ON THE FLY ESTIMATION OF THE SPARSITY DEGREE IN COMPRESSED SENSING USING SPARSE SENSING MATRICES

Valerio Bioglio, Tiziano Bianchi, Enrico Magli

Department of Electronics and Telecommunications, Politecnico di Torino

ABSTRACT

In this paper, we propose a mathematical model to estimate the sparsity degree k of exactly k-sparse signals acquired through Compressed Sensing (CS). Our method does not need to recover the signal to estimate its sparsity, and is based on the use of sparse sensing matrices. We exploit this model to propose a CS acquisition system where the number of measurements is calculated on-the-fly depending on the estimated signal sparsity. Experimental results on block-based CS acquisition of black and white images show that the proposed adaptive technique outperforms classical CS acquisition methods where the number of measurements is set a priori.

***Index Terms*—** Compressed Sensing, Sparsity Estimation, Sparse Sensing Matrices, Adaptive Sensing

1. INTRODUCTION

Compressed Sensing (CS) [1, 2], a novel signal acquisition technique that is emerging in the recent years, allows one to greatly reduce the number of measurement needed to acquire a signal. If a signal having dimension n is known to be sparse or compressible, instead of taking n samples of the signal, CS suggests to take only $m \ll n$ linear combinations of the signal entries. State-of-the-art CS systems are able to recover k-sparse signals from only $m = O(k \log(n/k))$ linear measurements [3]. Many recovery algorithms are based on linear programming techniques, such as Lasso [4] or Basis Pursuit (BP) [5]; moreover, faster greedy algorithms, such as Orthogonal Matching Pursuit (OMP) [6] or Compressive Sampling Matching Pursuit (CoSaMP) [7], can be exploited and usually provide comparable empirical performance.

Since the number of linear measurements required for the recovery depends on the sparsity degree of the signal, the knowledge of k is crucial for a CS system. If m is too small, the recovery algorithms do not guarantee the signal reconstruction; if it is too big, the acquisition turns out to be redundant. Moreover, k is needed as input to greedy recovery algorithms, like CoSaMP or OMP, where the number of iterations is bounded by k. Also, Lasso techniques require to tune a parameter λ, for which the knowledge of k is helpful [8].

In this paper, we propose a method to estimate the sparsity degree of exactly k-sparse signals directly from their linear measurements. The method exploits the fact that, in the case of a k-sparse signal, sparse sensing matrices produce sparse measurements whose sparsity degree depends on k. The use of sparse sensing matrices is feasible in CS systems; at the price of a small increase of the number of measurements needed for the recovery, a complexity reduction of the sensing process is obtained [9, 10]. The proposed method can be used to estimate the signal sparsity during acquisition, adaptively choosing the number of measurements according to the estimated k. Experimental results on block-based CS acquisition of images [11] confirm the validity of our model, showing that the proposed adaptive acquisition strategy outperforms CS systems in which the number of measurements is set a priori.

2. MOTIVATIONS AND BACKGROUND

A signal $s \in \mathbb{R}^n$ is called k-sparse if there exists a basis $\Phi \in \mathbb{R}^{n\times n}$ such that $s = \Phi x$ and x has only k nonzero entries. The matrix $A \in \mathbb{R}^{m\times n}$ is called sensing matrix and $y \in \mathbb{R}^m$, obtained through the multiplication $A \cdot x$, is called measurement vector. The (i, j) element of A is denoted as a_{ij}, while the i-th entry of vector x is denoted as x_i. The measurements vector can also be obtained by the original signal s through the multiplication $y = \Psi s$, where $\Psi = A\Phi^{-1}$ is the compression matrix. According to CS [1], k-sparse signal can be recovered even from $m \ll n$ measurements by solving the minimization problem

$$\widehat{x} = \arg\min_x ||x||_0 \quad \text{s.t.} \quad Ax = y. \tag{1}$$

A necessary condition for the uniqueness of the solution of (1) is that $m > 2k$. Moreover, this NP-complete ℓ_0-problem can be rewritten as a more tractable ℓ_1-problem, in which case at least $m = O(k \log(n/k))$ measurements are needed [3]. As a consequence, the knowledge of the sparsity of the signal is essential to calculate the number of measurements required to guarantee a high recovery probability. However, this sparsity is generally unknown.

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.

In the CS literature, many papers propose various techniques for the estimation of the positions of the nonzero entries of the signal [12–14]. In the sparsity estimation problem, instead, one wants to guess only the sparsity degree k of the signal. In practice, sparsity estimation can be seen as a subproblem of support estimation. In [15], 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. However, the numerical sparsity is only a lower bound for the sparsity: an upper bound would be more useful for choosing the right m. In [16], the authors propose to estimate the sparsity of an image before its acquisition, by calculating the image complexity. Even if this technique gives good results, the proposed metric is based on the image pixel values, forcing to calculate a separate estimation that does not depend on the measurements. Some papers propose sequential acquisition techniques, in which the number of measurements is dinamically adapted until a satisfactory reconstruction performance is achieved [17–20]. However, even if the reconstruction can take into account the previously recovered signal, these methods require to solve a new problem at each new acquired measurement. It is evident that a method for estimating the sparsity directly from the measurements will dramatically decrease the complexity of sequential acquisition.

Random Gaussian (RG) sensing matrices are usually exploited in CS systems. Even if RG sensing matrices achieve an optimal scaling of the number of measurements required for the recovery [21], with these matrices the measurements are uniformly distributed on the hypersphere of radius given by the energy of the signal [22], meaning that the only way to estimate k exploiting RG sensing matrices is through the recovery of the signal. In order to estimate k without recovering x, a different kind of sensing matrix is needed. Moreover, we need a family of sensing matrices that combine good CS recovery and good sparsity estimation properties. In this paper, we propose to exploit Sparse Random Gaussian (SRG) sensing matrices. Each entry a_{ij} of a SRG sensing matrix is set to zero with probability $1-\gamma$, where γ is a tunable parameter, while with probability γ the entry will be nonzero. In the latter case, the value is drawn over $\mathbb{R}$ according to a Gaussian distribution of mean zero and variance $\frac{1}{m\gamma}$. To summarize,

$$a_{ij} \sim \begin{cases} 0 & \text{with probability } 1-\gamma \\ \mathcal{N}\left(0, \frac{1}{\gamma m}\right) & \text{with probability } \gamma. \end{cases} \tag{2}$$

A trade-off between costs and efficiency of SRG matrices is presented in [9], which derives lower bounds on the number of measurements required for the recovery as a function of the sparsity γ of the SRG matrix. Similar results are achieved in [10], showing that SRG matrices lead to a small increase of the number of measurements needed for the recovery.

3. SPARSITY ESTIMATION

We propose to estimate the signal sparsity k exploiting the number of nonzero elements of the measurement vector y. In order to do that, we begin by calculating the probability that an entry of y is equal to zero. We call d_i the number of entries for which a_{ij} and x_j are both nonzero for $1 \leq j \leq n$, *i.e.* the number of collisions between the i-th row of A and the signal x. The probability that a measurement is equal to zero is

$$\mathbb{P}(y_i = 0) = \sum_{j=0}^{k} \mathbb{P}(y_i = 0|d_i = j)\mathbb{P}(d_i = j) = \mathbb{P}(d_i = 0),$$

since $\mathbb{P}(y_i = 0|d_i > 0) = 0$ and $\mathbb{P}(y_i = 0|d_i = 0) = 1$ in $\mathbb{R}$. If we call $P = \mathbb{P}(d_i = 0)$, the value of P depends on the sparsity k, hence it is possible to write $P = P(k)$. Since the rows of A are i.i.d., the ℓ_0 norm of the measurements vector follows a binomial distribution, *i.e.*, letting h be the number of nonzero elements of y, $h = ||y||_0$, then $\mathbb{P}(h \,|\, k) \sim \mathcal{B}(1-P, m)$. In fact, h is a random variable counting the number of nonzero entries of the measurements vector y. Due to the nature of A, each measurement is independent of the others, and it is equal to zero with probability P. As a consequence, h is the result of m Bernoulli processes, hence it is distributed according to the presented binomial distribution.
This can be exploited to estimate k from the knowledge of h. Indeed, it is possible to estimate the proportion parameter $1-P$ of the distribution of $\mathbb{P}(h \,|\, k)$. Since this parameter depends on k, we can use its knowledge to estimate the sparsity of the signal. In practice, since A is generated according to (2), d_i is distributed according to a binomial distribution, and more precisely $\mathbb{P}(d_i) \sim \mathcal{B}(k, \gamma)$. This follows from the fact that d_i is a random variable that counts the number of collisions between the i-th row and the signal. The signal has k nonzero entries, hence d_i depends on the number of nonzero entries of the i-th row of A in these k positions. Since each entry of A is independent of the others and assumes a nonzero value with probability γ, d_i is the result of k Bernoulli processes, and it is distributed according to the presented binomial distribution. As a consequence, we can calculate P as

$$P = P(k) = \mathbb{P}(d_i = 0) = \binom{k}{0}\gamma^0(1-\gamma)^k = (1-\gamma)^k.$$

Transposing this equation, the value of k can be calculated as

$$k = \log_{1-\gamma}(P) = \frac{\log(P)}{\log(1-\gamma)}. \tag{3}$$

We divide the study of good estimators of k in two cases: with and without a priori information on the distribution of k. In the former case, a maximum a posteriori probability (MAP) estimator can be obtained, whereas in the latter case, a maximum-likelihood (ML) estimator can be calculated.

3.1. ML Estimator

The measurements sparsity h takes on values according to a binomial distribution; the ML estimator for the parameter $1-P$ of its binomial distribution is given by

$$1-\hat{P}=\frac{||y||_0}{m} \quad \Rightarrow \quad \hat{P}=1-\frac{h}{m}. \tag{4}$$

Collecting Eqs. (3) and (4), a ML estimator for k can be calculated; however, k is an integer number, while P is a real number. As a consequence, the ML estimator for k would result in a real value. In order to get an integer value simple to calculate, we pick the closest integer to the real ML estimator, even if the solution found is suboptimal. Finally, the estimator is given by

$$\hat{k}_{ML}=\left\lfloor\frac{1}{\log(1-\gamma)}\log\left(1-\frac{h}{m}\right)\right\rceil, \tag{5}$$

where $\lfloor\cdot\rceil$ is the closest integer function. To calculate an upper bound for k the binomial proportion confidence interval for P [23] is used, *i.e*, with probability larger than β,

$$P\geq\hat{P}-\sqrt{\frac{\hat{P}(1-\hat{P})}{m}}z_\beta,$$

where z_β is the β percentile of a standard normal distribution; as an example, $z_{0.99}=2.57$. As a consequence, the upper bound for k with probability β is given by

$$k\leq\tilde{k}^\beta_{ML}=\frac{1}{\log(1-\gamma)}\cdot\log\left(1-\frac{h}{m}-\sqrt{\frac{\frac{h}{m}(1-\frac{h}{m})}{m}}z_\beta\right).$$

In this case, $\tilde{k}^\beta_{ML}$ can be a real value, since it is only an upper bound for an integer value.

3.2. MAP Estimator

If, in addition to $\mathbb{P}(h\,|\,k)$, also the sparsity distribution $\mathbb{P}(k)$ is known, we can exploit this to find a MAP estimator for k. In this case, however, it is not possible to exploit the properties of the binomial distribution to get a closed formula for the estimator, since it also depends on the distribution $\mathbb{P}(k)$. We suppose k to be bounded, *i.e.* $k_{min}\leq k\leq k_{max}$, obtaining the MAP estimator

$$\begin{aligned}\hat{k}_{MAP} &= \operatorname*{arg\,max}_{k_{min}\leq k\leq k_{max}} \mathbb{P}(k\,|\,h)= \\ &= \operatorname*{arg\,max}_{k_{min}\leq k\leq k_{max}}\left(\frac{\binom{m}{h}(1-P(k))^h(P(k))^{m-h}\cdot\mathbb{P}(k)}{\mathbb{P}(h)}\right)\end{aligned} \tag{6}$$

where $P(k)=P$ since k is the only free parameter that generated P. Moreover, an upper bound for this estimate with probability larger than β can be found as

$$k\leq\tilde{k}^\beta_{MAP}=w_\beta,$$

where w_β is the β percentile of $\mathbb{P}(k\,|\,h)$. This probability distribution can be calculated when the distribution of k is known, hence it is simple to evaluate its percentiles. However, in order to have a MAP estimate of k, we need to assume an a priori distribution for k. In the following, we will exploit two models for the distribution of k.

Uniform sparsity: k is uniformly distributed between k_{min} and k_{max} with probability $(k_{max}-k_{min}+1)^{-1}$.

Binomial sparsity: $k=k_{min}+B$, where B is distributed according to the Binomial distribution $B\sim\mathcal{B}(k_{max}-k_{min},\alpha)$, with α a tunable parameter; $k_{min}\leq k\leq k_{max}$.

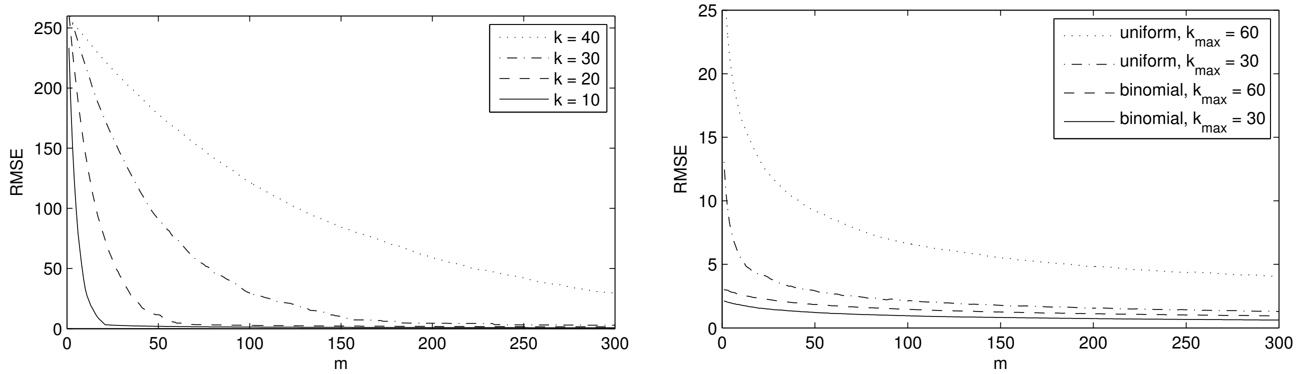
4. EXPERIMENTAL RESULTS

Here we describe the results of our experiments on the sparsity estimation, proving that a good approximation of k can be calculated through SRG matrices without any additional cost. In all the experiments, the sensing matrices are SRG created according to (2) with $\gamma=0.1$. Each point of the figures is obtained through a 10000 runs simulation. The signal is randomly drawn and has sparsity exactly k, *i.e.*, x has k nonzero entries drawn according to a Gaussian distribution $\mathcal{N}(0,1)$. First, to evaluate the correctness of the proposed estimators on the number of measurements collected, we calculate the root square mean error (RMSE) of an estimator $\hat{k}$ as

$$RMSE(\hat{k})=\sqrt{\frac{\sum_{i=1}^r(\hat{k}-k)^2}{r}},$$

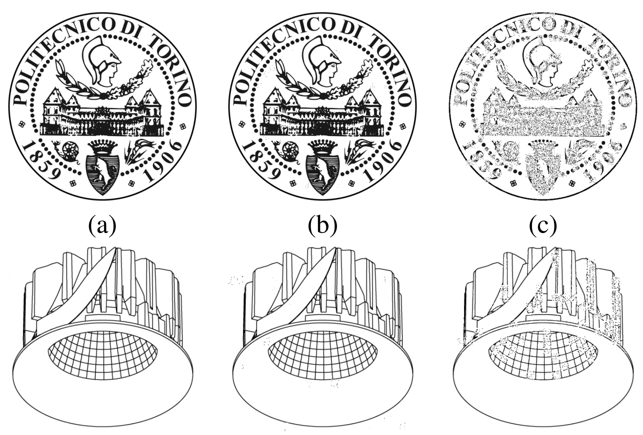
where r is the number of runs of the experiment. The RMSE is connected to the confidence interval calculated in the previous section, since $RMSE(\hat{k})\leq 2(\tilde{k}^\beta-\hat{k})$ with probability β. Moreover, it is possible to prove that $\tilde{k}^\beta-\hat{k}=O(1/\sqrt{m})$, i.e. the goodness of the estimation improves as the number of measurements increases. In Fig. 1-(a), the RMSE of the ML estimator (5) is presented as a function of the number of measurements m. Signal sparsity k is set to be constant, since the signal does not follow any known distribution. We can see that the accuracy of the estimator depends on both m and k: the more measurements we get, the more accurate the estimator is. Moreover, the estimation is better if the signal is sparser, since in this case the binomial confidence interval becomes tighter. In fact, for $k=40$ the ratio k/n is too high, and the confidence interval turns out to be too wide to ensure a fast convergence. The same discussion is valid for the MAP estimator (6) presented in Fig. 1-(b). In this case, the a priori distributions of k presented in the previous section are exploited: each curve represents a different value of k_{max}, that is the maximum value k can get, while k_{min} is always set to 1. Comparing the two plots of Fig. 1, we see that the estimation improves if the distribution of k is known, as expected.

In order to demonstrate the practical usefulness of the proposed sparsity estimation technique, we simulate a block-based [11,24,25] CS acquisition system that adjusts the num-



(a) (b)

Fig. 1. RMSE of different estimators vs. m for $n = 300, \gamma = 0.1$: (a) $\hat{k}_{ML}$; (b) $\hat{k}_{MAP}$, $\alpha = 0.2$ for binomial sparsity



(a) (b) (c)

(d) (e) (f)

Fig. 2. Recovery of sparse images under different acquisition settings: original images, (a) 512×512 pixels, (d) 352×352 pixels; proposed adaptive sensing strategy, (b) $m = 190012$, PSNR = 37.1, (e) $m = 76044$, PSNR = 34.6; fixed number of measurements per block, (c) $m = 190464$, PSNR = 9.6, (f) $m = 76472$, PSNR = 17.4.

ber of measurement for each block on-the-fly depending on the estimated signal sparsity. We assume that the measurements are taken one by one, and the sensor has to decide when to stop acquiring measurements. CoSaMP needs at least $m = 3k$ measurements in order to recover the correct signal [7]; in order to guarantee a sufficiently high probability of correct recovery, we propose that the sensor stops the acquisition when $m \geq 3\tilde{k}^{\beta}$, which ensures that the minimum requirements of CoSaMP are met with a probability equal to β. Namely, at each step the sensor acquires one measurement and calculates the upper bound $\tilde{k}^{\beta}$; if $m \geq 3\tilde{k}^{\beta}$, the sensor assumes that the number of measurements sensed is sufficient to guarantee the recovery of the signal, otherwise another step begins. The proposed system is used to sense two black and white images, with size 512×512 pixels and 352×352 pixels, respectively. Each 16×16 block is independently acquired using an adaptive number of measurements, obtained according to the strategy described above, and then recovered using CoSaMP. The block sparsity is estimated through the ML estimator and the corresponding upper bound satisfies $\beta = 0.999$. We also simulate a system in which each 16×16 block is independently acquired using a fixed number of measurements obtained through a dense sensing matrix. In order to have a fair comparison, in this second case we set m as close as possible to the average number of measurements acquired in the first experiment. The recovered images are shown in Fig. 2. It is evident that estimating the sparsity during acquisition enables significant performance gains: using the same number of measurements, the proposed system is able to achieve a 20-25 dB increase in the peak signal to noise ratio (PSNR) of the recovered images. Further experiments showed that, in order to achieve the same quality of the recovered image, the nonadaptive system should collect a number of measurements close to the image size. Hence, the ability to estimate on the fly the required number of measurements proves very useful when applying CS to signals with unknown and/or varying degrees of sparsity.

5. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a method to estimate the sparsity of a signal in a CS acquisition system directly from its measurements. The method is based on the properties of sparse sensing matrices and on the hypothesis of exactly sparse signals. The results show that a realistic upper bound on the signal sparsity can be obtained directly from the measurements, by applying simple estimators. Moreover, we show that the knowledge of the sparsity can be exploited to adjust the number of measurements to acquire, outperforming acquisition methods where the number of measurements is set a priori, especially when the signal sparsity varies widely.

The method is currently limited to exactly sparse signals, which are relevant in many important practical settings, such as the acquisition of images with few gray levels or the detection of anomalies in network traffic [26]. Moreover, we are currently working on the extension of the proposed method in the case of signals that are not exactly sparse.

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] 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.

[4] 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.

[5] S. Chen, D. L. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[6] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[7] D. Needell and J. A. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[8] 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.

[9] 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.

[10] 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.

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

[12] J. Haupt and R. Baraniuk, “Robust support recovery using sparse compressive sensing matrices,” in *Information Sciences and Systems (CISS), 45th Annual Conference on. IEEE*, 2011.

[13] J. Scarlett, J. S. Evans, and S. Dey, “Compressed sensing with prior information: Information-theoretic limits and practical decoders,” *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 427–439, 2013.

[14] A. Gilbert and P. Indyk, “Sparse recovery using sparse matrices,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 937–947, 2010.

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

[16] 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.

[17] D.M. Malioutov, S. Sanghavi, and A.S. Willsky, “Compressed sensing with sequential observations,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, March 2008, pp. 3357–3360.

[18] 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.

[19] 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.

[20] Yang You, Jian Jin, Wei Duan, Ningning Liu, Yuantao Gu, and Jian Yang, “Zero-point attracting projection algorithm for sequential compressive sensing,” *IEICE Electronics Express*, vol. 9, no. 4, pp. 314–319, 2012.

[21] M. J. Wainwright, “Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting,” *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5728–5741, 2009.

[22] T. Bianchi, V. Bioglio, and E. Magli, “On the security of random linear measurements,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2014, pp. 3992–3996.

[23] J. T. Morisette and S. Khorram, “Exact binomial confidence interval for proportions,” *Photogramm. Eng. and Remote Sensing*, vol. 64, no. 4, pp. 281–282, 1998.

[24] S. Mun and J. E. Fowler, “Block compressed sensing of images using directional transforms,” in *IEEE Int. Conf. Image Proc. (ICIP),*, 2009, pp. 3021–3024.

[25] G. Coluccia, D. Valsesia, and E. Magli, "Smoothness-constrained image recovery from block-based random projections," in *Multimedia Signal Processing (MMSP), 15th International Workshop on*, 2013, pp. 129–134.

[26] M. Mardani, G. Mateos, and G.B. Giannakis, "Dynamic anomalography: Tracking network anomalies via sparsity and low rank," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 1, pp. 50–66, Feb 2013.