On the Performance of Adaptive Sensing for Sparse Signal Inference

Rui M. Castro
Eindhoven University of Technology
The Netherlands
Email: rmcastro@tue.nl

***Abstract*—In this short paper we survey recent results characterizing the fundamental draws and limitations of adaptive sensing for sparse signal inference. We consider two different adaptive sensing paradigms, based either on single-entry or linear measurements. Signal magnitude requirements for reliable inference are shown for two different inference goals, namely signal detection and signal support estimation.**

I. Introduction

In this short paper we survey recent results characterizing the fundamental draws and limitations of adaptive sensing. One of the key aspects of adaptive sensing is that the data collection process is sequential and adaptive. In different fields these sensing/experimenting paradigms are known by different names, such as *sequential experimental design* in statistics and economics (see [1], [2], [3], [4], [5]), *active learning* or *adaptive sensing/sampling* in computer science, engineering and machine learning (see [6], [7], [8], [9], [10], [11], [12], [13], [14]). An essential aspect of adaptive sensing is the intricate coupling between data analysis and acquisition, which creates a powerful feedback structure. This is a double-edged sword: it is key to harness the power of sequential experimental design but also raises challenges in the analysis of such methodologies – indeed it creates complicated and strong dependencies in the data sequence.

We consider a model where the signal of interest is represented by a sparse vector $\boldsymbol{x} \in \mathbb{R}^n$, meaning that most entries of $\boldsymbol{x}$ are zero and only few of the entries are nonzero. Specifically let S be a subset of $\{1, \ldots, n\}$ of nonzero entries of $\boldsymbol{x}$, and assume that for all $i \in \{1, \ldots, n\}$ such that $i \notin S$ we have $x_i = 0$. We refer to S as the signal support set and this is our main object of interest. We consider two distinct classes of problems: (i) *signal detection*, where we want to test if S belongs to a particular class of subsets of $\{1, \ldots, n\}$, and (ii) *support estimation*, where we desire to actually estimate S. The signal $\boldsymbol{x}$ in naturally assumed to be unknown, but we can collect partial information about it through noisy measurements. In particular we consider generalizations of the normal means model allowing for multiple and sequential measurements, therefore enabling adaptive sensing strategies. Our focus is primarily on single-entry observations, but in Section III we discuss also a different (and statistically more powerful) sensing model which allows for linear measurements of the signal - in what is often referred to as Compressive Sensing (CS).

II. Single-entry Measurements

This sensing model was first proposed in [15]. Measurements are of the form

$$Y_k = x_{A_k} + \Gamma_k^{-1/2} W_k \ , \qquad k = 1, 2, \ldots \ ,$$

where A_k, Γ_k are taken to be functions of $\{Y_i, A_i, \Gamma_i\}_{i=1}^{k-1}$, and W_k are standard normal random variables, independent of $\{Y_i\}_{i=1}^{k-1}$ and also independent of $\{A_i, \Gamma_i\}_{i=1}^{k}$. In words, each measurement corresponds to a single signal entry corrupted with additive Gaussian noise, and the choice of entry and noise level can be controlled. However, there is a total sensing budget constraint that must be satisfied, namely

$$\sum_{k=1}^{\infty} \Gamma_k \leq m \ , \tag{1}$$

where $m > 0$. In the above model A_k should be viewed as the *sensing action* taken at time k, and Γ_k is the *precision* of the corresponding measurement. We have control over both quantities. Informally stated, measurements are collected sequentially, and for each measurement we can choose which entry of $\boldsymbol{x}$ to observe, and what is the precision (i.e. accuracy) of the measurement. We are allowed to collect as many measurements as desired provided the cumulative precision used satisfies the budget (1). Note that in this model we are allowed to collect an infinite (but countable) number of measurements, provided the precision Γ_k converges to zero as k grows. Although this might seem strange at first, it is not entirely unreasonable in practice - in many sensing modalities the precision is directly proportional to the amount of time necessary to collect a measurement, and therefore (1) can be viewed simply as a time constraint. This is the case in various imaging modalities (e.g. in astronomy) where long exposure times are used to reduce the noise level, which is inversely proportional to the exposure time. It is important to note that there are also settings where the actual number of measurements is limited, and there is little control on the precision level. In that case (1) might represent a constrain on the total number of measurements, provided Γ_k is not a function of k. The results in the latter setting are similar to the ones presented in the current paper, especially when studying asymptotics (when both n and m grow).

It is important to note that we can consider both deterministic sequential designs or random sequential designs. In the latter we allow the choices A_k and Γ_k to incorporate extraneous randomness, which is not explicitly described in the model. The collection of conditional distributions of A_k, Γ_k given $\{Y_i, A_i, \Gamma_i\}_{i=1}^{k-1}$ for all k is referred to as the *sensing strategy*. Note that, within the sensing paradigm above we can also consider non-adaptive sensing, meaning that the choice of sensing actions and corresponding precision is made before collecting any data. Formally this means that $\{A_k, \Gamma_k\}_{k\in\mathbb{N}}$ is statistically independent from $\{Y_k\}_{k\in\mathbb{N}}$. Note that a non-adaptive design can still be random.

The case $m = n$ is of particular interest, allowing a direct comparison between adaptive and non-adaptive sensing methodologies. When $m = n$ we allow on average one unit of precision per each of the signal entries. So, if there is no reason to give preference to any particular entry of $\boldsymbol{x}$, the natural optimal non-adaptive sensing strategy should simply measure each entry of $\boldsymbol{x}$ exactly once, with precision one. This corresponds to the well studied normal means model.

For simplicity of presentation we consider only signals of the form

$$x_i = \begin{cases} \mu & \text{if } i \in S \\ 0 & \text{if } i \notin S \end{cases} ,$$

where $\mu > 0$ is called the *signal amplitude*. This restriction is also considered in [16], [17] in the non-adaptive sensing context and does not substantially hinder the generality of the results presented in this manuscript.

As stated before we consider two different inference problems: (i) signal detection and (ii) support estimation. For the detection problem (i) the goal is to determine if a signal is present or absent. We formulate the problem as a binary hypothesis testing, and test a simple null hypothesis against a composite alternative. In particular the null hypothesis H_0 is simply $S = \emptyset$, and the alternative hypothesis H_1 is $S \in \mathcal{C}$, where $\mathcal{C}$ is some class of non-empty subsets of $\{1, \ldots, n\}$. For simplicity of presentation we assume that all the sets in $\mathcal{C}$ have the same cardinality s. A test procedure based on the (adaptive) measurements is described by a binary test function $\hat{\phi}\left(\{A_i, \Gamma_i, Y_i\}_{i=1}^{\infty}\right) \in \{0, 1\}$, and a natural way to measure the performance of such a test function is the *worst case* risk

$$R(\hat{\phi}) = \mathbb{P}_{\emptyset}(\hat{\phi} \neq 0) + \max_{S\in\mathcal{C}} \mathbb{P}_S(\hat{\phi} \neq 1) ,$$

where $\mathbb{P}_S$ denotes the joint probability distribution of $\{A_i, \Gamma_i, Y_i\}_{i=1}^{\infty}$ for a given support set S. Characterizing the relation between $R(\hat{\phi})$, n, m, μ, and $\mathcal{C}$ is our main objective.

The goal of the estimation problem (ii) is (statistically) more ambitious, as we seek to actually identify the support set S. An estimation procedure is a function $\hat{S}$ mapping $\{A_i, \Gamma_i, Y_i\}_{i=1}^{\infty}$ to a subset of $\{1, \ldots, n\}$. There are several sensible ways to measure "closeness" between $\hat{S}$ and the true support set S, for instance the worst case probability of making any errors

$$\max_{S\in\mathcal{C}} \mathbb{P}_S[\hat{S} \neq S] .$$

A somewhat more stringent metric is the worst case expected number of errors $\max_{S\in\mathcal{C}} \mathbb{E}_S[|\hat{S}\Delta S|]$, and clearly $\mathbb{P}_S[\hat{S} \neq S] \leq \mathbb{E}_S[|\hat{S}\Delta S|]$. We will focus mainly on the first metric in this manuscript, but remark that the two metrics are essentially equivalent in several cases.

A. Single-entry Measurements: Results

In this section we present the fundamental tradeoffs for the inference problems presented above. Clearly these results bear some dependency on the class of sets $\mathcal{C}$:

Definition II.1 (symmetric class)**.** *Let S be a random set, drawn uniformly at random from $\mathcal{C}$. If for all $i \in \{1, \ldots, n\}$ we have $\mathbb{P}(i \in S) = s/n$ the class $\mathcal{C}$ is said to be symmetric.*

In words, in a symmetric class of sets there is no reason to give a priori preference to any individual entry. Many classes $\mathcal{C}$ of interest satisfy this mild symmetry, for instance all the classes in [16]. Of particular interest is the maximal class of all the subsets of $\{1, \ldots, n\}$ with cardinality s, which corresponds to *lack of structure* in the sparsity pattern S. If the class $\mathcal{C}$ is smaller then we say the sparsity patterns S have *structure*. An example of a structured class is presented later.

Theorem II.1 ([18])**.** *Let $\mathcal{C}$ be a symmetric class, and let $\hat{\Phi}$ be an arbitrary adaptive sensing testing procedure. For any $0 < \epsilon < 1$, if $R(\hat{\Phi}) \leq \epsilon$ then necessarily*

$$\mu \geq \sqrt{\frac{2n}{sm} \log \frac{1}{2\epsilon}} .$$

As argued before, the case $m = n$ is of particular interest, as it allows for comparison between adaptive and non-adaptive sensing performance: in that case the above bound is of the order $\sqrt{2/s}$. It is remarkable that the extrinsic signal dimension n plays no role in this bound, and only the intrinsic signal dimension s is relevant. This is in stark contrast to what is known for the same problem if one restricts to the classical setting of non-adaptive sensing, as in [19], [20], [17]. For instance, for the class of all subsets with cardinality s the non-adaptive sensing lower bound is of the order $\sqrt{\log(n/s^2)}$ if $s < o(\sqrt{n})$. Therefore signals need to be much stronger in order to be reliably detected when using non-adaptive sensing.

The above adaptive sensing lower bound is valid for any symmetric class, and in particular for the maximal class of all subsets S with cardinality s. For this class there is a adaptive sensing methodology able to nearly achieve the lower bound.

Proposition II.1 ([18])**.** *Let $s_n > \log\log\log n$ and consider the class $\mathcal{C}$ of all subsets with cardinality s_n. Furthermore let $\mu > \sqrt{\frac{32 \log\log\log n}{s_n}}$. There is an adaptive sensing testing strategy for which*

$$R(\hat{\Phi}) \to 0 ,$$

as $n \to \infty$.

The mentioned procedure is based on the idea of distilled sensing [15], but it does require some simple modifications to attain the desired bound (see [18]). Note that the order of the bound matches the one of the lower bound up to a factor $\log\log\log n$. It is conjectured that this is an artifact of the specific procedure, however, there are currently no known procedures able to tighten this gap. Perhaps more noteworthy is the fact that extra structure in the class $\mathcal{C}$ is not helpful in the adaptive sensing detection scenario! This is quite different than in the non-adaptive sensing case, where the structure of the set $\mathcal{C}$ can play a very prominent role as well documented in [16], [21], [22], for instance.

The estimation problem exhibit similar trends, but structure of the set $\mathcal{C}$ can give important cues on the design of adaptive sensing methodologies. We focus first on the unstructured case where $\mathcal{C}$ is the class of all subsets of $\{1,\ldots,n\}$ with cardinality s.

Theorem II.2 ([18])**.** *Let $\mathcal{C}$ be the class of all subsets with cardinality s, and let $\hat{S}$ be an arbitrary adaptive sensing support estimator. For any $0<\epsilon<1$, if $\max_{S\in\mathcal{C}} \mathbb{P}_S[\hat{S}\neq S] \leq \max_{S\in\mathcal{C}} \mathbb{E}_S[|\hat{S}\Delta S|] \leq \epsilon$ then necessarily*

$$\mu^2 \geq \sqrt{\frac{2n}{m}\left(\log s + \log\frac{n-s}{n+1} + \log\frac{1}{2\epsilon}\right)} \ .$$

Again, focusing on the case $m=n$ and assuming also the signal is sufficiently sparse (meaning $s_n = o(n)$), we see that μ needs to be on the order of $\sqrt{2\frac{n}{m}\log(s_n)}$ to ensure the probability of making any errors goes to zero as n increases. This result is again in stark contrast with what is possible with non-adaptive sensing, where the signal magnitude μ needs to be on the order of $\sqrt{2\log n}$ to ensure the probability of error goes to zero. Furthermore the above lower bound is tight, as there is a procedure that allows for exact support recovery with probability approaching 1 provided the signal amplitude is of the order $2\sqrt{\log s_n + \log\log n}$ (see [23], [24]). The $\log\log n$ term and the "wrong" constant in the bound are artifacts of their method (which is parameter adaptive and agnostic about s_n), and can be avoided when considering a different approach - running in parallel n entry-wise properly calibrated sequential likelihood ratio tests, which require the knowledge of the sparsity level s_n. Such a procedure achieves precisely the lower bound in the theorem.

It is interesting to notice that, unlike for detection, structure in the class $\mathcal{C}$ can be extremely helpful for estimation. This is the case both for adaptive and non-adaptive sensing. Perhaps the simplest type of structure to consider is when the set S is an "interval", meaning all the entries of S are contiguous (e.g. $S=\{i,i+1,i+s_n-1\}$ for some i). Then adaptive sensing can successfully recover the support with probability approaching 1 provided the signal magnitude is of the order $\sqrt{2\log(s_n)/s_n}$, and this is the optimal rate (unpublished work). Adaptive sensing under other structural constrains (e.g., cliques in a complete graph, paths in a graph) have to the best of our knowledge not been thoroughly studied yet, and therefore remain an important direction for future work.

III. Linear Measurements and Compressed Sensing

The sensing model described in the previous section can be modified to allow for linear measurements, in lieu of single-entry samples. Formally the sensing model becomes

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{W} \ ,$$

where $\boldsymbol{Y}\in\mathbb{R}^l$ denotes the observations, $\boldsymbol{A}\in\mathbb{R}^{l\times n}$ is the design/sensing matrix, $\boldsymbol{x}\in\mathbb{R}^n$ is the unknown signal, and $\boldsymbol{W}\in\mathbb{R}^l$ is a normal multivariate vector with zero mean an identity covariance matrix. The rows of $\boldsymbol{A}$ can be designed sequentially, and the i^{th} row (denoted by $\boldsymbol{A}_{i\cdot}$) can depend explicitly on $\{Y_j,\boldsymbol{A}_{j\cdot}\}_{j=1}^{i-1}$. Note that W_i is a normal random variable independent of $\{Y_j,\boldsymbol{A}_{j\cdot},W_j\}_{j=1}^{i-1}$ and also independent of $\boldsymbol{A}_{i\cdot}$. This setting is particularly interesting when we impose norm constrains on $\boldsymbol{A}$, namely

$$\mathbb{E}\left[\|\boldsymbol{A}\|_F^2\right] \leq m \ , \qquad (2)$$

where $\|\cdot\|_F$ is the Frobenius matrix norm. Like (1), this sensing budget condition is very natural and the issue of noise is otherwise irrelevant. The norm of each row of $\boldsymbol{A}$ plays here the role of the precision parameters Γ_k in (1).

Inference based on linear measurements is at the heart of compressed sensing. Most existing literature focused on the non-adaptive sensing paradigm, and identified strategies to recover signals from a small number of measurements, see for instance [25], [26], [27]. In our setting this means l is chosen to be as small as possible, while making the restriction $l=m$. In the results described below we consider only the sensing budget restriction (2) and assume the number of measurements l can be potentially infinite.

As linear measurements are more powerful/general than entry-wise ones, we might expect some performance improvement in both the detection and estimation inference tasks. The detection problem was been carefully studied in [28] and the author has shown that for reliable detection it is necessary and sufficient for the signal magnitude to be of the order $\frac{1}{s_n}\sqrt{n/m}$. Although this result is somewhat similar to the one in Theorem II.1 we notice that the dependency on the sparsity level s_n is better, and therefore weaker signals can be detected using linear measurements. Perhaps surprisingly adaptive sensing is of no help in this scenario, and detection procedures achieving the optimal performance can be non-adaptive. Furthermore, the structure of the class $\mathcal{C}$ does not help, provided the class is symmetric. This means that, like in the single-entry measurement case, structure is of no use for detection. However, this statement is true both for adaptive and non-adaptive sensing paradigms, meaning that the extra flexibility of adaptive sensing provides no advantage for detection using linear measurements.

For the estimation problem the story is a bit different: adaptive sensing can exhibit an advantage over non-adaptive sensing, as documented in [29], [30], [31]. Furthermore structural information about S can be extremely helpful. In [18] it is shown that for the unstructured case the same lower bound as in Theorem II.2 applies in the context of linear measurements (although the proof of the result requires a few small modifications). Procedures achieving (or nearly achieving) this bound exist, namely [31], [32]. For the non-adaptive sensing paradigm information theoretical lower bounds have also been shown, namely the signal amplitude must exceed a constant times $\sqrt{\frac{n}{m}\sigma^2\log n}$, as shown for instance in [33].

The factor of n/m is the sensing energy per dimension and $\sqrt{\log n}$ is needed to ensure that the signal is larger than the largest noise contribution. Therefore adaptive sensing is advantageous, especially in the typical case when the signal dimension n is very large.

If the sparsity patterns exhibit some structure there are also results contrasting adaptive and non-adaptive sensing, but the story is far from complete. In [34] the authors devise an algorithm that can identify the support set S with high probability when S is an "interval" (see the last paragraph of Section II) provided the signal magnitude is of the order $\sqrt{(n/m)(\log(s_n)/s_n^2)}$. Furthermore they prove a lower bound of the form $\sqrt{(n/m)/s_n^2}$, which matches the upper bound apart from the $\sqrt{\log s_n}$ factor (which does not appear to be an artifact of the algorithm). Again, note that linear measurements are advantageous over entry-wise ones, for which signal magnitude must scale like $\sqrt{(n/m)(\log(s_n)/s_n)}$ for this problem.

IV. Final Remarks

In this brief note we surveyed existing results over adaptive sensing of sparse signals. We considered both entry-wise and linear measurements and clarified in which situations can adaptive sensing yield interesting gains over non-adaptive designs. A clear picture exists for the unstructured scenario, where one assumes only that the support set S is sparse. If in addition one can make structural assumptions over S than it is clear that support estimation is possible for even weaker signals. With so few results available along those lines this remains an interesting avenue for future research.

Acknowledgment

The author would like to thank the anonymous reviewers for their helpful pointers and remarks.

References

[1] A. Wald, *Sequential Analysis*. John Wiley & Sons, Inc., 1947.

[2] H. Chernoff, "Sequential design of experiments," *The Annals of Mathematical Statistics*, vol. 30, no. 3, pp. 755–770, September 1959.

[3] M. A. El-Gamal, "The role of priors in active Bayesian learning in the sequential statistical decision framework," in *Maximum Entropy and Bayesian Methods*, W. T. Grandy Jr. and L. H. Schick, Eds. Kluwer, 1991, pp. 33–38.

[4] P. Hall and I. Molchanov, "Sequential methods for design-adaptive estimation of discontinuities in regression curves and surfaces," *The Annals of Statistics*, vol. 31, no. 3, pp. 921–941, 2003.

[5] G. Blanchard and D. Geman, "Hierarchical testing designs for pattern recognition," *The Annals of Statistics*, vol. 33, no. 3, pp. 1155–1202, 2005.

[6] D. Cohn, Z. Ghahramani, and M. Jordan, "Active learning with statistical models," *Journal of Artificial Intelligence Research*, no. 4, pp. 129–145, 1996.

[7] Y. Freund, H. S. Seung, E. Shamir, and N. Tishby, "Selective sampling using the query by committee algorithm," *Machine Learning*, vol. 28, no. 2-3, pp. 133–168, August 1997.

[8] E. Novak, "On the power of adaption," *Journal of Complexity*, vol. 12, pp. 199–237, 1996.

[9] A. Korostelev and J.-C. Kim, "Rates of convergence for the sup-norm risk in image models under sequential designs," *Statistics & probability Letters*, vol. 46, pp. 391–399, 2000.

[10] S. Dasgupta, "Coarse sample complexity bounds for active learning," in *Advances in Neural Information Processing (NIPS)*, 2005.

[11] S. Hanneke, "Rates of convergence in active learning," *Annals of Statistics*, vol. 39, no. 1, pp. 333–361, 2010.

[12] V. Koltchiinskii, "Rademacher complexities and bounding the excess risk in active learning," *Journal of Machine Learning Research*, vol. 11, September 2010.

[13] N. Balcan, A. Beygelzimer, and J. Langford, "Agostic active learning," in *23rd International Conference on Machine Learning*, Pittsburgh, PA, 2006.

[14] R. Castro and R. Nowak, "Minimax bounds for active learning," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2339–2353, July 2008.

[15] J. Haupt, R. Castro, and R. Nowak, "Distilled sensing: Adaptive sampling for sparse detection and estimation," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6222 – 6235, September 2011.

[16] L. Addario-Berry, N. Broutin, L. Devroye, and G. Lugosi, "On combinatorial testing problems," *The Annals of Statistics*, vol. 38, no. 5, pp. 3063–3092, 2010.

[17] D. Donoho and J. Jin, "Higher criticsm for detecting sparse heterogenous mixtures," *Annals of Statistics*, vol. 32, no. 3, pp. 962–994, 2004.

[18] R. M. Castro, "Adaptive sensing performance lower bounds for sparse signal estimation and detection," *Preprint*, 2012, (available at http://arxiv.org/abs/1206.0648).

[19] Y. Ingster, "Some problem of hypothesis testing leading to infinitely divisible distributions," *Mathematical Methods of Statistics*, vol. 6, pp. 47–69, 1997.

[20] Y. Ingster and I. Suslina, *Nonparametric Goodness-of-Fit Testing under Gaussian Models*, ser. Lecture Notes in Statistics. Springer, 2003.

[21] E. Arias-Castro, E. J. Candès, H. Helgason, and O. Zeitouni, "Searching for a trail of evidence in a maze," *The Annals of Statistics*, vol. 36, no. 4, pp. 1726–1757, 2008.

[22] C. Butucea and Y. Ingster, "Detection of a sparse submatrix of a high-dimensional noisy matrix," *Preprint*, 2011, (available at http://arxiv.org/abs/1109.0898).

[23] M. Malloy and R. Nowak, "Sequential analysis in high-dimensional multiple testing and sparse recovery," in *The IEEE International Symposium on Information Theory*, Saint Petersburg, August 2011, (available at http://arxiv.org/abs/1103.5991v1).

[24] ——, "On the limits of sequential testing in high dimensions," in *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, California, 2011, (available at http://arxiv.org/abs/1105.4540).

[25] E. Candès and M. Davenport, "How well can we estimate a sparse vector?" *Applied and Computational Harmonic Analysis*, vol. 34, no. 2, pp. 317–323, 2013, (available at http://arxiv.org/abs/1111.4646).

[26] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[27] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[28] E. Arias-Castro, "Detecting a vector based on linear measurements," *Electronic Journal of Statistics*, vol. 6, pp. 547–558, 2012.

[29] E. Arias-Castro and M. Davenport, "Compressive binary search," in *The IEEE International Symposium on Information Theory*, Cambridge, Massachusetts, July 2012, (available at http://arxiv.org/abs/1202.0937).

[30] E. Arias-Castro, E. Candès, and M. Davenport, "On the fundamental limits of adaptive sensing," *Preprint*, 2011, (available at http://arxiv.org/abs/1111.4646).

[31] M. Malloy and R. Nowak, "Near-optimal adaptive compressed sensing," in *Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, California, 2012, (available at http://homepages.cae.wisc.edu/~mmalloy/papers/NearOptimalACS.pdf).

[32] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," August 2012, (available at http://www.win.tue.nl/~rmcastro/publications/SCS.pdf).

[33] M. J. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans Inform Theory*, vol. 55, no. 5, 2009.

[34] S. Balakrishnan, M. Kolar, A. Rinaldo, and A. Singh, "Recovering block-structured activations using compressive measurements," *Preprint*, 2012, (available at http://arxiv.org/abs/1209.3431).