Cosparsity in Compressed Sensing

Maryia Kabanava, Holger Rauhut

Abstract Analysis ℓ_1-recovery is a strategy of reconstructing a signal that is sparse in some transform domain from incomplete observations. In this chapter we give an overview of the analysis sparsity model and present theoretical conditions that guarantee successful recovery of corresponding signals from noisy measurements. We derive a bound on the number of Gaussian and subgaussian measurements by examining the provided theoretical guarantees under the additional assumption that the transform is generated by a frame, which means that there are just few non-zero inner products of the signal with the frame elements.

1 Introduction

As already outlined in this book (see in particular Chapter 1), compressed sensing aims at acquiring signals from undersampled and possibly corrupted measurements. In mathematical terms, the available data about a signal $x \in \mathbb{R}^n$ is given by a set of measurements

$$y = Ax + e, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ with $m \ll n$ is the sensing matrix and $e \in \mathbb{R}^m$ represents a noise vector. Since this system is undetermined it is hopeless to recover x from y without additional information. The key idea is to take into account prior knowledge about the structure of x.

Maryia Kabanava,
RWTH Aachen University, Lehrstuhl C für Mathematik (Analysis), Templergraben 55, 52062 Aachen, Germany, e-mail: kabanava@mathc.rwth-aachen.de

Holger Rauhut,
RWTH Aachen University, Lehrstuhl C für Mathematik (Analysis), Templergraben 55, 52062 Aachen, Germany, e-mail: rauhut@mathc.rwth-aachen.de

The standard assumption in compressed sensing is that the signal is sparse in some orthonormal basis (as outlined in Chapter 1), which means that it can be represented as a linear combination of only few basis elements. This setting corresponds to the *synthesis* sparsity model. This chapter is concerned with a more general sparsity model, where one assumes that the signal is sparse after a possibly redundant transform, see for instance [33, 34, 43, 6, 16] for initial papers on this subject. This analysis sparsity model – also called cosparsity model – leads to more flexibility in the modeling of sparse signals. Many sparse recovery methods can be adapted to this setting including convex relaxation leading to analysis ℓ_1-minimization (see below) and greedy-like methods [34, 22]. Relevant analysis operators can be generated by the discrete Fourier transform, wavelet [31, 36, 39], curvelet [5] or Gabor transforms [25]. The popular method of total variation minimization [38, 9, 35, 4] corresponds to analysis with respect to a difference operator.

This chapter gives an overview on the analysis sparsity model and its use in compressed sensing. We will present in particular recovery guarantees for ℓ_1-minimization including versions of the null space property and the restricted isometry property. Moreover, we give estimates on the number of measurements required for (approximate) recovery of signals using random Gaussian and subgaussian measurements. Parts of these results (or their proofs) are new and have not appeared elsewhere in the literature yet.

Notation: We use Ω_Λ to refer to a submatrix of Ω with the rows indexed by Λ (we emphasize that our notation differs from the general notation of the book, where this is rather the submatrix corresponding to the columns of Ω); α_Λ stands for the vector whose entries indexed by Λ coincide with the entries of α and the rest are filled by zeros. On some occasions with a slight abuse of notation we refer to α_Λ as an element of $\mathbb{R}^{|\Lambda|}$. We use $[p]$ to denote the set of all natural numbers not exceeding p, i.e., $[p] = \{1,2,\ldots,p\}$. The sign of a real number $r \neq 0$ is $\operatorname{sgn}(r) = \frac{r}{|r|}$. For a vector $\alpha \in \mathbb{R}^p$ we define its sign vector $\operatorname{sgn}(\alpha) \in \mathbb{R}^p$ by

$$(\operatorname{sgn}(\alpha))_i = \begin{cases} \frac{\alpha_i}{|\alpha_i|}, & \text{for all } i \text{ such that } \alpha_i \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The operator norm of a matrix A is given by $\|A\|_{2\to 2} := \sup_{\|x\|_2 \le 1} \|Ax\|_2$; A^T is the transpose of A. The orthogonal complement of a subspace $S \subset \mathbb{R}^p$ is denoted by $S^\perp$. The orthogonal projection onto S is performed by the operator $\mathcal{P}_S$. The notation B_2^p stands for the unit ball with respect to the ℓ_2-norm and S^{n-1} is the unit sphere in $\mathbb{R}^n$.

Acknowledgement

M. Kabanava and H. Rauhut acknowledge support by the European Research Council through the grant StG 258926.

2 Analysis vs. synthesis sparsity

We recall from Chapter 1 that a vector $x \in \mathbb{R}^n$ is k-sparse, if $\|x\|_0 = |\{\ell : x_\ell \neq 0\}| \leq k$. For signals arising in applications it is more common to have a sparse expansion in some basis or dictionary rather than being sparse themselves This means, that for a matrix $D \in \mathbb{R}^{n\times q}$, $q \geq n$, whose columns form a so-called dictionary of $\mathbb{R}^n$ (a spanning set), x can be represented as

$$x = D\alpha, \;\; \alpha \in \mathbb{R}^q,$$

where α is k-sparse. It is common to choose an orthogonal matrix $D \in \mathbb{R}^{n\times n}$, so that we have sparsity with respect to an orthonormal basis. However, also a redundant frame may be used. Here, we "synthesize" x from a few columns d_j of D, which is the reason why this is also called the synthesis sparsity model. The set of all k-sparse signals can be described as

$$\bigcup_{T\subset[n]:|T|=k} V_T = \bigcup_{T\subset[n]:|T|=k} \operatorname{span}\{d_j : j \in T\}, \tag{2}$$

i.e., it is a union of subspaces which are generated by k columns of D.

The analysis sparsity model assumes that Ωx is (approximately) sparse, where $\Omega \in \mathbb{R}^{p\times n}$ is a so-called analysis operator. Denoting the rows of Ω by $\omega_j \in \mathbb{R}^n$, $j = 1,\dots,p$, the entries of Ωx are given as $\langle \omega_j, x\rangle$, $j = 1,\dots,p$, i.e., we analyze x by taking inner products with the ω_j. If Ωx is k-sparse, then x is called *ℓ-cosparse*, where the number $\ell := p-k$ is referred to as *cosparsity* of x. The index set of the zero entries of Ωx is called the *cosupport* of x. The motivation to work with the cosupport rather than the support in the context of analysis sparsity is that it is the location of the zero-elements which define a corresponding subspace. In fact, if Λ is the cosupport of x, then

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

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

$$\bigcup_{\Lambda\subset[p]:\#\Lambda=\ell} W_\Lambda, \tag{3}$$

where W_Λ denotes the orthogonal complement of the linear span of $\{\omega_j : j \in \Lambda\}$. In this sense, the analysis sparsity model falls into the larger class of union of subspaces models [2].

When $\Omega \in \mathbb{R}^{n\times n}$ is an orthogonal matrix, then the analysis sparsity model coincides with the synthesis sparsity model. More precisely, if $x = \Omega^T\alpha$ for a k-sparse $\alpha \in \mathbb{R}^n$ (so taking $D = \Omega^T$ as the basis) then $\Omega x = \alpha$ is k-sparse as well, meaning that x is $(n-k)$-cosparse.

Taking $\Omega \in \mathbb{R}^{p\times n}$ with $p > n$, the analysis sparsity model is more general and offers more flexibility than synthesis sparsity. The following considerations on dimensionality and number of subspaces in (2) and (3) illustrate this point, see also [34, Section 2.3].

Let us first consider the generic case that both the rows $\omega_j \in \mathbb{R}^n$, $j = 1, \ldots, p$, of the analysis operator as well as the columns $d_j \in \mathbb{R}^n$, $j = 1, \ldots, q$, of the dictionary are in general linear position, i.e., any collection of at most n of these vectors are linearly independent. Then the following table comparing the s-sparse model (2) and the ℓ-cosparse model (3) applies.

model	subspaces	no. subspaces	subspace dim.
synthesis	$V_T := \operatorname{span}\{d_j, j \in t\}$	$\binom{q}{k}$	k
analysis	$W_\Lambda := \operatorname{span}\{\omega_j, j \in \Lambda\}^\perp$	$\binom{p}{\ell}$	$n - \ell$

As suggested in [34] one way of comparing the two models is to consider an ℓ-cosparse analysis model and an $(n-\ell)$-sparse synthesis model so that the corresponding subspaces have the same dimensions. If for instance $\ell = n-1$ so that the subspace dimension is $n - \ell = 1$, there are $\binom{q}{1} = q$ subspaces in the synthesis model, while there are $\binom{p}{n-1}$ subspaces in the analysis sparsity model. Typically p is somewhat larger than n, and if q is not extremely large, the number of subspaces of dimension 1 is much larger for the analysis sparsity model than for the synthesis sparsity model. Or in other words, if one is looking for a synthesis sparsity model having the same one-dimensional subspaces as in a given analysis sparsity model then one needs $q = \binom{p}{n-1}$ many dictionary elements – usually way too many to be handled efficiently. More generally speaking, the analysis sparsity model contains many more low-dimensional subspaces than the synthesis sparsity model, but the situation reverses for high-dimensional subspaces [34].

As another difference to the synthesis sparsity model where any value of the sparsity k between 1 and n can occur, the dimension of W_Λ is restricted to a value between 0 and n and in the case of "generic" analysis operator (i.e., any set of at most n rows of Ω is linearly independent) the cosparsity is restricted to values between $p-n$ and p so that the sparsity of Ωx must be at least $p-n$ for a non-trivial vector x. However, if one considers only approximate sparsity (see below), then also smaller values of the sparsity make sense.

Sometimes it is desired not too have too many low dimensional subspaces in the model and then it is beneficial if there are linear dependencies among the rows of the analysis operator Ω. In this case, the above table does no longer apply and the number of subspace may be significantly smaller. A particular situation where this happens is connected to the popular method of total variation. Here the analysis operator is a one or two dimensional difference operator (many linear dependencies appear for the two-dimensional case). In the one dimensional case, let $\Omega = \Omega_{DIF} \in \mathbb{R}^{(n-1)\times n}$ be the matrix with entries

$$\Omega_{DIF} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix}$$

Analysis ℓ_1-minimization with the analysis operator given by Ω_{DIF} promotes piecewise constant signals with sparse gradient. In the two dimensional case this operator is applied in the vertical and horizontal direction separately.

Another important case that we will consider in more detail in this Chapter appears when the rows ω_j of $\Omega \in \mathbb{R}^{p\times n}$ form a frame [14, 8, 12, 25], i.e., there exist constants $0 < a \le b < \infty$ such that

$$a\|x\|_2^2 \le \|\Omega x\|_2^2 = \sum_{j=1}^{p} |\langle \omega_j, x\rangle|^2 \le b\|x\|_2^2. \tag{4}$$

Clearly, in our finite dimensional case such constants always exist if the ω_j span $\mathbb{R}^n$. For simplicity, we will often refer to Ω itself as a frame. If $a = b$, then Ω is called a tight frame. Frames are more general than orthonormal bases and allow for stable expansions. They are useful, for instance, when orthonormal bases with certain properties do not exist (see e.g. the Balian-Low theorem in [1, 25]). Moreover, their redundancy can be useful for tasks like error corrections in transmission of information etc.

Any signal x is uniquely determined by its frame coefficients Ωx. To reconstruct x from Ωx we can make use of the canonical dual frame. Its elements are given by the columns of the Moore–Penrose pseudo inverse $\Omega^\dagger = (\Omega^T\Omega)^{-1}\Omega^T$ and for any x we have $x = \Omega^\dagger(\Omega x)$. Lower and upper frame bounds of the canonical dual frame are b^{-1} and a^{-1}, respectively.

Particular frames of importance include Gabor frames [25], wavelet frames [31, 36, 39], shearlet [26] and curvelet frames [5].

In practice, signals are usually not exactly sparse or cosparse. In order to measure the error of approximation we recall that the error of k-term approximation of $x \in \mathbb{R}^n$ in ℓ_1 is defined as

$$\sigma_k(x)_1 := \inf_{z:\|z\|_0 \le k} \|x - z\|_1.$$

In the cosparse case we use the quantity $\sigma_k(\Omega x)_1$ as a measure of how close x is to being $(p-k)$-cosparse. We remark that although for generic analysis operators $\Omega \in \mathbb{R}^{p\times n}$, the vector Ωx has at least $p-n$ nonzero entries (unless x is trivial), the approximation error $\sigma_k(\Omega x)_1$ may nevertheless become small for values of $k < p-n$. A particular case, where this occurs arises in the setting of localized frames [18], where the Gramian $\Omega\Omega^T$ has quick off-diagonal decay. It is shown in [6, Section 1.4] that Ωx has small approximation error $\sigma_k(\Omega x)_1$ if $x = \Omega^T\alpha$ for an (approximately) k-sparse vector α.

3 Recovery of cosparse signals

We now turn to the compressed sensing problem of recovering an (approximately) cosparse vector $x \in \mathbb{R}^n$ from underdetermined linear measurements

$$y = Ax,$$

where $A \in \mathbb{R}^{m \times n}$ is a measurement matrix with $m < n$. Let $\Omega \in \mathbb{R}^{p \times n}$ be the analysis operator generating the analysis cosparsity model.

In analogy with the standard sparsity case (synthesis sparsity model) outlined in Chapter 1, Section 1.3, one might start with the ℓ_0-minimization problem

$$\min_{z \in \mathbb{R}^n} \|\Omega z\|_0 \text{ subject to } Az = y. \tag{5}$$

However, this combinatorial optimization problem is again NP-hard in general. As an alternative, we may use its ℓ_1-relaxation

$$\min_{z \in \mathbb{R}^n} \|\Omega z\|_1 \text{ subject to } Az = y \tag{6}$$

or in the noisy case

$$\min_{z \in \mathbb{R}^n} \|\Omega z\|_1 \text{ subject to } \|Az - y\|_2 \leq \varepsilon. \tag{7}$$

Alternative approaches include greedy-type algorithms such as Greedy Analysis Pursuit (GAP) [33, 34] or thresholding-based methods such as iterative hard thresholding, see [22].

We start by presenting conditions under which the solution of (6) coincides with the solution of (5), so that the original cosparse vector is recovered. We discuss versions of the null space property and the restricted isometry property. When the measurement matrix $A \in \mathbb{R}^{m \times n}$ is taken at random (as usual in compressed sensing), then an analysis of these concepts leads to so-called uniform recovery bounds stating that with a random draw of the measurement matrix one can with high probability recover all k-sparse vectors under a certain lower bound on the number of measurements. In contrast, nonuniform recovery results state that a given (fixed) cosparse vector can be recovered from a random draw of the measurement matrix with a certain probability. Such nonuniform guarantees can be derived with conditions that depend both on the matrix A and the vector x to be recovered. We will state such conditions later on in this section.

3.1 Analysis null space property

As in the standard synthesis sparsity case, the null space property of the measurement matrix A characterizes recovery via analysis ℓ_1-minimization. In analogy to the null space property described in Section 1.3.2 of the introductory chapter, we say that given an analysis operator $\Omega \in \mathbb{R}^{p\times n}$, a measurement matrix $A \in \mathbb{R}^{m\times n}$ satisfies the Ω-null space property of order k if, for all subsets $\Lambda \in [p]$ of cardinality $|\Lambda| \geq p-k$, it holds

$$\|\Omega_{\Lambda^c} v\|_1 < \|\Omega_\Lambda v\|_1 \quad \text{for all } v \in \ker A \setminus \{0\}.$$

Analogously to Theorem 1.2 of Chapter 1 it can be shown that every $(p-k)$-cosparse vector can be recovered exactly via analysis ℓ_1-minimization (6). In order to guarantee stable and robust recovery we use the following version of the null space property extending the corresponding notions from the standard synthesis sparsity case [20, Chapter 4].

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

$$\|\Omega_{\Lambda^c} v\|_2 < \frac{\rho}{\sqrt{k}}\|\Omega_\Lambda v\|_1 + \tau\|Av\|_2 \quad \text{for all } v \in \mathbb{R}^n. \tag{8}$$

The following theorem has been shown in [28], similarly to [20, Theorem 4.22]

Theorem 1. *Let $\Omega \in \mathbb{R}^{p\times n}$ be a frame with lower frame bound $a > 0$. Let $A \in \mathbb{R}^{m\times n}$ satisfy the robust ℓ_2-stable Ω-null space property of order k with constant $0 < \rho < 1$ and $\tau > 0$. Then for any $x \in \mathbb{R}^n$ the solution $\hat{x}$ of (7) with $y = Ax + e$, $\|e\|_2 \leq \varepsilon$, approximates the vector x with ℓ_2-error*

$$\|x - \hat{x}\|_2 \leq \frac{2(1+\rho)^2}{\sqrt{a}(1-\rho)}\frac{\sigma_k(\Omega x)_1}{\sqrt{k}} + \frac{2\tau(3+\rho)}{\sqrt{a}(1-\rho)}\varepsilon. \tag{9}$$

We will analyze the stable null space property for Gaussian random matrices A directly in Section 4.4.

3.2 Restricted isometry property

It is a by now classical approach to analyze sparse recovery algorithms via the restricted isometry property. A version for the analysis sparsity called the D-RIP (dictionary RIP) was introduced in [6].

Definition 2. A measurement matrix $A \in \mathbb{R}^{m\times n}$ satisfies the restricted isometry property adapted to $\Omega \in \mathbb{R}^{p\times n}$ (the D-RIP) with constant δ_k if

$$(1-\delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\|x\|_2^2 \tag{10}$$

holds for all $x = \Omega^T \alpha$ with $\|\alpha\|_0 \leq k$.

If $\Omega = \mathrm{Id}$, then we obtain the standard RIP.

In the important case that the rows of Ω form a frame, see (4), we have the following recovery result generalizing the one from the synthesis sparsity case. For the case of tight frames, it has been shown in [6], while the general case can be found in [19, Proposition 9].

Theorem 2. *Let $\Omega \in \mathbb{R}^{p\times n}$ be a frame with frame bounds $a,b > 0$ and $(\Omega^\dagger)^T$ its canonical dual frame. Suppose that the measurement matrix $A \in \mathbb{R}^{m\times n}$ obeys the restricted isometry property with respect to $(\Omega^\dagger)^T$ with constant $\delta_{2k} < \sqrt{a/b}/9$. Let $y = Ax + e$ with $\|e\|_2 \leq \varepsilon$. Then the solution $\hat{x}$ of (6) satisfies*

$$\sqrt{a}\|x-\hat{x}\|_2 \leq c_0 \frac{\sigma_k(\Omega x)_1}{\sqrt{k}} + c_1\varepsilon \tag{11}$$

for constants c_0, c_1 that depend only on δ_{2k}.

For the case that Ω is a one or two dimensional difference operator corresponding to total variation minimization, recovery guarantees have been provided in [35].

3.3 Recovery conditions via tangent cones

The null space property and the restricted isometry property of A guarantee that all cosparse vectors can be recovered via analysis ℓ_1-minimization from measurements obtained by applying A. It is also useful to have recovery conditions that not only depend on A but also on the vector to be recovered. In fact, such conditions are at the basis for the nonuniform recovery guarantees stated later.

This section follows the approach of [10] that works with tangent cones of $\|\Omega \cdot\|_1$ at the vector to be recovered. For fixed $x \in \mathbb{R}^n$ we define the convex cone

$$T(x) = \mathrm{cone}\{z - x : z \in \mathbb{R}^n,\ \|\Omega z\|_1 \leq \|\Omega x\|_1\}, \tag{12}$$

where the notation “cone” stands for the conic hull of the indicated set. The set $T(x)$ consists of the directions from x, which do not increase the value of $\|\Omega x\|_1$. The following result describes a geometric property, which guarantees exact recovery, see Figure 1. It was proved in [28] and is analogous to Proposition 2.1 in [10], see also [20, Theorem 4.35].

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

Proof. For convenience we prove that the condition $\ker A \cap T(x) = \{0\}$ implies recovery. For the other direction we refer to [28].

Suppose there is $z \in \mathbb{R}^n$ such that $Az = Ax$ and $\|\Omega z\|_1 \leq \|\Omega x\|_1$. Then $z - x \in T(x)$ and $z - x \in \ker A$. Since $\ker A \cap T(x) = \{0\}$, we conclude that $z - x = 0$, so that x is the unique minimizer. □

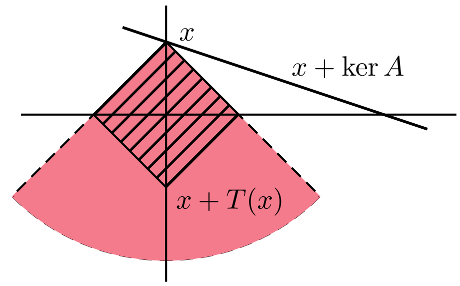

Fig. 1 Geometry of successful recovery. The dashed region corresponds to the set $\{z : \|\Omega z\|_1 \leq \|\Omega x\|_1\}$.

When the measurements are noisy, we use the following criteria for robust recovery [28].

Theorem 4. *Let $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $y = Ax + e$ with $\|e\|_2 \leq \varepsilon$. If*

$$\inf_{\substack{v \in T(x) \\ \|v\|_2 = 1}} \|Av\|_2 \geq \tau \tag{13}$$

for some $\tau > 0$, then a solution $\hat{x}$ of the analysis ℓ_1 minimization problem (7) satisfies

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

Proof. Since $\hat{x}$ is a minimizer of (7), we have $\|\Omega \hat{x}\|_1 \leq \|\Omega x\|_1$ and $\hat{x} - x \in T(x)$. Our assumption (13) implies

$$\|A(\hat{x} - x)\|_2 \geq \tau \|\hat{x} - x\|_2. \tag{14}$$

On the other hand, an upper bound for $\|A\hat{x} - Ax\|_2$ is given by

$$\|A\hat{x} - Ax\|_2 \leq \|A\hat{x} - y\| + \|Ax - y\|_2 \leq 2\varepsilon. \tag{15}$$

Combining (14) and (15) we get the desired estimate. □

3.4 Dual certificates

Another common approach for recovery conditions is based on duality. For the standard synthesis sparsity model corresponding results have been obtained for instance in [21, 41], see also [20, Theorem 4.26–4.33]. In fact, the first contribution to compressed sensing by Candès et al. [7] is based on such an approach.

Apparently, Haltmeier [27] first addressed the problem of robust recovery of a signal by analysis ℓ_1-minimization, when the analysis operator is given by a frame. His result has been generalized in [17], which for our particular case is stated in the following theorem. In order to formulate it, we recall that the subdifferential of a convex function $f:\mathbb{R}^n \to \mathbb{R}$ is defined as

$$\partial f(x) = \{v \in \mathbb{R}^n : f(z) \geq f(x) + \langle z-x, v\rangle \text{ for all } z \in \mathbb{R}^n\}.$$

The subdifferential of the ℓ_1-norm is given as $\partial\|\cdot\|(x) = \{(v_j)_j : v_j \in \partial|\cdot|(x_j)\}$, where the subdifferential of the absolute value function is

$$\partial|\cdot|(u) = \begin{cases} [-1,1] & \text{if } u = 0, \\ \operatorname{sgn}(u) & \text{if } u \neq 0. \end{cases}$$

Theorem 5. *Let $x \in \mathbb{R}^n$ be cosparse with cosupport Λ and let noisy measurements $y = Ax + e$ be given with $\|e\|_2 \leq \varepsilon$. Assume the following:*

1. *There exists $\eta \in \mathbb{R}^m$ (the dual vector) and $\alpha \in \partial\|\cdot\|_1(\Omega x)$ such that*
$$A^T\eta = \Omega^T\alpha, \ \ \textit{with } \|\alpha_\Lambda\|_\infty \leq \kappa < 1. \tag{16}$$
2. *The sensing matrix A is injective on $W_\Lambda = \ker \Omega_\Lambda$ implying that there exists $C_A > 0$ such that*
$$\|Ax\|_2 \geq C_A\|x\|_2, \ \ \textit{for any } \ x \in W_\Lambda. \tag{17}$$

Then any solution $\hat{x}$ of the analysis ℓ_1-minimization problem (7) approximates x with ℓ_2-error

$$\|x - \hat{x}\|_2 \leq C\varepsilon.$$

In the inverse problems community (16) is also referred to as source (range) condition, while (17) is called injectivity condition and they are commonly used to provide error estimates for the sparsity promoting regularizations [3, 24]. In [3] the error is measured by the Bregman distance. The work of Grasmair [24] provides a more general result, where the error is measured in terms of the regularization functional. The first results concerning the ℓ_2-error estimates (as stated above) are given in [27], where the analysis operator is assumed to be a frame. The result [17] extends it to general analysis operators and so-called decomposable norms.

We close this section by showing that the conditions in Theorem 5 are stronger than the tangent cone condition of Theorem 4.

Theorem 6. *Let $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$ with cosupport Λ satisfy (16) and (17) with vectors $\alpha \in \partial\|\cdot\|_1(\Omega x)$, $\|\alpha_\Lambda\|_\infty \leq \kappa < 1$, and $\eta \in \mathbb{R}^m$. Then*

$$\inf_{\substack{v \in T(x) \\ \|v\|_2 = 1}} \|Av\|_2 \geq \tau$$

with

$$\tau = \left(C_A^{-1} + \frac{(C_A + \|A\|_{2\to 2})\,\|\eta\|_2}{C_{\Omega,\Lambda} C_A (1-\kappa)}\right)^{-1},$$

where $C_{\Omega,\Lambda}$ depends only on Ω and Λ.

The proof follows the same lines as the proof of the main result in [17]. We start with the following lemma.

Lemma 1. *Let $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$ with cosupport Λ satisfy (16) and (17) with vectors $\alpha \in \partial\|\cdot\|_1(\Omega x)$, $\|\alpha_\Lambda\|_\infty \le \kappa < 1$, and $\eta \in \mathbb{R}^m$. Then for any $v \in T(x)$*

$$\|\Omega_\Lambda v\|_1 \le \frac{\|\eta\|_2}{1-\kappa}\|Av\|_2.$$

Proof. For $v \in T(x)$, there exists $\beta_\Lambda \in \mathbb{R}^p$ (that is, β_Λ is zero on Λ^c) such that $\|\beta_\Lambda\|_\infty \le 1$ and $\|(\Omega v)_\Lambda\|_1 = \langle(\Omega v)_\Lambda, \beta_\Lambda\rangle$. The subdifferential $\partial\|\cdot\|_1(\Omega x)$ of the ℓ_1-norm at the point Ωx consists of all vectors α, such that any $\beta \in \mathbb{R}^p$ satisfies

$$\|\beta\|_1 \ge \|\Omega x\|_1 + \langle\alpha, \beta - \Omega x\rangle.$$

Since $x \in \mathbb{R}^n$ is cosparse with cosupport Λ, we can explicitly write

$$\partial\|\cdot\|_1(\Omega x) = \{\alpha \in \mathbb{R}^p : \alpha_{\Lambda^c} = \operatorname{sgn}(\Omega x),\ \|\alpha_\Lambda\|_\infty \le 1\}.$$

Taking into account that the vector β_Λ has zero-entries on the index set Λ^c and $\|\beta_\Lambda\|_\infty \le 1$, it follows that $\operatorname{sgn}(\Omega x) + \beta_\Lambda \in \partial\|\cdot\|_1(\Omega x)$. Every $v \in T(x)$ is represented as

$$v = \sum_j t_j v_j,\ \ v_j = z_j - x,\ \ \|\Omega z_j\|_1 \le \|\Omega x\|_1,\ \ t_j \ge 0.$$

Let $\alpha \in \partial\|\cdot\|_1(\Omega x)$, so that $\alpha_{\Lambda^c} = \operatorname{sgn}(\Omega x)$. The definition of the subdifferential implies then that

$$\begin{aligned} 0 &\ge \|\Omega z_j\|_1 - \|\Omega x\|_1 \ge \langle\operatorname{sgn}(\Omega x) + \beta_\Lambda, \Omega(z_j - x)\rangle \\ &= \langle\operatorname{sgn}(\Omega x) + \beta_\Lambda - \alpha, \Omega v_j\rangle + \langle\alpha, \Omega v_j\rangle \ge \langle\beta_\Lambda - \alpha_\Lambda, \Omega_\Lambda v_j\rangle + \langle\alpha, \Omega v_j\rangle. \end{aligned}$$

Multiplying by $t_j \ge 0$ and summing up over all j gives

$$0 \ge \langle\beta_\Lambda - \alpha_\Lambda, \Omega_\Lambda(\sum_j t_j v_j)\rangle + \langle\alpha, \Omega(\sum_j t_j v_j)\rangle = \langle\beta_\Lambda - \alpha_\Lambda, (\Omega v)_\Lambda\rangle + \langle\alpha, \Omega v\rangle.$$

Due to the choice of β_Λ and duality of the ℓ_1-norm and ℓ_∞ norm we obtain

$$0 \ge \|\Omega_\Lambda v\|_1 - \|\alpha_\Lambda\|_\infty \|\Omega_\Lambda v\|_1 + \langle\alpha, \Omega v\rangle,$$

which together with (16) gives

$$\begin{aligned} \|\Omega_\Lambda v\|_1 &\le -\frac{\langle\alpha, \Omega v\rangle}{1 - \|\alpha_\Lambda\|_\infty} = -\frac{\langle\Omega^T\alpha, v\rangle}{1 - \|\alpha_\Lambda\|_\infty} = -\frac{\langle A^T\eta, v\rangle}{1 - \|\alpha_\Lambda\|_\infty} \\ &= -\frac{\langle\eta, Av\rangle}{1 - \|\alpha_\Lambda\|_\infty} \le \frac{\|\eta\|_2}{1-\kappa}\|Av\|_2. \end{aligned}$$

This concludes the proof. □

Proof (of Theorem 6). The idea is to split $v \in T(x)$ into its projections onto the subspace $W_\Lambda = \ker\Omega_\Lambda$ and its complement $W_\Lambda^\perp$. Since we are in finite dimensions, it follows that $\|\Omega_\Lambda w\|_2 \ge C_{\Omega,\Lambda}\|w\|_2$ for all $w \in W_\Lambda^\perp$ for some constant $C_{\Omega,\Lambda}$. Taking into account (17) we obtain

$$\begin{aligned}
\|v\|_2 &\leq \|\mathcal{P}_{W_\Lambda} v\|_2 + \|\mathcal{P}_{W_\Lambda^\perp} v\|_2 \leq C_A^{-1}\|A\mathcal{P}_{W_\Lambda} v\|_2 + \|\mathcal{P}_{W_\Lambda^\perp} v\|_2 \\
&= C_A^{-1}\|A(v - \mathcal{P}_{W_\Lambda^\perp} v)\|_2 + \|\mathcal{P}_{W_\Lambda^\perp} v\|_2 \\
&\leq C_A^{-1}\|Av\|_2 + C_A^{-1}\|A\mathcal{P}_{W_\Lambda^\perp} v\|_2 + \|\mathcal{P}_{W_\Lambda^\perp} v\|_2 \\
&\leq C_A^{-1}\|Av\|_2 + (1 + C_A^{-1}\|A\|_{2\to 2})\|\mathcal{P}_{W_\Lambda^\perp} v\|_2 \\
&\leq C_A^{-1}\|Av\|_2 + (1 + C_A^{-1}\|A\|_{2\to 2})C_{\Omega,\Lambda}^{-1}\|\Omega_\Lambda \mathcal{P}_{W_\Lambda^\perp} v\|_2.
\end{aligned}$$

Since $\Omega_\Lambda \mathcal{P}_{W_\Lambda} v = 0$ and $\Omega_\Lambda \mathcal{P}_{W_\Lambda^\perp} v = \Omega_\Lambda(v - \mathcal{P}_{W_\Lambda} v) = \Omega_\Lambda v$ by definition of W_Λ, the estimate above can be continued to obtain

$$\|v\|_2 \leq C_A^{-1}\|Av\|_2 + \frac{C_A + \|A\|_{2\to 2}}{C_{\Omega,\Lambda} C_A}\|\Omega_\Lambda v\|_2 \leq C_A^{-1}\|Av\|_2 + \frac{C_A + \|A\|_{2\to 2}}{C_{\Omega,\Lambda} C_A}\|\Omega_\Lambda v\|_1.$$

As a final step we apply Lemma 1. □

4 Recovery from random measurements

A main task in compressed sensing is to obtain bounds for the minimal number of linear measurements required to recover a (co-)sparse vector via certain recovery methods, say analysis ℓ_1-minimization. It is up till now open to rigorously prove such guarantees – and in particular, verify the conditions of the previous sections – for deterministic sensing matrix constructions in the optimal parameter regime, see for instance [20, Chapter 6.1] for a discussion. Therefore, we pass to random matrices.

A matrix that is populated with independent standard normal distributed entries (see also Chapter 1) is called a *Gaussian matrix*. We will also consider subgaussian matrices. To this end, we introduce the the ψ_2-norm of a random variable X which is defined as

$$\|X\|_{\psi_2} := \inf\left\{c > 0 : \mathbb{E}\exp\left(|X|^2/c^2\right) \leq 2\right\}.$$

A random variable X is called *subgaussian*, if $\|X\|_{\psi_2} < \infty$. Boundedness of the ψ_2-norm of a random variable X is equivalent to the fact that its tail satisfies $\mathbb{P}(|X| > t) \leq 2e^{-ct^2}$ and its moments $(\mathbb{E}|X|^p)^{1/p}$ do not grow faster than $\sqrt{p}$. Standard examples of subgaussian random variables are Gaussian, Bernoulli and bounded random variables. A matrix with independent mean-zero and variance one subgaussian entries is called a *subgaussian matrix*.

Definition 3. A random vector X in $\mathbb{R}^n$ is called *isotropic*, if $\mathbb{E}|\langle X, x\rangle|^2 = \|x\|_2^2$ for every $x \in \mathbb{R}^n$. A random vector $X \in \mathbb{R}^n$ is *subgaussian*, if $\langle X, x\rangle$ are subgaussian random variables for all $x \in \mathbb{R}^n$. The ψ_2 norm of X is defined as

$$\|X\|_{\psi_2} = \sup_{u \in S^{n-1}} \|\langle X, u \rangle\|_{\psi_2}.$$

A random matrix whose rows are independent, isotropic and subgaussian is called an isotropic subgaussian ensemble. Subgaussian random matrices are examples of such matrices.

We will give an overview on recovery results for analysis ℓ_1-minimization where the analysis operator is a frame and the measurement matrix is Gaussian or more generally comes from an isotropic subgaussian ensemble. We cover both uniform recovery, which is studied via the Ω-null space property or via the D-RIP, as well as nonuniform recovery bounds, where especially for the Gaussian case, it is possible to derive explicit bounds with small constants. For some of these results, new proofs are provided.

4.1 Uniform recovery via the restricted isometry property

We recall from Chapter 1 (Theorem 1.5) that a rescaled Gaussian random matrix $A \in \mathbb{R}^{m\times n}$ satisfies the standard restricted isometry property, i.e. $\delta_k \le \delta$ with probability at least $1-\theta$ provided that $m \ge C\delta^{-2}(k\ln(eN/k) + \ln(2\theta^{-1}))$. A similar estimate, derived for the first time in [6], holds for the D-RIP in (2).

Theorem 7. *Let $A \in \mathbb{R}^{m\times n}$ be a draw of a Gaussian random matrix and let $\Omega \in \mathbb{R}^{p\times n}$ be an analysis operator. If*

$$m \ge C\delta^{-2}(k\ln(ep/k) + \ln(2\theta^{-1})) \tag{18}$$

then with probability at least $1-\theta$ the matrix $\frac{1}{\sqrt{m}}A$ satisfies the restricted isometry property adapted to Ω with constant $\delta_k \le \delta$.

Proof. This is a generalization of the standard restricted isometry property of Gaussian matrices. The proof relies on the concentration of measure phenomenon formulated in Theorem 1.4 of Chapt. 1 and the covering argument presented in the same chapter in Lemma 1.3. The only difference in comparison to the proof of Theorem 1.5 in Chapt. 1 occurs in the step of taking a union bound with respect to all k-dimensional subspaces. In our case, there are $\binom{p}{k} \le \left(\frac{ep}{k}\right)^k$ subspaces, which is reflected in the term $k\ln(ep/k)$. □

The above result including its proof extends to isotropic subgaussian random matrices. An alternative proof approach may be based on Theorem 14 below.

Theorem 8 (Corollary 3.1 of [13]). *Let $A \in \mathbb{R}^{m\times n}$ be a draw of an isotropic subgaussian ensemble and let $\Omega \in \mathbb{R}^{p\times n}$ be an analysis operator. If*

$$m \ge C\delta^{-2}(k\ln(ep/k) + \ln(2\theta^{-1})$$

then with probability at least $1-\theta$ *the matrix* $\frac{1}{\sqrt{m}}A$ *satisfies the restricted isometry property adapted to* Ω *with constant* $\delta_k \leq \delta$.

An extension of the above bound to Weibull matrices has been shown in [19].

Applying the above results for the canonical dual frame $(\Omega^\dagger)^T$ of a frame Ω in combination with Theorem 2 shows that the analysis ℓ_1-program

$$\min \|\Omega z\|_1 \quad \text{subject to } Az = Ax$$

with a random draw of a (sub-)gaussian matrix $A \in \mathbb{R}^{m\times n}$ recovers every ℓ-cosparse vector x with $\ell = p-k$ exactly with high probability provided

$$m \geq \frac{Cb}{a} k \ln(ep/k). \tag{19}$$

The difference to the standard synthesis sparsity case is merely the appearance of p instead of n inside the logarithmic factor as well as the ratio b/a of the frame bounds. Clearly, this ratio is one for a tight frame.

We will return to uniform recovery with Gaussian measurements in Subsection 4.4, where we will study the Ω-null space property directly which allows to give an explicit and small constant in the bound (18) on the number of measurements. The approach relies on techniques that are introduced in the next section concerning nonuniform recovery. (These methods are easier to apply in the nonuniform setting which is the reason why we postpone an analysis of the Ω-null space property to later.)

4.2 Nonuniform recovery from Gaussian measurements

We now turn to nonuniform results for recovery of cosparse signals with respect to a frame being the analysis operator and using Gaussian measurement matrices, which state that a given (fixed) cosparse vector can be recovered with high probability under a certain bound on the number of measurements. We will not qualitatively improve over (19), but we will obtain a very good constant, which is in fact optimal in a certain "asymptotic" sense. The main result stated next appeared in [28], but we give a slightly different proof here, which allows us later to extend this approach also to the subgaussian case.

Theorem 9. *Let* $\Omega \in \mathbb{R}^{p\times n}$ *be a frame with frame bounds* a, $b > 0$ *and* $x \in \mathbb{R}^n$ *be an* ℓ*-cosparse vector and* $k = p-\ell$. *Let* $A \in \mathbb{R}^{m\times n}$ *be a Gaussian random matrix and let noisy measurements* $y = Ax+e$ *be taken with* $\|e\|_2 \leq \varepsilon$. *If for* $0 < \theta < 1$ *and some* $\tau > 0$

$$\frac{m^2}{m+1} \geq 2k\left(\sqrt{\frac{b}{a}\ln\left(\frac{ep}{k}\right)} + \sqrt{\frac{\ln(\theta^{-1})}{k}} + \tau\sqrt{\frac{1}{2k}}\right)^2, \tag{20}$$

then with probability at least $1-\theta$*, every minimizer* $\hat{x}$ *of (7) satisfies*

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

Setting $\varepsilon = 0$ yields exact recovery via (6). Roughly speaking, i.e., for rather large k,m,p the bound (20) reads

$$m > 2\frac{b}{a}k\ln(ep/k)$$

for having recovery with "high probability".

Our proof relies on the recovery condition of Theorem 4 based on tangent cones as well as on convex geometry and Gordon's escape through a mesh theorem [23] and is inspired by [10], see also [20, Chapter 9].

According to (13) the successful recovery of a signal is achieved, when the minimal gain of the measurement matrix over the tangent cone is greater than some positive constant. For Gaussian matrices the probability of this event can be estimated by Gordon's escape through a mesh theorem [23], [20, Theorem 9.21]. To present it formally we introduce some notation. For a set $T \subset \mathbb{R}^n$ we define its Gaussian width by

$$\ell(T) := \mathbb{E} \sup_{x\in T} \langle x, g\rangle, \tag{21}$$

where $g \in \mathbb{R}^n$ is a standard Gaussian random vector. Due to rotation invariance (21) can be written as

$$\ell(T) = \mathbb{E}\,\|g\|_2 \cdot \mathbb{E} \sup_{x\in T} \langle x, u\rangle,$$

where u is uniformly distributed on S^{n-1}. We recall [20, Theorem 8.1] that

$$E_n := \mathbb{E}\,\|g\|_2 = \sqrt{2}\,\frac{\Gamma\left((n+1)/2\right)}{\Gamma\left(n/2\right)}$$

satisfies

$$\frac{n}{\sqrt{n+1}} \leq E_n \leq \sqrt{n},$$

so that up to some factor of order $\sqrt{n}$ the Gaussian width is basically equivalent to the mean width of a set, see Figure 2.

In order to gain more intuition about the Gaussian width we remark that a d-dimensional subspace $U \subset \mathbb{R}^n$ intersected with the sphere S^{n-1} satisfies $\ell(U \cap S^{n-1}) \sim \sqrt{d}$ so that for a subset T of the sphere the quantity $\ell(T)^2$ can somehow be interpreted as its dimension (although this interpretation should be handled with care.)

Next we state the version of Gordon's escape through a mesh theorem from [20, Theorem 9.21].

Theorem 10 (Gordon's escape through a mesh). *Let* $A \in \mathbb{R}^{m\times n}$ *be a Gaussian random matrix and* T *be a subset of the unit sphere* $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.

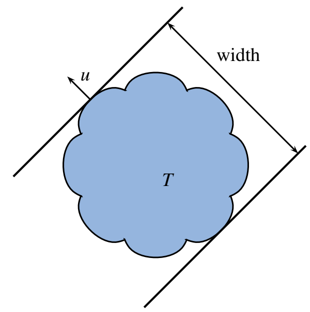

Fig. 2 The mean width of a set T in the direction u. When T is symmetric, $2\sup_{x\in T}\langle x,u\rangle = \sup_{x\in T}\langle x,u\rangle - \inf_{z\in T}\langle z,u\rangle$, which corresponds to the smallest distance between two hyperplanes orthogonal to the direction u, such that T is contained between them.

Then, for $t>0$, it holds

$$\mathbb{P}\left(\inf_{x\in T}\|Ax\|_2 > E_m - \ell(T) - t\right) \geq 1 - e^{-\frac{t^2}{2}}. \tag{22}$$

Recall the set

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

from (12) and set $T := T(x)\cap S^{n-1}$. To provide a bound on the number of Gaussian measurements, we compute the Gaussian width of T. We start with establishing a connection between $\ell(T)$ and $\ell(\Omega(T))$, where $\Omega(T)$ is the set obtained by applying the operator Ω to each element T.

Theorem 11. *Let $\Omega\in\mathbb{R}^{p\times n}$ be a frame with a lower frame bound $a>0$ and $T\subset\mathbb{R}^n$. Then*

$$\ell(T) \leq a^{-1/2}\ell(\Omega(T)).$$

Before giving the proof of Theorem 11, we recall Slepian's inequality, see [29] or [20, Chapter 8.7] for details. For a random variable X we define $\|X\|_2 = \left(\mathbb{E}|X|^2\right)^{1/2}$.

Theorem 12. *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be Gaussian centered processes. If for all s, $t\in T$*

$$\|X_t - X_s\|_2 \leq \|Y_t - Y_s\|_2,$$

then

$$\mathbb{E}\sup_{t\in T} X_t \leq \mathbb{E}\sup_{t\in T} Y_t.$$

Proof (of Theorem 11). By the definition of the Gaussian width

$$\ell(T) = \mathbb{E}\sup_{v\in T}\langle g,v\rangle = \mathbb{E}\sup_{v\in T}\langle g,\Omega^\dagger\Omega v\rangle$$

$$= \mathbb{E}\sup_{w\in\Omega(T)}\langle(\Omega^\dagger)^T g, w\rangle. \tag{23}$$

We consider two Gaussian processes,

$$X_w = \langle (\Omega^\dagger)^T g, w\rangle \text{ and } Y_w = \|(\Omega^\dagger)^T\|_{2\to 2}\langle h, w\rangle,$$

where $w \in \Omega(T)$ and $g \in \mathbb{R}^n$, $h \in \mathbb{R}^p$ are both standard Gaussian vectors. For any standard Gaussian vector h it holds $\mathbb{E}|\langle h,w\rangle|^2 = \|w\|_2^2$. So if w, $w' \in \Omega(T)$, then

$$\|X_w - X_{w'}\|_2 \leq \|(\Omega^\dagger)^T\|_{2\to 2}\|w-w'\|_2 = \|Y_w - Y_{w'}\|_2.$$

It follows from Theorem 12 that

$$\mathbb{E} \sup_{w\in\Omega(T)} \langle (\Omega^\dagger)^T g, w\rangle \leq \|(\Omega^\dagger)^T\|_{2\to 2}\,\mathbb{E} \sup_{w\in\Omega(T)} \langle h, w\rangle. \tag{24}$$

An upper bound of the canonical dual frame is $(\Omega^\dagger)^T$ is a^{-1}. Hence, $\|(\Omega^\dagger)^T\|_{2\to 2} \leq a^{-1/2}$. Together with (23) and (24) this gives

$$\ell(T) \leq a^{-1/2}\,\mathbb{E} \sup_{w\in\Omega(T)} \langle h, w\rangle = a^{-1/2}\ell(\Omega(T)).$$

□

The next theorem from [28, Section 2.2] provides a good bound on $\ell(\Omega(T))$.

Theorem 13. *Let $\Omega \in \mathbb{R}^{p\times n}$ be a frame with upper frame bound $b > 0$ and $x \in \mathbb{R}^n$ be ℓ-cosparse with $\ell = p-k$. For $T := T(x)\cap S^{n-1}$, it holds*

$$\ell(\Omega(T))^2 \leq 2bk\ln\left(\frac{ep}{k}\right). \tag{25}$$

Proof. Since Ω is a frame with an upper frame constant b, we have

$$\Omega(T) \subset \Omega(T(x))\cap\Omega(S^{n-1}) \subset K(\Omega x)\cap\left(\sqrt{b}B_2^p\right),$$

where

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

The supremum over a larger set can only increase, hence

$$\ell(\Omega(T)) \leq \sqrt{b}\,\ell\left(K(\Omega x)\cap B_2^p\right). \tag{26}$$

An upper bound for the Gaussian width $\ell\left(K(\Omega x)\cap B_2^p\right)$ can be given in terms of the the polar cone $\mathcal{N}(\Omega x) = K(\Omega x)^\circ$ defined by

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

By duality of convex programming, see [10] or [20, Proposition 9.22], we have

$$\ell\left(K(\Omega x)\cap B_2^p\right) \leq \mathbb{E}\min_{z\in\mathcal{N}(\Omega x)} \|g - z\|_2$$

and by Hölder's inequality

$$\ell\left(K(\Omega x)\cap B_2^p\right)^2 \leq \left(\mathbb{E}\min_{z\in\mathcal{N}(\Omega x)}\|g-z\|_2\right)^2 \leq \mathbb{E}\min_{z\in\mathcal{N}(\Omega x)}\|g-z\|_2^2. \tag{27}$$

Let Λ be the cosupport of x. Then one can verify that

$$\mathcal{N}(\Omega x) = \bigcup_{t\geq 0}\left\{z\in\mathbb{R}^p : z_i = t\,\mathrm{sgn}(\Omega x)_i,\ i\in\Lambda^c,\ |z_i|\leq t,\ i\in\Lambda\right\}. \tag{28}$$

To proceed, we fix t, minimize $\|g-z\|_2^2$ over all possible entries z_j, take the expectation of the obtained expression and finally optimize over t. Taking into account (28), we have

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

where S_t is the soft-thresholding operator given by

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

Taking expectation we arrive at

$$\begin{aligned}\mathbb{E}\min_{z\in\mathcal{N}(\Omega x)}\|g-z\|_2^2 &\leq \mathbb{E}\left[\sum_{i\in\Lambda^c}(g_i - t\,\mathrm{sgn}(\Omega x)_i)^2\right] + \mathbb{E}\left[\sum_{i\in\Lambda}S_t(g_i)^2\right]\\ &= k(1+t^2) + (p-k)\,\mathbb{E}\,S_t(g)^2,\end{aligned} \tag{29}$$

where g is a univariate standard Gaussian random variable. The expectation of $S_t(g)^2$ is estimated by integration,

$$\begin{aligned}\mathbb{E}\,S_t(g)^2 &= \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{-t}(x+t)^2 e^{-\frac{x^2}{2}}\,dx + \int_t^{\infty}(x-t)^2 e^{-\frac{x^2}{2}}\,dx\right]\\ &= \frac{2}{\sqrt{2\pi}}\int_0^{\infty}x^2 e^{-\frac{(x+t)^2}{2}}\,dx = \frac{2e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\int_0^{\infty}x^2 e^{-\frac{x^2}{2}}e^{-xt}\,dx\\ &\leq e^{-\frac{t^2}{2}}\sqrt{\frac{2}{\pi}}\int_0^{\infty}x^2 e^{-\frac{x^2}{2}}\,dx = e^{-\frac{t^2}{2}}.\end{aligned} \tag{30}$$

Substituting the estimate (30) into (29) gives

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

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

$$\ell\left(K(\Omega x)\cap B_2^p\right)^2 \leq k\left(1+2\ln(p/k)\right)+k = 2k\ln(ep/k). \tag{31}$$

By combining inequalities (26) and (31) we obtain

$$\ell(\Omega(T))^2 \leq 2bk\ln\frac{ep}{k}.$$

□

By Theorem 11 and Theorem 13 we obtain

$$\ell(T) \leq \sqrt{\frac{2bk}{a}\ln\frac{ep}{k}}.$$

At this point we are ready to prove our main result (Theorem 9) concerning the number of Gaussian measurements that guarantee the robust recovery of a fixed cosparse vector.

Set $t=\sqrt{2\ln(\theta^{-1})}$. The choice of m in (20) guarantees that

$$E_m-\ell(T)-t \geq \frac{m}{\sqrt{m+1}}-\sqrt{\frac{bk}{a}\ln\frac{ep}{k}}-\sqrt{2\ln(\theta^{-1})} \geq \tau.$$

Theorem 10 yields

$$\mathbb{P}\left(\inf_{x\in T}\|Ax\|_2 \geq \tau\right) \geq \mathbb{P}\left(\inf_{x\in T}\|Ax\|_2 \geq E_m-\ell(T)-t\right) \geq 1-\theta.$$

This completes the proof.

4.3 Nonuniform recovery from subgaussian measurements

Based on the estimates of the previous section in combination with the following result due to Mendelson et al. [32], we may extend the nonuniform recovery result also to subgaussian matrices. We remark, however, that due to an unspecified constant in (32) this technique does not necessarily improve over the uniform estimate (19) derived via the restricted isometry property. Nevertheless, we feel that this proof method is interesting.

Theorem 14 (Corollary 2.7 in [32]). *Let $T \subset S^{n-1}$, $X_i \in \mathbb{R}^n$, $i=1,\ldots,m$ be independent isotropic subgaussian random vectors with $\|X_i\|_{\psi_2} \leq \alpha$ and $0<\delta<1$. Define $A\in\mathbb{R}^{m\times n}$ as $A=(X_1,\ldots,X_m)^T$. If*

$$m \geq \frac{c_1 \alpha^4}{\delta^2} \ell(T)^2, \tag{32}$$

then with probability at least $1 - \exp\left(-c_2 \delta^2 m / \alpha^4\right)$ *for all* $x \in T$ *it holds*

$$1 - \delta \leq \frac{\|Ax\|_2^2}{m} \leq 1 + \delta, \tag{33}$$

where c_1, c_2 *are absolute constants.*

Theorem 15. *Let* $\Omega \in \mathbb{R}^{p \times n}$ *be a frame with frame bounds* a, $b > 0$, $x \in \mathbb{R}^n$ *be an* ℓ*-cosparse vector and* $k = p - l$. *Let* $X_i \in \mathbb{R}^n$, $i = 1, \ldots, m$, *be independent isotropic subgaussian random vectors with* $\|X_i\|_{\psi_2} \leq \alpha$ *and* $0 < \delta < 1$. *Define* $A \in \mathbb{R}^{m \times n}$ *as* $A = (X_1, \ldots, X_m)^T$. *If*

$$m \geq \frac{c_1 b \alpha^4}{a \delta^2} s \ln \frac{ep}{k},$$

then with probability at least $1 - \exp\left(-c_2 \delta^2 m / \alpha^4\right)$ *every minimizer* $\hat{x}$ *of (7) approximates* x *with the following* ℓ_2*-error*

$$\|x - \hat{x}\|_2 \leq \frac{2\varepsilon}{\sqrt{1 - \delta}}.$$

Proof. Inserting the estimate of Gaussian width (25) and (37) in (32) provides the above bound on the number of subgaussian measurements that guarantees successful nonuniform recovery via (6). □

4.4 Uniform recovery via the Ω-null space property

Let us now consider the stable Ω-null space property introduced in (8), which implies (uniform) recovery of all cosparse vectors via analysis ℓ_1-minimization, see Theorem 1. Since the restricted isometry property adapted to $(\Omega^\dagger)^T$ implies the Ω-null space property, see Theorem 2, we already know that Gaussian (and subgaussian) measurement matrices satisfy the Ω-null space property with high probability under Condition (18). The constant C in (18), however, is unspecified and inspecting the proof would reveal a rather large value. We follow now a different path based on convex geometry and Gordon's escape through a mesh theorem that leads to a direct estimate for the Ω-null space property and yields an explicit and small constant. This approach is inspired by [10] and has been applied in [20, Chapter 9.4] for the synthesis sparsity model for the first time. The next theorem was shown in [28] using additionally some ideas of [37].

Theorem 16. *Let* $A \in \mathbb{R}^{m \times n}$ *be a Gaussian random matrix,* $0 < \rho < 1$, $0 < \theta < 1$ *and* $\tau > 0$. *If*

$$\frac{m^2}{m+1} \geq 2k \left((1+\rho^{-1}) \sqrt{\frac{b}{a} \ln \frac{ep}{k}} + \frac{b(1+\rho^{-1})}{a\sqrt{2}} + \sqrt{\frac{\ln(\theta^{-1})}{k}} + \frac{1}{\tau} \sqrt{\frac{1}{2k}} \right)^2,$$

then with probability at least $1-\theta$ *for every vector* $x \in \mathbb{R}^n$ *and perturbed measurements* $y = Ax + e$ *with* $\|e\|_2 \leq \varepsilon$ *a minimizer* $\hat{x}$ *of (7) approximates* x *with* ℓ_2*-error*

$$\|x - \hat{x}\|_2 \leq \frac{2(1+\rho)^2}{\sqrt{a}(1-\rho)} \frac{\sigma_k(\Omega x)_1}{\sqrt{k}} + \frac{2\tau\sqrt{b}(3+\rho)}{\sqrt{a}(1-\rho)} \varepsilon.$$

Proof. (Sketch) We verify that A satisfies the ℓ_2-stable Ω-null space property (8). To this end we introduce the set

$$W_{\rho,k} := \left\{ w \in \mathbb{R}^n : \|\Omega_{\Lambda^c} w\|_2 \geq \rho/\sqrt{k} \|\Omega_\Lambda w\|_1 \text{ for some } \Lambda \subset [p],\ |\Lambda| = p-k \right\}.$$

If

$$\inf\left\{ \|Aw\|_2 : w \in W_{\rho,k} \cap S^{n-1} \right\} > \frac{1}{\tau}, \tag{34}$$

then for any $w \in \mathbb{R}^n$ such that $\|Aw\|_2 \leq \frac{1}{\tau}\|w\|_2$ and any set $\Lambda \subset [p]$ with $|\Lambda| \geq p-k$ it holds

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

For the remaining vectors $w \in \mathbb{R}^n$, we have $\|Aw\|_2 > \frac{1}{\tau}\|w\|_2$, which together with the fact that Ω is a frame with upper frame bound b leads to

$$\|\Omega_{\Lambda^c} w\|_2 \leq \|\Omega w\|_2 \leq \sqrt{b}\|w\|_2 < \tau\sqrt{b}\|Aw\|_2.$$

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

$$\|\Omega_{\Lambda^c} w\|_2 < \frac{\rho}{\sqrt{k}} \|\Omega_\Lambda w\|_1 + \tau\sqrt{b}\|Aw\|_2.$$

To show (34), we have to study the Gaussian width of the set $W_{\rho,k} \cap S^{n-1}$. According to Theorem 11

$$\ell(W_{\rho,k} \cap S^{n-1}) \leq a^{-1/2} \ell(\Omega(W_{\rho,k} \cap S^{n-1}). \tag{35}$$

Since Ω is a frame with an upper frame bound b, we have

$$\Omega\left(W_{\rho,k} \cap S^{n-1}\right) \subset \Omega\left(W_{\rho,k}\right) \cap \left(\sqrt{b}B_2^p\right) \subset T_{\rho,k} \cap \left(\sqrt{b}B_2^p\right) = \sqrt{b}\left(T_{\rho,k} \cap B_2^p\right), \tag{36}$$

with

$$T_{\rho,k} = \left\{ u \in \mathbb{R}^p : \|u_S\|_2 \geq \rho/\sqrt{k} \|u_{S^c}\|_1 \text{ for some } S \subset [p],\ |S| = k \right\}.$$

Inspired by [37] it was shown in [28, Section 3.2] that

$$\ell\left(\Omega\left(W_{\rho,k}\cap S^{n-1}\right)\right)\leq\sqrt{b}\left(1+\rho^{-1}\right)\left(\sqrt{2k\ln\frac{ep}{k}}+\sqrt{k}\right).$$

Together with (35) this gives

$$\ell(W_{\rho,k}\cap S^{n-1})\leq\sqrt{\frac{b}{a}}\left(1+\rho^{-1}\right)\left(\sqrt{2k\ln\frac{ep}{k}}+\sqrt{k}\right). \qquad (37)$$

An application of Theorem 10 and inequality (37) complete the proof. □

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, for $k=p-\ell$,

$$m>8(b/a)k\ln(ep/k). \qquad (38)$$

Moreover, the recovery is stable under passing to approximately cosparse vectors when adding slightly more measurements.

With the approach outlined in Section 4.3 it is also possible to extend the above bound on the number of measurements to subgaussian random matrices. This follows from a combination of Theorem 14 with the estimate (37) of the Gaussian width of the set $W_{\rho,k}\cap S^{n-1}$.

References

1. J.J. Benedetto, C. Heil, D.F. Walnut: Differentiation and the Balian–Low theorem. Journal of Fourier Analysis and Applications **1(4)**, 355–402 (1994)
2. T. Blumensath: Sampling and reconstructing signals from a union of linear subspaces. IEEE Transactions on Information Theory **57(7)**, 4660 – 4671 (2011)
3. M. Burger, S. Osher: Convergence rates of convex variational regularization. Inverse Problems **20(5)**, 1411–1421 (2004)
4. J.-F. Cai, W. Xu: Guarantees of total variation minimization for signal recovery. In Proceedings of 51st Annual Allerton Conference on Communication, Control, and Computing, 1266–1271 (2013)
5. E. J. Candès, D. L. Donoho: 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. E. J. Candès, Y. C. Eldar, D. Needell, P. Randall: Compressed sensing with coherent and redundant dictionaries. Appl. Comput. Harmon. Anal. **31(1)**, 59–73 (2011)
7. E. J. Candès, J., T. Tao, and J. K. Romberg: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory **52(2)**, 489–509 (2006)
8. P. G. Casazza, G. Kutyniok (Eds.): Finite frames. Theory and applications. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York (2013)
9. T. Chan, J. Shen: Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods. SIAM, (2005)
10. V. Chandrasekaran, B. Recht, P. A. Parrilo, A. S. Willsky: The Convex Geometry of Linear Inverse Problems. Found. Comput. Math. **12(6)**, 805–849 (2012)
11. S.S. Chen, D.L. Donoho, M.A. Saunders: Atomic decomposition by basis pursuit. IAM Journal on Scientific Computing **20(1)**, 33–61 (1998)

12. O. Christensen: An Introduction to Frames and Riesz Bases. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston (2003)
13. M. Davenport, M. Wakin: Analysis of orthogonal matching pursuit using the restricted isometry property. IEEE Trans. Inform. Theory **56(9)**, 4395–4401 (2010)
14. R.J. Duffin, A.C. Schaeffer. A class of nonharmonic Fourier series. Trans. Amer. Math. Soc. **72(2)**, 341–366 (1952)
15. R. Dudley: Uniform central limit theorems. Cambridge Studies in Advanced Mathematics **63**, Cabridge University Press (1999)
16. M. Elad, P. Milanfar, R. Rubinstein: Analysis versus synthesis in signal priors. Inverse Problems **23(3)**, 947–968 (2007)
17. J. Fadili, Gabriel Peyré, S. Vaiter, C.-A. Deledalle, J. Salmon: Stable recovery with analysis decomposable priors. In 10th international conference on Sampling Theory and Applications (SampTA 2013), 113–116, Bremen, Germany (2013)
18. M. Fornasier and K. Gröchenig: Intrinsic localization of frames. Constr. Approx., **22(3)**, 395–415 (2005)
19. S. Foucart: Stability and robustness of ℓ_1-minimizations with Weibull matrices and redundant dictionaries. Linear Algebra and its Applications **441**, 4 – 21 (2014)
20. S. Foucart and H. Rauhut: *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser (2013)
21. J. J. Fuchs: On sparse representations in arbitrary redundant bases. IEEE Trans. Inform. Theory, **50(6)**:1341–1344 (2004)
22. R. Giryes, S. Nam, M. Elad, R. Gribonval, M.E. Davies: Greedy-Like Algorithms for the Cosparse Analysis Model. Linear Algebra and its Applications **441**, 22 – 60 (2014)
23. Y. Gordon: On Milman's inequality and random subspaces which escape through a mesh in $\mathbb{R}^n$. In Geometric aspects of functional analysis (1986/87) **1317** of Lecture Notes in Math., 84–106, Springer, Berlin (1988)
24. M. Grasmair: Linear convergence rates for Tikhonov regularization with positively homogeneous functionals. Inverse Problems **27(7)**, 075014 (2011)
25. K. Gröchenig: Foundations of time-frequency analysis. Appl. Numer. Harmon. Anal. Birkhäuser Boston, Boston, MA (2001)
26. K. Guo, G. Kutyniok, D. Labate: Sparse multidimensional representations using anisotropic dilation and shear operators in: G. Chen, M. Lai, (Eds.), Wavelets and Splines: Athens 2005, Proceedings of the International Conference on the Interactions between Wavelets and Splines, Athens, GA, May 16–19 (2005)
27. M. Haltmeier: Stable signal reconstruction via ℓ^1-minimization in redundant, non-tight frames. IEEE Transactions on Signal Processing **61(2)**, 420–426 (2013)
28. M. Kabanava, H. Rauhut: Analysis ℓ_1-recovery with frames and Gaussian measurements.
29. M. Ledoux, M. Talagrand: Probability in Banach Spaces. Springer-Verlag, Berlin, Heidelberg, NewYork (1991)
30. Y. Liu, T. Mi, Sh. Li: Compressed sensing with general frames via optimal-dual-based ℓ_1-analysis. IEEE Transactions on information theory **58(7)**, 4201–4214 (2012)
31. S. Mallat: A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press (2008)
32. S. Mendelson, A. Pajor, N. Tomczak-Jaegermann: Reconstruction and subgaussian operators in asymptotic geometric analysis. Geom. Funct. Anal. **17(4)**, 1248–1282 (2007)
33. S. Nam, M.E. Davies, M. Elad, R. Gribonval: Cosparse analysis modeling – uniqueness and algorithms. in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2011)
34. S. Nam, M.E. Davies, M. Elad, R. Gribonval: The cosparse analysis model and algorithms. Appl. Comput. Harmon. Anal. **34(1)**, 30–56 (2013)
35. D. Needell, R. Ward: Stable image reconstruction using total variation minimization. http://arxiv.org/abs/1202.6429
36. A. Ron, Z. Shen: Affine systems in $L_2(R^d)$: the analysis of the analysis operator. J. Funct. Anal. **148(2)**, 408–447 (1997)
37. M. Rudelson, R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008

38. L.I. Rudin, S. Osher, E. Fatemi: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
39. I. Selesnick, M. Figueiredo: Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. In Proceedings of SPIE **7446** (2009)
40. M.Talagrand: The Generic Chaining. Springer Monographs in Mathematics, Springer-Verlag (2005)
41. J. A. Tropp: Recovery of short, complex linear combinations via l_1 minimization. IEEE Trans. Inform. Theory **51(4)**, 1568–1570 (2005)
42. J.A. Tropp: Just relax: Convex programming methods for identifying sparse signals in noise. IEEE Transactions on information theory **52(3)**, 1030–1051 (2006)
43. S. Vaiter, G. Peyré, Ch. Dossal, J. Fadili: Robust sparse analysis regularization. IEEE Transactions on information theory **59(4)**, 2001–2016 (2013)