LAPLACE MIXTURES MODELS FOR EFFICIENT COMPRESSED SENSING WITH SIDE INFORMATION

Chiara Ravazzi

Institute of Electronics,
Computer and Telecommunication Engineering,
National Research Council, Italy

Enrico Magli

Department of Electronics
and Telecommunications (DET),
Politecnico di Torino, Italy

ABSTRACT

In this paper, we propose a new method for the recovery of a sparse signal from few linear measurements using a reference signal as side information. Modeling the signal coefficients with a double Laplace mixture model, and assuming that the distribution of the components of the prior information differs slightly from the unknown signal, the problem is formulated as a weighted ℓ_1 minimization problem.

We derive sufficient conditions for perfect recovery and we show that our method is able to reduce significantly the number of measurements required for reconstruction. Numerical experiments demonstrate that the proposed approach outperforms the best algorithms for compressed sensing with prior information and is robust in imperfect scenarios.

***Index Terms*—** Compressed sensing, mixture models, side information, sparse recovery, weighted ℓ_1 minimization.

1. INTRODUCTION

The theory of compressed sensing (CS) has proved that a k-sparse signal $x^\star \in \mathbb{R}^n$ (i.e., it has at most $k \ll n$ nonzero entries) can be recovered from a small collection of linear measurements $y = Ax^\star \in \mathbb{R}^m$ ($m \ll n$) via the constrained ℓ_1 minimization, that consists in selecting the element which is compatible with the observations which has minimal ℓ_1-norm.

In this paper, we consider the problem of compressed sensing with side information as addressed in [1]. More precisely, we are interested in recovering the high dimensional signal $x^\star$ from y, with the additional information that $x^\star$ is similar to a reference signal w. This problem arises in several situations, as in compressive image sampling [2, 3], where the spatial and temporal correlation within image/video is exploited. Also in sensor/camera networks [4, 5, 6], the signals acquired by close sensors are similar and can be used as prior information to reduce the number of measurements needed for reconstruction. We refer to [1] for an overview of the applications.

In [1], the authors propose to solve the following optimization problem

$$\min_{x\in\mathbb{R}^n} \|x\|_1 + \gamma\|x - w\|_p^p \quad \text{s.t. } y = Ax \tag{1}$$

with $p \in \{1, 2\}$ and $\gamma > 0$, referred as ℓ_1-ℓ_1 minimization, and ℓ_1-ℓ_2 minimization, respectively. Moreover, sufficient conditions on the number of measurements for perfect reconstruction are derived. In particular, it is shown that the number of measurements required by ℓ_1-ℓ_1 minimization is much smaller than that obtained using classical CS.

The use of prior information as a tool to reduce the number of measurements required for signal reconstruction has appeared in CS literature [1, 7, 8] also with different assumptions. In [7], the authors employ as prior information an estimate T of the support of $x^\star$ and propose a truncated ℓ_1-minimization problem, i.e. the minimization of

$$\min \|x_{T^c}\|_1 \quad \text{s.t. } y = Ax^\star. \tag{2}$$

It should be noticed that (2) can be adapted to our problem using $T = \text{supp}(w)$ (Mod-CS, [7]). Another piece of literature [8, 9] considers a weighted ℓ_1-minimization with weights $w_i = -\log p_i$ where p_i is the probability that $x_i = 0$. It should be remarked that in our setting p_i is not available.

In this paper, we propose a new weighted ℓ_1 minimization, which we call 2LMM-CS. The fundamental idea is to use a *good generative model* for sparse and compressible vectors [10]. For this purpose, we use a Laplace mixture model (2LMM) as the parametric representation of the prior distribution of the signal coefficients. Because of the partial symmetry of the signal sparsity, we know that each coefficient should have one out of only two distributions: a Laplace with small variance with high probability and a Laplace with large variance with low probability. This model has been shown effective to represent sparse signals or compressible signals in [11]. Then, we cast the estimation problem as a non convex optimization problem that incorporates the parametric representation of the signal. However, the optimization problem turns out to be computationally hard. Assuming that

This work is supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n.279848

the distribution of the nonzero coefficients is similar to that of unknown signal, the estimation problem is simplified to a weighted ℓ_1-minimization.

We show that under certain conditions the number of measurements required for reconstruction can be significantly reduced compared to the techniques used in literature before. Finally, numerical experiments show that 2LMM-CS achieves excellent performance in several situations and outperforms the state of the art on this subject.

2. SUPPORT DETECTION AND SPARSE SIGNAL ESTIMATION VIA 2-LMM

2.1. Modeling sparse or compressible vectors

We consider a two-state mixture model as a prior that describes our knowledge about the sparsity of the signal $x^\star$. Because of the partial symmetry of the signal sparsity, we consider the case in which x is a random variable of the form

$$x_i = z_i u_i + (1 - z_i) v_i \quad i \in [n]$$

where u_i are identically and independently distributed (i.i.d.) according to $\mathsf{Laplace}(0, \alpha)$, v_i are i.i.d. as $\mathsf{Laplace}(0, \beta)$ and z_i are i.i.d. Bernoulli random variables with probability mass function $f(z_i = 1) = 1 - p$, with $p = K/n \ll 1/2$, $\alpha \approx 0$, $\beta \gg 0$, and $K \geq k$ is an estimate of the signal sparsity, in order to ensure that we have few large coefficients. This mixture model is completely described by three parameters: the sparsity ratio $p \ll 1/2$ (or K, equivalently), α that is expected to be small and $\beta > \alpha$ if the signal is sparse. It should be noticed that vectors generated from this distribution are typically compressible, according to definition [10].

2.2. Estimation using 2-LMM generative model

Let $\Theta = (\alpha, \beta)$ and consider the logarithm of the conditional distribution: $L(x; \Theta) := \log[f(x|y; \Theta)]$

Proposition 1. *Given y, A, Θ,*

$$-L(x;\Theta) = \begin{cases} J(x, \pi; \Theta) - \sum_{i=1}^n H(\pi_i) & \text{if } y = Ax \\ +\infty & \text{if } y \neq Ax \end{cases} \tag{3}$$

where

$$J(x, \pi; \Theta) = \sum_{i=1}^n \left[\frac{\pi_i |x_i|}{\alpha} + \pi_i \log \alpha - \pi_i \log(1-p) + \frac{(1-\pi_i)|x_i|}{\beta} + (1-\pi_i)\log\beta - (1-\pi_i)\log p \right], \tag{4}$$

$\pi_i = \pi_i(x) = \mathbb{E}[z_i | x; \Theta] = f(z_i = 1 | x; \Theta)$ *and H is the natural entropy function.*

The proof is obtained as a simple consequence of Jensen's inequality and using the logarithm properties.

Corollary 1. *The following optimization problems are equivalent*

$$\max_{x \in \mathbb{R}^n} L(x; \Theta) \tag{5}$$

$$\min_{x \in \mathbb{R}^n} \min_{\pi \in \mathbb{R}^n} J(x, \pi; \Theta) - \sum_{i=1}^n H(\pi_i) \quad \textit{s.t. } Ax = y \tag{6}$$

Given y, A, we consider the following modified optimization problem:

$$\min_{x \in \mathbb{R}^n} \min_{\pi \in \Sigma_{n-K}} J(x, \pi; \Theta) - \sum_{i=1}^n H(\pi_i) \quad \text{s.t. } Ax = y \tag{7}$$

that introduces the constraint $\pi \in \Sigma_{n-K}$, which allows to take into account that we seek a sparse solution with a guess of the sparsity level K. It should be noted that there is not a closed form solution to problem (7). However, partial minimization of function in (7) with respect to π leads to the following expression.

Lemma 1. *Let*

$$\widehat{\pi} = \widehat{\pi}(x, \Theta) = \operatorname*{arg\,min}_{\pi \in \Sigma_{n-K}} J(x, \pi; \Theta) - \sum_{i=1}^n H(\pi_i)$$

then

$$\widehat{\pi} = \sigma_{n-K}\left(\frac{1}{1 + \frac{\alpha}{\beta}\frac{p}{1-p} \mathrm{e}^{|x|\left(\frac{1}{\alpha} - \frac{1}{\beta}\right)}} \right). \tag{8}$$

where

$$\sigma_j(v) = \min_{\{z \in \mathbb{R}^n:\ |\mathrm{supp}(z)| \leq j\}} \|z - v\|_2$$

is a thresholding operator which acts on v by keeping the j biggest elements in absolute value and setting the others to zero.

3. COMPRESSED SENSING WITH PRIOR INFORMATION VIA 2LMM

Let us consider the optimization problem in (7) and suppose α and β are fixed and $0 \approx \alpha < \beta$. Assuming that the distribution of the signal coefficients of w is similar to that of $x^\star$, we fix

$$\pi_i = \pi_i(w) = f(z_i = 1 | w, \alpha, \beta) = \frac{1}{1 + \frac{\alpha}{\beta}\frac{p}{1-p}\mathrm{e}^{|w_i|\left(\frac{1}{\alpha} - \frac{1}{\beta}\right)}}. \tag{9}$$

and $\widehat{\pi}(w) = \sigma_{n-K}(\pi_i(w))$. Let $T^{\mathrm{c}} = \mathrm{supp}(\widehat{\pi})$ and $T = \{i \in [n] : \widehat{\pi}_i = 0\}$. We have $\widehat{\pi}_i = \pi_i, \forall i \in T^{\mathrm{c}}$ and $\widehat{\pi}_i = 0, \forall i \in T$.

It should be noticed that, given α, β, π, minimization of (7) over x is equivalent to computation of

$$\min_{x \in \mathbb{R}^n} \sum_{i \in T} \omega |x_i| + \sum_{i \in T^{\mathrm{c}}} \left(\pi_i + (1 - \pi_i)\omega\right) |x_i|, \tag{10}$$

s.t. $Ax = y$ with $\omega = \alpha/\beta$.

Definition 1. *Let $\Lambda \subset \{1\dots,n\}$ with $|\Lambda| = k \leq K$, $\omega \in [0,1]$, $\pi \in [0,1]^n$, and $T = \{i \in [n] : [\sigma_{n-K}(\pi)]_i = 0\}$. A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the non-uniform weighted (ω,π,K,Λ)-NSP if for any $h \in \mathrm{Ker}(A) \setminus \{0\}$, we have*

$$\omega\|h_\Lambda\|_1 + (1-\omega)\sum_{i\in S}\pi_i|h_i| < \sum_{i\in\Lambda^c}(\pi_i + (1-\pi_i)\omega)|h_i|$$

where $S = (\Lambda \cap T^c) \cup (\Lambda^c \cap T)$.

Definition 1 is non-uniform and depends on a fixed set Λ. As will be clear in next results, this condition is necessary and sufficient for the recovery of a sparse vector supported on Λ using (10). The following definition, instead, considers a weighted uniform null space property, which is a necessary and sufficient condition for the recovery of all k-sparse vectors from compressive measurements via (10).

Definition 2. *Let $\omega \in [0,1]$ and $\pi \in [0,1]^n$. A matrix $A \in \mathbb{R}^{m\times n}$ is said to satisfy the weighted (ω,π,K)-Null Space Property (NSP) of order k if it satisfies the non-uniform weighted (ω,π,K,Λ)-Null Space Property (NSP) for all $\Lambda \subset [n]$ with $|\Lambda| \leq k$.*

Theorem 1. *The weighted ℓ_1-minimization in* (10) *uniquely recovers every k-sparse vector $x^\star$ from measurements $y = Ax^\star$ if and only if A satisfies the (ω,π,K)-NSP of order k.*

Theorem 2. *Let $\Lambda \subset [n]$ with $|\Lambda| \leq k \leq K$, $\omega \in [0,1]$, $\pi \in [0,1]^n$, and $T = \{i \in [n] : [\sigma_{n-K}(\pi)]_i = 0\}$ and $A \in \mathbb{R}^{m\times n}$ be a matrix whose entries are i.i.d. Gaussian random variables with zero-mean and unit variance. Then A satisfies non-uniform weighted (ω,π,K,Λ)-NSP with probability greater than $1-\epsilon$ if the following condition holds*

$$\frac{m}{\sqrt{m+1}} \geq \sqrt{k+|S|} + c_1\sqrt{\frac{k}{\ln(\mathrm{e}n/k)}} + \sqrt{2\ln\epsilon^{-1}} + c_2\sqrt{\left[(1-\omega)^2\sum_{i\in S}\pi_i^2 + \omega^2 k + \sum_{i\in\Lambda\cap T^c}2\pi_i\omega(1-\omega)\right]\ln\left(\frac{\mathrm{e}n}{k}\right)}. \quad (11)$$

where $S = (\Lambda \cap T^c) \cup (\Lambda^c \cap T)$ and c_1, c_2 are constants independent of $k, n, |S|$, and δ.

The proof is obtained by modifying analogous proofs for NSP for Gaussian matrices provided in [12, 13].

Corollary 2. *Let $\Lambda \subset [n]$ with $|\Lambda| \leq k$, $\omega \in [0,1]$, $\pi \in [0,1]^n$, $T = \{i \in [n] : [\sigma_{n-K}(\pi)]_i = 0\}$ and $A \in \mathbb{R}^{m\times n}$ be a matrix whose entries are i.i.d. Gaussian random variables with zero-mean and unit variance. The weighted ℓ_1-minimization in* (10) *uniquely recovers $x^\star$ supported on Λ with $|\Lambda| \leq k \leq K$ from measurements $y = Ax^\star$ if it holds condition* (11).

The following corollary shows that in the large system limit, as n is large enough and for a sufficiently small $\alpha \approx 0$ and the prior information has good enough quality, then the number of measurements sufficient for perfect reconstruction of a sparse vector supported on Λ with $|\Lambda| \leq k$ can be significantly reduced.

Corollary 3. *Let $\alpha \approx 0$, $k \leq K$, $\pi = \pi(w)$ as defined in* (9) *and $A \in \mathbb{R}^{m\times n}$ be a matrix whose entries are i.i.d. Gaussian random variables with zero-mean and unit variance. If $\|w - x^\star\| \leq \frac{1}{2}\min_{i\in\Lambda}|x_i^\star|$ then the weighted ℓ_1-minimization in* (10) *uniquely recovers $x^\star$ from measurements $y = Ax^\star$ if it holds condition*

$$m \geq K + O\left((K-k)\frac{1}{1+\frac{\alpha}{\beta}\mathrm{e}^{|r(w)_K|(1/\alpha - 1/\beta)}}\ln(\mathrm{e}n/k)\right),$$

where $r(w)$ the non increasing rearrangement of w, i.e. $r(w) = (|w_{i_1},\dots,w_{i_n}|)$ with $|w_{i_\ell} \geq w_{i_{\ell+1}}|$ for all $\ell = 1,\dots,n-1$.

It should be noticed if $|r(w)_K| \neq 0$, being $\alpha \approx 0$ the second term is very small, suggesting that just K measurements are sufficient for recovery.

4. NUMERICAL EXPERIMENTS

We compare 2LMM-CS with classical CS and the best algorithms for CS with side information known in literature: Mod-CS, ℓ_1-ℓ_1 and ℓ_1-ℓ_2 minimization (see the Sec. I for an overview of these methods).

As a first experiment, we employ the same setting analyzed in [1]. A signal $x^\star$ of length $n = 1000$ is generated with sparsity $k = 70$. The nonzero elements of $x^\star$ are drawn from a standard Gaussian distribution. The prior information w is obtained $w = x^\star + z$, where z is a 28-sparse signal, whose nonzero elements are drawn from a Gaussian distribution with standard deviation 0.8. The vector z is such that $|\mathrm{supp}(z) \cap \mathrm{supp}(x^\star)| = 22$ and $|\mathrm{supp}(z) \cap \mathrm{supp}(x^\star)^c| = 6$. The resulting vector w differs significantly from the true vector $x^\star$ and the relative distance is $\|x^\star - w\|_2/\|x^\star\|_2 = 0.502$. The sensing matrix A with m rows and n columns is sampled from the Gaussian ensemble with zero mean and variance $1/m$. In ℓ_1-ℓ_1 and ℓ_1-ℓ_2 minimization $\gamma = 1$, as employed in [1]. For CS-2LMM the mixture parameters have been set as follows: $\alpha = 10^{-4}$, $\beta = 10$, $K = |\mathrm{supp}(w)| = 76$. Instead Mod-CS uses as prior information $T = \mathrm{supp}(w)$. Fig. 1 shows the empirical recovery success rate, averaged over 50 experiments, as a function of the number of measurements m. For a fixed m, we mean the success when a given algorithm reconstructs the signal $x^\star$ with a relative error smaller than 10^{-2}. It should be noticed that ℓ_1-ℓ_1 minimization achieves the best performance, if compared with CS, Mod-CS and ℓ_1-ℓ_2 minimization. It requires $m \geq 140$ measurements to recover perfectly the signal with a probability larger than 0.95. It should be appreciated that 2LMM-CS reduces this number to 80.

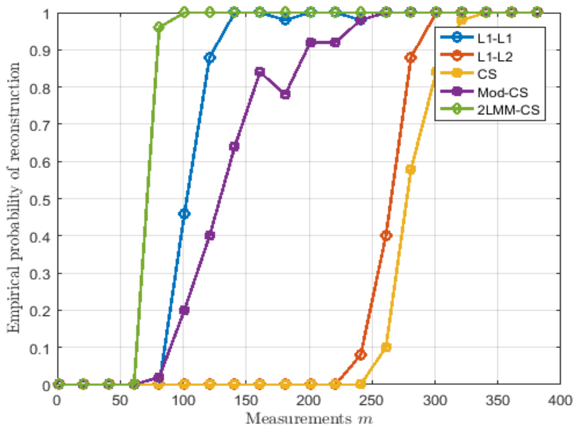

Fig. 1. Empirical probability of reconstruction of classical CS (BP), ℓ_1-ℓ_1 minimization, ℓ_1-ℓ_2 minimization, and 2LMM-CS.

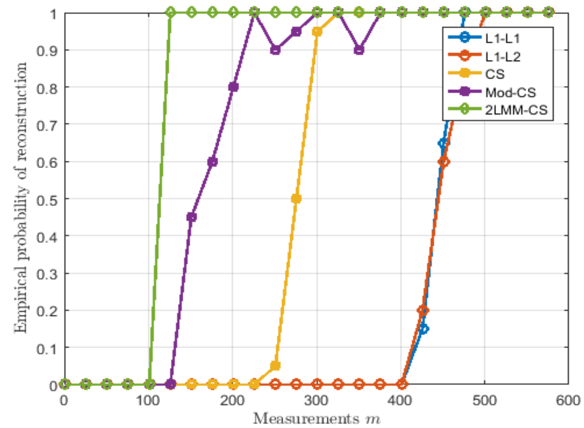

Fig. 2. Empirical probability of reconstruction of classical CS (BP), ℓ_1-ℓ_1 minimization, ℓ_1-ℓ_2 minimization, and 2LMM-CS.

We now investigate the performance of the algorithms in imperfect scenarios. We consider signal $x^\star$ and z generated as in the previous experiment. The prior information w is obtained by $w = x^\star + z + \eta$, where η is a gaussian noise with standard deviation 10^{-3}. The resulting relative error is $\|w - x\|/\|x\| = 0.6489$. The sensing matrix A is sampled from the Gaussian ensemble with zero mean and variance $1/m$. Mod-CS uses as prior information the set T of the 123 largest components in absolute value of vector w. For CS-2LMM the mixture parameters have been set as follows: $\alpha = 10^{-4}$, $\beta = 10$, and $K = 123$. Fig. 2 depicts the empirical recovery success rate, averaged over 50 experiments, as a function of the number of measurements m. It should be noticed that 2LMM-CS achieves the best performance and the condition for perfect reconstruction is $m \geq 125$. Mod-CS has the second best performance requiring $m \geq 225$ measurements, followed by classical CS with $m \geq 325$. In this setting, ℓ_1-ℓ_1 minimization and ℓ_1-ℓ_2 minimization require $m \approx 500$, behaving poorly in this context.

Finally, we consider the case in which $x^\star$ is generated as above and the prior information is a blurred version of $x^\star$. More precisely, w is obtained from $x^\star$ by applying a blur filter of order 3: $w_i = \left(x^\star_{i-1} + x^\star_i + x^\star_{i+1}\right)/3$. The relative error is $\|w-x\|/\|x\| = 0.8147$. Figure 3 emphasizes that 2LMM-CS achieves the best performance also in this setting, requiring about 175 measurements for reconstruction. Instead, Mod-CS, CS and ℓ_1-ℓ_1 minimization need a number of measurements for reconstruction dramatically larger ($m \geq 325$) than 2LMM-CS. Moreover ℓ_1-ℓ_1 has the same performance of CS, bringing no significant benefits in this case. Instead, the ℓ_1-ℓ_2 minimization is not able to perform the recovery with a number of measurements smaller than 425.

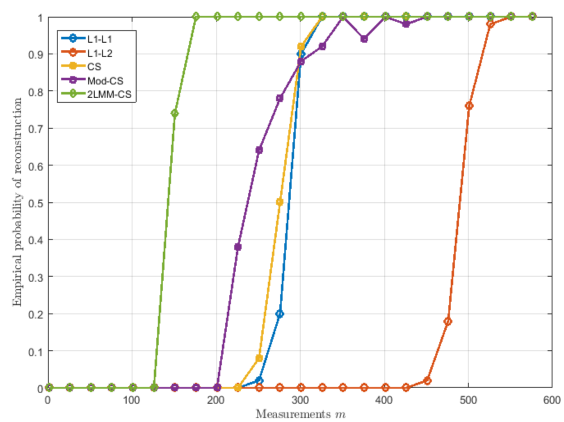

Fig. 3. Empirical probability of reconstruction of classical CS (BP), ℓ_1-ℓ_1 minimization, ℓ_1-ℓ_2 minimization, and 2LMM-CS.

5. CONCLUDING REMARKS

In this paper, we have shown a new method to efficiently perform sparse recovery in presence of side information. Combining MAP estimation with the parametric representation of the signal with a Laplace mixture model, we have formulated the problem as a weighted ℓ_1-minimization. The main theoretical contribution includes the derivation of sufficient conditions for perfect recovery. Numerical simulations show that these new algorithms reduce the number of measurements required for reconstruction and are robust for several models of prior information.

6. REFERENCES

[1] J. F. C. Mota, N. Deligiannis, and M. R. D. Rodrigues, "Compressed sensing with prior information: Optimal strategies, geometry, and bounds," Preprint: http://dx.doi.org/10.1109/18.382009.

[2] L. W. Kang and C. S. Lu, "Distributed compressive video sensing," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 1169–1172.

[3] V. Stanković, L. Stanković, and S. Cheng, "Compressive image sampling with side information," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov 2009, pp. 3037–3040.

[4] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 92–101, March 2008.

[5] V. Cevher, A. Sankaranarayanan, M. F. Duarte, D. Reddy, R. G. Baraniuk, and R. Chellappa, *Compressive Sensing for Background Subtraction*, pp. 155–168, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[6] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," http://arxiv.org/abs/0901.3403, 2005 (revised 2009).

[7] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.

[8] M. Amin Khajehnejad, Weiyu Xu, A. Salman Avestimehr, and Babak Hassibi, "Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models," 2011.

[9] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing Sparsity by Reweighted L1 Minimization," Tech. Rep.

[10] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5016–5034, Aug 2012.

[11] C. Ravazzi and E. Magli, "Bayesian tuning for support detection and sparse signal estimation via iterative shrinkage-thresholding," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4628–4632.

[12] Mark Rudelson and Roman Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Comm. Pure Appl. Math*, p. 1045.

[13] M. P. Friedlander, H. Mansour, R. Saab, and Ö. Yilmaz, "Recovering compressively sampled signals using partial support information.," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1122–1134, 2012.