Signal sparsity estimation from compressive noisy projections via -sparsified random matrices

SIGNAL SPARSITY ESTIMATION FROM COMPRESSIVE NOISY PROJECTIONS VIA γ-SPARSIFIED RANDOM MATRICES

C. Ravazzi, S. M. Fosson, T. Bianchi, and E. Magli

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

ABSTRACT

In this paper, we propose a method for estimating the sparsity of a signal from its noisy linear projections without recovering it. The method exploits the property that linear projections acquired using a sparse sensing matrix are distributed according to a mixture distribution whose parameters depend on the signal sparsity. Due to the complexity of the exact mixture model, we introduce an approximate two-component Gaussian mixture model whose parameters can be estimated via expectation-maximization techniques. We demonstrate that the above model is accurate in the large system limit for a proper choice of the sensing matrix sparsifying parameter. Moreover, experimental results demonstrate that the method is robust under different signal-to-noise ratios and outperforms existing sparsity estimation techniques.

Index Terms— Compressed sensing, Gaussian mixture models, sparse matrices, sparsity

1. INTRODUCTION

Compressed Sensing (CS) [1, 2] is a novel signal acquisition technique based on the recovery of an unknown signal from a small set of linear measurements. The main result of CS is that if a signal having dimension n is known to be sparse, i.e., it can be well approximated by only $k \ll n$ nonzero entries in a suitable basis, then it can be efficiently recovered using only $m \ll n$ linear combinations of the signal entries.

Most of the applications of CS usually assume that the sparsity k is known before acquiring the signal. However, in many practical settings this is not always the case. Some signals may have a time-varying sparsity, as in spectrum sensing [3], or spatially-varying sparsity, as in the case of block based acquisition of images [4]. Since the number of linear measurements required for the recovery depends on the sparsity degree of the signal [5], the knowledge of k is crucial for a CS system. Also, many recovery algorithms require to know the sparsity of the signal for an optimal tuning of parameters. For example, Lasso techniques [6] require to choose a parameter λ which is related to k [7], whereas for greedy algorithms, such as Orthogonal Matching Pursuit (OMP) [8] or Compressive Sampling Matching Pursuit (CoSaMP) [9], the number of iterations is bounded by k. Another problem is that in some cases may not be immediately clear whether the signal is actually sparse and which is the correct sparsifying basis [10].

Due to the deployment of many practical CS systems, the problem of estimating the sparsity degree has begun to be recognized as a major gap between theory and practice [11, 12] and the literature on the subject is very recent [10, 13, 14]. In [10], the sparsity of the signal is lower-bounded through the numerical sparsity, i.e., the ratio between the ℓ_1 and ℓ_2 norms of the signal, where these quantities can be estimated from random projections obtained using Cauchy distributed and Gaussian distributed matrices, respectively. However, measurement taken with Cauchy distributed matrices cannot be used later for signal reconstruction. In [15], the authors propose to estimate the sparsity of an image before its acquisition, by calculating the image complexity. However, the proposed metric is based on the image pixel values, forcing to calculate a separate estimation that does not depend on the measurements.

In order to deal with an unknown sparsity degree, some authors propose sequential acquisition techniques, in which the number of measurements is dynamically adapted until a satisfactory reconstruction performance is achieved [16, 17, 18, 19]. Since these methods require to solve a minimization problem at each newly acquired measurement, they may prove too complex when the underlying signal is not sparse, or if one is only interested in assessing the sparsity degree of a signal under a certain basis without reconstructing it.

In this paper, we propose a method for estimating the sparsity degree of a signal directly from its linear measurements. The method is based on projections obtained using a γ-sparsified random matrix and exploits the fact that, under this kind of matrices, the measurements are distributed according to a mixture whose parameters depend on k. The proposed method extends the algorithm in [13], which works only in the case of exactly k-sparse signals, to the case of noisy sparse signals or only compressible signals. The fundamental idea is to approximate the measurements model by a two-component Gaussian mixture model (2-GMM),

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.

whose parameters can be easily estimated via expectation-maximization techniques. As a major contribution, we prove that there is a regime of behavior, defined by the scaling of the measurement sparsity γ and the sparsity k, where this approximation is accurate. An interesting property of the proposed method is that measurements acquired using a γ-sparsified random matrix also enable signal reconstruction, with only a slight performance degradation with respect to dense matrices [20, 21].

2. PROBLEM FORMULATION

In this paper, we consider the following affine system

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

where $x \in \Sigma_k := \{x \in \mathbb{R}^n : |\mathrm{supp}(x)| = k \ll n\}$ is an unknown deterministic signal $x \in \mathbb{R}^n$ with exactly k nonzero entries (we refer to k as the signal sparsity), $y \in \mathbb{R}^m$ is the observation vector, $A \in \mathbb{R}^{m\times n}$ is the sensing matrix, and $\eta \in \mathbb{R}^m$ is a Gaussian noise $\mathcal{N}(0, \sigma^2 I)$. Our goal is to estimate k from the measurements y without recovering the signal x.

Since we are trying to estimate the signal sparsity from noisy measurements, we expect that the performance of any sparsity estimator will depend on $\lambda = \min_{i\in\mathrm{supp}(x)} |x_i|$. Our analysis considers γ-sparsified matrices [20], in which the entries of the matrix A are independently and identically distributed according to

$$A_{ij} = \begin{cases} \mathcal{N}(0, \frac{1}{\gamma}) & \text{w.p. } \gamma, \\ 0 & \text{w.p. } 1-\gamma. \end{cases} \tag{2}$$

It is well known [20] that if $\gamma k = \Theta(1)$ and $\lambda^2 k = \Theta(1)$, then at least $\max\{\Theta(k\log(n/k)), \Theta(k\log(p-k)/\log k)\}$ measurements are necessary for estimating the support of the signal.

For simplicity of exposition and in order to emphasize the relevant parameters in the estimation problem, we consider the restricted class of signals introduced in [20] (Restricted ensemble A): $\mathcal{X} = \{x \in \Sigma_k : |x_i| = \lambda, \forall i \in \mathrm{supp}(x)\}$. However, our analysis can be extended to any k-sparse signal. With the above assumptions, any measurement $y_i = \sum_{j\in[n]} A_{ij}x_j + \eta_i$ is a random variable whose distribution is a mixture of $k+1$ Gaussians

$$\begin{aligned} &y_i \sim \mathcal{N}(0, \alpha_s) \quad \text{w.p. } p_s = \binom{k}{s}\gamma^s(1-\gamma)^{k-s}, \\ &\alpha_s = \lambda^2 s/\gamma + \sigma^2 = s\alpha_1 - (s-1)\alpha_0, \end{aligned} \tag{3}$$

with $s \in \{0, \dots, k\}$ denoting the number of nonzero entries of i-th row in A colliding with the signal support. Given the set of m independent and identically distributed samples $y = (y_1, \dots, y_m)^\top$, the corresponding log-likelihood function is given by

$$\log f(y|\alpha, k) = \sum_{i=1}^m \log \sum_{s=0}^k p_s \phi(y_i|\alpha_s)$$

where $\phi(y_i|\alpha_s) = \frac{1}{\sqrt{2\pi\alpha_s}} \exp\left(-x^2/(2\alpha_s)\right)$.

In this paper the sparsity evaluation is recast into the problem of estimating the number of mixture components and parameters from the collection of samples $(y_1, \dots, y_m)$. A natural solution to our estimation problem would be to consider a ML estimation (see [22]). Let $z = (z_{is})_{i\in[m], s\in\{0\}\cup[k]}$ be the matrix whose entry z_{is} is the hidden variable, which is equal to one if and only if component s produced measurement i and zero otherwise. Note that each column contains exactly one entry equal to one. Let $f(y, z)$ be the joint distribution of y and hidden variables z and consider the log-likelihood function

$$\begin{aligned} L(y, z|\alpha, k) &:= \log f(y, z|\alpha, k) \\ &= \sum_{i=1}^m \sum_{s=0}^k z_{is} \log\left(p_s \phi(y_i|\alpha_s)\right), \end{aligned}$$

where the last equality is true since z_{is} is zero for all but one term in the inner sum. The ML solution prescribes to choose $k \in \{0, \dots, n\}$ and $\alpha \in \mathbb{R}_+^2$ such that

$$(\widehat{\alpha}_{\mathrm{ML}}, \widehat{k}_{\mathrm{ML}}) = \arg\max_{k,\alpha} \mathbb{E}_z \left[L(y, z|\alpha, k)\right]. \tag{4}$$

As already noted, even if the signal sparsity k were known, the maximum likelihood problem (4) would be a typical estimation problem for a finite mixture distribution and would not admit a closed-form solution, and the computational complexity of (4) would be practically unfeasible. One possible approach is to resort to the well known Expectation-Maximization (EM) algorithm in [22]. This is a powerful tool for finding ML solutions to problems involving observed and hidden variables which is known to converge to a local maximum of the likelihood.

3. ITERATIVE SPARSITY ESTIMATION BASED ON EXPECTATION-MAXIMIZATION

The EM algorithm can be adapted to estimate also the number of mixture components and, consequently, the signal sparsity. For brevity we shall only report the final equations. Starting from initial values $k(0), \alpha_0(0), \alpha_1(0)$, and computing for $s = 2, \dots, k(0)$ the parameters $p_s(0), \alpha_s(0)$ by (3), at each iteration we need to update the posterior probabilities

$$\pi_{is}(t+1) = p_s(t)\phi(y_i|\alpha_s(t)) / \sum_{\ell=0}^{k(t)} p_\ell(t)\phi(y_i|\alpha_\ell(t))$$

and the mixture parameters $p_0(t+1) = \sum_{i=1}^m \pi_{i0}(t+1)/m$, $k(t+1) = \lceil \log(p_0(t+1))/\log(1-\gamma) \rceil$, and $\alpha_0(t+1)$, $\alpha_0(t+1)$ by solving

$$\min_{\alpha_0, \alpha_1} \sum_{s=0}^{k(t+1)} \sum_{i=1}^m \left(\frac{\pi_{is}(t+1)y_i^2}{\alpha_s(\alpha_0, \alpha_1)} + \pi_{is}(t+1) \log \alpha_s(\alpha_0, \alpha_1), \right)$$

where $\alpha_s(\alpha_0, \alpha_1) = s\alpha_1 - (s-1)\alpha_0$. Finally $p_s(t+1), \alpha_s(t+1)$ are given by (3) with $s = 2, \dots, k(t+1)$.

However, the EM implementation suffers from the following limitations. (a) There is not a closed formula to update the variances, due to the fact that the variances of the mixture components are dependent. (b) The computational complexity of the algorithm depends on the signal sparsity: as k increases, the number of parameters to update gets larger. Moreover we are not interested in a complete classification/clustering of the data but only to discriminate the measurements that are produced by noise in order to estimate the signal sparsity.

3.1. 2-GMM approximation for large system limit

Our main goal is to show that the EM presented in the previous section can be simplified in the large system limit as $n, k \to \infty$. The next theorem reveals that there is a regime of behavior, defined by the scaling of the measurement sparsity γ and the signal sparsity k, where the measurements can be approximately described by a two-component Gaussian mixture model. We state this fact formally below. Recall that, given two distributions f, g their Kolmogorov distance is defined by

$$\|f - g\|_{\mathrm{K}} = \sup_{a \in \mathbb{R}} \left| \int_{-\infty}^{a} f(x)\mathrm{d}x - \int_{-\infty}^{a} g(x)\mathrm{d}x \right|.$$

Theorem 1. *Let us consider the sequence of distribution functions $f_k(y) = \sum_{s=0}^{k} p_s \phi(y|\alpha_s)$. If $\gamma k = \tau$ and $\lambda^2 k = \chi$, then*

$$\lim_{k \to \infty} \|f_k - f_\star\|_{\mathrm{K}} \leq C\tau$$

where C is a constant (independent of parameters τ) and

$$f_\star(y) = \mathrm{e}^{-\tau} \phi\left(y|\sigma^2\right) + (1 - \mathrm{e}^{-\tau})\phi\left(y \middle| \sigma^2 + \frac{\chi}{1 - \mathrm{e}^{-\tau}}\right)$$

The proof of Theorem 1 can be obtained by re-working on the Jensen's inequality and is omitted for brevity. The quantile-quantile plot in Figure 1 compares a sample of data generated by k-GMM (vertical axis) with parameter $\tau = \gamma k$ to a data sample generated by 2-GMM (horizontal axis). It should be noticed that when τ is small (left), the points in the picture approximately lie on 45 degree reference line. This linearity suggests that the compared distributions are similar. If τ increases then the points follow a different pattern, suggesting that the data samples from k-GMM are not well approximated by the 2-GMM probability model. This confirms that sparsifying the measurement ensemble has asymptotic effect in the accuracy of the approximation as derived in Theorem 1.

In the regime $\gamma k = \Theta(1)$ and $\lambda^2 k = \Theta(1)$, using the approximation in Theorem 1, we recast the problem of inferring the signal sparsity as the problem of estimating the parameters of a two-component Gaussian mixture whose joint distribution of y and hidden variable z is given by $f(y, z|\alpha, \beta, p) = pz\phi(y|\alpha) + (1-p)(1-z)\phi(y|\beta)$ with $z \in \{0, 1\}$. Starting from an initial guess of mixture parameters $\alpha(0), \beta(0), p(0)$, Algorithm 1 computes, at each iteration, the posterior distribution $\pi_i(t) = \mathbb{P}(z_i = 1|\alpha(t), \beta(t), p(t))$ (E-Step) and re-estimate the mixture parameters (M-Step) until a stopping criterion is satisfied. Finally the estimation of the signal sparsity is provided by $\widehat{k} = \log(p_{\text{final}})/\log(1-\gamma)$.

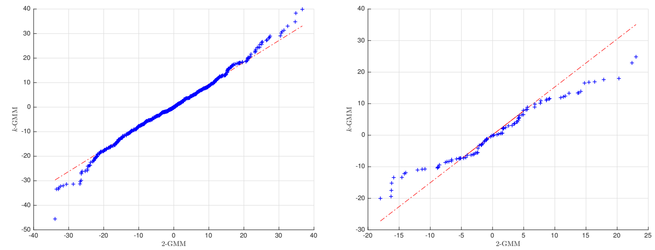

Fig. 1. Illustration of the 2-GMM approximation: quantile-quantile plot for k-GMM against 2-GMM data samples for $\tau = 1$ (left) and $\tau = 5$ (right).

Algorithm 1 Sparsity estimation via 2-GMM approximation

Require: Measurements $y \in \mathbb{R}^n$, parameters γ, k

Initialization: $\alpha(0), \beta(0), p(0)$

for $t = 0, 1, \dots, StopIter$ **do**

E-step: compute the posterior probabilities

$$\pi_i(t+1) = \frac{\frac{p(t)}{\sqrt{\alpha(t)}} \mathrm{e}^{-y_i^2/(2\alpha(t))}}{\frac{p(t)}{\sqrt{\alpha(t)}} \mathrm{e}^{-y_i^2/(2\alpha(t))} + \frac{1-p(t)}{\sqrt{\beta(t)}} \mathrm{e}^{-y_i^2/(2\beta(t))}}$$

M-Step: compute the mixture parameters

$$p(t+1) = \frac{\sum_{i=1}^{m} \pi_i(t+1)}{m}, \; k(t+1) = \frac{\log\left(\frac{\sum_{i=1}^{m} \pi_i(t+1)}{m}\right)}{\log(1-\gamma)}$$

$$\alpha(t+1) = \frac{\sum_{i=1}^{m} \pi_i y_i^2}{\sum_{i=1}^{m} \pi_i}, \; \beta(t+1) = \frac{\sum_{i=1}^{m} (1-\pi_i) y_i^2}{\sum_{i=1}^{m} (1-\pi_i)}.$$

end for

The following convergence theorem can be proved.

Theorem 2. *The sequence of signal sparsity estimations $k(t)$ generated by Algorithm 1 converges to a limit point.*

For brevity, we omit the proof, which can be readily derived from standard convergence arguments for dynamical systems [22].

4. SIMULATIONS

In this section we test Algorithm 1 in different settings and compare it to the algorithm proposed by Lopes in [10], and to the sparsity estimator in the noise-free case proposed in [13], referred to as oracle estimator.

In the first experiment, we set $k = 200$, $\frac{k}{m} \in [0.1, 1]$, $\sigma = 1$. We choose the nonzero values of x uniformly at ran-

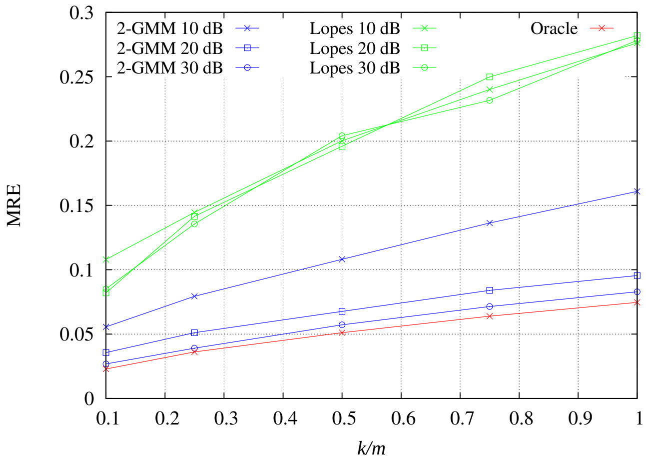

Fig. 2. Mean relative error for $k = 200$, $x_i \in \{0, \pm\lambda\}$

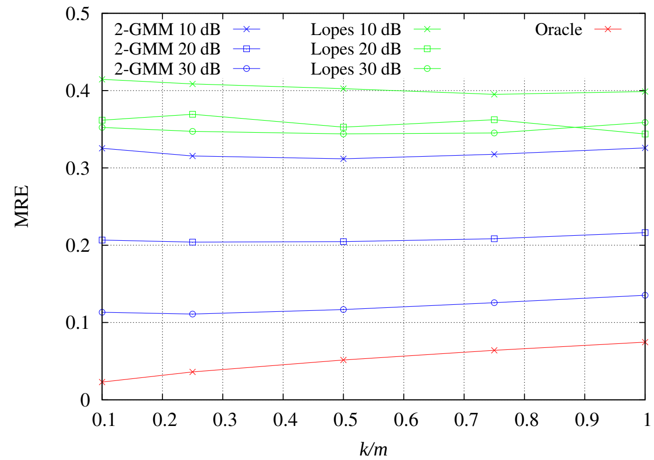

Fig. 3. Mean relative error for $k = 200$, x Gaussian

dom in $\{-\lambda, \lambda\}$. The energy of x is then $\lambda^2 k$, which is equal to the mean energy of y with sensing matrix as in (2). We then define the mean signal-to-noise ratio as SNR$= \lambda^2 k/\sigma^2$, and we set $\gamma = \frac{\text{SNR(dB)}}{5 \cdot 10^3}$, where SNR(dB) indicates the SNR expressed in dB. This choice is based on empirical evidence suggesting a more stable behavior when γ depends on the SNR.

We evaluate the performance in terms of mean relative error, defined as: MRE $= |\widehat{k} - k|/k$, where $\widehat{k}$ is the estimated sparsity. We start 2-GMM by an M-step on $\pi_i(0) = \frac{1}{2}$ for any $i = 1, \ldots, m$. The procedure is stopped at the first time t_0 such that $|k(t_0) - k(t_0 - 1)| < 10^{-2}$, and $\widehat{k}$ taken as the closest integer to $k(t_0)$ (in all our simulations, a few tens of iterations were sufficient to get convergence). In Figure 2, we plot the average results over 500 different runs. The performance curves, obtained with SNRs from 10 to 30 dB, show that 2-GMM performs better than [10], and, as a difference from it, gets closer to the oracle estimator as the SNR increases.

We remark that the signal-to-noise ratio for Lopes's method in principle cannot be evaluated, while in practice it is always very high. This is due to the use of a sensing matrix which is partially generated according to a Cauchy distribution, for which variance is infinite. Such observation explains the behavior of Lopes's performance, which is substantially invariant to the SNR. On the one hand, this means more robustness to noise, on the other hand, it causes a remarkable distance from the oracle even for high SNRs.

In Figure 3, in order to evaluate the algorithm robustness to different signal distributions, we repeat the first experiment using Gaussian signals, that is, the nonzero x_is are generated according to a zero-mean Gaussian distribution, and normalized so that their energy is still equal to $\lambda^2 k$. In this case, we set $\gamma = \frac{\text{SNR(dB)}}{2 \cdot 10^4}$. In this Gaussian framework, accuracy is lower as x might have some small entries easily confused with noise. However, also in this experiment we can appreciate a gain with respect to Lopes's algorithm and a natural improvement as the SNR increases. Hence, numerical results suggest that the proposed method can be employed also for signals not in $\{0, \pm\lambda\}$.

We underline that the performance gain with respect to Lopes's method is obtained with no substantial increase of complexity: Lopes's algorithm consists in the computation of a mean and a median over the m-length measurement vector; each EM step in 2-GMM requires a number of computations of order m, and, as already said, we observed that few tens of iterations were sufficient to converge. In conclusion, the order of complexity for Lopes algorithm and 2-GMM is comparable.

Regarding the choice of γ, we remark that in many practical situations the SNR of the system is approximately known, hence it makes sense to design γ based on it. On the other hand, γ should be rescaled with k when k increases, to keep τ finite (see Theorem 1). This seems contradictory as the estimate of k is our ultimate goal; however, one can solve the issue by rescaling γ with respect to the length n, which in most realistic settings is expected to scale with k.

5. CONCLUSIONS

This paper has proposed an iterative algorithm for estimation of the signal sparsity starting from compressive and noisy projections via γ-sparsified random matrices. This iterative procedure is obtained by modeling the projections using an approximate 2-GMM. The main theoretical contribution includes the precise characterization of 2-GMM for a specific choice of the parameter γ. Numerical results confirm that 2-GMM outperforms methods known in the literature, with no substantial increase of complexity, and works for different signals models. Future work will be devoted to better understand the relationship between γ and other parameters in the system, like the SNR, and to derive theoretical guarantees for more general signal models.

6. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. Candes and T. Tao, "Near optimal signal recovery from projection: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] F. Zeng, L. Chen, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 37–48, 2011.

[4] L. Gan, "Block compressed sensing of natural images," in *IEEE Int. Conf. DSP*, 2007, pp. 403–406.

[5] E. Arias-Castro, E. J. Candes, and M. A. Davenport, "On the fundamental limits of adaptive sensing," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 472–481, 2013.

[6] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, 2007.

[7] A. Mousavi, A. Maleki, and R. G. Baraniuk, "Asymptotic analysis of LASSOs solution path with implications for approximate message passing," in *arXiv preprint*, 2013, p. arXiv:1309.5979.

[8] Joel A. Tropp, Anna, and C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4655–4666, 2007.

[9] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmonic Anal.*, vol. 26, no. 3, pp. 301–321, Apr. 2008.

[10] M. E. Lopes, "Estimating unknown sparsity in compressed sensing," in *Proc. 30th Int. Conf. Machine Learning*, Atlanta, Georgia, USA, 2013, vol. 28.

[11] R. Ward, "Compressed sensing with cross validation," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, Dec 2009.

[12] Y. C. Eldar, "Generalized sure for exponential families: Applications to regularization.," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, 2009.

[13] V. Bioglio, T. Bianchi, and E. Magli, "On the fly estimation of the sparsity degree in compressed sensing using sparse sensing matrices," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2015, pp. 3801–3805.

[14] Miles E. Lopes, "Compressed sensing without sparsity assumptions," *CoRR*, vol. abs/1507.07094, 2015.

[15] S. Lan, Q. Zhang, X. Zhang, and Z. Guo, "Sparsity estimation in image compressive sensing," in *IEEE Int. Symp. on Circuits and Systems (ISCAS)*, 2012.

[16] D.M. Malioutov, S. Sanghavi, and A.S. Willsky, "Compressed sensing with sequential observations," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2008, pp. 3357–3360.

[17] M.S. Asif and J. Romberg, "Streaming measurements in compressive sensing: ℓ_1 filtering," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, Oct 2008, pp. 1051–1058.

[18] P.J. Garrigues and L. El Ghaoui, "An homotopy algorithm for the Lasso with online observations," in *Neural Information Processing Systems (NIPS)*, 2008, vol. 21.

[19] Y. You, J. Jin, W. Duan, N. Liu, Y. Gu, and J. Yang, "Zero-point attracting projection algorithm for sequential compressive sensing," *IEICE Electronics Express*, vol. 9, no. 4, pp. 314–319, 2012.

[20] W. Wang, M. J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2967–2979, 2010.

[21] D. Omidiran and M. J. Wainwright, "High-dimensional variable selection with sparse random projections: Measurement sparsity and statistical efficiency," *JMLR*, vol. 11, pp. 2361–2386, 2010.

[22] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.