \ln \frac{ep}{s}} + \sqrt{s}\right). \end{aligned} \quad (38)$$

Set $t = \sqrt{2 \ln(\varepsilon^{-1})}$. The fact that $E_m \geq m/\sqrt{m+1}$ along with condition (7) yields

$$E_m \geq \frac{1}{\sqrt{A}} \ell\left(\Omega\left(W_{\rho,s} \cap \mathbb{S}^{d-1}\right)\right) + t.$$

The monotonicity of probability and Theorem 5 imply

$$\mathbb{P}\left(\inf \|\mathbf{M}\mathbf{w}\|_2 > 0 : \mathbf{w} \in W_{\rho,s} \cap \mathbb{S}^{d-1}\right) \geq 1 - e^{-\frac{t^2}{2}} = 1 - \varepsilon,$$

which guarantees that with probability at least $1 - \varepsilon$

$$\|\Omega_{\Lambda^c} \mathbf{w}\|_2 < \frac{\rho}{\sqrt{s}} \|\Omega_{\Lambda} \mathbf{w}\|_1$$

for all $\mathbf{w} \in \ker \mathbf{M} \setminus \{\mathbf{0}\}$ and any $\Lambda \subset [p]$ with $\#\Lambda = p - s$, see (32). This means that \mathbf{M} satisfies the ℓ_2 -stable Ω -null space property of order s . Finally, we apply Theorem 9. \square

Finally, we extend to robustness of the recovery with respect to perturbations of the measurements.

Theorem 11. *Let $\mathbf{M} \in \mathbb{R}^{m \times d}$ be a Gaussian random matrix, $0 < \rho < 1$, $0 < \varepsilon < 1$ and $\tau > 1$. If*

$$\frac{m^2}{m+1} \geq \frac{2(1+(1+\rho^{-1})^2)\tau^2 B}{(\tau-1)^2 A} s \left(\sqrt{\ln \frac{ep}{s}} + \frac{1}{\sqrt{2}} + \sqrt{\frac{A \ln(\varepsilon^{-1})}{Bs(1+(1+\rho^{-1})^2)}} \right)^2, \quad (39)$$

then with probability at least $1 - \varepsilon$ for every vector $\mathbf{x} \in \mathbb{R}^d$ and perturbed measurements $\mathbf{y} = \mathbf{\Omega}\mathbf{x} + \mathbf{w}$ with $\|\mathbf{w}\|_2 \leq \eta$ a minimizer $\hat{\mathbf{x}}$ of (5) approximates \mathbf{x} with ℓ_2 -error

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{2(1 + \rho)^2}{\sqrt{A}(1 - \rho)} \frac{\sigma_s(\mathbf{\Omega}\mathbf{x})_1}{\sqrt{s}} + \frac{2\tau\sqrt{2B}(3 + \rho)}{\sqrt{m}\sqrt{A}(1 - \rho)}\eta.$$

Proof. Condition (39) together with $E_m \geq m/\sqrt{m+1}$ imply

$$E_m \left(1 - \frac{1}{\tau}\right) \geq \sqrt{\frac{B}{A}(1 + (1 + \rho^{-1})^2)} \left(\sqrt{2s \ln \frac{ep}{s}} + \sqrt{s}\right) + \sqrt{2 \ln(\varepsilon^{-1})},$$

which is equivalent to

$$E_m - \sqrt{\frac{B}{A}(1 + (1 + \rho^{-1})^2)} \left(\sqrt{2s \ln \frac{ep}{s}} + \sqrt{s}\right) - \sqrt{2 \ln(\varepsilon^{-1})} \geq \frac{E_m}{\tau}.$$

Taking into account (38) we may conclude

$$E_m - \frac{1}{\sqrt{A}} \ell \left(\mathbf{\Omega} \left(W_{\rho,s} \cap \mathbb{S}^{d-1} \right) \right) - \sqrt{2 \ln(\varepsilon^{-1})} \geq \frac{E_m}{\tau} \geq \frac{1}{\tau} \sqrt{\frac{m}{2}}.$$

Then according to Theorem 5

$$\mathbb{P} \left(\inf \|\mathbf{M}\mathbf{w}\|_2 > \frac{\sqrt{m}}{\tau\sqrt{2}} : \mathbf{w} \in W_{\rho,s} \cap \mathbb{S}^{d-1} \right) \geq 1 - \varepsilon.$$

This means that for any $\mathbf{w} \in \mathbb{R}^d$ such that $\|\mathbf{M}\mathbf{w}\|_2 \leq \frac{\sqrt{m}}{\tau\sqrt{2}}\|\mathbf{w}\|_2$ and any set $\Lambda \subset [p]$ with $\#\Lambda \geq p - s$ it holds with probability at least $1 - \varepsilon$

$$\|\mathbf{\Omega}_{\Lambda^c}\mathbf{w}\|_2 < \frac{\rho}{\sqrt{s}} \|\mathbf{\Omega}_{\Lambda}\mathbf{w}\|_1.$$

For the remaining vectors $\mathbf{w} \in \mathbb{R}^d$, we have $\|\mathbf{M}\mathbf{w}\|_2 > \frac{\sqrt{m}}{\tau\sqrt{2}}\|\mathbf{w}\|_2$, which together with the fact that $\mathbf{\Omega}$ is a frame with upper frame bound B leads to

$$\|\mathbf{\Omega}_{\Lambda^c}\mathbf{w}\|_2 \leq \|\mathbf{\Omega}\mathbf{w}\|_2 \leq \sqrt{B}\|\mathbf{w}\|_2 < \frac{\tau\sqrt{2B}}{\sqrt{m}}\|\mathbf{M}\mathbf{w}\|_2.$$

Thus, for any $\mathbf{w} \in \mathbb{R}^d$,

$$\|\mathbf{\Omega}_{\Lambda^c}\mathbf{w}\|_2 < \frac{\rho}{\sqrt{s}} \|\mathbf{\Omega}_{\Lambda}\mathbf{w}\|_1 + \frac{\tau\sqrt{2B}}{\sqrt{m}} \|\mathbf{M}\mathbf{w}\|_2.$$

Finally, we apply Theorem 10. □

4 Numerical experiments

In this section we present the results of numerical experiments on synthetic data performed in Matlab using the `cvx` package.

For the first set of experiments we constructed tight frames $\mathbf{\Omega}$ as an orthonormal basis of the range of the matrix the rows of which were drawn randomly and independently from \mathbb{S}^{d-1} . In order

to obtain also non-tight frames we simply varied the norms of the rows of $\mathbf{\Omega}$. As dimensions for the analysis operator, we have chosen $d = 200$ and $p = 250$. The maximal number of zeros that can be achieved in the analysis representation $\mathbf{\Omega}\mathbf{x}$ is less than d , since otherwise $\mathbf{x} = 0$. Therefore, the sparsity level of $\mathbf{\Omega}\mathbf{x}$ was always greater than 50. For each trial we fixed a cosparsity l (resulting in the sparsity $s = p - l$) and selected at random l rows of an analysis operator $\mathbf{\Omega}$ that constitute the cosupport Λ of the signal. To produce a signal \mathbf{x} we constructed a basis \mathbf{B} of $\ker \mathbf{\Omega}_\Lambda$, drew a coefficient vector \mathbf{c} from a normalized standard Gaussian distribution and set $\mathbf{x} = \mathbf{B}\mathbf{c}$. We ran the algorithm and counted the number of times the signal was recovered correctly out of 70 trials. A reconstruction error of less than 10^{-5} was considered as a successful recovery. The curves in Figure 1 depict the relation between the number of measurements and the sparsity level such that the recovery was successful at least 98% of the time. Each point on the line corresponds to the maximal sparsity level that could be achieved for the given number of measurements.

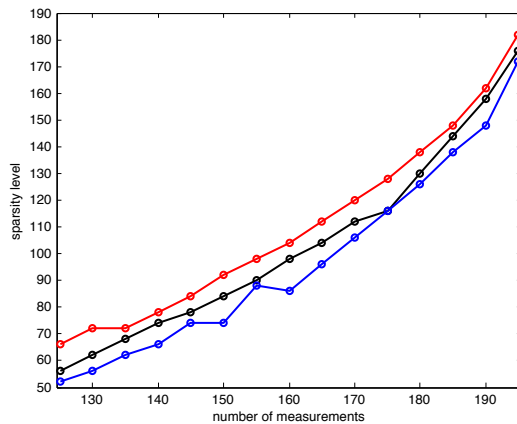


Figure 1: Recovery for different analysis operators. The red curve corresponds to a tight frame, the black one has frame bound ratio B/A of 13.1254 and for the blue one $B/A = 45.7716$.

The experiments clearly show that analysis ℓ_1 -minimization works very well for recovering cosparse signals from Gaussian measurements. (Note that only a comparison of the experiments with the nonuniform recovery guarantees make sense.) The frame bound ratio B/A indeed influences the performance of the recovery algorithm (4) – although the degradation with increasing value of B/A is less dramatic than indicated by our theorems. The reason for this may be that the theorems give estimates for the worst case, while the experiments can only reflect the typical behavior.

Acknowledgements

M. Kabanava and H. Rauhut acknowledge support by the Hausdorff Center for Mathematics, University of Bonn, and by the European Research Council through the grant StG 258926.

References

- [1] Aldroubi A., Chen X., Powell A. M.: Perturbations of measurement matrices and dictionaries in compressed sensing. *Appl. Comput. Harmon. Anal.* 33(2), 282–291 (2012)

- [2] Amelunxen D., Lotz M., McCoy M. B., Tropp J. A.: Living on the edge: A geometric theory of phase transitions in convex optimization. to appear in Inform. Inference
- [3] Boyd S., Vandenberghe L.: Convex optimization. Cambridge University Press, Cambridge (2004)
- [4] Cai J.-F., Osher S., Shen Z.: Split Bregman methods and frame based image restoration. Multiscale Model. Simul. 8(2), 337–369 (2009/10)
- [5] Candès E. J., Donoho D. L.: New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. Comm. Pure Appl. Math. 57(2), 219–266 (2004)
- [6] Candès E. J., Tao J., T., Romberg J. K.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory 52(2), 489–509 (2006)
- [7] Candès E. J., Eldar Y. C., Needell D., Randall P.: Compressed sensing with coherent and redundant dictionaries. Appl. Comput. Harmon. Anal. 31(1), 59–73 (2011)
- [8] Chan T., Shen J.: Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods. SIAM (2005)
- [9] Chandrasekaran V., Recht B., Parrilo P. A., Willsky A. S.: The Convex Geometry of Linear Inverse Problems. Found. Comput. Math. 12(6) 805–849 (2012)
- [10] Chen S.S., Donoho D.L., Saunders M.A.: Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* 20(1), 33–61 (1998)
- [11] Cohen A., Dahmen W., DeVore R.A.: Compressed sensing and best k-term approximation J. Amer. Math. Soc. 22(1), 211–231 (2009)
- [12] Donoho D.L.: Compressed sensing. IEEE Trans. Inform. Theory 52(4), 1289–1306 (2006)
- [13] Donoho D. L., Tanner J.: Counting faces of randomly-projected polytopes when the projection radically lowers dimension. J. Amer. Math. Soc. 22(1), 1–53 (2009)
- [14] Donoho D. L., Tanner J.: Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 367(1906), 4273–4293 (2009)
- [15] Elad M., Milanfar P., Rubinstein R.: Analysis versus synthesis in signal priors. Inverse Problems 23(3), 947–968 (2007)
- [16] Foucart S.: Stability and robustness of ℓ_1 -minimizations with Weibull matrices and redundant dictionaries. Linear Algebra Appl. 441, 4–21 (2014)
- [17] Foucart S., Rauhut H.: A Mathematical Introduction to Compressive Sensing. Appl. Numer. Harmon. Anal. Birkhäuser, Boston (2013)
- [18] Giryes R., Nam S., Elad M., Gribonval R., Davies M.E.: Greedy-Like Algorithms for the Cospase Analysis Model. Lin. Alg. Appl. 441, 22–60 (2014)

- [19] Gordon Y.: On Milman’s inequality and random subspaces which escape through a mesh in \mathbf{R}^n . In Geometric aspects of functional analysis (1986/87), volume 1317 of Lecture Notes in Math. 84–106. Springer, Berlin (1988)
- [20] Gribonval R., Nielsen M.: Sparse representations in unions of bases IEEE Trans. Inform. Theory 49(12), 3320–3325 (2003)
- [21] Gröchenig K.: Foundations of Time-Frequency Analysis. Appl. Numer. Harmon. Anal. Birkhäuser Boston, Boston, MA (2001)
- [22] Kabanava M., Rauhut H., Zhang H.: Robust analysis ℓ_1 -recovery from Gaussian measurements and total variation minimization. <http://arxiv.org/abs/1407.7402>
- [23] Ledoux M.: The Concentration of Measure Phenomenon. AMS (2001)
- [24] Ledoux M., Talagrand M.: Probability in Banach Spaces. Springer-Verlag, Berlin, Heidelberg, NewYork (1991)
- [25] Liu Y., Mi T., Li Sh.: Compressed sensing with general frames via optimal-dual-based ℓ_1 -analysis. IEEE Transactions on information theory 58(7), 4201–4214 (2012)
- [26] Mallat S.: A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press (2008)
- [27] Massart P.: Concentration Inequalities and Model Selection. volume 1896 of Lecture Notes in Mathematics, Springer (2007)
- [28] Mendelson S., Pajor A., Tomczak-Jaegermann N.: Reconstruction and subgaussian operators in asymptotic geometric analysis. Geom. Funct. Anal. 17(4), 1248–1282 (2007)
- [29] Nam S., Davies M.E., Elad M., Gribonval R.: The cospase analysis model and algorithms. Appl. Comput. Harmon. Anal. 34(1), 30–56 (2013)
- [30] Needell D., Ward R.: Stable image reconstruction using total variation minimization. <http://arxiv.org/abs/1202.6429>
- [31] Rao N., Recht B., Nowak R.: Tight measurement bounds for exact recovery of structured sparse signals. In Proceedings of AISTATS (2012)
- [32] Ron A., Shen Z.: Affine systems in $L_2(\mathbf{R}^d)$: the analysis of the analysis operator. J. Funct. Anal. 148(2), 408–447 (1997)
- [33] Rudelson M., Vershynin R.: On sparse reconstruction from Fourier and Gaussian measurements. Comm. Pure Appl. Math. 61(8), 1025–1045 (2008)
- [34] Selesnick I., Figueiredo M.: Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. Proceedings of SPIE 7446, p. 74460D (2009)
- [35] Vaiter S., Peyré G., Dossal Ch., Fadili J.: Robust sparse analysis regularization. IEEE Transactions on information theory 59(4), 2001–2016 (2013)