1. REPORT DATE **21 DEC 2009**	2. REPORT TYPE	3. DATES COVERED **01-09-2008 to 31-08-2009**
4. TITLE AND SUBTITLE **Recovery of Clustered Sparse Signals from Compressive Measurements**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Rice University,ECE,Houston,TX,77005**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Note: In the following paper, there is a typo in the Acknowledgements that was not caught until after publication. After consultation with the grant reporting office (Karen McCauley via aror.reports@us.army.mil on 12/10/09) and with the technical monitor (Dr. John Lavery on 12/16/09), it was agreed to note the correction with the statement:

Robert C. Sharpley (co-PI ARO MURI}
"Model Classes, Approximation, and Metrics for Dynamic Processing of Urban Terrain Data"
December 20, 2009

Recovery of Clustered Sparse Signals from Compressive Measurements

Volkan Cevher[1], Piotr Indyk[1,2], Chinmay Hegde[1], and Richard G. Baraniuk[1]

(1) Electrical and Computer Engineering, Rice University, Houston, TX
(2) Computer Science and Artificial Intelligence Lab, MIT, Cambridge, MA

Abstract:

We introduce a new signal model, called (K, C)-sparse, to capture K-sparse signals in N dimensions whose nonzero coefficients are contained within at most C clusters, with $C < K \ll N$. In contrast to the existing work in the sparse approximation and compressive sensing literature on block sparsity, no prior knowledge of the locations and sizes of the clusters is assumed. We prove that $\mathcal{O}\left(K + C \log(N/C)\right)$ random projections are sufficient for (K, C)-model sparse signal recovery based on subspace enumeration. We also provide a robust polynomial-time recovery algorithm for (K, C)-model sparse signals with provable estimation guarantees.

1. Introduction

Compressive sensing (CS) is an alternative to Shannon/Nyquist sampling for the acquisition of sparse or compressible signals in an appropriate basis [1, 2]. By sparse, we mean that only K of the N basis coefficients are nonzero, where $K \ll N$. By compressible, we mean the basis coefficients, when sorted, decay rapidly enough to zero so that they can be well-approximated as K-sparse. Instead of taking periodic samples of a signal, CS measures inner products with random vectors and then recovers the signal via a sparsity-seeking convex optimization or greedy algorithm. The number of compressive measurements M necessary to recover a sparse signal under this framework grows as $M = \mathcal{O}\left(K \log(N/K)\right)$

In many applications, including imaging systems and high-speed analog-to-digital converters, such a saving can be dramatic; however, the dimensionality reduction from N to M is still not on par with state-of-the-art transform coding systems. While many natural and manmade signals can be described to a first-order as sparse or compressible, their sparse supports often have an underlying domain specific structure [3–6]. Exploiting this structure in CS recovery has two immediate benefits. First, the number of compressive measurements required for stable recovery decreases due to the reduction in the degrees of freedom of a sparse or compressible signal. Second, true signal information can be better differentiated from recovery artifacts during signal recovery, which increases recovery robustness. Only by exploiting a priori information on coefficient structure in addition to signal sparsity, can CS hope to be competitive with the state-of-the-art transform cod-

ing algorithms for dimensionality reduction.

Fortunately, it is possible to design CS recovery algorithms that exploit the knowledge of structured sparsity models with provable performance guarantees [3, 5, 6]. In particular, the model-based CS recovery framework in [3] generalizes to any structured-sparsity model that has a tractable model-based approximation algorithm. This framework has been applied productively to two structured signal models: block sparsity and wavelet trees with robust recovery guarantees from $\mathcal{O}\left(K\right)$ measurements [3]. To recover signals that have structured sparsity, problem-specific convex relaxation approaches are also used in the literature with recovery guarantees similar to those in [3]; e.g., for block sparse signals, see [5, 6].

In this paper, we introduce a new structured sparsity model, called the (K, C)-model, that constrains the K-sparse signal coefficients to be contained within at most C-clusters. In contrast to the block sparsity model in [5, 6], our proposed model does not assume prior knowledge of the locations and sizes of the coefficient clusters. We show that $\mathcal{O}\left(K + C \log(N/C)\right)$ random projections are sufficient for (K, C)-model signal recovery using a subspace counting argument. We also provide a polynomial-time model-based approximation algorithm based on dynamic programming and a CS recovery algorithm based on the model-based recovery framework of [3]. In contrast to the clustered sparse recovery algorithm based on the probabilistic Ising model in [7], the (K, C)-model has provable performance guarantees.

The paper is organized as follows. Section 2 provides the necessary theoretical and algorithmic background on model-based CS. Section 3 introduces the (K, C)-model, derives its sampling bound for CS recovery, and describes a dynamic programming solution for optimal (K, C)-model approximation. Section 4 discusses the aspect of compressibility and highlights some connections to the block sparsity model. Simulation results are given in Section 5 to demonstrate the effectiveness of the (K, C)-model. Section 6 provides our conclusions.

2. Model-based CS Background

A K-sparse signal vector x lives in $\Sigma_K \subset \mathbb{R}^N$, which is a union of $\binom{N}{K}$ subspaces of dimension K. Other than its K-sparsity, there are no further constraints on the support or values of its coefficients. A *union-of-subspaces*

signal model (a *signal model* in the sequel for brevity) endows the K-sparse signal x with additional structure that allows certain K-dimensional subspaces in Σ_K and disallows others [4, 8].

More formally, let $x_{|\Omega}$ represent the entries of x corresponding to the set of indices $\Omega \subseteq \{1, \ldots, N\}$, and let Ω^C denote the complement of the set Ω. A signal model $\mathcal{M}_K$ is then defined as the union of m_K canonical K-dimensional subspaces

$$\mathcal{M}_K = \bigcup_{m=1}^{m_K} \mathcal{X}_m, \ \mathcal{X}_m := \{x : x_{|\Omega_m} \in \mathbb{R}^K, x_{|\Omega_m^C} = 0\}.$$

Each subspace $\mathcal{X}_m$ contains all signals x with $\text{supp}(x) \in \Omega_m$. Thus, the signal model $\mathcal{M}_K$ is defined by the set of possible supports $\{\Omega_1, \ldots, \Omega_{m_K}\}$. Signals from $\mathcal{M}_K$ are called K-*model sparse*. Likewise, we may define $\mathcal{M}_K^c$ to be the set of c-wise differences of signals belonging to $\mathcal{M}_K$. Clearly, $\mathcal{M}_K \subseteq \Sigma_K$ and $\mathcal{M}_K^4 \subseteq \Sigma_{4K}$. In the sequel, we will use an algorithm $\mathbb{M}(x; K)$ that returns the best K-term approximation of the signal x under the model $\mathcal{M}_K$.

If we know that the signal x being acquired is K-model sparse, then we can relax the standard restricted isometry property (RIP) [1] of the CS measurement matrix Φ and still achieve stable recovery from the compressive measurements $y = \Phi x$. The *model-based* RIP $\mathcal{M}_K$-RIP requires that

$$(1 - \delta_{\mathcal{M}_K})\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1 + \delta_{\mathcal{M}_K})\|x\|_2^2 \quad (1)$$

hold for signals $x \in \mathcal{M}_K$ [4, 8], where $\delta_{\mathcal{M}_K}$ is the model-based RIP constant.

Blumensath and Davies [4] have quantified the number of measurements M necessary for a subgaussian CS matrix to have the $\mathcal{M}_K$-RIP with constant $\delta_{\mathcal{M}_K}$ and with probability $1 - e^{-t}$ to be

$$M \ge \frac{2}{c\delta_{\mathcal{M}_K}^2}\left(\ln(2m_K) + K\ln\frac{12}{\delta_{\mathcal{M}_K}} + t\right). \quad (2)$$

This bound can be used to recover the conventional CS result by substituting $m_K = \binom{N}{K} \approx (Ne/K)^K$.

To take practical advantage of signal models in CS, we can integrate them into a standard CS recovery algorithm based on iterative greedy approximation. The key modification is surprisingly simple [3]: we merely replace the best K-term approximation step with the best K-term model-based approximation $\mathbb{M}(x; K)$. For example, in the CoSaMP algorithm [9], the best LK-term approximation (with L a small integer) is modified to incorporate a best LK-term model-based approximation. The resulting algorithm (see [3]) then inherits the following model-based CS recovery guarantee at each iteration i, when the measurement matrix Φ has the $\mathcal{M}_K^4$-RIP with $\delta_{\mathcal{M}_K^4} \le 0.1$:

$$\|x - \hat{x}_i\|_2 \le 2^{-i}\|x\|_2 + 20\left(\|x - x_{\mathcal{M}_K}\|_2 \right.$$

$$\left. + \frac{1}{\sqrt{K}}\|x - x_{\mathcal{M}_K}\|_1 + \|n\|_2\right),$$

where $x_{\mathcal{M}_K} = \mathbb{M}(x; K)$ is the best model-based approximation of x within $\mathcal{M}_K$.

3. The (K, C)-Model

Motivation: The *block sparsity model* is used in applications where the significant coefficients of a sparse signal appear in designated blocks on the ambient signal dimension, e.g., group sparse regression problems, DNA microarrays, MIMO channel equalization, source localization in sensor networks, and magnetoencephalography [3, 5, 6, 10–14]. It has been shown that recovery algorithms provably improve standard CS recovery by exploiting this block-sparse structure [3, 5].

The (K, C)-model generalizes the block sparsity model by allowing the significant coefficients of a sparse signal to appear in at most C clusters of unknown size and location (Figure 1(a)). This way, the (K, C)-model further accommodates additional applications in, e.g., neuroscience problems that are involved with decoding of natural images in the primary visual cortex (V1) or understanding the statistical behavior of groups of neurons in the retina [15]. In this section, we formulate the (K, C)-model as a union of subspaces and pose an approximation algorithm on this union of subspaces.

To define the set of (K, C)-sparse signals, without loss of generality, we focus on canonically sparse signals in $N + 2$ dimensions whose first and last coefficients are zero. Consider expressing the support of such signals via run-length coding with a vector $\beta = (\beta_1, \ldots, \beta_{2C+1})$ $(\beta_j \ne 0)$, where β_{odd} counts the number of continuous zero-signal values and β_{even} counts the number of continuous nonzero-signal values (i.e., clusters).

Definition: The (K, C)-sparse signal model $\mathcal{M}_{(K,C)}$ is defined as

$$\mathcal{M}_{(K,C)} = \left\{x \in R^{N+2} \left| \sum_{i=1}^{2C+1} \beta_i = N + 2, \sum_{i=1}^{C} \beta_{2i} = K\right.\right\}. \quad (3)$$

Sampling Bound: The number of subspaces $m_{(K,C)}$ in $\mathcal{M}_{(K,C)}$ can be obtained by counting the number of *positive* solutions to the following integer equations:

$$\begin{aligned}
\beta_1 + \beta_2 + \ldots + \beta_{2C+1} &= N + 2, \\
\beta_2 + \beta_4 + \ldots + \beta_{2C} &= K,
\end{aligned}$$

which can be rewritten as

$$\begin{aligned}
\beta_1 + \beta_3 + \ldots + \beta_{2C+1} &= N + 2 - K, \\
\beta_2 + \beta_4 + \ldots + \beta_{2C} &= K.
\end{aligned} \quad (4)$$

Note that the number of positive integer solutions to the following problem:

$$\beta_1 + \beta_2 + \beta_3 + \ldots + \beta_n = N,$$

is given by $\binom{N-1}{n-1}$. Then, we can count the solutions to the two of decoupled problems in (4) and multiply the number of solutions to obtain $m_{(K,C)}$:

$$m_{(K,C)} = \binom{N + 1 - K}{C}\binom{K - 1}{C - 1}. \quad (5)$$

Plugging (5) into (2), we obtain the sampling bound for $\mathcal{M}_{(K,C)}$:

$$M = \mathcal{O}\left(K + C \log \frac{N}{C}\right). \quad (6)$$

Note that the (K,C)-sampling bound (6) becomes the standard CS bound of $M = \mathcal{O}\left(K \log \frac{N}{K}\right)$ when $C \approx K$.

Model Approximation Algorithm: In this section we focus on designing an algorithm $\mathbb{M}(x; K, C)$ for finding the best (K,C)-model approximation to a given signal x. The algorithm uses the principle of dynamic programming [16]. For simplicity, we focus on the problem of finding the *cost* of the best (K,C)-clustered signal approximation in ℓ_2. This solution generalizes to the best (K,C)-clustered signal approximation in ℓ_p for $p \geq 1$. The actual sparsity pattern can be then recovered using standard back-tracing techniques; see [16] for the details.

The algorithm $\mathbb{M}(x; K, C)$ computes an array cost$[i, j, k, c]$, where $1 \leq i \leq j \leq N$, $0 \leq k \leq K$, and $0 \leq c \leq C$. At the end of the algorithm, each entry cost$[i, j, k, c]$ contains the smallest cost of approximating $x_{i:j}$, the signal vector restricted to the index set $[i, \ldots, j]$, using at most k non-zero entries that span at most c clusters. $\mathbb{M}(x; K, C)$ performs the following operations.

(*Initialization*) When either $c = 0$ or $k = 0$, the signal approximation costs can be computed directly, since cost$[i, j, 0, c] = \|x_{i:j}\|_2^2$ and cost$[i, j, k, 0] = \|x_{i:j}\|_2^2$, for all valid indices i, j, k, c. Moreover, for all entries i, j, k, c such that $c > 0$ and $j - i + 1 \leq k$, we have cost$[i, j, k, c] = 0$ since we can include all $j - i + 1$ coordinates of the vector $x_{i:j}$ in the approximation.

(*Main loop*) All other cost entries can then be computed using the following recursion:

$$\text{cost}[i, j, k, c] = \min_{c^\star = 0 \ldots c} \min_{k^\star = 0 \ldots k} \min_{j^\star = i \ldots j-1}$$
$$\left\{ \text{cost}[i, j^\star, k^\star, c^\star] \times \text{cost}[j^\star + 1, j, k - k^\star, c - c^\star] \right\}.$$

The correctness of the algorithm follows from the following observation. Let $\overline{v}$ be the best (k, c)-clustered approximation of $x_{i:j}$. Unless all entries of $x_{i:j}$ can be included in the approximation $\overline{v}$ (in which case $j - i + 1 \geq k$ and the entry has been already computed during initialization), then there must exist an index $l \in [i, \ldots, j]$ such that x_l is not included in $\overline{v}$. Let $l^\star = l$ if $l < j$, and $l^\star = j - 1$ otherwise. Let $k^\star$ be the number of non-zero entries present in the *left segment* of $\overline{v}_{i:l^\star}$, and let $c^\star$ be the number of clusters present in that left segment. Then, it must be the case that $\overline{v}_{i:l^\star}$ is the best $(k^\star, c^\star)$-approximation to $x_{i:l}$, and $\overline{v}_{l+1:j}$ is the best $(k - k^\star, c - c^\star)$-approximation to $x_{(l^\star+1):j}$. Otherwise, those better approximations could have been concatenated together to yield an even better (k, c)-approximation of $x_{i:j}$. Thus, the recursive formula will identify the optimal split and compute the optimal approximation cost.

The cost table contains $\mathcal{O}\left(N^2 K C\right)$ entries. Each entry can be computed in $\mathcal{O}(NKC)$ time. Thus, the running time of the algorithm is $\mathcal{O}\left(N^3 K^2 C^2\right)$.

4. Additional Remarks

Compressibility: Just as compressible signals are *nearly K-sparse* and live close to the union of subspaces Σ_K in $\mathbb{R}^N$, (K, C)-compressible signals are *nearly (K, C)-model sparse* and live close to the restricted union of subspaces $\mathcal{M}_{(K,C)}$. Here, we rigorously introduce a (K, C)-compressible signal model in terms of the decay of their (K, C)-model approximation error.

We first define the ℓ_2 error incurred by approximating $x \in \mathbb{R}^N$ by the best approximation in $\mathcal{M}_{(K,C)}$:

$$\sigma_{\mathcal{M}_{(K,C)}}(x) \triangleq \inf_{\bar{x} \in \mathcal{M}_{(K,C)}} \|x - \bar{x}\|_2 = \|x - \mathbb{M}(x; K, C)\|_2.$$

The decay of the (K, C)-model approximation error in (7) defines the (K, C)-compressibility of a signal. Then, a set of (K, C)-*model s-compressible signals* is given by

$$\mathfrak{M}_s = \left\{ x \in \mathbb{R}^N : \sigma_{\mathcal{M}_{j(K,C)}}(x) \leq S(jK)^{-1/s}, \right.$$
$$\left. 1 \leq K \leq N, S < \infty, j = 1, \ldots, \left\lfloor \frac{N}{K} \right\rfloor \right\}. \quad (7)$$

Define $S_{\mathfrak{M}}$ as the smallest value of S for which this condition holds for x and s.

We use the restricted amplification property (RAmP) and the nested approximation property (NAP) in [3] to ensure that the (K, C)-model based CoSaMP recovery possesses the following guarantee for (K, C)-model s-compressible signals at each iteration i:

$$\|x - \widehat{x}_i\|_2 \leq 2^{-i} \|x\|_2 + 35 \left(\|n\|_2 + \frac{S_{\mathfrak{M}}}{K^s}(1 + \ln\lceil N/K \rceil) \right), \quad (8)$$

when Φ has the $\mathcal{M}_{(K,C)}^4$-RIP with $\delta_{\mathcal{M}_{(K,C)}^4} \leq 0.1$ and the (ϵ_K, r)-RAmP with $\epsilon_K \leq 0.1$ and $r = s - 1$.

Simulation via Block Sparsity: It is possible to recover (K, C)-sparse signals by using the block sparsity model if we are willing to pay an added penalty in terms of the number of measurements. To demonstrate this, we define uniform blocks of size K/C (e.g., average cluster length) on the signal space. Then, it is straightforward to see that the number of active blocks B in the block sparse model is upper-bounded by

$$B \leq 2(C - 1) + \frac{K - 2(C - 1)}{K/C} \leq 3C. \quad (9)$$

To reach this upper bound, we first construct a (K, C)-sparse signal that has $(C - 1)$-clusters with 2 coefficients and a single cluster with the remaining sparse coefficients. We then place the clusters with two coefficients at the boundary of the block sparse model so that each cluster activate two blocks in the block sparse model to arrive at (9). Then, the (K, C)-equivalent block sparse model requires $M = \mathcal{O}\left(BK/C + B \log \frac{N}{B}\right)$ samples, where $B = 3C$.

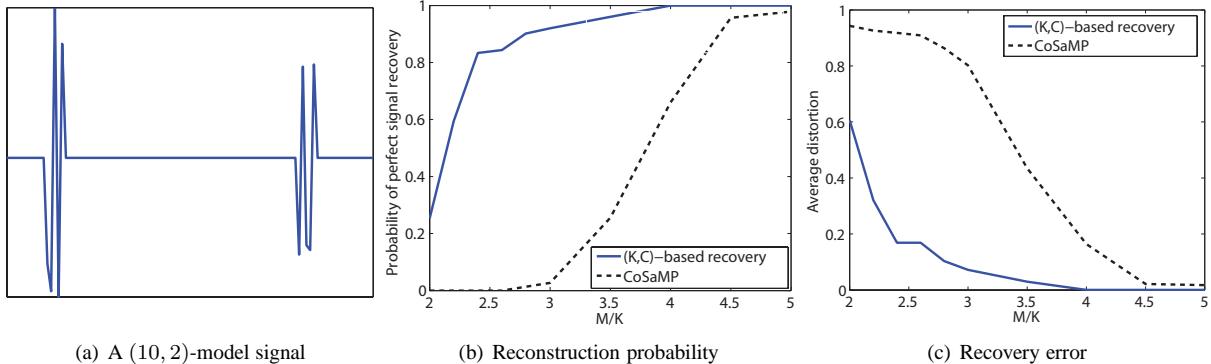(a) A $(10,2)$-model signal   (b) Reconstruction probability   (c) Recovery error

Figure 1: *Monte Carlo simulation results for (K,C)-model based recovery with $K = 10, C = 2.$.*

5. Experiments

In this section we demonstrate the performance of (K,C)-model based recovery. Our test signals are the class of length-100 clustered-sparse signals with $K = 10, C = 2$. We run both the CoSaMP algorithm as well as (K,C)-model based CoSaMP algorithm [3] until convergence for 1000 independent trials. In Fig. 1(a), a sample realization of the signal is displayed. It is evident from Figs. 1(b) and (c) that enforcing the structured sparsity model in the recovery process significantly improves CS reconstruction performance. In particular, Fig. 1(b) demonstrates that approximately 85% of the signals are almost perfectly recovered at $M = 2.5K$, whereas CoSaMP fails to recover any signals at this level of measurements. Instead, traditional sparsity-based recovery requires $M \geq 4.5K$ to attain comparable performance. Similarly, Figure. 1(c) displays the rapid decrease in average recovery distortion of our proposed method, as compared to the conventional approach. The (K,C)-sparse approximation algorithm codes are available at dsp.rice.edu/software/KC.

6. Conclusions

In this paper, we have introduced a new sparse signal model that generalizes the block-sparsity model used in the CS literature. To exploit the provable model-based CS recovery framework of [3], we developed a dynamic programming algorithm that computes, for any given signal, its optimal ℓ_2-approximation within our clustered sparsity model. We then demonstrated that significant performance gains can be made by exploiting the clustered signal model beyond the simplistic sparse model that are prevalent the CS literature.

References:

[1] E. J. Candès, "Compressive sampling," in *Proc. International Congress of Mathematicians*, vol. 3, (Madrid, Spain), pp. 1433–1452, 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, pp. 1289–1306, Sept. 2006.

[3] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," 2008. Preprint. Available at http://dsp.rice.edu/cs.

[4] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Info. Theory*, Dec. 2008.

[5] Y. Eldar and M. Mishali, "Robust recovery of signals from a union of subspaces," 2008. Preprint.

[6] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," Mar. 2008. Preprint.

[7] V. Cevher, M. F. Duarte, C. Hegde, and R. G. Baraniuk, "Sparse signal recovery using Markov Random Fields," in *Proc. Workshop on Neural Info. Proc. Sys. (NIPS)*, (Vancouver, Canada), Dec. 2008.

[8] Y. M. Lu and M. N. Do, "Sampling signals from a union of subspaces," *IEEE Signal Processing Mag.*, vol. 25, pp. 41–47, Mar. 2008.

[9] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, June 2008.

[10] J. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Processing*, vol. 86, pp. 572–588, Apr. 2006.

[11] Y. Kim, J. Kim, and Y. Kim, "Blockwise sparse regression," *Statistica Sinica*, vol. 16, no. 2, p. 375, 2006.

[12] L. Meier, S. van de Geer, and P. Buhlmann, "The group lasso for logistic regression," *Journal of Royal Stat. Society: Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 53–71, 2008.

[13] F. Parvaresh, H. Vikalo, S. Misra, and B. Hassibi, "Recovering Sparse Signals Using Sparse Measurement Matrices in Compressed DNA Microarrays," *IEEE Journal of Selected Topics in Sig. Proc.*, vol. 2, no. 3, pp. 275–285, 2008.

[14] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroenceph. and Clin. Neurophys.*, vol. 95, no. 4, pp. 231–251, 1995.

[15] P. J. Garrigues and B. A. Olshausen, "Learning Horizontal Connections in a Sparse Coding Model of Natural Images," in *Advances in Neural Info. Proc. Sys. (NIPS)*, 2008.

[16] T. H. Corman, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. MIT Press and McGraw-Hill, New York, USA, 2001.