Robust analysis ℓ_1-recovery from Gaussian measurements and total variation minimization

M. Kabanava[*], H. Rauhut[†], H. Zhang[‡]

July 28, 2014

Abstract

Analysis ℓ_1-recovery refers to a technique of recovering a signal that is sparse in some transform domain from incomplete corrupted measurements. This includes total variation minimization as an important special case when the transform domain is generated by a difference operator. In the present paper we provide a bound on the number of Gaussian measurements required for successful recovery for total variation and for the case that the analysis operator is a frame.

1 Introduction

Compressive sensing is a recent field in signal processing that predicts that sparse vectors can be stably reconstructed from incomplete measurements via efficient algorithms [5, 4]. Traditionally, the synthesis sparsity model is used in this context, where it is assumed that the signal can be written as a linear combination of a few elements of an appropriate basis. Recently, the analysis sparsity model (or cosparsity model) has gained significant interest as well [2, 11, 7, 9], where it is assumed that the signal is sparse after a transformation. In the case of a basis transformation, the synthesis and analysis models coincide, but if the transform is redundant then the analysis sparsity model is different, and in fact the class of analysis sparsity models is richer than the class of synthesis sparsity models. While the naive recovery approach of ℓ_0-minimization is NP-hard, one may use convex relaxations, i.e., ℓ_1-minimization in both the synthesis and analysis sparsity cases. Especially in the synthesis sparsity case, ℓ_1-minimization is by now rather well understood, see e.g. [5]. Despite recent progress in the theory of analysis ℓ_1-minimization [2, 11, 7, 9], there still remain a number of questions to be explored. In particular, in the important special case of total variation minimization which is ubiquitious in image processing only a few contributions analyzing bounds for recovery from underdetermined measurements are available [12]. This article contributes to this topic by providing a bound on the number of necessary Gaussian random measurements in order to recover a gradient sparse signal via total variation minimization. Moreover, we also provide an

[*]RWTH Aachen University, Chair for Mathematics C (Analysis), Templergraben 55, 52062 Aachen Germany, kabanava@mathc.rwth-aachen.de

[†]RWTH Aachen University, Chair for Mathematics C (Analysis), Templergraben 55, 52062 Aachen Germany, rauhut@mathc.rwth-aachen.de

[‡]College of Science, National University of Defense Technology, Changsha, Hunan, 410073 China hhuuii.zhang@gmail.com

alternative bound to the one in [9], on the number of required Gaussian measurements for recovery of analysis-sparse signals with respect to a frame.

Our main results concern so-called nonuniform recovery using a Gaussian random measurement matrix, i.e., we fix a sparse (or rather cosparse) vector (with respect to a given analysis operator) and provide bounds that guarantee that the given vector is recovered via analysis ℓ_1-minimization with high probability. Our bounds are particularly good when the sparsity is not very small compared to the ambient dimension. In fact, in the case of analysis-sparsity with respect to a very redundant frame, there are natural lower bounds for the sparsity, so that in this context the bounds in [9] may turn out to be trivial in certain situations as they require more measurements than the ambient dimension. In contrast, the bounds derived in this paper are always non-trivial in the sense that the number of required measurements is always lower than the ambient dimension.

Following [3], our analysis is based on estimating widths of tangent cones of a transformed ℓ_1-norm at the (co-)sparse vector to be recovered, similarly to [9] or [5, Chapter 9.2]. This is in contrast to uniform recovery bounds which are often based on the restricted isometry property or the null space property [5], see also [9] for precise bounds for an analysis sparsity version of the null space property for Gaussian random measurements.

In mathematical terms, we wish to recover a signal $x \in \mathbb{R}^d$ from measurements

$$y = Mx + w, \tag{1}$$

where $M \in \mathbb{R}^{m \times d}$ is a measurement matrix and $w \in \mathbb{R}^m$ with $\|w\|_2 \leq \eta$ corresponds to noise. When $m \ll d$, there are infinitely many solutions to (1). However, the prior sparsity knowledge about the underlying signal x makes its recovery possible.

We assume that x possesses a structure generated by a matrix $\Omega \in \mathbb{R}^{p \times d}$, called the analysis operator, in the sense that the application of Ω to x produces a vector with a small number of non-zero entries. If Ωx has s non-zero entries, then x is called ℓ-*cosparse*, where the number $\ell := p - s$ is refered to as *cosparsity* of x (with respect to Ω). The index set of the zero entries of Ωx is called the *cosupport* of x. Analysis ℓ_1-minimization tries to recover the signal by computing the minimizer of

$$\min_{z \in \mathbb{R}^d} \|\Omega z\|_1 \quad \text{subject to} \quad \|Mz - y\|_2 \leq \eta. \tag{2}$$

In this paper we consider two prominent examples of the analysis operator. The method of total variation corresponds to the program (2), when Ω is a difference operator. In the one-dimensional case it is defined by the matrix

$$\Omega = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}. \tag{3}$$

In this setting (2) promotes piecewise constant signals with sparse gradient. Another important example of the analysis operator appears when the rows ω_j of Ω form a frame, i.e., if there exist constants $0 < A \leq B < \infty$ such that

$$A\|x\|_2^2 \leq \|\Omega x\|_2^2 = \sum_{j=1}^{p} |\langle \omega_j, x \rangle|^2 \leq B\|x\|_2^2. \tag{4}$$

2

We are interested in the minimal number of measurements in terms of the sparsity (or cosparsity) required to recover a cosparse vector from its measurements y in (1) when the matrix M is a Gaussian random matrix, i.e., its entries are independent standard normal distributed random variables. We rely on a recent result of Foygel and Mackey in [6] which is in spirit to the geometric approach of [3] in order to provide such a bound on the number of Gaussian measurements needed to recover x via (2), when Ω satisfies either (3) or (4).

1.1 Main Results

We first provide a general bound for the required number m of Gaussian measurements in order to (stably) recover an analysis-sparse signal $x \in \mathbb{R}^d$ via analysis ℓ_1-minimization, see Theorem 3. An important feature of the result is that m is always (essentially) less than the ambient dimension d. Based on this, we show that a signal $x \in \mathbb{R}^d$ which is s-sparse with respect to the difference operator Ω in (3) (that is, whose gradient is s-sparse) can be recovered with high probability if "roughly", i.e., ignoring terms of lower order

$$m > d \left(1 - \frac{1}{\pi} (1 - \frac{s+1}{d})^2 \right), \tag{5}$$

see Theorem 4.

As the second case, we consider $\Omega \in \mathbb{R}^{p \times d}$ to be a frame with frame bounds $A, B > 0$ in (4). Theorem 6 below shows that a signal $x \in \mathbb{R}^d$ which is ℓ-cosparse with cosupport Λ, i.e., $\operatorname{supp} \Omega x = \Lambda^c$ can be recovered from m Gaussian measurements via analysis ℓ_1-minimization with high probability if, roughly speaking,

$$m > d - \frac{2}{\pi p B} \left(\sum_{\ell \in \Lambda} \|\omega_\ell\|_2 \right)^2 \tag{6}$$

In both cases (5) and (6), the number of measurements is clearly less than d. Note that the usual bounds in compressive sensing require

$$m > cs \log(d/s) \tag{7}$$

for recovery of s-sparse vectors via ℓ_1-minimization from Gaussian random measurements [5, Chapter 9], [3]. Morever, the bound derived in [9] for analysis sparse recovery with respect to a frame roughly requires

$$m \geq \frac{2B}{A} s \ln(ep/s). \tag{8}$$

see also [10]. One realizes a structural difference of these bounds to the new ones stated above. In fact, even for sparsity $s = 1$, (5) roughly requires $m > (1 - \pi^{-1})d \approx 0.6817\,d$, while (7) requires $m \geq c \log(d)$, which is of much smaller order (but note that such a bound is so far not available for TV-minimization and gradient sparse signals). However, if s is proportional to d, i.e., $s = \alpha d$ for some $\alpha \in (0, 1)$ then (5) requires $m \geq c_\alpha d$ with $c_\alpha = 1 - \frac{1}{\pi}(1 - \alpha)^2$ which is always less than 1, while (7) gives $m \geq c'_\alpha d$ with $c'_\alpha = c\alpha \ln(\alpha^{-1})$ which may become larger than 1 because the available estimates for c are larger than 2 [5].

In order to place the bound (6) of Theorem 6 for the frame case into context, we recall some facts on the analysis sparsity model. Consider the generic case, that the rows of the analysis operator Ω, that is, the frame elements ω_j, $j = 1, \ldots, p$, are in general linear position so that a subcollection

of d frame elements is always linearly independent. Then any subspace $W_\Lambda = \text{span}\{\omega_j, j \in \Lambda\}^\perp$ with $\#\Lambda = \ell$, of ℓ-cosparse vectors, where $\perp$ denotes the orthogonal complement, has dimension $d - \ell$, which means that the smallest value that the sparsity parameter $s = p - \ell$ can take for a nontrivial vector x is $p - d + 1$, see also [11, 10]. Therefore, if $p = \kappa d$ for some $\kappa > 1$, then the sparsity is always proportional to the dimension d, i.e., $s \sim \alpha d$ which means that the bound (6) reads $m \geq c_{\kappa,\alpha} d$ with $c_{\kappa,\alpha} = 1 - \frac{2(\kappa-\alpha)^2}{\pi\kappa B}$, assuming that the frame is normalized. If $\kappa > 2$, then the sparsity is even always at least as large as the ambient dimension, so that (7) becomes a trivial bound. Similar considerations apply to the previous bound in [9] on analysis-sparse recovery with respect to frames. In contrast, let us consider the new bound (6) for a tight unit norm frame, i.e., $\|\omega_j\|_2 = 1$ and $A = B = p/d$. For cosparsity $\#\Omega = \ell = p - s$, (6) yields then

$$m > d - \frac{2d}{\pi p^2}(p - s)^2 = \left(1 - \frac{2(p - s)^2}{\pi p^2}\right)d.$$

For sparsity proportional to p, the number of measurements clearly scales like βd where β is always strictly smaller than 1, which shows that our new bound is better in the natural regime $p = \kappa d$ with $\kappa \geq 2$, say. In contrast, the previous bound (8) requires m to be larger than d for a frame in general position with $p = \kappa d$, see Figure 1.
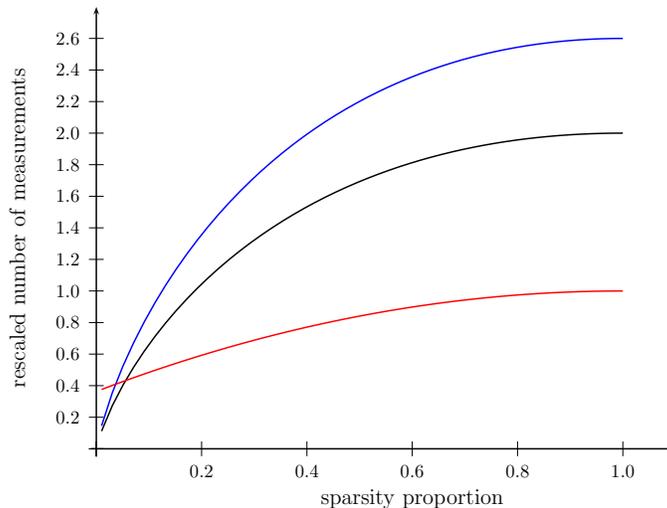

Figure 1: Comparison of the standard recovery bounds and the new one for the setting of a normalized tight frame with frame bounds $A = B = \frac{p}{d}$. The red line corresponds to the new bound on the rescaled number of measurements $\frac{m}{d}$ as a function of sparsity fraction $\frac{s}{p}$. The black and blue lines represent the bound given by (8) for $p = d$ and $p = 1.3d$ respectively.

1.2   Notation

We denote the rows of $\Omega \in \mathbb{R}^{p\times d}$ by ω_j, $j = 1, \ldots, p$. We use Ω_Λ to refer to a submatrix of Ω with the rows indexed by Λ; Ω^T is the transpose 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. The sign of a real number $r \neq 0$

is $\text{sgn}(r) = \frac{r}{|r|}$. For a vector $\alpha \in \mathbb{R}^p$ we define its sign vector $\text{sgn}(\alpha) \in \mathbb{R}^p$ by

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

We write S^{d-1} for the unit sphere in $\mathbb{R}^d$.

Acknowledgements

2 Recovery from Gaussian measurements

2.1 Theoretical guarantees

Our analysis is based on work of Chandrasekaran et al. [3], where sufficient conditions for robust recovery make use of the tangent cone (also called descent cone) of a convex function (e.g., the ℓ_1-norm) at the signal to be recovered, see also [14]. In the case of analysis ℓ_1-minimization these are formulated explicitly in [9]. For a fixed $x \in \mathbb{R}^d$ we define a tangent cone as

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

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 $\|\Omega x\|_1$.

Theorem 1. *Let $x \in \mathbb{R}^d$. If*

$$\inf_{\substack{v \in T(x) \\ \|v\|_2 \leq 1}} \|Mv\|_2 \geq \tau, \tag{9}$$

for some $\tau > 0$, then a minimizer $\hat{x}$ of (2) satisfies

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

2.2 Gordon's escape through a mesh theorem

According to (9), 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 [8], see also [5, Theorem 9.21]. In order to present Gordon's result formally, we introduce some notation.

Let $g \in \mathbb{R}^m$ be a standard Gaussian random vector. Then

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

satisfies

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

For a set $T \subset \mathbb{R}^d$ we define its Gaussian width by

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

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

Theorem 2. *Let $M \in \mathbb{R}^{m \times d}$ be a Gaussian random matrix and T be a subset of the unit sphere $S^{d-1} = \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$. Then, for $t > 0$, it holds*

$$\mathbb{P}\left(\inf_{x \in T} \|Mx\|_2 > E_m - \ell(T) - t\right) \geq 1 - e^{-\frac{t^2}{2}}. \tag{10}$$

Thus, to provide a bound on the number of Gaussian measurements, we estimate the Gaussian width of the tangent cone $T(x)$ intersected with the unit sphere.

2.3   Estimates for the Gaussian width

A basic technique to estimate the Gaussian width of the tangent cone was developed in the work of Stojnic [13] and was later refined in a series of papers [1, 3], see also [14]. The result states that the Gaussian width of the tangent cone can be bounded by the Euclidean distance of the standard normal vector to the scaled cone generated by the subdifferential.

Recall that the Euclidean distance to a set $T \subset \mathbb{R}^d$ is the function defined by

$$\mathrm{dist}(x, T) := \inf\{\|x - u\|_2 : u \in T\}.$$

The subdifferential of a convex function $f : \mathbb{R}^d \to \mathbb{R}$ at a point $x \in \mathbb{R}^d$ is the set

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

It is shown in [13, 3] that

$$\ell(T(x) \cap \mathbb{S}^{d-1})^2 \leq \inf_{t \geq 0} \mathbb{E}\left[\mathrm{dist}(g, t \cdot \partial \|\Omega \cdot \|_1(x))\right]^2, \tag{11}$$

where $g \in \mathbb{R}^d$ is the standard normal vector.

To provide a bound on the expected squared distance from (11) we generalize Proposition 3 in [6] valid for Ω being a basis to the following result.

Lemma 1. *Let $\Omega \in \mathbb{R}^{p \times d}$ be an analysis operator, $g \in \mathbb{R}^d$ be a standard normal random vector and $x \in \mathbb{R}^d$. Then*

$$\inf_{t \geq 0} \mathbb{E}\left[\mathrm{dist}(g, t \cdot \partial \|\Omega \cdot \|_1(x))\right]^2 \leq d - \frac{[\ell(\partial \|\Omega \cdot \|_1(x))]^2}{\max_{\|z\|_2 \leq 1} \|\Omega z\|_1^2}. \tag{12}$$

Proof. Since $\partial \|\Omega \cdot \|_1(x)$ is a compact set, there exists $w_0 \in \partial \|\Omega \cdot \|_1(x)$ such that

$$\langle g, w_0 \rangle = \max_{w \in \partial \|\Omega \cdot \|_1(x)} \langle g, w \rangle.$$

Then

$$\mathrm{dist}(g, t \cdot \partial \|\Omega \cdot \|_1(x))^2 \leq \|g - tw_0\|_2^2 = \|g\|_2^2 - 2t\langle g, w_0 \rangle + t^2 \|w_0\|_2^2.$$

6

Moreover,

$$\partial\|\Omega\cdot\|_1(x) = \Omega^T\left(\partial\|\cdot\|_1(\Omega x)\right)$$

and

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

which implies that $w_0 = \Omega^T\alpha_0$ for some $\alpha_0 \in \partial\|\cdot\|_1(\Omega x)$. Thus

$$\mathrm{dist}(g, t\cdot\partial\|\Omega\cdot\|_1(x))^2 \le \|g\|_2^2 - 2t \max_{w\in\partial\|\Omega\cdot\|_1(x)}\langle g, w\rangle + t^2\|\Omega^T\alpha_0\|_2^2.$$

Due to duality and the fact that $\|\alpha_0\|_\infty \le 1$ we obtain

$$\|\Omega^T\alpha_0\|_2 = \max_{\|z\|_2\le 1}\langle z, \Omega^T\alpha_0\rangle = \max_{\|z\|_2\le 1}\langle \Omega z, \alpha_0\rangle \le \max_{\|z\|_2\le 1}\|\Omega z\|_1\|\alpha_0\|_\infty \le \max_{\|z\|_2\le 1}\|\Omega z\|_1.$$

So altogether it holds

$$\mathrm{dist}(g, t\cdot\|\Omega\cdot\|_1(x))^2 \le \|g\|_2^2 - 2t\max_{w\in\partial\|\Omega\cdot\|_1(x)}\langle g, w\rangle + t^2\max_{\|z\|_2\le 1}\|\Omega z\|_1^2$$

and by taking the expectation of both sides we obtain

$$\mathbb{E}\,\mathrm{dist}(g, t\cdot\partial\|\Omega\cdot\|_1(x))^2 \le d - 2t\ell(\partial\|\Omega\cdot\|_1(x)) + t^2\max_{\|z\|_2\le 1}\|\Omega z\|_1^2.$$

Setting $t = \frac{\ell(\partial\|\Omega\cdot\|_1(x))}{\max_{\|z_2\|\le 1}\|\Omega z\|_1^2}$ yields

$$\inf_{t\ge 0}\mathbb{E}\left[\mathrm{dist}(g, t\cdot\partial\|\Omega\cdot\|_1(x))\right]^2 \le d - \frac{[\ell(\partial\|\Omega\cdot\|_1(x))]^2}{\max_{\|z\|_2\le 1}\|\Omega z\|_1^2}.$$

$\square$

2.4 Number of Gaussian measurements

Gordon's escape through a mesh, Theorem 2, together with the estimates (11) and (12) leads to the next result.

Theorem 3. *Let $\Omega \in \mathbb{R}^{p\times d}$ be an analysis operator and x be cosparse with cosupport Λ. For a random draw $M \in \mathbb{R}^{m\times d}$ of a Gaussian matrix, let noisy measurements $y = Mx + w$ be given with $\|w\|_2 \le \eta$ and $0 < \varepsilon < 1$. If*

$$\frac{m^2}{m+1} \ge \left(\sqrt{d - \frac{(\mathbb{E}\,\|\Omega_\Lambda g\|_1)^2}{\max_{\|z\|_2\le 1}\|\Omega z\|_1^2}} + \sqrt{2\ln(\varepsilon^{-1})} + \tau\right)^2, \tag{13}$$

then with probability at least $1 - \varepsilon$, any minimizer $\hat{x}$ of (2) satisfies

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

Proof. Estimates (11) and (12) imply that

$$E_m - \ell\left(T(x) \cap \mathbb{S}^{d-1}\right) - t \geq E_m - \sqrt{d - \frac{[\ell(\partial\|\Omega \cdot \|_1(x))]^2}{\max_{\|z\|_2 \leq 1} \|\Omega z\|_1^2}} - t. \tag{14}$$

For any $z \in \partial\|\Omega \cdot \|_1(x)$ there is $\alpha \in \mathbb{R}^p$ with $\|\alpha_\Lambda\|_\infty \leq 1$ such that

$$z = \Omega^T \alpha = \Omega^T \operatorname{sgn}(\Omega x) + \Omega^T \alpha_\Lambda.$$

Therefore,

$$\ell(\partial\|\Omega \cdot \|_1(x)) = \mathbb{E} \max_{z \in \partial\|\Omega \cdot \|_1(x)} \langle g, z \rangle = \mathbb{E}\langle g, \Omega^T \operatorname{sgn}(\Omega x)\rangle + \mathbb{E} \max_{\|\alpha_\Lambda\|_1 \leq 1} \langle g, \Omega^T \alpha_\Lambda \rangle$$

$$= \mathbb{E} \max_{\|\alpha_\Lambda\|_1 \leq 1} \langle (\Omega g)_\Lambda, \alpha_\Lambda \rangle = \mathbb{E} \|\Omega_\Lambda g\|_1.$$

Plugging this into (14) gives

$$E_m - \ell\left(T(x) \cap \mathbb{S}^{d-1}\right) - t \geq E_m - \sqrt{d - \frac{(\mathbb{E} \|\Omega_\Lambda g\|_1)^2}{\max_{\|z\|_2 \leq 1} \|\Omega z\|_1^2}} - t.$$

Setting $t = \sqrt{2\ln(\varepsilon^{-1})}$, the choice of m in (13) guarantees that

$$E_m - \ell\left(T(x) \cap \mathbb{S}^{d-1}\right) - t \geq \frac{m}{\sqrt{m+1}} - \sqrt{d - \frac{(\mathbb{E} \|\Omega_\Lambda g\|_1)^2}{\max_{\|z\|_2 \leq 1} \|\Omega z\|_1^2}} - \sqrt{2\ln(\varepsilon^{-1})} \geq \tau.$$

The monotonicity of probability together with Theorems 1 and 2 yields

$$\mathbb{P}\left(\inf_{x \in T(x) \cap \mathbb{S}^{d-1}} \|Mx\|_2 \geq \tau\right) \geq \mathbb{P}\left(\inf_{x \in T(x) \cap \mathbb{S}^{d-1}} \|Mx\|_2 \geq E_m - \ell\left(T(x) \cap \mathbb{S}^{d-1}\right) - t\right) \geq 1 - \varepsilon.$$

This concludes the proof. $\qquad\square$

3 Explicit choice of the analysis operator

Theorem 3 can be further refined for special choices of the operator Ω. We start with the one-dimensional difference operator.

Theorem 4. *Let $\Omega : \mathbb{R}^d \to \mathbb{R}^{d-1}$ be a one-dimensional difference operator. Let $x \in \mathbb{R}^d$ be ℓ-cosparse with respect to Ω and $s = d - 1 - \ell$. For a random draw $M \in \mathbb{R}^{m \times d}$ of a Gaussian matrix, let noisy measurements $y = Mx + w$ be given with $\|w\|_2 \leq \eta$ and $0 < \varepsilon < 1$. If*

$$\frac{m^2}{m+1} \geq \left(\sqrt{d\left(1 - \frac{1}{\pi}\left(1 - \frac{s+1}{d}\right)^2\right)} + \sqrt{2\ln(\varepsilon^{-1})} + \tau\right)^2, \tag{15}$$

then with probability at least $1 - \varepsilon$, any minimizer $\hat{x}$ of (2) satisfies

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

Proof. According to the definition of Ω, for any $z \in \mathbb{R}^d$ it holds

$$\|\Omega z\|_1 = \sum_{i=1}^{d-1} |z_{i+1} - z_i| \le 2\|z\|_1 \le 2\sqrt{d}\|z\|_2. \tag{16}$$

Hence, $\max_{\|z\|_2 \le 1} \|\Omega z\|_1^2 \le 4d$. The rows ω_j of $\Omega \in \mathbb{R}^{(d-1)\times d}$ satisfy $\|\omega_i\|_2 = \sqrt{2}$ and the properties of the Gaussian distribution imply

$$\mathbb{E} \|\Omega_\Lambda g\|_1 = \mathbb{E} \sum_{i\in\Lambda} |\langle g, \omega_i\rangle| = \sqrt{\frac{2}{\pi}} \sum_{i\in\Lambda} \|\omega_i\|_2 = \frac{2}{\sqrt{\pi}}(d-1-s). \tag{17}$$

Plugging (16) and (17) into (13) implies the desired result. $\qquad\square$

A comparison of the theoretical bound to the bound obtained from the numerical experiments is shown on Figure 2. We considered signals in $\mathbb{R}^{200}$. We ran the algorithm and counted the number of times the signal was recovered correctly out of 200 trials. A reconstruction error of 10^{-5} was considered as a successful recovery.
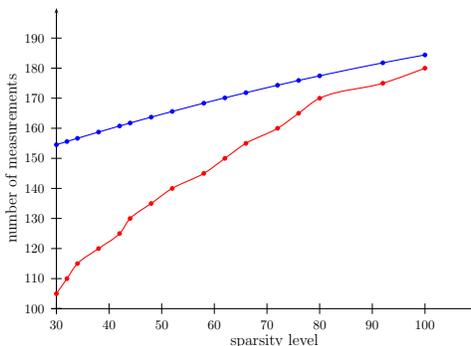

Figure 2: The red line depicts the relation between the number of measurements and the sparsity level from the numerical experiments. The blue line corresponds to the theoretical bound.

Theorem 4 can be extended to two dimensions. Let $X \in \mathbb{R}^{d\times d}$. The two-dimensional difference operator collects all vertical and horizontal derivatives of X into a single vector. If we concatenate the columns of X into the vector $x \in \mathbb{R}^{d^2}$, then we can represent this operator by the matrix $\Omega \in \mathbb{R}^{2d(d-1)\times d^2}$, whose action is given by

$$\Omega x = (X_{21} - X_{11}, \ldots, X_{dd} - X_{d-1d}, X_{12} - X_{11}, \ldots, X_{dd} - X_{dd-1})^T.$$

Each row ω_i of Ω has exactly two non-zero entries with values -1 and 1 at the proper locations.

Theorem 5. *Let $\Omega \in \mathbb{R}^{2d(d-1)\times d^2}$ define a two-dimensional difference operator. Let $x \in \mathbb{R}^{d^2}$ be ℓ-cosparse with respect to Ω and $s = 2d(d-1) - \ell$. For a random draw $M \in \mathbb{R}^{m\times d^2}$ of a Gaussian matrix, let noisy measurements $y = Mx + w$ be given with $\|w\|_2 \le \eta$ and $0 < \varepsilon < 1$. If*

$$\frac{m^2}{m+1} \ge \left(\sqrt{d^2 \left(1 - \frac{1}{\pi}\left(1 - \frac{1}{d} - \frac{s}{2d^2}\right)^2\right)} + \sqrt{2\ln(\varepsilon^{-1})} + \tau \right)^2, \tag{18}$$

9

then with probability at least $1 - \varepsilon$, any minimizer $\hat{x}$ of (2) satisfies

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

Proof. Let $z \in \mathbb{R}^{d^2}$. Then

$$\|\Omega z\|_1 = \sum_{j=1}^{d} \sum_{i=1}^{d-1} |z_{i+1j} - z_{ij}| + \sum_{i=1}^{d} \sum_{j=1}^{d-1} |z_{ij+1} - z_{ij}| \le 4\|z\|_1 \le 4d\|z\|_2$$

and it follows

$$\max_{\|z\|_2 \le 1} \|\Omega z\|_1^2 \le 16d^2.$$

As in the one-dimensional case

$$\mathbb{E} \|\Omega_\Lambda g\|_1 = \sqrt{\frac{2}{\pi}} \sum_{i \in \Lambda} \|\omega_i\|_2 = \frac{2}{\sqrt{\pi}} (2d(d-1) - s).$$

As the final step we apply formula (13). $\qquad\square$

For a Ω being a frame, we obtain the following bound on the required number of measurements.

Theorem 6. *Let $\Omega : \mathbb{R}^d \to \mathbb{R}^p$ be a frame with an upper frame bound B. Let $x \in \mathbb{R}^d$ be cosparse with cosupport Λ. For a random draw $M \in \mathbb{R}^{m \times d}$ of a Gaussian matrix, let noisy measurements $y = Mx + w$ be given with $\|w\|_2 \le \eta$ and $0 < \varepsilon < 1$. If*

$$\frac{m^2}{m+1} \ge \left(\sqrt{d - \frac{2}{\pi} \frac{\left(\sum_{i \in \Lambda} \|\omega_i\|_2\right)^2}{pB}} + \sqrt{2\ln(\varepsilon^{-1})} + \tau \right)^2, \tag{19}$$

then with probability at least $1 - \varepsilon$, any minimizer $\hat{x}$ of (2) satisfies

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

Proof. The only difference to the proof of Theorem 4 is the following estimate, which is due to the Cauchy inequality and the fact that Ω is a frame:

$$\max_{\|z\|_2 \le 1} \|\Omega z\|_1^2 \le p \max_{\|z\|_2 \le 1} \|\Omega z\|_2^2 \le p \max_{\|z\|_2 \le 1} B\|z\|_2^2 \le pB.$$

$\qquad\square$

The bound on the number of measurements (19) does not have an explicit dependence on the ratio of the frame bounds. So it can not be directly compared to the results provided in [9], see also (8). However, an important observation is that the right hand side in (19) is strictly less than d for any p and any number of elements in the cosupport of the signal (provided ε^{-1} and τ are not too large).

4 Conclusions

We have presented results on the nonuniform recovery from Gaussian random measurements of analysis-sparse signals with respect to the one- and two-dimensional difference operators or with respect to a frame. The derived bound on required measurements is always smaller than the ambient dimension of a signal and it is particularly suitable for the case when the sparsity of the analysis representation of the signal is not very small.

References

[1] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: Phase transitions in convex programs with random data. *preprint*, 2013.

[2] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. Anal.*, 31(1):59 – 73, 2011.

[3] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[4] Y. Eldar and G. Kutyniok, editors. *Compressed Sensing - Theory and Applications*. Cambridge Univ. Press, 2012.

[5] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.

[6] R. Foygel and L. Mackey. Corrupted sensing: Novel guarantees for separating structured signals. *IEEE Trans. Inform. Theory*, 60(2):1223–1247, Feb 2014.

[7] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. Davies. Greedy-like algorithms for the cosparse analysis model. *Linear Algebra and Appl.*, 441(0):22 – 60, 2014.

[8] 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)*, volume 1317 of *Lecture Notes in Math.*, pages 84–106. Springer, 1988.

[9] M. Kabanava and H. Rauhut. Analysis ℓ_1-recovery with frames and Gaussian measurements. *preprint*, 2013.

[10] M. Kabanava and H. Rauhut. Cosparsity in compressed sensing. *preprint*, 2014.

[11] S. Nam, M. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Appl. Comput. Harmon. Anal.*, 34(1):30–56, 2013.

[12] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. *SIAM J. Imag. Sci.*, 6(2):1035–1058, 2013.

[13] M. Stojnic. Various thresholds for ℓ_1-optimization in compressed sensing. *ArXiv e-prints*, 2009.

[14] J. A. Tropp. Convex recovery of a structured signal from independent random linear measurements. *ArXiv e-prints*, 2014.