Recovery of cosparse signals with Gaussian measurements

Holger Rauhut
Hausdorff Center for Mathematics
Endenicher Allee 60
53115 Bonn, Germany
Email: rauhut@hcm.uni-bonn.de

Maryia Kabanava
Hausdorff Center for Mathematics
Endenicher Allee 60
53115 Bonn, Germany
Email: maryia.kabanava@hcm.uni-bonn.de

***Abstract*—This paper provides theoretical guarantees for the recovery of signals from undersampled measurements based on ℓ_1-analysis regularization. We provide both nonuniform and stable uniform recovery guarantees for Gaussian random measurement matrices when the rows of the analysis operator form a frame. The nonuniform result relies on a recovery condition via tangent cones and the case of uniform recovery is based on an analysis version of the null space property.**

I. Introduction

Compressed sensing is a recent field of mathematical signal processing that exploits the sparsity of a signal in order to reconstruct it from incomplete and possibly corrupted measurements. A signal $x \in \mathbb{R}^d$ is sparse, if the number of nonzero entries of x, denoted by $\|x\|_0$, is relatively small. The information about $x \in \mathbb{R}^d$ is provided by $m \ll d$ linear measurements

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

where $M \in \mathbb{R}^{m\times d}$ is a measurement matrix and ε corresponds to noise. Since this system is underdetermined it is impossible to recover x from y without additional information.

The most common approach for recovering x is to use regularization. This leads to an optimization problem of the form

$$\min_{z\in\mathbb{R}^d} \|Mz - y\|_2^2 + \lambda R(z).$$

The second term penalizes large values of $R(z)$ and reflects our prior knowledge on the signal to be recovered. In case of noiseless observations $\varepsilon = 0$ we rather use

$$\min_{z\in\mathbb{R}^d} R(z) \quad \text{subject to} \quad Mz = y.$$

The *analysis sparsity prior* assumes that x is sparse in some transform domain, that is, given an analysis operator $\Omega \in \mathbb{R}^{p\times d}$, the vector Ωx is sparse. Such operators can be generated by the discrete Fourier transform, the finite difference operator (related to total variation), wavelet [11], [17], [19] or curvelet transforms [3]. Then the signal is reconstructed by solving

$$\min_{z\in\mathbb{R}^d} \|\Omega z\|_1 \quad \text{subject to} \quad Mz = y. \tag{P_1}$$

Problem (P_1) often appears in image processing [2], [5]. Theoretical guarantees for the successful recovery of x via (P_1) were studied in [4], [7], [10], [13], [14], [20]. In the present paper we assume that the analysis operator is given by a frame. Put formally, let $\{\omega_i\}_{i=1}^p$, $\omega_i \in \mathbb{R}^d$, be a frame, i.e., there exist positive constants A, $B > 0$ such that for all $x \in \mathbb{R}^d$

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

Its elements are collected as rows of the matrix $\Omega \in \mathbb{R}^{p\times d}$. The analysis representation of a signal x is given by the vector $\Omega x = \{\langle \omega_i, x\rangle\}_{i=1}^p \in \mathbb{R}^p$. Cosparsity is then defined as follows.

Definition 1: Let $x \in \mathbb{R}^d$, $\Omega \in \mathbb{R}^{p\times d}$ and $s = \|\Omega x\|_0$. The cosparsity of x with respect to Ω is defined as

$$l := p - s. \tag{2}$$

The index set of the zero entries of Ωx is called the cosupport of x. If x is l-cosparse, then Ωx is s-sparse with $l = p - s$. From Definition 1 it follows, that if Λ is the cosupport of x, then

$$\langle \omega_j, x\rangle = 0, \quad \forall j \in \Lambda.$$

Hence, the set of l-cosparse signals can be written as $\cup_{\#\Lambda=l} W_\Lambda$, where W_Λ is the orthogonal complement of the linear span of $\{\omega_j : j \in \Lambda\}$.

We formulate theoretical guarantees for recovery of cosparse signals (P_1) via tangent cones that are similar to the conditions stated in [6], [12]. Based on this, we are able to provide the following bound on the number of Gaussian measurements required for nonuniform recovery.

Theorem 1: Let x be l-cosparse with $l = p - s$, that is, Ωx is s-sparse. Let $M \in \mathbb{R}^{m\times d}$ be a Gaussian random matrix and $0 < \varepsilon < 1$. If

$$\frac{m^2}{m+1} \ge \frac{2Bs}{A}\left(\sqrt{\ln\frac{ep}{s}} + \sqrt{\frac{A\ln(\varepsilon^{-1})}{Bs}}\right)^2, \tag{3}$$

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

Roughly speaking, a fixed l-cosparse vector is recovered with high probability from $m > 2(B/A)s\ln(ep/s)$ Gaussian measurements. For $\Omega = \mathrm{Id}$, this bound strengthens a result in [6]. We can also incorporate the case of noisy measurements (1). But for the ease of presentation, we omit it here.

Usually, the signals to be recovered are only approximately cosparse. The quantity

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

describes the ℓ_1-best approximation error to Ωx by s-sparse vectors. The Ω-null space property of M to be defined below ensures stability of reconstruction. Analyzing it for Gaussian random matrices leads to the following stable and uniform recovery result.

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

$$\frac{m^2}{m+1} \geq \frac{2Bs\left(1+\rho^{-1}\right)^2}{A}\left(\sqrt{\ln\frac{ep}{s}} + \frac{1}{\sqrt{2}} + \sqrt{\frac{A\ln(\varepsilon^{-1})}{Bs}}\right)^2, \tag{4}$$

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

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

For the standard case $\Omega = \mathrm{Id}$, this theorem improves the main result in [18] with respect to the constant and adds stability in ℓ_2.

We will give proof sketches here. Detailed arguments will be contained in [15].

We use the notation Ω_Λ to refer to a submatrix of Ω with the rows indexed by Λ. $(\Omega x)_S$ stands for the vector in $\mathbb{R}^p$ whose entries indexed by S coincide with the entries of Ωx and the rest are filled by 0. Let B_2^p denote a unit ball in $\mathbb{R}^p$ with respect to the ℓ_2-norm.

II. Nonuniform recovery from Gaussian measurements

In the present section we provide bounds on the number of measurements required for exact recovery of x by (P_1), where $M \in \mathbb{R}^{m\times d}$ is a Gaussian random matrix. We use the idea presented in [6], that requires to calculate the Gaussian widths of tangent cones.

For fixed $x \in \mathbb{R}^d$, we define the convex cone

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

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

Proof: First assume that $\ker M \cap T(x) = \{0\}$. Let $z \in \mathbb{R}^d$ be a vector that satisfies

$$\|\Omega z\|_1 \leq \|\Omega x\|_1 \text{ subject to } Mz = Mx.$$

This means that $z - x \in T(x)$ and $z - x \in \ker M$. According to our assumption we conclude that $z - x = 0$, so that x is the unique minimizer.

On the other hand, if x is the unique minimizer of (P_1), then $\|\Omega(x+v)\|_1 > \|\Omega x\|_1$ for all $v \in \ker M \setminus \{0\}$, which implies that $v \notin T(x)$. This means that

$$(\ker M \setminus \{0\}) \cap T(x) = \emptyset$$

or equivalently $\ker M \cap T(x) = \{0\}$. ∎

To prove Theorem 1 we rely on Theorem 3, which requires that the null space of the measurement matrix M misses the set $T(x)$. The next ingredient of the proof is a variation of Gordon's escape through the mesh theorem [9], which was first used in the context of compressed sensing in [18]. To formulate this theorem whose proof will be present in a journal paper in preparation, we introduce some notation.

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

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

we have

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

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 4: Let $\Omega \in \mathbb{R}^{p\times d}$ be a frame with constants A, $B > 0$. 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

$$\mathbb{P}\left(\inf_{x\in T}\|Mx\|_2 > E_m - \frac{1}{\sqrt{A}}\ell\left(\Omega(T)\right) - t\right) \geq 1 - e^{-\frac{t^2}{2}}, \tag{5}$$

where $\Omega(T)$ corresponds to the set of elements produced by applying Ω on elements from T.

With $T := T(x)\cap S^{d-1}$ the number of Gaussian measurements required to guarantee the exact reconstruction of x with probability $1 - e^{-t^2/2}$ is determined by

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

If Ω is a frame, then

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

where

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

The supremum over a larger set can only increase, hence

$$\ell(\Omega(T)) \leq \sqrt{B}\,\ell\left(K(\Omega x) \cap B_2^p\right). \tag{6}$$

We next give an upper bound for $\ell\left(K(\Omega x) \cap B_2^p\right)$ involving the polar cone $\mathcal{N}(\Omega x) = K(\Omega x)^\circ$ defined by

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

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

$$\ell\left(K(\Omega x) \cap B_2^p\right) \leq \mathbb{E}\min_{z\in\mathcal{N}(\Omega x)}\|g - z\|_2. \tag{7}$$

The proof is an application of convex analysis, see [1], [6]. Now the problem of estimating $\ell(\Omega(T))$ is reduced to bounding $\mathbb{E} \min_{z\in N(\Omega(x))} \|g-z\|_2$, where Ωx is an s-sparse vector. By Hölder's inequality

$$\left(\mathbb{E} \min_{z\in\mathcal{N}(\Omega x)} \|g-z\|_2\right)^2 \leq \mathbb{E} \min_{z\in\mathcal{N}(\Omega x)} \|g-z\|_2^2 \tag{8}$$

and with some extra calculation (improving slightly over a bound in [6]) we can show that

$$\mathbb{E} \min_{z\in\mathcal{N}(\Omega x)} \|g-z\|_2^2 \leq 2s \ln \frac{ep}{s}.$$

Together with inequalities (6) and (7) this gives

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

Proof of Theorem 1: Set $t = \sqrt{2\ln(\varepsilon^{-1})}$. The fact that $E_m \geq m/\sqrt{m+1}$ along with condition (3) yields

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

Theorem 4 implies

$$\mathbb{P}\left(\inf_{x\in T} \|Mx\|_2 > 0\right) \geq 1 - e^{-\frac{t^2}{2}} = 1 - \varepsilon,$$

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

III. Ω-NULL SPACE PROPERTY

The proof of Theorem 2 is based on the following concept.

Definition 2: A matrix $M \in \mathbb{R}^{m\times d}$ is said to satisfy the ℓ_2-stable Ω-null space property of order s with $0 < \rho < 1$, if for any set $\Lambda \subset [p]$ with $\#\Lambda \geq p - s$ it holds

$$\|\Omega_{\Lambda^c} v\|_2 < \frac{\rho}{\sqrt{s}} \|\Omega_\Lambda v\|_1 \quad \text{for all } \; v \in \ker M \setminus \{0\}. \tag{9}$$

This is the strengthened version of the recovery condition stated in [13]. If Ω is the identity map $\mathrm{Id} : \mathbb{R}^d \to \mathbb{R}^d$, then condition (9) becomes the standard ℓ_2-stable null space property [8].

Theorem 5: Let $\Omega \in \mathbb{R}^{p\times d}$ be a frame and $M \in \mathbb{R}^{m\times d}$ satisfy the ℓ_2-stable Ω-null space property of order s with constant $0 < \rho < 1$. Then for any $x \in \mathbb{R}^d$ the solution $\hat{x}$ of (P_1) with $y = Mx$ approximates the vector x with ℓ_2-error

$$\|x - \hat{x}\|_2 \leq \frac{2(1+\rho)^2}{\sqrt{A}(1-\rho)} \frac{\sigma_s(\Omega x)_1}{\sqrt{s}}. \tag{10}$$

Inequality (10) means that l-cosparse vectors are exactly recovered by (P_1) and vectors $x \in \mathbb{R}^d$, such that Ωx is close to an s-sparse vector in ℓ_1, can be well approximated in ℓ_2 by a solution of (P_1). The proof goes along the same lines as in the standard case. For the sake of brevity we omit it here.

IV. UNIFORM RECOVERY FROM GAUSSIAN MEASUREMENTS

The ℓ_2-stable Ω-null space property of order s of the measurement matrix $M \in \mathbb{R}^{m\times d}$ ensures the exact recovery of any l-cosparse vector by solving (P_1). The same strategy as in the Section II allows us to give the bound on the number of Gaussian measurements required for the ℓ_2-stable Ω-null space property to hold.

To prove Theorem 2 let us introduce the set

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

If

$$\min\{\|Mw\|_2 : w \in W_{\rho,s} \cap S^{d-1}\} > 0, \tag{11}$$

then for all $w \in \ker M \setminus \{0\}$ and any $\Lambda \subset [p]$ with $\#\Lambda = p-s$ we have

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

which implies that M satisfies the ℓ_2-stable Ω-null space property of order s. To show (11) we apply Theorem 4, according to which we have to study the Gaussian width of the set $\Omega\left(W_{\rho,s} \cap S^{d-1}\right)$. Since Ω is a frame, we have

$$\Omega\left(W_{\rho,s} \cap S^{d-1}\right) \subset \Omega\left(W_{\rho,s}\right) \cap \left(\sqrt{B}B_2^p\right) \subset T_{\rho,s} \cap \left(\sqrt{B}B_2^p\right),$$

with

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

Then

$$T_{\rho,s} \cap \left(\sqrt{B}B_2^p\right) = \bigcup_{\#S=s} \left\{w \in \mathbb{R}^p : \|w\|_2 \leq \sqrt{B}, \right. \\ \left. \|w_S\|_2 \geq \frac{\rho}{\sqrt{s}}\|w_{S^c}\|_1\right\}.$$

Lemma 1: Let the set D be defined by

$$D := \mathrm{conv}\{x \in S^{p-1} : \#\,\mathrm{supp}\, x \leq s\}.$$

Then

$$T_{\rho,s} \cap \left(\sqrt{B}B_2^p\right) \subset \left(1 + \rho^{-1}\right)\left(\sqrt{B}D\right). \tag{12}$$

A similar result was stated as Lemma 4.5 in [18], so we omit the proof.

Lemma 1 implies that

$$\ell\left(T_{\rho,s} \cap \left(\sqrt{B}B_2^p\right)\right) \leq \sqrt{B}\left(1+\rho^{-1}\right)\ell(D).$$

Lemma 2: The Gaussian width of D satisfies

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

Proof: Due to the definition of the Gaussian width

$$\ell(D) = \mathbb{E} \sup_{x\in D} \langle g, x\rangle = \mathbb{E} \sup_{\substack{\|x\|_2=1,\\ \#\,\mathrm{supp}\, x\leq s}} \langle g, x\rangle, \tag{13}$$

where $g \in \mathbb{R}^p$ is a standard Gaussian random vector. Hölder's inequality applied to (13) and an estimate on the maximum squared ℓ_2-norm of a sequence of standard Gaussian random vectors (see e.g. [16, Lemma 3.2]) give

$$\ell(D) \leq \mathbb{E} \max_{S\subset[p],\#S=s} \|g_S\|_2 \leq \sqrt{\mathbb{E} \max_{S\subset[p],\#S=s} \|g_S\|_2^2}$$

$$\leq \sqrt{2\ln\binom{p}{s}} + \sqrt{s} \leq \sqrt{2s\ln\frac{ep}{s}} + \sqrt{s}.$$

The last inequality follows from the fact that

$$\binom{p}{s} \leq \left(\frac{ep}{s}\right)^s.$$

■

Proof of Theorem 2: The reasoning above shows that

$$\ell\left(\Omega\left(W_{\rho,s}\cap S^{d-1}\right)\right) \leq \sqrt{B}\left(1+\rho^{-1}\right)\ell(D)$$
$$\leq \sqrt{B}\left(1+\rho^{-1}\right)\left(\sqrt{2s\ln\frac{ep}{s}}+\sqrt{s}\right).$$

Set $t = \sqrt{2\ln(\varepsilon^{-1})}$. The fact that $E_m \geq m/\sqrt{m+1}$ along with condition (4) yields

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

Theorem 4 implies

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

$$\geq 1 - e^{-\frac{t^2}{2}} = 1-\varepsilon,$$

which guarantees

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

for all $w \in \ker M\backslash\{0\}$ and any $\Lambda \subset [p]$ with $\#\Lambda = p-s$. This means that M satisfies the ℓ_2-stable Ω-null space property of order s. Finally, apply Theorem 5. ■

V. Uniform recovery from Gaussian measurements

In this work we provided conditions that guarantee the uniqueness of the solution of the optimization problem (P_1), when the analysis operator is given by a frame. The modification of the Gordon's escape through the mesh theorem allowed to derive a bound on the number of Gaussian random measurements needed to satisfy these conditions.

Acknowledgment

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

References

[1] S. Boyd, L. Vandenberghe. Convex optimization. Cambridge University Press, Cambridge, 2004.
[2] J.-F. Cai, S. Osher, Z. Shen. Split Bregman methods and frame based image restoration. *Multiscale Model. Simul.*, 8(2):337–369, 2009/10.
[3] 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.
[4] 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.
[5] T. Chan, J. Shen. Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods. SIAM, 2005.
[6] 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.
[7] M. Elad, P. Milanfar, R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23(3):947–968, 2007.
[8] S. Foucart, H. Rauhut. A Mathematical Introduction to Compressive Sensing. Birkhäuser, to appear.
[9] Y. Gordon. 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.*, pages 84–106. Springer, Berlin, 1988.
[10] 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.
[11] S. Mallat. A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press, 2008.
[12] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.*, 17(4):1248–1282, 2007.
[13] S. Nam, M.E. Davies, M. Elad, R. Gribonval. The cosparse analysis model and algorithms. *Appl. Comput. Harmon. Anal.*, 34(1):30–56, 2013.
[14] D. Needell, R. Ward. Stable image reconstruction using total variation minimization. http://arxiv.org/abs/1202.6429
[15] H.Rauhut, M. Kabanava. Analysis ℓ_1–recovery with frames and Gaussian measurements. In preparation.
[16] N. Rao, B. Recht, R. Nowak. Tight measurement bounds for exact recovery of structured sparse signals. In Proceedings of AISTATS, 2012.
[17] 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.
[18] M. Rudelson, R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.
[19] I. Selesnick, M. Figueiredo. Signal restoration with overcomplete wavelet transforms: comparison of analysis and synthesis priors. Proceedings of SPIE, vol. 7446, 2009, p. 74460D.
[20] S. Vaiter, G. Peyré, Ch. Dossal, J. Fadili. Robust sparse analysis regularization. *IEEE Transactions on information theory*, 59(4):2001–2016, 2013.