LEARNING-BASED DESIGN OF MEASUREMENT MATRIX WITH INTER-COLUMN CORRELATION FOR COMPRESSIVE SENSING

Mahdi Parchami † *Hamidreza Amindavar* † *Wei-Ping Zhu* ⋆

† Department of Electrical Engineering
Amirkabir University of Technology, Tehran, Iran
{mparchami,hamidami}@aut.ac.ir

⋆ Department of Electrical and Computer Engineering
Concordia University, Montreal, Quebec, Canada
weiping@ece.concordia.ca

ABSTRACT

In this paper, a new approach for the design of measurement matrix, $\mathbf{\Phi}$, for compressive sensing (CS) in a generic context is proposed. In accordance with well-known classical CS theory, we take the elements of $\mathbf{\Phi}$ to be random, yet, we include correlations within the elements of the individual columns of $\mathbf{\Phi}$. To this end, a new structure for $\mathbf{\Phi}$ is proposed where the correlations of interest are controlled by a selectable parameter. We aim at optimizing the proposed $\mathbf{\Phi}$ with respect to the latter correlation parameter by leveraging an appropriate criterion in a learning-based framework. We evaluate the performance of the proposed $\mathbf{\Phi}$ and compare it with the state-of-the-art literature including random $\mathbf{\Phi}$ with independent and identically distributed (i.i.d.) elements. Performance advantage of the proposed approach is validated in different CS scenarios.

***Index Terms*—** Compressive sensing, learning, measurement matrix, sparse reconstruction.

1. INTRODUCTION

Over the past two decades, there has been a large amount of research on the problem of recovering a structurally sparse signal, $\mathbf{x} \in \mathbb{C}^m$, i.e. one with k non-zero elements, from a set of compressed measurements, $\mathbf{y}$, of the form

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{v}, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{C}^{n \times m}$ is a known measurement matrix and $\mathbf{v} \in \mathbb{C}^n$ represents the measurement noise [1]. This problem, often referred to as compressive sensing (CS), stems from a wide range of applications including medical resonance imaging (MRI), spectroscopy, radar, Fourier optics, shortwave-infrared cameras, facial recognition and network tomography [2, 3].

The fundamental challenges in the revolution of CS theory can be identified as the following: development of efficient and practical sparse signal recovery algorithms stemmed from the groundbreaking methods in [4, 5, 6], investigation of suitable structures such as sparsity and coherence in the signal of interest, $\mathbf{x}$, e.g. [7, 8, 9], taking into account the additive noise, $\mathbf{v}$, especially for low-SNR conditions in CS, as in [10, 11], and finally, design of an appropriate measurement matrix, $\mathbf{\Phi}$, as followed in [12, 13, 14]. In this paper, we basically focus on the latter topic.

Regarding the measurement matrix design problem, in [12], a framework to construct deterministic measurement matrices and optimize them along with a sparsifying basis is proposed. This approach, basically developed for image signals, employs a set of training images to learn the sensing matrix and the basis. Within the same line of work, in particular for multiple-input multiple-output (MIMO) radar signals, the authors in [13] consider a measurement matrix design which can improve the receiver SNR performance. Further in [15], a deterministic yet structurally random matrix (SRM) to obtain CS measurements is developed, wherein the sensing signal is first pre-randomized and then fast-transformed and subsampled to provide the measured samples. Therein, it is shown that the SRM has comparable reconstruction performance to that of a completely random sensing matrix. More recently in [14], a subsampled structure of the form $\mathbf{P}_\Omega\mathbf{\Psi}$ is considered for $\mathbf{\Phi}$ with $\mathbf{\Psi}$ and $\mathbf{P}_\Omega$ respectively being the basis matrix and the subsampling operator. Therein, a data-driven optimization is sought to obtain the index set, Ω, of the projector, $\mathbf{P}_\Omega$.

All aforementioned methods tend to encounter the problem of measurement matrix design from a deterministic viewpoint, which makes sense due to imposing less computational burden. Yet, fully random measurement matrices with their appealing properties such as the ease of creation, good reconstruction power, inherent incoherence and satisfying well the restricted isometry property (RIP) [16] are desirable in practice. In addition, there are still a few fundamental remarks that can delineate the employment of the traditional i.i.d. Gaussian measurement matrix for CS[1]. One such remark is the fact that the independent columns of a randomly generated $\mathbf{\Phi}$ may have correlations within their own entries. In this work, towards posing a more generic structure for the random $\mathbf{\Phi}$, we capitalize on employing the inter-column correlation (ICC) in designing $\mathbf{\Phi}$. To this end, we propose a two-step algorithm where in the first step, an initial estimate for $\mathbf{\Phi}$ is obtained by considering it in the form of $\mathbf{P}\mathbf{\Psi}$ and optimizing the suggested structure for $\mathbf{P}$ with respect to a measure of ICC using an appropriate criterion from [14]. The columns of the initial estimate of $\mathbf{\Phi}$ are used in the next step to estimate the ICC matrix which is ultimately exploited to generate the independent columns of a Gaussian measurement matrix. The proposed work is evaluated and compared against the state-of-the-art literature where no such correlation is considered. Our experiments reveal the advantage of using the ICC-based $\mathbf{\Phi}$ in different CS scenarios.

2. PROPOSED MEASUREMENT MATRIX

In this section, the proposed approach for the design of CS measurement matrix is presented. First, we discuss the structure of the

[1] See Section III.A of [4] for a discussion of this. Such random measurement matrices are referred to as admissible measurement matrices.

initial $\mathbf{\Phi}$ used for the estimation of the ICC. Next, we explain our method of choosing one of the suggested structure's parameters that controls the amount of ICC. The columns of the resulting initial $\mathbf{\Phi}$ are ultimately exploited to estimate the ICC in the proposed Gaussian measurement matrix.

2.1. Structure of the initial $\mathbf{\Phi}$

Measurement matrices of the form $\mathbf{\Phi} = \mathbf{P\Psi}$ are of direct interest in CS, where one is able to choose the projector matrix, $\mathbf{P}$, given a proper basis, $\mathbf{\Psi}$ [14][2]. We employ the foregoing structure as an initial suggestion for the measurement matrix denoted by $\mathbf{\Phi}^{(0)}$, yet, in contrast to [14], where $\mathbf{P}_\Omega$ is designed only to select n out of m rows of $\mathbf{\Psi}$ in order to generate $\mathbf{\Phi}^{(0)}$, we make use of linear combinations of the rows of $\mathbf{\Psi}$ to define $\mathbf{P}$. To this end, we consider the structure presented in (2) at the bottom of this page for the projector matrix. As seen, there exist L non-zero elements at each row of the suggested $\mathbf{P}$, namely, $\mathbf{w} = \{w_1, w_1, \cdots, w_L\}$. This is equivalent to selecting a linear combination of L rows of $\mathbf{\Psi}$, weighted by w_l's, as each row of the matrix $\mathbf{P}$. Assuming that the increment ℓ_{inc} in the starting position of the window, $\mathbf{w}$, is smaller than its length, L, it is easily seen that each row of the resulting $\mathbf{P}$ has $\ell_{corr} = L - \ell_{inc}$ rows of $\mathbf{\Psi}$ in common with its adjacent rows (and fewer with other rows). Given that each row of the basis matrix, $\mathbf{\Psi}$, corresponds to a distinct frequency or character in the transform domain, and subsequently, each row is orthogonal to the other, this way of defining the matrix $\mathbf{P}$ imposes correlations across its rows, or equivalently, within each of its columns. This correlation is to be used in the considered measure for ICC in Subsection 2.2.

2.2. Selection of the ICC measure

Even though it seems that the choice of parameters L and ℓ_{inc} can control the amount of ICC to a large extent, this is not convenient in practice. In this sense, it should be noted that the choice of n and m, i.e., the measurement and sparse signal lengths, respectively, is often implied by the problem of interest. As well, taking into account the structure in (2), it can be shown that $L + (n-1)\ell_{inc}$ shall be equal to m. This limits the available choices for L and ℓ_{inc}, and therefore, these parameters are not considered as degrees of freedom here. Rather, in order to adjust the amount of ICC, we rely on the inherent shape of the window $\mathbf{w}$. In general, different window types can be assumed for $\mathbf{w}$, due to its being a matter of choice in the current problem. In essence, any bell-shaped window with decaying sidelobes and controllable decay rate would properly satisfy this purpose. In this regard, we take the window $\mathbf{w}$ as L equispaced samples from a zero-mean Gaussian distribution function with adjustable variance, v^2, or standard deviation, v, over an interval, I. To have non-negligible values for $\mathbf{w}$, we assume $I \in [-3, 3]$ in compliance with a standard Gaussian function, i.e., one with unit variance, and tend to select the appropriate values of v from the range $[0, v_{max}]$. An illustration of our discussion is shown in Fig. 1. As observed, the two adjacent Gaussian windows of length L over the interval I have ℓ_{corr} overlapping taps in both cases (a) and (b). Yet, in case of (a), due to a comparatively high value for the deviation v, the resulting correlation, and therefore ICC, between the two windows is larger with respect to case (b) where such variance is visibly smaller. We make use of this fact and take the deviation v of the Gaussian window $\mathbf{w}$ as our measure of the ICC. In this respect, we take advantage of the maximum-energy criterion in [14][3] and tend to maximize the energy of the observed measurements with respect to the above-mentioned variance. Employing this learning-based criterion, we have for the optimum deviation parameter

$$v_{opt_j} = \underset{v}{\operatorname{argmax}} \ \|\mathrm{P}(v)\mathbf{\Psi}\mathbf{x}^{(j)}\|_2^2, \tag{3}$$

where $\|.\|_2$ denotes the Euclidean norm, $\mathbf{x}^{(j)}$ is the j-th training sparse signal with $j \in \{1, 2, \cdots, J\}$, and the projector matrix $\mathrm{P}(v)$ is subsequently a function of the deviation v of the taps in $\mathbf{w}$. Note that the optimal variance, v_{opt_j}, is obtained for each training signal $\mathbf{x}^{(j)}$ separately, resulting in a total of J projector matrices, $\{\mathrm{P}(v_{opt_j})\}$, and thus, J deterministic measurement matrices, namely, $\{\mathbf{\Phi}_j^{(0)}\}$. The latter set of initial measurement matrices, if trained under a proper set of sparse signals, bury important information about the ICC within each of their columns and will be used in

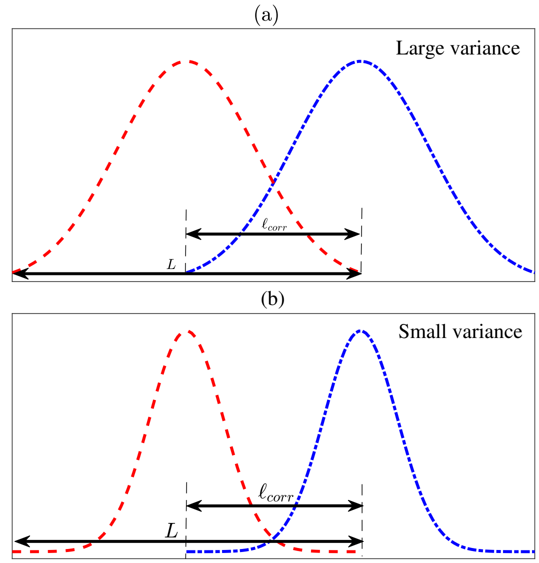

Fig. 1: Comparison between the existing correlation of two Gaussian windows with the same values of L and ℓ_{corr}, (a): case of large variance, (b): case of small variance.

[2]The so-called sparsifying basis matrix $\mathbf{\Psi}$ is often taken as some transform matrix such as DFT or DCT.

[3]This heuristic yet highly efficient criterion amounts to choosing the structure in the matrix P that preserves the highest energy in the observations given by $\mathbf{y} = \mathrm{P}\mathbf{\Psi}\mathbf{x}$.

$$\mathbf{P}_{n\times m} = \begin{bmatrix} \underbrace{w_1 \ \ w_2 \ \ \dots \ \ \dots \ \ \dots \ \ \dots \ \ \dots \ \ w_L}_{L} & 0 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \underbrace{0 \ \ 0 \ \ \dots \ \ 0}_{\ell_{inc}} \ \ w_1 \ \ w_2 \ \ \dots \ \ \dots & \dots & \dots & \dots & w_L & 0 & 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots \ \ \vdots \ \ \vdots \ \ \vdots \ \ \vdots \ \ \vdots \ \ \vdots \ \ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 \ \ 0 \ \ \dots \ \ \dots \ \ \dots \ \ \dots \ \ \dots \ \ \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & w_1 & w_2 & \dots & w_L \end{bmatrix}, \tag{2}$$

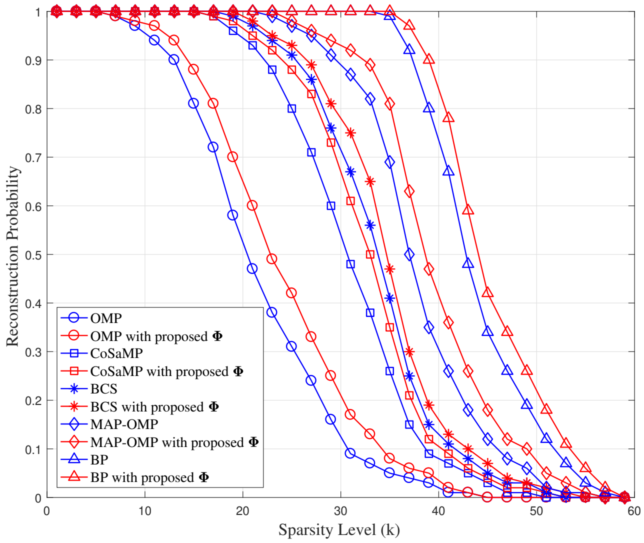

Fig. 2: Reconstruction probability versus sparsity level for a randomly generated binary signal (Bernoulli process) with noise-free measurements using the parameter set $\{n = 120, m = 250, L = 12, \ell_{inc} = 2\}$.

the following subsection to estimate the ICC of the proposed measurement matrix.

2.3. Generation of random $\mathbf{\Phi}$ with ICC

The observations $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ can be written in the following form

$$\mathbf{y} = \sum_{i=1}^{m} \boldsymbol{\phi}_i \, x_i, \tag{4}$$

with $\boldsymbol{\phi}_i$ and x_i respectively being the i-th column of $\mathbf{\Phi}$ and the i-th element of $\mathbf{x}$. It is observed that each column of the measurement matrix, $\mathbf{\Phi}$, corresponds to one element of the sparse signal, $\mathbf{x}$. Hence, in its most general form, the term ICC can be actually viewed as m correlation matrices each corresponding to one of the m columns of $\mathbf{\Phi}$. To estimate the aforementioned correlation matrices, an enough number of ensembles of the measurement matrix columns can be used. To do so, we exploit the columns of the J measurement matrices $\{\mathbf{\Phi}_j^{(0)}\}$ obtained in the previous subsection as a means of estimating the ICC. In this sense, the following estimator is used for the ICC within $\mathbf{\Phi}$

$$E\{\boldsymbol{\phi}_i \boldsymbol{\phi}_i^H\} = \frac{1}{J}\sum_{j=1}^{J} \boldsymbol{\phi}_{i_j}^{(0)} \, \boldsymbol{\phi}_{i_j}^{(0)\,H}, \quad 1 \le i \le m. \tag{5}$$

where $\boldsymbol{\phi}_{i_j}^{(0)}$ denotes the i-th column of $\mathbf{\Phi}_j^{(0)}$ and $\{.\}^H$ is the matrix Hermitian. Now, the proposed measurement matrix can be generated by incorporating the $n \times n$ correlation matrices given by (5) into an appropriate random vector generator independently for each column.

3. NUMERICAL EXPERIMENTS

In this section, we present a summary of the obtained results in different numerical experiments, illustrating the advantage of the proposed measurement matrix for CS. To implement the suggested approach, we handle the optimization in (2) by choosing v_{opt} from

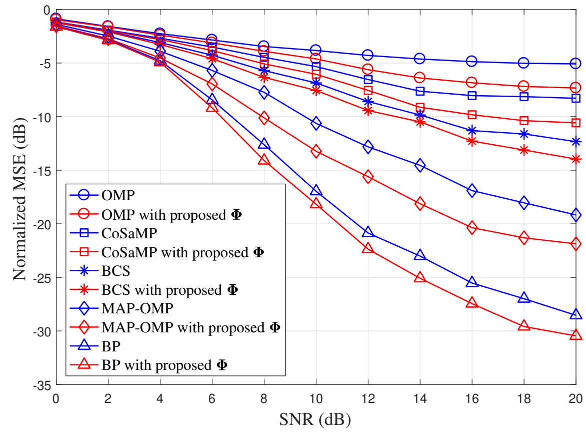

Fig. 3: Normalized MSE versus SNR (dB) for the randomly generated Gaussian signal using the parameter set $\{n = 120, m = 250, k = 30, L = 12, \ell_{inc} = 2\}$.

a discrete set of 100 points spanning linearly the range of $[0, v_{max}]$ with v_{max}=1.25. In this way, in addition to having less burden, we avoid the need for convexity guarantees for the objective function. In order for m to be equal to $L + (n-1)\ell_{inc}$, if required, adequate zero-padding is assumed at the end of the signal vector $\mathbf{x}$. Choosing the number of training signals, i.e. J, as 10 provided sufficient performance in all tested scenarios, without the restriction of using the same training and test signals.

To demonstrate the robustness of the proposed approach, we use the suggested $\mathbf{\Phi}$ in different basic CS methods and compare the results to the case of using an i.i.d. measurement matrix. To this end, we consider the orthogonal matching pursuit (OMP) [4], the compressive sampling matching pursuit (CoSaMP) [17], the Bayesian compressive sensing [5], the recently proposed MAP-OMP [18], and the basis pursuit (BP) methods [1]. The latter is actually the optimal solution of the CS problem in the sense of ℓ_1-norm minimization. We distinct two scenarios where in the former, a binary signal of length m, consisting of k ones with the rest of its elements being zero, is to be recovered from n measurements[4]. In the latter scenario, we consider to recover randomly generated Gaussian signals contaminated with white noise at different SNR's. Both synthetic signals are generated using Matlab. The reconstruction probability[5] for different k is calculated empirically for the former case, as shown in Fig. 2. As well, the normalized mean squared error (MSE) between the true and reconstructed signals versus the SNR level is illustrated in Fig. 3 using the same methodology. It is observed that in both scenarios, using the proposed ICC-based measurement matrix helps improving the reconstruction performance, while this advantage is more pronounced for adverse conditions where the number of measurements, n, is not large enough compared to the sparsity level, k. Also, it is seen that the gap in MSE performance of the proposed and the traditional measurement matrices is increased monotonically with increment of SNR. This improvement is in the order of 3 dBs at best for the scenario under test.

Next, to compare the performance of the proposed measurement matrix design method with that of the other methods, we consider the same aforementioned scenarios but employ different measurement matrices in the MAP-OMP method to recover the sparse sig-

[4]Note that in this case, the CS method is employed only to retrieve the unknown indices of the non-zero signal elements.

[5]This is in fact the probability that the entire binary signal is recovered perfectly.

Table 1: Reconstruction probability in various sparsity levels using different measurement matrix design methods.

Measurement matrix design method	k=30	k=35	k=40	k=45
Measurement matrix with i.i.d. elements	0.89	0.69	0.30	0.12
Measurement matrix with fixed ICC	0.89	0.71	0.33	0.14
Measurement matrix in [14] with f_{avg}	0.91	0.74	0.36	0.15
Proposed Measurement matrix	0.93	0.81	0.41	0.18

Table 2: Normalized MSE (dB) versus SNR (dB) using different measurement matrix design methods.

Measurement matrix design method	5 dB	10 dB	15 dB	20 dB
Measurement matrix with i.i.d. elements	-4.70	-10.62	-15.55	-19.17
Measurement matrix with fixed ICC	-4.79	-11.16	-17.03	-19.74
Measurement matrix in [14] with f_{avg}	-5.02	-11.92	-17.98	-20.96
Proposed Measurement matrix	-5.57	-13.23	-19.10	-21.87

nals. The measurement matrices include the classical i.i.d. one, a Gaussian random matrix with some correlation structure as the ICC, and the learning-based approach in [14] with its objective function being f_{avg}. As for the correlation structure within each column of the Gaussian random matrix, we use an exponentially decaying correlation model, namely, $e^{-\alpha\tau}$, with α and τ respectively equal to a fixed adjustable constant and the lag between each two entries of a column. In contrast with the proposed approach where the ICC is adjusted in a learning-based framework, we call the latter method the measurement matrix with fixed ICC. Also, the objective function f_{avg} suggests an averaging over training signals, i.e. j, before optimizing for the index set Ω used in $\mathbf{P}_\Omega$ [14]. The corresponding results are shown in Table 1 and Table 2, respectively for the case of binary and Gaussian sparse signals. It is inferred that the proposed method, due to making use of the learning-based procedure to update the ICC, outperforms the method with fixed ICC. Also, compared to the deterministic learning-based measurement matrix in [14], the random structure of the proposed measurement matrix is observed to provide better reconstruction accuracy.

Finally, to investigate the relative performance of the proposed approach in real-world scenarios, we implement the proposed approach along with other methods on a data set consisting of MRI images of knee, available in mridata [19]. For this purpose, we use the discrete cosine transform (DCT) as the sparse domain in which the MRI images are represented. Also, to be able to evaluate the performance for different image sizes, we downsample the original images. In Fig. 4 parts (a) and (b), the normalized MSE in the reconstruction of the MRI images versus the image resolution and the compression ratio (namely, m/n) has been shown, in respect. The visibly smaller reconstruction error in different conditions confirms the advantage of using the proposed approach for measurement matrix design in practical scenarios.

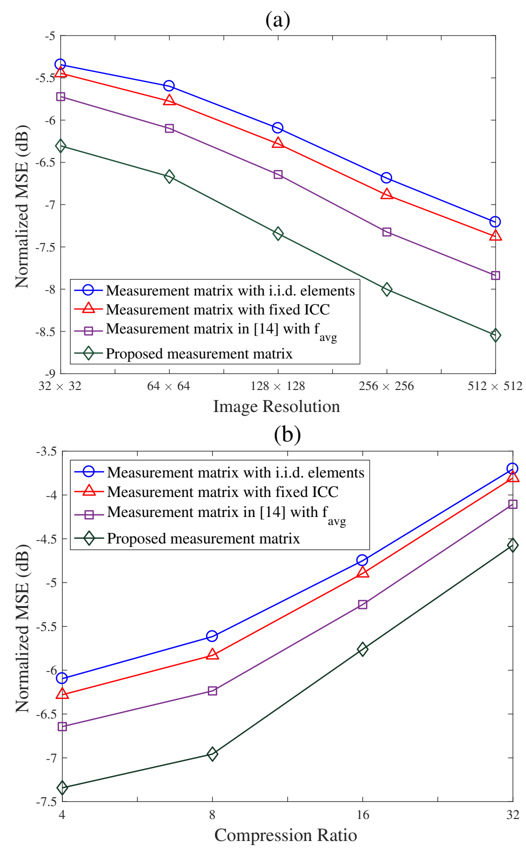

Fig. 4: (a): Normalized MSE versus image resolution for a compression ratio of 4, (b): Normalized MSE versus compression ratio for images of resolution 128×128.

4. CONCLUDING REMARKS

We presented a new approach to design CS random measurement matrix, $\mathbf{\Phi}$, for CS by generalizing the conventional Gaussian matrix with i.i.d. entries. Our contribution is to consider the correlation within the entries of each column, namely, the ICC. The latter is related to a variance parameter, v, and is obtained by expressing the measurement matrix as $\mathbf{P}(v)\mathbf{\Psi}$ and using the maximum-energy criterion in a learning-based framework to obtain the optimal value for v. Experimentations show the advantage of integrating the proposed training-based $\mathbf{\Phi}$ into various basic CS methods as well as its superior performance relative to a few previous measurement matrix design methods. Future directions involve the employment of different correlation structures in $\mathbf{\Phi}$ based on the features of the sparse signal.

5. REFERENCES

[1] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Springer New York, 2013.

[2] E. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[3] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[4] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[5] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[6] C. Hegde, A. C. Sankaranarayanan, W. Yin, and R. G. Baraniuk, "NuMax: A convex approach for learning near-isometric linear embeddings," *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 6109–6121, Nov 2015.

[7] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4053–4085, Sept 2011.

[8] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4997–5008, Oct 2014.

[9] J. Fang, J. Li, Y. Shen, H. Li, and S. Li, "Super-resolution compressed sensing: An iterative reweighted algorithm for joint parameter learning and sparse signal recovery," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 761–765, June 2014.

[10] M. L. Malloy and R. D. Nowak, "Near-optimal adaptive compressed sensing," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 4001–4012, July 2014.

[11] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 5117–5144, Sept 2016.

[12] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1395–1408, July 2009.

[13] Y. Yu, A. P. Petropulu, and H. V. Poor, "Measurement matrix design for compressive sensing-based MIMO radar," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5338–5352, Nov 2011.

[14] L. Baldassarre, Y. H. Li, J. Scarlett, B. Gözcü, I. Bogunovic, and V. Cevher, "Learning-based compressive subsampling," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 4, pp. 809–822, June 2016.

[15] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 139–154, Jan 2012.

[16] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, April 2007.

[17] 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.

[18] N. Lee, "MAP support detection for greedy sparse signal recovery algorithms in compressive sensing," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 4987–4999, Oct 2016.

[19] M. Lustig and S. Vasanawala, "MR datasets for compressed sensing," Last accessed on September 2017, available at `http://www.mridata.org/`.