2018 5th International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing, and Imaging (CoSeRa)

Weighted sparse recovery with expanders

Bubacarr Bah
African Institute for Mathematical Sciences (AIMS) South Africa and
Division of Applied Mathematics, Stellenbosch University
Cape Town, South Africa
Email: bubacarr@aims.ac.za

***Abstract*—Weighted sparse recovery refers to sparse recovery where the signals have "weighted sparsity" and weighted l1 minimization is used for reconstruction. We propose novel weighted sparse recovery guarantees with expander matrices, using a weighted versions of the null space property to derive these guarantees. These expender matrices are very sparse, hence they have favourable computational benefits compared to their dense counterparts. Moreover, we show that it is possible to achieve sample complexities that are linear in the weighted sparsity of the signal, where the sampling rates can be smaller than those of standard sparse recovery. Furthermore, these results reduce to known results in standard sparse recovery and sparse recovery with prior information.**

I. Introduction

A. Problem statement and overview

In using scaled binary matrices as sensing matrices, the weighted sparse recovery problem, introduced in [17], is modified in the following way. Firstly, let $\mathbf{x} \in \mathbb{R}^n$ be the target signal, which is k-weighted sparse, i.e. $\sum_{i\in\mathcal{S}} \omega_i \leq s$, where $\mathcal{S}$ is the support of $\mathbf{x}$ and $1 \leq \omega_i < \infty$ for $i \in [n]$. We sense $\mathbf{x}$ using a measurement (sensing) matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ to get observations $\mathbf{y} \in \mathbb{R}^m$, precisely $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ is a noise vector, which is bounded, i.e. $\|\mathbf{e}\|_1 \leq \eta < \infty$. To recover $\mathbf{x}$ from $\mathbf{y}$ we solve the following modified weighted ℓ_1-minimization problem.

$$\min_{\mathbf{z}\in\mathbb{R}^n} \|\mathbf{z}\|_{\omega,1} \quad \text{subject to} \quad \|\mathbf{A}\mathbf{z}-\mathbf{y}\|_1 \leq \eta, \quad (1)$$

where $\|\mathbf{z}\|_{\omega,1} = \sum_{i=1}^n \omega_i |z_i|$ with weights $\omega_i \geq 1$. The difference in (1) from the standard weighted ℓ_1-minimization problem is the replacement of the ℓ_2-norm by the ℓ_1-norm in the data fidelity term (constraints). The scaled binary matrices, $\mathbf{A}$, used in this work are adjacency matrices of expander graphs defined below.

Definition I.1. *Let $G = ([N],[n],E)$ be a left-regular bipartite graph with N left vertices, n right vertices, a set of edges E and left degree d. If, for any $\epsilon \in (0,1/2)$ and any $\mathcal{S} \subset [N]$ of size $|\mathcal{S}| \leq k$, we have that $|\Gamma(\mathcal{S})| \geq (1-\epsilon)d|\mathcal{S}|$, then G is referred to as a (k,d,ϵ)*-expander graph.

These matrices can extremely sparse having only d ones (non-zeros entries) per column. This make them more attractive for computational purposes. Moreover, we used weighted sparsity instead of standard sparsity due to the fact we assume to have prior knowledge of the structure of the class of signals we consider. Due to this prior knowledge we assign weights that control the likelihood of the inclusion of certain indices in the support of our signal. More precisely, in the above set-up smaller weights are assigned to those indices which are deemed "more likely" to belong to the true underlying support.

In the standard sparse recovery setting, binary matrices are known to possess what is referred to as the *square-root bottleneck*, that is they require $m = \Omega\left(k^2\right)$ rows instead of the *optimal* $\mathcal{O}\left(k\log\left(N/k\right)\right)$ rows to be "good" compressed sensing matrices with respect to optimal recovery guarantees in the ℓ_2 norm, see [7], [8]. Yet, in [4], the authors show that such sparse matrices achieve optimal sample complexity (*optimally* few rows of $\mathcal{O}\left(k\log\left(N/k\right)\right)$) if one instead considers error guarantees in the ℓ_1 norm. This manuscript develops comparable results for sparse binary matrices in the setting of weighted ℓ_1 minimization.

The contributions of this work include (i) the introduction of the weighted robust null space property, satisfied by adjacency matrices of (k,d,ϵ)-expander graphs, see Definition II.1 in Section II-A; (ii) the characterization of weighted sparse recovery guarantees for (1) using these matrices, see Theorem II.2, in Section II-A; (iii) the derivation of sampling rates that are linear in the weighted sparsity of the signals using such matrices, see Theorem II.3 in Section II-B. Numerical experiments support the theoretical results, see Section III.

The author was partially supported by Rachel Wards NSF CAREER Grant, award #1255631, and by the Alexander von Humboldt Foundation, under the programme financed by the BMBF entitled "German Research Chairs".

B. Related work

Firstly, the use of sparse sensing matrices from expander graphs in standard sparse recovery is not new, see [4], which is motivated by computational advantages of these matrices over their dense counterparts. However, this work is the first to apply these matrices to weighted sparse recovery.

Secondly, weighted sparse recovery which was introduced in [17] was using structured matrices and bounded orthonormal systems, which was dictated by the polynomial interpolation problem they authors were looking at. The results derived in [17] were extended to the use of Gaussian matrices in [3]. This work is a further extension to the use of expander matrices.

Standard sparse recovery with weighted ℓ_1 minimization has a bit of a longer history, see [1], [3], [5], [6], [10]–[13], [15]–[20]. Unlike this work, all of these prior works considered dense matrices, either subgaussian matrices or structured random matrices.

An earlier version of this work is on arXiv, [2]. Due to space issues a lot of details are left out in this paper, in particular we skip all the proofs in this paper and refer the interested reader to the longer version of the paper on arXiv, i.e. [2].

II. Theoretical results

The main results of this work give recovery guarantees for weighted ℓ_1 minimization (1) when the sampling operators are adjacency matrices of expander graphs for the class of weighted sparse signals. We characterize the appropriate weighted robust null space property and expansion condition that the adjacency matrix of a (k, d, ϵ)-expander graph needs to satisfy to guarantee robust weighted sparse recovery. These results reduce to the standard sparsity and unweighted ℓ_1 minimization results when the weights are all chosen to be equal to one. We derive sample complexities, in terms of the weighted sparsity s, of weighted sparse recovery using weighted ℓ_1 minimization compared to unweighted ℓ_1 minimization with adjacency matrices of a (k, d, ϵ)-expander graphs. These sample complexities are linear in $\omega(\mathcal{S})$ and reduce to known results of standard sparse recovery and sparse recovery with prior information.

A. Robust weighted sparse recovery guarantees

The *weighted null space property* ($\boldsymbol{\omega}$-NSP) has been used to give sparse recovery guarantees [12]–[14], [17] with two schemes for choice of weights. In [12], [17] the weights $\boldsymbol{\omega} \geq 1$; whilst in [13], [14] the weights $\boldsymbol{\omega} \leq 1$. Similar to [17], we consider the *weighted robust NSP* ($\boldsymbol{\omega}$-RNSP) for the type of matrices we focus on, which is the robust version of the NSP in the weighted case and follows from the unweighted RNSP proposed in [9] for such matrices.

Definition II.1 ($\boldsymbol{\omega}$-RNSP). *Given a weight vector $\boldsymbol{\omega}$, a matrix $\mathbf{A} \in \mathbb{R}^{n\times N}$ is said to have the robust $\boldsymbol{\omega}$-RNSP of order s with constants $\rho < 1$ and $\tau > 0$, if*

$$\|\mathbf{v}_\mathcal{S}\|_{\omega,1} \leq \rho \|\mathbf{v}_{\mathcal{S}^c}\|_{\omega,1} + \tau\sqrt{s}\|\mathbf{A}\mathbf{v}\|_1, \qquad (2)$$

for all $\mathbf{v} \in \mathbb{R}^N$ and all $\mathcal{S} \subset [N]$ with $\omega(\mathcal{S}) \leq s$.

We will derive conditions under which an expander matrix satisfies the $\boldsymbol{\omega}$-RNSP to deduce error guarantees for weighted ℓ_1 minimization (1). This is formalized in the following theorem.

Theorem II.1 (Theorem 3.1, [2]). *Let the matrix $\mathbf{A} \in \{0,1\}^{n\times N}$ be the adjacency matrix of a (k, d, ϵ)-expander graph. If $\epsilon_{2k} < 1/6$, then $\mathbf{A}$ satisfies the $\boldsymbol{\omega}$-RNSP* (2) *with*

$$\rho = \frac{2\epsilon_{2k}}{1-4\epsilon_{2k}}, \text{ and } \tau = \frac{1}{\sqrt{d}(1-4\epsilon_{2k})}. \qquad (3)$$

Based on Theorem II.1 we provide reconstruction guarantees in the following theorem.

Theorem II.2 (Theorem 3.2, [2]). *Let $\mathbf{A}$ be the adjacency matrix of a (k, d, ϵ)-expander graph with $\epsilon_{2k} < 1/6$. Given any $\mathbf{x} \in \mathbb{R}^N$, if $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_1 \leq \eta$, a solution $\widehat{\mathbf{x}}$ of* (1) *is an approximation of $\mathbf{x}$ with the following error bounds*

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_{\omega,1} \leq C_1\sigma_s(\mathbf{x})_{\omega,1} + C_2\sqrt{s}\eta, \qquad (4)$$

where the constants $C_1, C_2 > 0$ depend only on d and ϵ.

B. Sample complexity

Here we derive sample complexities in terms of the weighted sparsity, s, of weighted sparse recovery using weighted ℓ_1-minimization with sparse adjacency matrices of (k, d, ϵ)-expander graphs. These sample complexity bounds are linear in the weighted sparsity of the signal and can be smaller than sample complexities of standard sparse recovery using unweighted ℓ_1-minimization with and sparse adjacency matrices of (k, d, ϵ)-expander graphs. Moreover, these results recover known results for the settings of a) uniform weights, b) polynomially growing weights, c) sparse recovery with prior support estimates, and d) known support. In particular, in the setting of *sparse recovery with prior support estimates*, depending on mild assumptions on the growth of the weights and how well is the support estimate aligned with the true support will lead to a reduction in sample complexity. The following derivations, without loss of generality, assume an ordering of the entries of the signal in order of magnitude such that $\mathcal{S}$ has the first k largest in magnitude entries of the signal.

Theorem II.3 (Theorem 3.3, [2])**.** *Fix weights $\omega_j \geq 1$. Suppose that $\gamma > 0$ depending on the choice of weights, and $0 \leq \delta < 1$. Consider an adjacency matrix of a (k, d, ϵ)-expander $\mathbf{A} \in \{0,1\}^{n \times N}$, and a signal $\mathbf{x} \in \mathbb{R}^N$ supported on $\mathcal{S} \subset [N]$ with $|\mathcal{S}| \leq k$ and $\sum_{i \in \mathcal{S}} \omega_i \leq s$. Assume that noisy measurements are taken, $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_1 \leq \eta$ and $\epsilon_{2k} < 1/6$. Then with probability at least $1 - \delta$, any solution $\widehat{\mathbf{x}}$ of* (1) *satisfies* (4)*, if*

$$n = \mathcal{O}\left(s/(\epsilon^2\gamma)\right), \text{ and } d = \mathcal{O}\left(\epsilon n/k\right). \quad (5)$$

III. Experimental Results

In these experiments we consider the class of *weighted sparse signals* modeled in [3]. Precisely, the probability for an index to be in the support of the signal is proportional to the reciprocal of the square of the weights assigned to that index. We also considered *polynomially growing weights*. In particular, we assign weights $\omega_j = j^{1/5}$ where the indices are ordered such that the support corresponds to the smallest in magnitude set of weights. The goal of the experiments was to compare the performance of weighted sparse recovery using weighted ℓ_1-minimization and standard sparse recovery using unweighted ℓ_1-minimization using both Gaussian sensing matrices and sensing matrices that are sparse binary adjacency matrices of expander graphs (hence forth referred to as expander matrices) in terms of a) sample complexity b) computational runtimes, and c) accuracy of reconstruction. The $m \times N$ Gaussian matrices have i.i.d. standard normal entries scaled by $\sqrt{m}$ while the expander matrices are generated by putting d ones at uniformly at random locations in each column. We draw signals of dimension N from the above mentioned model, where the nonzero values are randomly generated as scaled sums of Gaussian and uniformly random variables without any normalization. We encode the signals using these matrices and add Gaussian white noise with noise level $\|\mathbf{e}\|_2 \leq 10^{-6} =: \eta_2$ and define η_1 such that $\|\mathbf{e}\|_1 \leq \eta_1$. For the weighted sparse reconstruction, we use (1) with expanders and use a modified version of (1), replacing the ℓ_1 by ℓ_2 and η_1 by η_2 in the data fidelity term of (1), with Gaussian matrices; while the standard sparse reconstruction used

$$\min_{\mathbf{z} \in \mathbb{R}^N} \|\mathbf{z}\|_{\omega,1} \text{ subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_p \leq \eta_p, \quad (6)$$

with $p = 1$ for expanders and $p = 2$ for Gaussian matrices.

The following results are averaged over many realizations for each problem instance (s, m, N). The dimension of the signal is $N = 2^{10}$. For the expander matrices we fixed $d = \lceil 2\log(N) \rceil$ and we vary the number of measurements m such that $m/N \in [\max(2d/N, 0.05), 0.35]$; and for each m we vary the weighted sparsity of the supp($\mathbf{x}$), $\mathcal{S}$, such that $\omega(\mathcal{S})/m = s/m \in [1/\min(m), 2.5]$. Then we record k as the largest $|\mathcal{S}|$ for a given s. We consider a reconstruction successful if the recovery error in the ℓ_2-norm is below $10\eta_1$ or $10\eta_2$ for expander or Gaussian matrices respectively and a failure otherwise. Then we compute the *empirical probabilities* as the ratio of the number of successful reconstructions to the number of realizations.

A. Sample complexities via phase transitions

We present below sample complexity comparisons using the phase transition framework in the phase space of $(s/m, m/N)$. Note that in all the figures we normalized (standardized) the values of s/m in such a way that the normalized s/m is between 0 and 1 for fair comparison. Figure 1 shows phase transition curves in the form of contours of empirical probabilities of 50% (solid curves) and 95% (dashed curves) for expander and Gaussian matrices using either ℓ_1 or $\ell_{\omega,1}$ minimization. Both matrices have similar performance and by having larger area under the contours, weighted sparse recovery using (1), outperforms standard sparse recovery using (6).

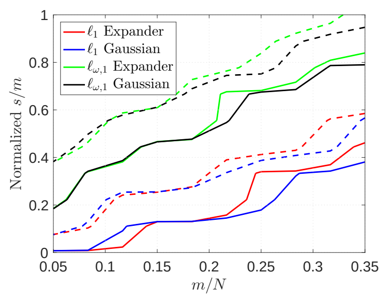

Fig. 1. Contour plots depicting phase transitions of 50% and 95% recovery probabilities (dashed and solid curves respectively).

The result in Figure 1 is further elucidated by the plots in Figure 2 and Figure 3. In the latter we show a snap shot for fixed $s/m = 1.25$ and varying m while in the former we show a snap shot for fixed $m/N = 0.1625$ and varying s. Both plots confirm the comparative performance of expanders to Gaussian matrices and the superiority of weighted ℓ_1 minimization over unweighted ℓ_1 minimization.

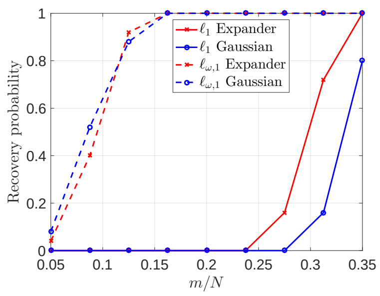

Fig. 2. Recovery probabilities for a fixed $s/m = 1.25$ and varying m.

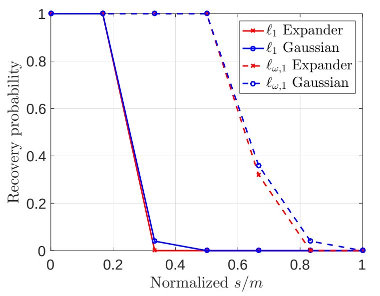

Fig. 3. Recovery probabilities for a varying s and fixed $m/N = 0.1625$.

B. Computational runtimes

To compare runtimes we sum the generation time of $\mathbf{A}$ (Gaussian or expander), encoding time of the signal using $\mathbf{A}$, and the reconstruction time, with weighted ℓ_1 minimization over unweighted ℓ_1 minimization, and we average this over the number of realizations. In Figure 4 we plot average runtimes for varying m/N. This clearly shows that expanders have small runtimes.

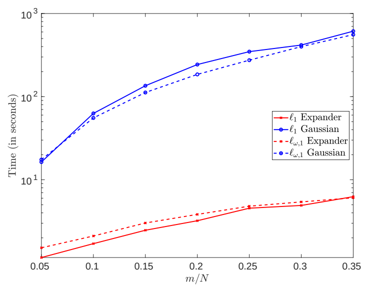

Fig. 4. Runtime comparisons.

C. Accuracy of reconstructions

In Figures 5 and 6 we plot relative approximation errors in the $\ell_{\omega,1}$ norm. Figure 5 is for a fixed $s/m = 1.25$ and varying m; while Figure 6 for fixed $m/N = 0.1625$ and varying s. In Figures 7 and 8 we plot relative approximation errors in the ℓ_2 norm. Similarly, Figure 7 is for a fixed $s/m = 1.25$ and varying m; while Figure 8 is for fixed $m/N = 0.1625$ and varying s. In both sets of figures we see that weighted ℓ_1 minimization converges faster with smaller number of measurements than unweighted ℓ_1 minimization; but also we see that Gaussian sensing matrices have smaller approximation errors than the expanders.

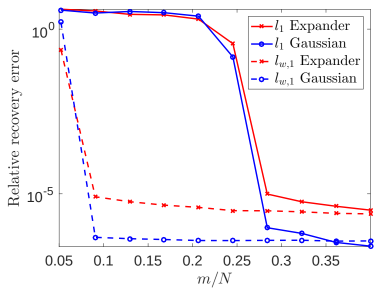

Fig. 5. Relative errors in the $\ell_{\omega,1}$ norm for a fixed $s/m = 1.25$ and varying m.

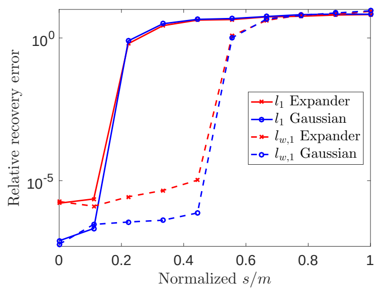

Fig. 6. Relative errors in the $\ell_{\omega,1}$ norm for a fixed $m/N = 0.1625$ and varying s.

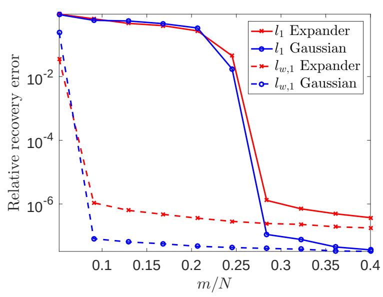

Fig. 7. Relative errors in the ℓ_2 norm for a varying m and fixed $s/m = 1.25$.

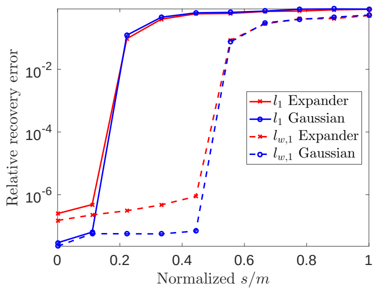

Fig. 8. Relative errors in the ℓ_2 norm for a varying s and fixed $m/N = 0.1625$.

IV. Conclusion

We give the first rigorous error guarantees for weighted ℓ_1 minimization with sparse measurement matrices and weighted sparse signals. The matrices are computationally efficient considering their fast application and low storage and generation complexities. The derivation of these error guarantees uses the weighted robust null space property proposed for the more general setting of weighted sparse recovery. We also derived sampling rates for weighted sparse recovery using these matrices. These sampling bounds are linear in s and can be smaller than sampling rates for standard sparse recovery depending on the choice of weights.

References

[1] Ben Adcock. Infinite-dimensional ℓ_1 minimization and function approximation from pointwise data. *arXiv preprint arXiv:1503.02352*, 2015.

[2] Bubacarr Bah. Weighted sparse recovery with expanders. *arXiv preprint arXiv:1604.01067*, 2016.

[3] Bubacarr Bah and Rachel Ward. The sample complexity of weighted sparse approximation. *arXiv preprint arXiv:1507.06736*, 2015.

[4] R. Berinde, A.C. Gilbert, P. Indyk, H. Karloff, and M.J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 798–805. IEEE, 2008.

[5] Jean-Luc Bouchot, Benjamin Bykowski, Holger Rauhut, and Christoph Schwab. Compressed sensing Petrov-Galerkin approximations for parametric PDEs. 2015.

[6] Emmanuel J Candès, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008.

[7] V.t Chandar. A negative result concerning explicit matrices with the restricted isometry property. *preprint*, 2008.

[8] R.A. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23(4):918–925, 2007.

[9] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[10] Michael P Friedlander, Hassan Mansour, Rayan Saab, and Oezguer Yilmaz. Recovering compressively sampled signals using partial support information. *Information Theory, IEEE Transactions on*, 58(2):1122–1134, 2012.

[11] Laurent Jacques. A short note on compressed sensing with partially known signal support. *Signal Processing*, 90(12):3308–3312, 2010.

[12] Amin M Khajehnejad, Weiyu Xu, Amir S Avestimehr, and Babak Hassibi. Weighted ℓ_1 minimization for sparse recovery with prior information. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pages 483–487. IEEE, 2009.

[13] Hassan Mansour and Rayan Saab. Recovery analysis for weighted ℓ_1-minimization using a null space property. *arXiv preprint arXiv:1412.1565*, 2014.

[14] Hassan Mansour and Özgür Yilmaz. Weighted-ℓ_1 minimization with multiple weighting sets. In *SPIE Optical Engineering + Applications*, pages 813809–813809. International Society for Optics and Photonics, 2011.

[15] Samet Oymak, M Amin Khajehnejad, and Babak Hassibi. Recovery threshold for optimal weight ℓ_1 minimization. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2032–2036. IEEE, 2012.

[16] Ji Peng, Jerrad Hampton, and Alireza Doostan. A weighted ℓ_1-minimization approach for sparse polynomial chaos expansions. *Journal of Computational Physics*, 267:92–111, 2014.

[17] Holger Rauhut and Rachel Ward. Interpolation via weighted ℓ_1 minimization. *Applied and Computational Harmonic Analysis*, 2015.

[18] Namrata Vaswani and Wei Lu. Modified-CS: Modifying compressive sensing for problems with partially known support. *Signal Processing, IEEE Transactions on*, 58(9):4595–4607, 2010.

[19] R Von Borries, Jacques C Miosso, and Cristhian M Potes. Compressed sensing using prior information. In *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*, pages 121–124. IEEE, 2007.

[20] Weiyu Xu, M Amin Khajehnejad, Amir Salman Avestimehr, and Babak Hassibi. Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the grassmann angle framework. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 5498–5501. IEEE, 2010.